Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 42 (2007) 81–86
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

QSAR analysis of tyrosine kinase inhibitor using modified ant colony optimization and multiple linear regression

Wei-min Shi [a], Qi Shen [a,b,*], Wei Kong [a], Bao-xian Ye [a]

[a] *Chemistry Department, Zhengzhou University, Zhengzhou 450052, China*
[b] *State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China*

Received 21 December 2005; received in revised form 19 July 2006; accepted 11 August 2006
Available online 22 September 2006

Abstract

Quantitative structure–activity relationship (QSAR) models of inhibiting action of some analogues of 4-(3-bromoanilino)-6,7-dimethoxyquinazoline on epidermal growth factor receptor tyrosine kinase were constructed using modified ant colony optimization (ACO) method. As a comparison to this method, the evolutionary algorithm (EA) was also tested. It has been demonstrated that the modified ACO is a useful tool for variable selection comparable to EA. In the selected descriptors, electronic descriptor σ_Y^- is the most important descriptor in predicting EGFR inhibitory activity. Electron-donating groups such as Y-substituents enhance the activity as evident by negative σ_Y^-. In addition, for quinazoline substituents, nitro group has a large deactivating effect.

Keywords: Epidermal growth factor receptor (EGFR); QSAR; Ant colony optimization; Tyrosine kinase

1. Introduction

Many of the tyrosine kinase enzymes are involved in cellular signaling pathways and regulate key cell functions such as proliferation, differentiation, anti-apoptotic signaling and neurite outgrowth. Unregulated activation of these enzymes, through mechanisms such as point mutations or over-expression, can lead to a large percentage of clinical cancers [1,2]. The importance of tyrosine kinase enzymes in health and disease is further underscored by the existence of aberrations in tyrosine kinase enzymes signaling occurring in inflammatory diseases and diabetes. Inhibitors of tyrosine kinase as a new kind of effective anticancer drug are important mediators of cellular signal transduction that effects growth factors and oncogenes on cell proliferation [3,4]. The development of tyrosine kinase inhibitors has therefore become an active area of research in pharmaceutical science.

Epidermal growth factor receptor (EGFR) which plays a vital role as a regulator of cell growth is one of the intensely studied tyrosine kinase targets of inhibitors. EGFR is overexpressed in numerous tumors, including those derived from brain, lung, bladder, colon, breast, head and neck. EGFR hyperactivation has also been implicated in other diseases including polycystic kidney disease, psoriasis and asthma [5–7]. Since the hyperactivation of EGFR has been associated with these diseases, inhibitor of EGFR has potential therapeutic value and it has been extensively studied in the pharmaceutical industry.

One could not, however, confirm that the compounds designed would always possess good inhibitory activity to EGFR, while experimental assessments of inhibitory activity of these compounds are time-consuming and expensive. Consequently, it is of interest to develop a prediction method for biological activities before the synthesis. Quantitative structure–activity relationship (QSAR) searches information

* Corresponding author. Chemistry Department, Zhengzhou University, Zhengzhou 450052, China. Tel.: +86 371 67767957; fax: +86 371 67763220.
E-mail address: shenqi@zzu.edu.cn (Q. Shen).

0223-5234/$ - see front matter © 2006 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2006.08.001

relating chemical structure to biological and other activities by developing a QSAR model. Using such an approach one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be really synthesized and tested.

QSAR analysis is the analysis of the quantitative relationship between the experimental activity of a set of compounds and their structural properties using statistical methods. The experimental information may associate with biological properties, such as activity, toxicity or bioavailability, which are taken as dependent variables in building a model. The parameters to be calculated are numerous descriptors that are indicative of molecular structures. In QSARs, the number of compounds with the biological activity values available is usually small compared to the number of structural descriptors. This can lead either to possible overfitting or even to a complete failure in building a meaningful regression model. The selection of variables that are really indicative of the biological activity concerned is necessary for producing a useful predictive model and becoming one of the key steps in QSAR studies.

There have been many variable selection methods existing, the mostly used ones are stepwise regression, simulated annealing [8], evolutionary algorithms (EAs) [9,10] and genetic algorithms (GAs) [11,12] and so on. Here, ant colony optimization (ACO) algorithms [13–15] introduced to QSAR by present authors were used to perform the variable selection [16]. ACO has emerged recently as a stochastic optimization approach, which originated as a simulation of ant colony systems. ACO algorithms developed by Dorigo et al. have been inspired by colonies of real ants, which deposit a chemical substance called pheromone on the ground. This substance influences the choices the ants make: the larger the amount of pheromone deposited on a particular path, the larger the probability an ant selects that path. As a novel computational approach, ACO algorithms have attracted attention of researchers in many fields [17–23].

For the ease of multiple linear regression (MLR) implementation and the interpretability of the resulting equations, MLR techniques are used for building QSAR models. In the present work, we employed modified ACO algorithm for variable selection in MLR analysis of EGFR inhibitory activity and compared it to EA. It has been demonstrated that the modified ACO is a useful tool for variable selection comparable to EA.

2. Algorithms and data sets

2.1. Ant colony optimization algorithm

The ACO algorithms developed by Dorigo et al. have been inspired by the behavior of real ant colonies [13–15]. ACO can best be described by using the traveling salesman problem (TSP). Given a set of n cities with known distances between each pair of them, the aim of the TSP is to find the shortest path to travel all the cities exactly once and return to the starting city.

ACO was applied to TSP in the following way. Let $b_i(t)$ be the number of ants in city i at time t, and let $m = \sum_{i=1}^{n} b_i(t)$ be the total number of ants; let $\tau_{ij}(t)$ be the intensity of pheromone trail on connection (i,j) at time t. $\tau_{ij}(0) = \tau_0$ which is the initial amount of pheromone deposited on each of the edges. A certain amount of pheromone is dropped on connection (i,j) that ants move on it. As time goes on, the pheromone left gradually vanished. The pheromone level on the selected edge is updated according to the updating rule.

$$\tau_{ij}(t+n) = \rho\tau_{ij}(t) + \Delta\tau_{ij} \tag{1}$$

where $0 < \rho < 1$ and ρ is a coefficient which represents the extent of the pheromone retained on the path.

$$\Delta\tau_{ij} = \sum_{k=1}^{m} \Delta\tau_{ij}^{(k)} \tag{2}$$

$\Delta\tau_{ij}$ presented the increment of pheromone left on the path ij at this circle. $\Delta\tau_{ij}^{(k)}$ showed the pheromone that ant k left on the path ij at this circle. For each edge, the intensity of trail at time 0 i.e. $\tau_{ij}(0)$ is set to 0.

$$\Delta\tau_{ij}^{(k)} = \begin{cases} \dfrac{Q}{L_K} & \text{if ant } k \text{ goes from connection}(i,j) \\ 0 & \text{else} \end{cases} \tag{3}$$

where Q is a constant and L_K is the tour length found by the kth ant. Ant k makes its decision to move according to the pheromone amount on each path. The moving probability is

$$p_{ij}^{(k)} = \begin{cases} \dfrac{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}{\sum_{s\in \text{allowed}_k} \tau_{is}^{\alpha}\eta_{is}^{\beta}} & j \in \text{allowed}_k \\ 0 & \text{else} \end{cases} \tag{4}$$

here $\text{allowed}_k = \{0, 1, \cdots n-1\} - \text{tabu}_k$ is the set of cities available for ant k to select the next step, tabu_k records all the cities that the ant had ever passed. Visibility η_{ij} represents a local heuristic function. For the TSP, $\eta_{ij} = d(c_i, c_j)^{-1}$. α and β are parameters that allow a user to control the relative importance of pheromone trail versus visibility.

In other words the algorithm works as follows: m ants are positioned on a certain city, then every ant moves to another town which was chosen with a probability given by Eq. (4). The pheromone left on each path was updated according to Eqs. (1)–(3). This process is iterated until the minimum error criterion is attained or the number of iterations reaches a user-defined limit.

2.2. Modified ant colony optimization

According to the information positive feedback and the indirect communication mechanism of ACO, a modified ACO for variable selection is proposed as follows. For a variable selection problem expressed in a binary notation, an ant moves in an N-dimensional search space of N variables, it's motion is restricted to 0 or 1 on each dimension. State "1" represents the selection of this variable and state "0" represents the

reverse. In binary variable selection problem, the motion of ants is determined by moving probability of 0 or 1. The pheromone levels on each dimension (variable) rather than on a path are divided into two kinds: τ_{i0} and τ_{i1}, which represent the pheromone of a dimension i taking the values 1 and 0, respectively. The pheromone levels corresponding to a dimension taking the value 1 or 0 are updated according to the updating rule.

$$\tau_{i0}(\text{new}) = \rho\tau_{i0}(\text{old}) + \Delta\tau_{i0} \tag{5}$$

$$\Delta\tau_{i0} = \sum_{k=1}^{m} \Delta\tau_{i0}^{(k)} \tag{6}$$

$$\tau_{i1}(\text{new}) = \rho\tau_{i1}(\text{old}) + \Delta\tau_{i1} \tag{7}$$

$$\Delta\tau_{i1} = \sum_{k=1}^{m} \Delta\tau_{i1}^{(k)} \tag{8}$$

where $\Delta\tau_{i0}$ and $\Delta\tau_{i1}$ presented the increment of pheromone corresponding to i dimension taking the value 1 or 0 at this circle. $\Delta\tau_{i0}^{(k)}$ and $\Delta\tau_{i1}^{(k)}$ showed the amount of pheromone that ant k left on the variable i at this circle. For each dimension, the intensity of pheromone at time 0(τ_{i0} and τ_{i1}) is set to 0.

$$\Delta\tau_{i1}^{(k)} = F + F_{\text{H}} \quad \text{if } k\text{th ant selected variable } i \text{ both in the current iteration and in it's global best solution} \tag{9}$$

$$\Delta\tau_{i1}^{(k)} = F \quad \text{if } k\text{th ant selected variable } i \text{ only in the current iteration} \tag{10}$$

$$\Delta\tau_{i1}^{(k)} = F_{\text{H}} \quad \text{if } k\text{th ant selected variable } i \text{ only in it's historical global best solution} \tag{11}$$

$$\Delta\tau_{i0}^{(k)} = F + F_{\text{H}} \quad \text{if variable } i \text{ was not selected by ant } k \text{ in the current iteration either in it's historical global best solution} \tag{12}$$

$$\Delta\tau_{i0}^{(k)} = F \quad \text{if variable } i \text{ was not selected by ant } k \text{ in the current iteration} \tag{13}$$

$$\Delta\tau_{i0}^{(k)} = F_{\text{H}} \quad \text{if variable } i \text{ was not selected by ant } k \text{ in it's historical global best solution} \tag{14}$$

where F and F_{H} are defined by fitness function. For improving the convergence velocity, the information F_{H} which corresponds to the historical global best result of the ith ant was introduced to the increment of pheromone ($\Delta\tau_{i0}^{(k)}$ and $\Delta\tau_{i1}^{(k)}$). Ant k makes decision concerning the variable selection according to the pheromone amount. The moving probability is

$$p_i^{(k)} = \frac{\tau_{i1}}{\tau_{i1} + \tau_{i0}} \tag{15}$$

In the modified ACO, m ants select variables from all N variables according to the probability defined by Eq. (15). After one selection, the amount of pheromone is updated according to Eqs. (5)–(14). This process is iterated until the minimum error criterion is attained or the number of iteration reaches a user-defined limit.

In the modified ACO, the pheromone levels were updated not only by current individual's information but also by each ant's previous or historical global best performance, so the information positive feedback in the modified ACO was really different from that in the conventional ACO. Using each ant's previous best information, the modified ACO converges quite quickly towards the optimal position with satisfactory converging characteristics. The detail of the modified ACO has been described elsewhere [11].

2.3. Fitness function

In the modified ACO, the increment of pheromone left on a certain variable is measured according to a pre-defined fitness function. The following objective function is applied to variable selection in the modified ACO:

$$F = -\lg\left(\text{RSS}_p/\widehat{\sigma}^2_{\text{PLS}} + 2p\right) \tag{16}$$

here p is the number of dependent variables, RSS_p is the residual sum of squares of p-variable model, $\widehat{\sigma}^2_{\text{PLS}}$ is defined as the value of RSS corresponding to the minimum number of principal components in conventional PLS analysis of the original data set when further increase of the number of principal components does not cause a significant reduction in RSS. The smaller the residual sum of model is and the fewer variables are involved in the model, the larger the fitness function is and the higher the probability that the model is being selected.

2.4. Tyrosine kinase inhibitor data

Sixty-one analogues of 4-(3-bromoanilino)-6,7-dimethoxyquinazoline with the corresponding inhibitory activities taken from the study by Bridges et al. [5] were used as a data set for variable selection and QSAR analysis. The chemical structures are represented in Fig. 1. A list of inhibitory activities is given in Table 1. We randomly divided the data into two subsets, a training set of 47 compounds, and a predicting set of 14 compounds.

Over 100 descriptors were calculated, which encoded different aspects of the molecular structure and consist of electronic, thermodynamic, spatial, and structural descriptors. The descriptor analysis involves the detection and removal of those structural descriptors which exhibit high pair-wise correlations with other descriptors, or which contain little discriminatory information. Pairs of descriptors that are highly correlated ($r \geq 0.90$) encoded similar information, and one of them should be removed. Descriptors that contain a high percentage (≥90%) of identical values are also discarded. Thus, only 50 of total descriptors were used which are listed

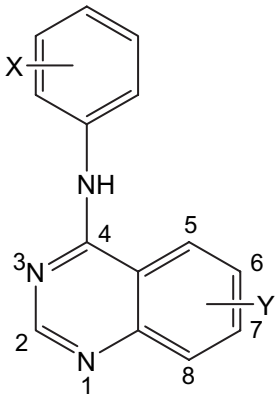

Fig. 1. Structure of 4-(X-bromoanilino)-Y-quinazolines.

in Table 2. All these molecular descriptors were generated using Cerius2[3.5] software on Silicon Graphics R3000 workstation. The modified ACO, EA and MLR algorithms were written in Matlab 5.3 and run on a personal computer (Intel Pentium processor 4/1.5 GHz 256 MB RAM).

3. Result and discussion

For the selection of the most important descriptors the modified ACO was used, which contained a population of 100 individuals and evolved for 200 iterations. Then the selected descriptors were taken as independent variables to build QSAR models with MLR method (ACO–MLR). The names of selected descriptors by modified ACO and statistical parameters obtained by these models for the training and prediction set are shown in Table 3. The best model with the largest fitness value includes seven descriptors and the second best one includes six descriptors. It was found that by using six descriptors the correlation coefficients (R^2) for the training and the test sets were 0.7011 and 0.7223, respectively. The standard deviation for the training set was 0.8917. The leave-one-out cross-validated correlation coefficient (R^2_{cv}) and standard deviation were 0.6347 and 0.9884, respectively, for the six-descriptor model. In these equations, negative coefficient of PMI-*X* indicates that increasing the principal moments of inertia about the *X*-axis of a molecule causes inhibitory activity decreasing. Y-substituents at 7-position enhance the activity by a positive steric effect shown by $B1_{Y,7}$. The positive coefficient of $B1_{X,3}$ indicates that the X-substituents at the 3-position of the 4-phenylamino moiety is beneficial for the activity. The negative coefficient of descriptor LUMO implied that molecules with low-energy LUMOs would promote the inhibitory activity. Shadow-*XY*frac [24] is a fraction of area of molecular shadow in the *XY* plane over an area of enclosing rectangle. Its large coefficient being negative shows that the selectivity increased with the decreased value of shadow-*XY*-frac. That is to say the smaller area of molecular shadow in the enclosing rectangle will benefit the activity. σ_Y^- is the most frequently appearing descriptor. Electron-donating groups such as Y-substituents enhance the activity as evident by negative σ_Y^-. The σ values are parameterized with respect

Table 1
Summary of experimental and calculated inhibitor data for analogues of 4-(3-bromoanilino)-6,7-dimethoxyquinazoline along with their structures used in QSAR study

No.	Substituent		lg(1/IC_{50})		
	X	Y	Observed	Calculated[a]	Series[b]
1	H	H	6.46	6.05	1
2	3-F	H	7.25	6.46	1
3	3-CL	H	7.64	7.47	1
4	3-Br	H	7.57	7.51	1
5	3-I	H	7.10	8.44	2
6	3-CF_3	H	6.24	6.45	2
7	H	6-OMe	7.26	6.42	1
8	3-Br	6-OMe	7.52	6.97	2
9	H	6-NH_2	6.11	7.02	1
10	3-CF_3	6-NH_2	6.24	6.72	1
11	3-Br	6-NH_2	9.11	7.85	2
12	H	6-NO_2	5.30	4.88	1
13	3-Br	6-NO_2	6.05	6.92	1
14	H	7-OMe	6.92	6.90	1
15	3-Br	7-OMe	8.00	8.33	1
16	H	7-NH_2	7.00	7.79	1
17	3-F	7-NH_2	8.70	8.31	1
18	3-Cl	7-NH_2	9.60	9.40	1
19	3-Br	7-NH_2	10.00	9.31	1
20	3-I	7-NH_2	9.46	10.02	1
21	3-CF_3	7-NH_2	8.48	8.56	1
22	H	7-NO_2	4.92	4.13	1
23	3-F	7-NO_2	5.22	4.90	2
24	3-Cl	7-NO_2	6.09	6.24	1
25	3-Br	7-NO_2	6.00	6.28	1
26	3-I	7-NO_2	6.27	6.69	1
27	H	6,7-Di-OMe	7.54	8.10	1
28	3-F	6,7-Di-OMe	8.42	9.19	2
29	3-Cl	6,7-Di-OMe	9.51	10.12	2
30	3-Br	6,7-Di-OMe	10.60	8.75	1
31	3-I	6,7-Di-OMe	9.05	8.31	1
32	3-CF_3	6,7-Di-OMe	9.62	8.47	1
33	3-Br	6-NHMe	8.40	7.56	2
34	3-Br	6-NMe_2	7.08	8.08	1
35	3-Br	6-NHCOOMe	7.92	7.90	2
36	3-Br	7-OH	8.33	8.72	2
37	3-Br	7-NHCOMe	7.40	8.34	1
38	3-Br	7-NHMe	8.16	7.82	1
39	3-Br	7-NHC_2H_5	7.92	8.67	2
40	3-Br	7-NMe_2	7.96	8.01	2
41	3-Br	6,7-Di-NH_2	9.92	9.67	1
42	3-Br	6-NH_2, 7-NHMe	9.16	7.78	1
43	3-Br	6-NH_2, 7-NMe_2	6.80	8.01	1
44	3-Br	6-NH_2, 7-OMe	8.42	9.31	1
45	3-Br	6-NH_2, 7-Cl	8.19	7.83	1
46	3-Br	6-NO_2, 7-NH_2	7.28	7.40	1
47	3-Br	6-NO_2, 7-NHMe	7.17	7.37	1
48	3-Br	6-NO_2, 7-NMe_2	5.70	6.04	1
49	3-Br	6-NO_2, 7-NHCOMe	7.55	7.25	1
50	3-Br	6-NO_2, 7-OMe	7.82	7.39	1
51	3-Br	6-NO_2, 7-Cl	7.60	8.09	2
52	3-Br	6,7-Di-OH	9.77	9.05	1
53	3-Br	6,7-Di-OC_2H_5	11.22	9.31	1
54	3-Br	6,7-Di-OC_3H_7	9.77	8.46	1
55	3-Br	6,7-Di-OC_4H_9	6.98	7.34	1
56	3-Br	6,7-Di-OMe	5.86	8.51	1
57	3-Br	5,6,7-Tri-OMe	9.17	9.55	2
58	2-Br	6,7-Di-OMe	6.89	7.95	1
59	4-Br	6,7-Di-OMe	9.02	9.03	1
60	3,4-di-Br	6,7-Di-OMe	10.14	9.96	1
61	3,5-di-Br	6,7-Di-OMe	6.95	7.42	1

[a] Calculated by the second equation in Table 3.
[b] Randomly selected as the member of training (1) and validating (2) sets.

Table 2
List of molecular descriptors for 4-(X-bromoanilino)-Y-quinazolines as candidate variables

Functional families of descriptors	Descriptors
Spatial descriptors	RadOfGyration (radius of gyration), shadow indices (surface area projections) (shadow-*XY*, shadow-*XZ*, shadow-*YZ*, shadow-*XY*frac, shadow-*XZ*frac, shadow-*YZ*frac, shadow-nu, shadow-*X*length, shadow-*Y*length, shadow-*Z*length), Vm (molecular volume), density, area (molecular surface area), PMI (principal moment of inertia) (PMI-mag-*X*, PMI-mag-*Y*, PMI-mag-*Z*) $B1_{Y,7}$, $B1_{X,3}$ (Veloop's sterimol parameter).
Structural descriptors	MW (molecular weight), Hbond acceptor (number of hydrogen bond acceptors), Hbond donor (number of hydrogen bond donors), Rotbonds (number of rotatable bonds).
Electronic descriptors	Apol (sum of atomic polarizabilities), dipole (dipole-mag, dipole-*X*, dipole-*Y*, dipole-*Z*) Sr (superdelocalizability), σ_Y^- (the Y-substituent effect on aromatic system).
Quantum mechanical descriptors	HOMO (highest occupied molecular orbital energy), LUMO (lowest unoccupied molecular orbital energy).
Thermodynamic descriptors	*C* log *P*, log *P* (the octanol/water partition coefficient), Fh2o(desolvation free energy for water), Foct (desolvation free energy for octanol), $MR_{CM^{**}\text{-}3}$, MolRef (molar refractivity).
E-state index	S-ssCH_3, S-aaCH, S-aasC, S-aaaC, S-NH_2, S-ssNH, S-aaN, S-sF, S-sCl, S-sBr
Indicator variable	*I*

to the 1,3-*N*- in the quinazoline ring. This indicates that the increased electron density on these nitrogens improves the binding of the molecules to the receptor. That is in accordance with the conclusions given by Bridges et al. [5] that for quinazoline substituents, nitro group has a large deactivating effect. It merits attention that descriptors relevant *X* coordinate axis such as shadow-*XY*frac and PMI-*X* has deactivating effect and shadow-*YZ*frac independent of *X*-axis has activating effect. The other descriptors are *E*-state indices. *E*-state indices [25,26] combine the electronic state of the bonded atom with its topological nature in the context of the whole molecule. In the symbol S-aaCH, 'S' represents electronic topological state of carbon atoms connected with hydrogen in the phenyl ring, 'a' stands for the bond in an aromatic ring, and 'C' represents the formula of the carbon. The nature of aromatic rings is structurally important for inhibitory potency. Negative coefficient of the descriptor 'S-aaCH' suggests that an increase in the value of S-aaCH would reduce the activity of molecules. Hydrophobicity, a factor usually much considered in the development of QSAR in biochemistry, seems to have a small effect on the activity as *A* log *P* is excluded in these equations and in the most frequently appearing descriptors. This may be due to values of "*A* log *P*" of most compounds that lie between 4 and 6. The correlation between the calculated and experimental values of lg $1/IC_{50}$ of six-descriptor model is shown in Fig. 2. The correlation coefficient (R^2) for the training set is 0.7011, and Rp for the prediction set is 0.7223. In Fig. 2, there is an obvious outlier with rather high deviation of calculated activity from the experimentally measured value of compound 56. This value for this compound is an outlier also in all equations shown in Table 3, no matter whether it was placed in the training sets or predicted sets.

Compared to the modified ACO, EAs were used for variable selection and the selected descriptors were also used for constructing QSAR models (EA–MLR). The minimum fitness value can be obtained in the searching process not only by the modified ACO but also by EAs. But fitness value drops more quickly in the modified ACO algorithm. The best fitness can be obtained after 35 iterations during the modified ACO searching, but it needs 55 iterations during EAs search.

4. Conclusions

The modified ACO algorithm has been employed in variable selection and satisfactory results have been obtained. In the selected descriptors, electronic descriptor σ_Y^- is the most important descriptor in predicting EGFR inhibitory activity. Electron-donating groups such as Y-substituents enhance the activity as evident by negative σ_Y^-. In addition, for quinazoline substituents, nitro group has a large deactivating effect. The modified ACO has been testified to be an effective method for variable selection comparable to EA. The statistical criteria of Table 3 feel relatively weak by looking to R2. Bridges et al. [5] hypothesized that certain substituted analogues possess the ability to induce a change in the conformation of the receptor when they bind with it, and the mechanism of these compounds may be different. Are the weak statistical criteria of Table 3 due to the method, or due to quality of experimental

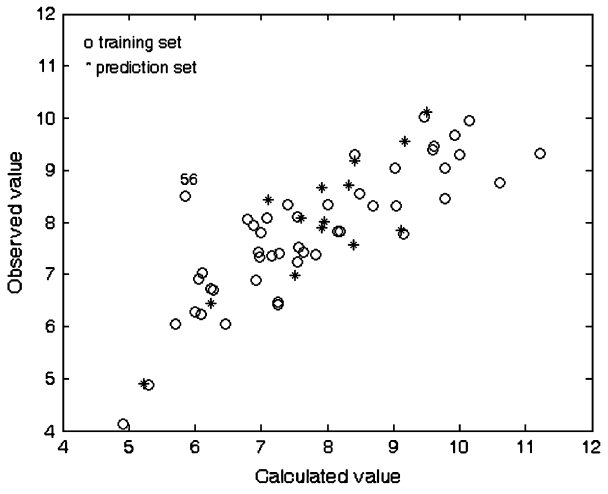

Fig. 2. Calculated versus observed lg $1/IC_{50}$ of six-variable model using MLR modeling.

Table 3
Results of variable selection by the modified ACO and MLR modelings

No.	Equation	R^{2a}	$R^{2\ b}_{cv}$	S^{a}	$S_{cv}{}^{c}$	F^{a}	$Rp^{2\ d}$
1	$\lg 1/IC_{50} = -18.5080 \times$ shadow-*XY*frac $- 0.0098 \times$ PMI-*X* $- 1.9034 \times \sigma^{-}_{Y} + 1.8556 \times B1_{Y,7} + 3.5000 \times B1I_{X,3} + 12.6039$	0.6704	0.5894	0.9248	1.0374	16.6859	0.5006
2	$\lg 1/IC_{50} = -20.9847 \times$ shadow-*XY*frac $- 0.0155 \times$ PMI-*X* $- 1.8850 \times \sigma^{-}_{Y} + 3.7443 \times B1_{X,3} - 0.2962 \times$ S-aaCH $- 0.0466 \times$ S-sF $+ 21.4143$	0.7011	0.6347	0.8917	0.9884	15.6408	0.7223
3	$\lg 1/IC_{50} = -22.5296 \times$ shadow-*XY*frac $- 0.4380 \times$ LUMO $+ 8.1485 \times$ shadow-*YZ*frac $- 0.0120 \times$ PMI-*X* $- 2.0967 \times \sigma^{-}_{Y} + 1.7580 \times B1_{Y,7} + 3.0909 \times B1_{X,3} + 11.2022$	0.7172	0.6247	0.8785	1.0187	14.1275	0.6422

[a] R^2: square of correlation coefficient; *S*: standard deviation; *F*: *F*-statistics.
[b] R^2_{cv}: square of the leave-one-out cross-validated correlation coefficient.
[c] S_{cv}: leave-one-out cross-validated standard deviation.
[d] Rp^2: square of correlation coefficient of prediction set.

data or due to the hypothesis? It cannot be said without more extensive work and benchmarks. Work is underway to face this interrogation.

Acknowledgements

The work was financially supported by the National natural Science Foundation of China (Grant No. 20505015, 20475050).

References

[1] A. Kurup, R. Garg, C. Hansch, Chem. Rev. 101 (2001) 2573–2600.
[2] B.D. Palmer, A.J. Kraker, B.G. Hartl, A.D. Panopoulos, R.L. Panek, B.L. Batley, G.H. Lu, T.K. Susanne, H.D.H. Showalter, W.A. Denny, J. Med. Chem. 42 (1999) 2373–2382.
[3] M. Oblak, M. Randic, T. Solmajer, J. Chem. Inf. Comput. Sci. 40 (2000) 994–1001.
[4] T. Naumann, T. Matter, J. Med. Chem. 45 (2002) 2366–2378.
[5] A.J. Bridges, H. Zhou, D.R. Cody, G.W. Rewcastle, A. McMichael, H.D.H. Showalter, D.W. Fry, A.J. Kraker, W.A. Denny, J. Med. Chem. 39 (1996) 267–276.
[6] C. Ma, K.A. Bower, H. Lin, G. Chen, C. Huang, X. Shi, J. Luo, Biochem. Pharmacol. 69 (2005) 1785–1794.
[7] R. Albuschat, W. Lowe, M. Weber, P. Luger, V. Jendrossek, Eur. J. Med. Chem. 39 (2004) 1001–1011.
[8] Jon M. Sutter, Steve L. Dixon, Peter C. Jurs, J. Chem. Inf. Comput. Sci. 35 (1995) 77–84.
[9] D. Rogers, A.J. Hopfinger, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[10] A. Yasri, D. Hartsough, J. Chem. Inf. Comput. Sci. 41 (2001) 1218–1227.
[11] Hugo Kubinyi, J. Chemom. 10 (1996) 119–133.
[12] F. Dieterle, B. Kieser, G. Gauglitz, Chemom. Intell. Lab. Syst. 65 (2003) 67–81.
[13] A. Colorni, M. Dorigo, V. Maniezzo, in: F. Varela, P. Bourgine (Eds.), Proceedings of the 1st European Conference on Artificial Life, Elsevier, Paris, France, 1991, pp. 134–142.
[14] E. Bonabeau, M. Dorigo, G. Theraulaz, Nature 406 (2000) 39–42.
[15] M. Dorigo, L.M. Gambardella, BioSystems 43 (1997) 73–81.
[16] Q. Shen, J.H. Jiang, J.C. Tao, G.L. Shen, R.Q. Yu, J. Chem. Inf. Model. 45 (2005) 1024–1029.
[17] Y. Wang, J.Y. Xie, in: IEEE: The 2000 IEEE Asia–Pacific Conference, 2000, pp. 54–57.
[18] A. Baue, B. Bullnheimer, R.F. Hartl, et al., in: CEC 99, Proceedings of the 1999 Congress on Evolutionary Computation, 1999, vol. 2, pp. 1450–1455.
[19] M. Dorigo, L.M. Gambardella, IEEE Trans. Evol. Comput. 1 (1997) 53–66.
[20] A.R.M. Silva, G.L. Ramalho, in: 2001 IEEE International Conference on Man and Cybernetics, 2001, vol. 5, pp. 3129–3133.
[21] W.Q. Xiong, P. Wei, in: International Conference on Proceedings, 2002, vol. 1, pp. 552–555.
[22] Y.M. Li, Z.B. Xu, in: Fifth International Conference on Computational Intelligence and Multimedia Applications, 2003, pp. 206–211.
[23] J.F. Gomez, H.M. Khodr, P.M. DeOliveira, et al., IEEE Trans. Power Syst. 19 (2004) 996–1004.
[24] R.H. Roxburgh, P.C. Jurs, Anal. Chim. Acta 199 (1987) 99–109.
[25] L.H. Hall, L.B. Kier, J. Chem. Inf. Comput. Sci. 31 (1991) 76–78.
[26] L.H. Hall, L.B. Kier, J. Chem. Inf. Comput. Sci. 35 (1995) 1039–1045.