Modified Ant Colony Optimization Algorithm for Variable Selection in QSAR Modeling: QSAR Studies of Cyclooxygenase Inhibitors

Qi Shen,[†,‡] Jian-Hui Jiang,[†] Jing-chao Tao,[†,‡] Guo-li Shen,[†] and Ru-Qin Yu*[,†]

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

Received December 29, 2004

A new version of an ant colony optimization (ACO) algorithm has been proposed. A modified ACO algorithm is proposed to select variables in QSAR modeling and to predict inhibiting action of some diarylimidazole derivatives on cyclooxygenase (COX) enzyme. As a comparison to this method, the evolution algorithm (EA) was also tested. Experimental results have demonstrated that the modified ACO is a useful tool for variable selection that needs few parameters to be adjusted and converges quickly toward the optimal position.

1. INTRODUCTION

Quantitative structure−activity relationship (QSAR) searches information relating chemical structure to biological and other activities by developing a QSAR model. Building a QSAR model begins with collecting experimental data and calculating theoretical parameters for the compounds involved. The experimental information may associate with biological properties, such as activity, toxicity, or bioavailability, which are taken as dependent variables in building a model. Chemical structure is represented by a variety of descriptors including spatial, electronic, topological, information-content, thermodynamic, conformational, quantum mechanical, and shape descriptors. Several hundreds even thousands of descriptors could be generated in QSAR studies. But only a part of them is statistically significant in terms of correlation with biological activity for a particular analysis, and variable selection is necessary for producing a useful predictive model. In QSARs, the number of compounds with the biological activity values available is usually small compared with 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 becoming one of the key steps in QSAR studies. The benefit gained from variable selection in QSAR is not only the stability of the model but also the interpretability of relationship between the descriptors and biological activity.

There have been many variable selection methods existing, the mostly used ones are stepwise regression, simulated annealing,[1] evolutionary algorithms,[2] and genetic algorithms (GAs),[3,4] to name just a few. Here, ant colony optimization (ACO) algorithms were introduced to perform the variable selection. ACO has emerged recently as a stochastic optimization approach, which originated as a simulation of ant colony systems. ACO algorithms developed by Dorigo[5−7] 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 the path. The ACO is a population-based approach and has been successfully applied to several combinational optimization problems,[8−14] such as traveling salesman problem (TSP), quadratic assignment problems, job-shop scheduling problem, and adaptive routing.

As a novel computational approach, ACO algorithms have attracted the attention of researchers in many fields. But there are some limitations of the ACO such as long searching time and the tendency to be trapped into local optima. Although ACO has been used for the TSP problem, the time required to find results of large size TSP problems is hardly bearable. As for many optimization methods, there are many parameters needed to be adjusted, and the ACO model is quite sensitive to these parameters. In the present paper, the ACO algorithm is modified to adopt the discrete combinatorial optimization problem of variable selection in QSAR. A modified ACO algorithm is proposed to select variables in MLR modeling of the cyclooxygenase inhibitor. The results were compared to those obtained by EAs. It has been demonstrated that the modified ACO is a useful tool for variable selection, which converges quickly toward the optimal position.

Nonsteroidal antiinflammatory drugs (NSAIDs) such as aspirin are widely utilized agents for the treatment of inflammation, pain, and fever. They exhibit their effect by inhibiting both distinct COX enzymes COX-1 and COX-2. However, decreased renal function, gastrointestinal ulceration, or bleeding often company the use of NSAIDs for the treatment of inflammation and pain.[15−18] As the level of COX-2 is very often significantly higher during periods of acute and chronic inflammation, developing selective COX-2 inhibitor is hopefully for finding antiinflammatory agents without the side effects of the current available NSAIDs. Experimental assessment of inhibitory ability can be expensive and time-consuming. QSAR can be used to predict inhibitory activity of COX-2 and helpful in the design of new COX-2 inhibitors.

* Corresponding author phone: +86-731-8821577; fax: 86-731-8822577; e-mail: rqyu@hnu.net.cn.
[†] Hunan University.
[‡] Zhengzhou University.

10.1021/ci049610z CCC: $30.25
Published on Web 04/28/2005

2. THEORY

2.1. Ant Colony Optimization Algorithm. The ACO algorithms have been inspired by the behavior of real ant colonies.[5−7] Real ants are capable of finding the shortest path from a food source to their colony (nest). They lay some pheromone on the ground, thus marking the path by the trail of the substance. The pheromone trail can be observed by other ants and motivates them to follow the path. A moving ant will follow the pheromone trail with a probability which is proportional to the number of ants choosing that path as each ant would reinforce the trail by depositing its own pheromone. The process is thus characterized by a positive feedback loop.

ACO can best be described by using the TSP problem. 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 to all cities exactly once and return to the starting city.

ACO was applied to the TSP problem in the following way. Let $b_i(t)$ be the number of ants in city i at time t, let $m = \sum_{i=1}^{n} b_i(t)$ be the total number of ants, and let $\tau_{ij}(t)$ be the intensity of the pheromone trail on connection (i,j) at time t. $\tau_{ij}(t) = \tau_0$ 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 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.

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

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

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

Here $allowed_k = \{0,1,\cdots n\text{-}1\} - tabu_k$ is the set of cities available for ant k to select the next step, and $tabu_k$ 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 control of the relative importance of the pheromone trail versus visibility.

In other words the algorithm works as follows: m ants are positioned on a certain city, and then every ant moves to another town which was chosen with a probability 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. In ACO ants use a chemical substance called a pheromone to communicate with each other, and this process is characterized by a positive feedback loop: the more ants that use a particular path, the more pheromone is deposited on that path, and the more it becomes attractive to other ants. Communication and positive feedback are the basic mechanism of the ACO algorithm. Pheromone evaporation is needed to avoid a too rapid convergence of the algorithm toward a suboptimal region and is in favor of the exploration of new areas of the search space. An enhanced positive feedback mechanism which means the large extent of the pheromone retained can not only speed up convergence speed but also make the algorithm possibly converging to local optima. A communication mechanism is beneficial to cooperation between individuals and helps the algorithm jumping out from local optima.

The conventional ACO algorithms were designed for solving ordering problems such as the traveling salesman problem. Variable selection in QSAR is a subset selection problem. Subset selection problems are quite different from ordering problems. Subset selection means to select the best subset out of a whole set. There is no concept of a path here, so it is difficult to apply the conventional ACO directly to variable selection in QSAR.

According to information positive feedback and the indirect communication mechanism of ACO, a modified ACO for variable selection is proposed. For a variable selection problem expressed in a binary notation, an ant moves in an N-dimensional search space of N variables, its 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, every ant selects the variables which were determined by a 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 value 1 and 0, respectively. The pheromone levels corresponding to a dimension taking the value 1 or 0 are updated according to the updating rule.

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

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

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

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

where $\Delta\tau_{i0}$ and $\Delta\tau_{i0}$ presented the increment of pheromone

corresponding to the 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 the pheromone at time 0 (τ_{i0} and τ_{i1}) is set to 0.

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

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

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

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

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

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

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 a 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.

There was no conception of path in the modified ACO. The moving probability with the value 1 or 0 was referred to each dimension rather than to each path. The pheromone levels corresponding to a dimension taking the value 1 or 0 were computed, and then the moving probability was calculated according to the pheromone amount to determine whether the variable was selected in the next iteration. In the modified ACO, the more the pheromone trail τ_{i1} was left on variable i, the more probable that variable would be selected. The information sharing mechanism is the same in the modified ACO and the conventional ACO. In the two ACO versions, the individual is updated according to information positive feedback and the indirect communication mechanism.

Using each ant's previous best information, the modified ACO converges quite quickly toward the optimal position with satisfactory converging characteristics. In the modified ACO, the pheromone levels were updated not only by the 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.

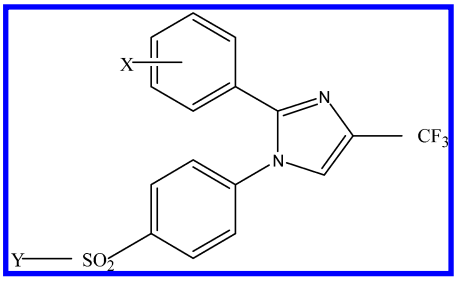

Figure 1. Structure of 1,2-diarylimidazole.

2.3. Fitness Function. In the modified ACO, the increment of pheromone left on a certain variable is measured according to a predefined fitness function. The following objective function is applied to variable selection in the modified ACO:

$$F = -\log\left(\mathrm{RSS}_p/\hat{\sigma}_{PLS}^{\,2} + 2{*}p\right) \tag{16}$$

Here p is the number of dependent variables, RSS_p is the residual sum of squares of p-variable model, and $\hat{\sigma}_{PLS}^2$ 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 a further increase of the number of principal components does not cause a significant reduction in RSS. The first term in the logarithmic part of eq 16 is defined as the accuracy of model, and the dimension of the model is restricted by the second term under the logarithm. The smaller the residual sum of the 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.

3. COX-2 INHIBITOR DATA

Forty-two 1,2-diarylimidazole derivatives with the corresponding inhibitory activities[19] were used to test the performance of the modified ACO in variable selection of QSAR analysis. The chemical structures are represented in Figure 1. A list of inhibitory activities is given in Table 1. The activity is expressed as IC_{50}, i.e., the molar concentration of the compound causing 50% inhibition of enzyme. We randomly divided the data taken from the study by Khanna et al. into two subsets, a training set of 34 compounds, and a predicting set of eight compounds.

Over 100 descriptors were calculated, which encoded different aspects of the molecular structure and consist of electronic, thermodynamic, spatial, and structural descriptors. After elimination of descriptors with very low variances, only 85 of total descriptors were used which are listed in Table 2. All these molecular descriptors were generated using Cerius2[3,5] software on Silicon Graphics R3000 workstation. The modified ACO, EAs, and MLR algorithms were written in Matlab 5.3 and run on a personal computer. (Intel Pentium processor 4/1.5G Hz 256 MB RAM).

4. RESULTS AND DISCUSSION

The modified ACO was first used for variable selection, which contained a population of 100 ants, and this process was iterated 200 cycles. Then the selected descriptors were

Table 1. Compounds with the Observed Bioactivities

compd	X	Y	Ln(1/IC_{50}) exp.
1	4-Cl	Me	2.2073
2	4-F	Me	2.3026
3	H	Me	2.1203
4[a]	4-Me	Me	1.8326
5	4-OMe	Me	0.5621
6	4-NHMe	Me	−0.3853
7	4-NMe_2	Me	0.3567
8[a]	4-SMe	Me	1.8326
9	4-Cl	NH_2	4.6052
10	4-F	NH_2	4.6052
11	H	NH_2	3.2189
12	4-Me	NH_2	3.2189
13	3-Cl	Me	2.8134
14[a]	3-Me	Me	2.8134
15	3-NMe_2	M	−1.1632
16	3-Cl	NH_2	4.8283
17	3-F	NH_2	3.5066
18[a]	3-Br	NH_2	4.9618
19[a]	3-Me	NH_2	3.5066
20	2-Me	Me	0.2231
21	2-F	NH_2	2.3026
22	2-Me	NH_2	1.6094
23	3-F-4-OMe	Me	1.8971
24	3-Cl-4-OMe	Me	2.0402
25[a]	3-Cl-4-NMe_2	Me	1.1394
26	3-F-4-NMe_2	Me	1.1087
27	3-Cl-4-Me	Me	3.5066
28	3-Me-4-F	Me	1.7720
29	3-Me-4-Cl	Me	2.4079
30[a]	3,4-OCH_2O-	Me	1.7720
31	3,4-Me_2	Me	1.1087
32	3-F-4-OMe	NH_2	3.5066
33	3-Cl-4-OMe	NH_2	3.9120
34	3-Br-4-OMe	NH_2	3.5066
35[a]	3-Cl-4-SMe	NH_2	4.6052
36	3-Cl-4-Me	NH_2	5.8091
37	3-OMe-4-Cl	NH_2	3.9120
38	3,4-F_2	NH_2	3.5066
39	3-Me-5-F	NH_2	3.5066
40	3,5-Me_2−4-OMe	Me	0.3285
41	3,5-F_2−4-OMe	NH_2	3.5066
42	4-Cl	Me	2.2073

[a] The compounds used for prediction.

taken as dependent variables to build QSAR models. The best model with a maximum fitness value contains 3 variables as given by the modified ACO search. The three variables are Jurs-TASA, Shadow-XZ, and ClogP. A summary of the performance variables of the model is provided in Table 3. The correlation between the experimental and calculated values of lg1/IC_{50} of eq 1 is shown in Figure 2.

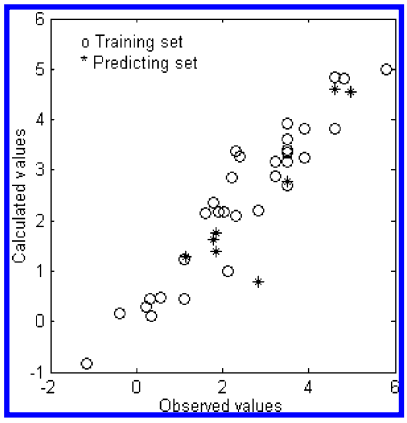

Figure 2. Calculated versus observed lg1/IC_{50} of three-variable model using MLR modeling.

The R^2 value for the training set is 0.8917, and Rp^2 value for the validation set is 0.7878. In the model presented in Table 3, Jurs-TASA denotes the sum of the solvent-accessible surface areas of atoms with an absolute value of partial charges less than 0.2. The negative coefficient of Jurs-TASA indicates that increasing the solvent-accessible surface areas causes inhibitory ability decreasing. The positive coefficient of ClogP indicates that the large hydrophobic character of the molecule is beneficial for the bioactivity. Shadow-XZ belongs to shadow indices descriptors. This set of descriptors helps to characterize the shape of the molecules.[20] They are calculated by projecting the molecular surface on three mutually perpendicular planes, *XY*, *YZ*, and *XZ*. They depend not only on the conformation but also on the orientation of the molecule. Among them shadow-*XZ* is a measure of the area of the projection of the "shadow" of the structure in the XZ plane. Its negative coefficient implied molecules with the large projecting shadow area in the *X*Z plane are detrimental to bioactivity.

To compare with the modified ACO, variable selection by EAs based MLR modeling was also performed. For comparison, 200 iterations are limited in both EAs and the modified ACO. The best model by EAs contains 3 descriptors and the three variables are Jurs-PNSA-3, Jurs-FNSA-2, and S-sNH2. The R^2 value for the training set is 0.8921, and the Rp^2 value for the validation set is 0.6665. Compared with ACO the correlation coefficient for the prediction set is rather low, though the correlation coefficient for the training set obtained by the equation is acceptable. This is a symptom of overfitting which seems to be related with the

Table 2. Descriptors Used for Selection in This Study

functional families of descriptors	descriptors
conformational descriptors	energy
electronic descriptors	Apol (sum of atomic polarizabilities), Dipole (dipole moment, including dipole-mag, dipole-X, dipole-Y, dipole-Z), HOMO (highest occupied molecular orbital energy), LUMO (lowest unoccupied molecular orbital energy), Sr (superdelocalizability)
spatial descriptors	RadOfGyration (radius of gyration), Jurs descriptors (Jurs charged partial surface area descriptors), Shadow indices (surface area projections), Area (molecular surface area), Density, PMI (principal moment of inertia, including PMI-mag, PMI-X, PMI-Y, PMI-Z), V_m (molecular volume)
structural descriptors	MW (molecular weight), Rotlbonds (number of rotatable bonds), Hbond acceptor (number of hydrogen bond acceptors), Hbond donor (number of hydrogen bond donors)
thermodynamic descriptors	AlogP (log of the partition coefficient), Fh2o (desolvation free energy for water), Foct (desolvation free energy for octanol), MolRef (molar refractivity)
E-state index	S_sCH_3, S_$ssCH_2$, S_aaCH, S_aasC, S_ssssC, S_sNH_2, S_aaN, S_aasN, S_dO, S_ssO, S_ssS, S_ddssS, S_sF, S_sCl
indicator variable	Iy, Lx,2

Table 3. Summary of the Performance Variables of the Best Model Given by the Modified ACO Algorithm[a]

descriptor	regression coeff	*t*-value
Jurs-TASA	−0.0195	0
Shadow-XZ	−0.1985	−0.0468
ClogP	1.5147	1.07946
Y-intercept	15.5984	0.9128

[a] R^2: square of correlation coefficient, $R^2 = 0.8917$; S: standard deviation, $S = 0.5494$; F: F-statistics, $F = 82.3132$; Rp^2: square of correlation coefficient of prediction set, $Rp^2 = 0.7878$; R_{max}: the maximum correlation coefficient among the variables, $R_{max} = 0.5831$.

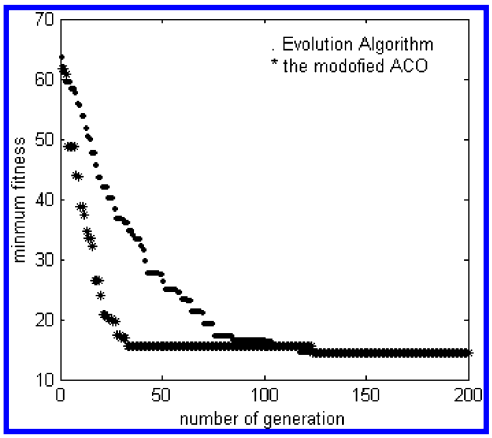

Figure 3. Minimum fitness value versus number of generation in EAs and the modified ACO algorithm.

relatively high correlation among variables (R_{max}=0.9221) involved in the equation. The best 2-variable model obtained by EAs is similar to that by ACO, and the two variables are Jurs-TASA and ClogP. The convergence results are listed in Figure 3. For the convenience of comparison, the ordinate of the graph represents ($RSS_p/\hat{\sigma}_{PLS}^2 + 2*p$) in Figure 3. From Figure 3 one can see that the minimum fitness value can be obtained in the searching process not only by the modified ACO but also by EAs. But the 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 70 iterations during the EAs search. A conventional ACO algorithm is relatively complicated with a slow searching speed and tends to fall into the local optima. But experimental results demonstrated that the modified ACO converges to the global best solution quickly.

When the modified ACO search terminates, 100 variable combinations are obtained by the population of 100 ants. One may count the number of times a particular molecular descriptor appears in 100 individual combinations. When one lists the descriptors by order of decreasing numbers of times of appearance, the most frequently appeared feathers are identified as the important descriptors. Jurs descriptors occupy an important position, and six of the 10 preferred descriptors are of the Jurs descriptor type. It indicated that Jurs descriptors play the most important role in association of COX-2 inhibitory activity. Jurs descriptors (charged partial surface area descriptors) encode features responsible for polar[21] and hydrophobic[22] interactions between molecules. This set of descriptors, combining the shape and electronic information to characterize the molecules, are calculated by mapping atomic partial charges on solvent-accessible surface areas of individual atoms. Therefore, one can infer that the intermolecular interactions play key roles in inhibiting COX-2 enzyme activity. The amino group at the Y-position, which has a smaller Jurs-TASA value than the methyl substituting group, enhances the activity of inhibitors, while the CH_3 substituting group at the Y-position reduces the inhibitory activity. The chloro substituting group at the X-position, which has a larger Jurs-TASA value than fluorine and methyl groups, reduces the activity of inhibitors, while fluorine and methyl groups at the X-position enhance the inhibitory activity. ClogP which related to the hydrophobic character of the molecule belongs to the factors with much attention being paid to the development of QSAR models in biochemistry. They are shown to be important in QSAR of inhibiting the COX-2 enzyme. Besides these variables, descriptors Shadow-XZ, density, and indicator variable Iy also have proved to be important for inhibiting activity. An indicator variable for the Y-substituted derivatives, Iy is assigned the value of 1 when an amino substituting group is present at the Y-position, and a value of 0 is assigned otherwise.

A parameter needed to adjust in the modified ACO is coefficient ρ which represents the extent to which the pheromone is retained, and $(1 - \rho)$ represents the pheromone evaporation. Pheromone evaporation $(1 - \rho)$ is really related to the accuracy of algorithm and convergence velocity. The coefficient (ρ) as defined enables a greater exploration of the search space and minimizes the chance of premature convergence to local optimal solutions. The strength of the pheromone determines the balance between the exploration of new points in the state space and the exploitation of accumulated knowledge. A too large value of parameter ρ is unfavorable for information positive feedback and makes the convergence speed slow. Setting ρ small is favorable for converging more easily but possibly causes overfitting. Accordingly, ρ is set to 0.7 by experience to keep balance between the accuracy and convergence speed.

5. CONCLUSION

A conventional ACO algorithm is relatively complicated, and there are many parameters needed to regulate it. Its searching speed is slow and tends to fall into the local optima. In the present study, the ant colony optimization algorithm is modified to be used in variable selection in QSAR modeling for predicting COX-2 inhibition activity. Comparing with EAs, the modified ACO converges to the best solution quickly. It has been demonstrated that the modified ACO is a useful tool for variable selection with nice performance and the ability to select preferred variables with satisfactory convergence rates.

ACKNOWLEDGMENT

The work was financially supported by the National Natural Science Foundation of China (Grant No. 20375012, 20372059, 20205005).

REFERENCES AND NOTES

(1) Shen, M.; LeTiran, A.; Xiao, Y.; et al. Quantitative Structure−Activity Relationship Analysis of Functionalized Amino Acid Anticonvulsant Agents Using *k* Nearest Neighbor and Simulated Annealing PLS Methods. *J. Med. Chem*. **2002**, *45*, 2811−2823.

(2) Brian, T. L. Evolutionary Programming Applied to the Development of Quantitative Structure−Activity Relationships and Quantitative

Structure−Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1279−128.
(3) Cho, S. J.; Hermsmeier, M. A. Genetic Algorithm Guided Selection: Variable Selection and Subset Selection. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 927−936.
(4) Yasri, A.; Hartsough, D. Toward an Optimal Procedure for Variable Selection and QSAR Model Building**.** *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1218−1227.
(5) Colorni, A.; Dorigo, M.; Maniezzo, V. Distributed optimization by ant colonies. *Proceedings of the 1st European Conference on Artificial Life;* Varela, F., Bourgine, P., Eds.; Elsevier: Paris, France, 1991; pp 134−142.
(6) Bonabeau, E.; Dorigo, M.; Theraulaz, G. Inspiration for optimization from social insect behavior. *Nature* **2000**, *406*, 39−42.
(7) Krieger, M. J.; Billeter, J. B.; Keller, L. Ant-like task allocation and recruitment in cooperative robots. *Nature* **2000**, *406*, 992−995.
(8) Wang, Y.; Xie, J. Y. Ant colony optimization for multicast routing in Circuits and Systems. IEEE: *The 2000 IEEE Asia-Pacific Conference;* 2000; pp 54−57.
(9) Baue, A.; Bullnheimer, B.; Hartl, R. F.; et al. An ant colony optimization approach for the single machine total tardiness problem. CEC 99, *Proceedings of the 1999 Congress on Evolutionary Computation;* 1999; Vol. 2, pp 1450−1456.
(10) Dorigo, M.; Gambardella, L. M. Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Trans. Evolutionary Comput.* **1997**, *1*, 53−66.
(11) Silva, A. R. M.; Ramalho, G. L. Ant system for the set covering problem Systems. *2001 IEEE International Conference on Man and Cybernetics;* 2001; Vol. 5, pp 3129−3133.
(12) Xiong, W. Q.; Wei, P. A kind of ant colony algorithm for function optimization. *Machine Learning and Cybernetics. International Conference on Proceedings;* 2002; Vol. 1, pp 552−555.
(13) Li, Y. M.; Xu, Z. B. An ant colony optimization heuristic for solving maximum independent set problems. *Fifth International Conference on Computational Intelligence and Multimedia Applications*; 2003; pp 206−211.
(14) Gomez, J. F.; Khodr, H. M.; DeOliveira, P. M.; et al. Ant Colony System Algorithm for the Planning of Primary Distribution Circuits Power Systems. *IEEE Transactions on Power Systems* **2004**, *19*, 996−1004.
(15) Gierse, J. K.; McDonald, J. J.; Hauser, S. D.; et al. A Single Amino Acid Difference Between Cyclooxygenase-1 (COX-1) and -2 (COX-2) Reverses the Selectivity of COX-2 Specific Inhibitors. *J. Biol. Chem.* **1996**, *271*, 15810−15814.
(16) Seibert, K.; Zhang, Y.; Leahy, K.; et al. Pharmacological and biochemical demonstration of the role of cyclooxygenase-2 in inflammation and pain. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12013−12017.
(17) O'Neill, G. P.; Ford-Hutchinson, A. W. Expression of mRNA for cyclooxyenase-1 and cyclooxygenase-2 in human tissues. *FEBS Lett.* **1993**, *330*, 156−160.
(18) Kargman, S.; Charleson, S.; Cartwright, M.; et al. Characterization of prostaglandin G/H synthase 1 and 2 in rat, dog, monkey, and human gastrointestinal tracts. *Gastroenterology* **1996**, *111*, 445−454.
(19) Khanna, L. K.; Weier, R. M.; Yu, Y.; et al. 1,2-Diarylimidazoles as Potent, Cyclooxygenase-2 Selective, and Orally Active Antiinflammatory Agents. *J. Med. Chem.* **1997**, *40*, 1634−1647.
(20) Rohrbaugh, R. H.; Jurs, P. C. Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Anal. Chim. Acta* **1987**, *199*, 99−109.
(21) Stanton, D. T.; Jurs, P. C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure−property relationship studies. *Anal. Chem.* **1990**, *62*, 2323−2329.
(22) Stanton, D. T.; Madhav, P. J.; Wilson, L. J.; et al. Development of a Quantitative Structure−Activity Relationship Model for Inhibition of Gram-positive Bacterial Cell Growth by Biarylamides. *J. Chem. Inf. Comput. Sci.* **2004,** *44*, 221−229.

CI049610Z