Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis
35 (2004) 679–687
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

QSAR analysis of cyclooxygenase inhibitor using particle swarm optimization and multiple linear regression

Jian-Xia Lü, Qi Shen, Jian-Hui Jiang, Guo-Li Shen, Ru-Qin Yu*

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

Received 2 January 2004; received in revised form 16 February 2004; accepted 20 February 2004

Abstract

Quantitative structure–activity relationship (QSAR) models of inhibiting action of some diarylimidazole derivatives on cyclooxygenase (COX) enzyme were constructed using modified particle swarm optimization (PSO) method. As a comparison to this method, the genetic algorithm (GA) was also tested. It has been demonstrated that the modified PSO is a useful tool for variable selection comparable to GA and even superior to GA. QSAR models are constructed separately for COX-2 inhibitory activity and selectivity of COX-2 inhibition over COX-1. The spatial descriptors play a key role in the compounds' activity and selectivity to COX-2, especially Jurs descriptors. Polar interactions are the principal binding strength between compounds and COX-2 enzyme. In addition, the aqueous desolvation free energy (F_{H_2O}) value of substituent will affect the COX-2 inhibitory activity, while the charge distribution can affect the selectivity to COX-2.

Keywords: Quantitative structure–activity relationship; Cyclooxygenase inhibitors; Particle swarm optimization; Variable selection

1. Introduction

Cyclooxygenase (COX) performs the first step in the creation of prostaglandins from a common fatty acid, arachidonic acid [1,2]. It adds two oxygen molecules to arachidonic acid, initiating a set of reactions that will ultimately create a host of unusual molecules. Two distinct isoforms have been identified separately termed COX-1 and COX-2 [3]. COX-1 is a constitutive enzyme in most mammalian tissues and is responsible for keeping the stomach lining intact and maintaining functional kidneys. However, COX-2 is expressed only after an inflammatory stimulus, releasing metabolites that are used to induce inflammation and pain [4–6]. During normal physiology

Abbreviations: COX, cyclooxygenase; F_{H_2O}, aqueous desolvation free energy; GA, genetic algorithm; GA–MLR, QSAR modeling by MLR method using descriptors selected by GA; Jurs-FNSA-1, fractional charged partial surface areas; Jurs-FPSA-1, fractional charged partial surface areas; Jurs-PNSA-2, total charge weighted negative surface area; Jurs-RNCG, relative negative charge; Jurs-RPCG, relative positive charge; MLR, multiple linear regression; PMI-mag, principal moments of inertia about the principal axes of a molecule; PSO, particle swarm optimization; PSO–MLR, QSAR modeling by MLR method using descriptors selected by PSO; QSAR, quantitative structure–activity relationship; Shadow-*X*length, length of molecule in the *X* dimension; Shadow-*XZ*frac, fraction of area of molecular shadow in the *XZ* plane over area of enclosing rectangle

* Corresponding author. Tel.: +86-731-882-2782; fax: +86-731-882-2782.
E-mail address: rqyu@hnu.net.cn (R.-Q. Yu).

0731-7085/$ – see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.jpba.2004.02.026

COX-2 levels are undetectable in most tissues. During periods of acute and chronic inflammation, the level of COX-2 is very often significantly higher.

Non-steriodal anti-inflammatory drugs, for example, aspirin, can block the binding of arachidonic acid in the COX active site. The normal messages are not delivered, so people do not feel the pain and not launch an inflammatory response. Hence, this kind of drugs, as COX inhibitors, is widely utilized agents for the treatment of inflammation, pain, and fever. However, the chronic usage of these drugs has been associated to the induction of gastrointestinal mucosal lesions, perforations, and bleeding in part of the population [7]. Decreased renal functions have been observed in some patients [8]. The reason is that these COX inhibitors also inhibit COX-1 to produce the necessary prostaglandins. The association of COX-2 with inflammation led to the designing of the selective COX-2 inhibitors which block the prostaglandin's production in inflammatory cells while not interfering with the homeostatic (COX-1) production of prostaglandins in the gastrointestinal tract [9–12]. One could not, however, confirm that the compounds designed would always possess good selectivity to COX-2, while the synthesis and testing of these compounds on human COX-1 and -2 enzymes 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) models have been built using the experimental data accumulated [13]. 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.

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 is taken as dependent variable in building a model. The parameters to be calculated are numerous descriptors that are indicative of molecular structures. Since hundreds of molecular descriptors are available for QSAR analysis and only a part of them is statistically significant in terms of correlation with biological activity for a particular analysis, variable selection is necessary for producing a useful predictive model. There have been many variable selection methods exist, the mostly used ones are stepwise regression, simulated annealing, evolutionary algorithms and genetic algorithms (GAs) and so on. Here, particle swarm optimization (PSO) algorithms introduced to chemometrics by present authors were used to perform the variable selection [14]. Particle swarm optimization is an evolutionary computation technique developed by Kennedy and Eberhart [15], by simulating social behavior of bird flocking or fish schooling. Similar to GA, PSO is a population based optimization tool. The system is initialized with a population of random solutions, and searches for optima by updating generations. Unlike GA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, are "flown" through the problem space by following the current optimum particles. Compared to GA, the advantages of PSO are that PSO is easy to implement and there are few parameters to adjust.

Multiple linear regression (MLR) techniques are used for building QSAR models rather than other methods such as factor analysis or principle components regression, for the ease of MLR implementation and the interpretability of the resulting equations. In the present work, we employed modified PSO algorithm for variable selection in MLR analysis and compared it to GA. It has been demonstrated that the modified PSO is a useful tool for variable selection comparable to GA and even superior to GA. QSAR models are constructed separately for COX-2 inhibitory activity and COX-2 selectivity. As far as we are aware, this is the first QSAR study using a hybrid method to predict the compounds' selectivity to COX-2.

2. Algorithms and data sets

2.1. Particle swarm optimization

Particle swarm optimization (PSO) involves simulating social behavior among individuals (particles) "flying" through a multi-dimensional search space, each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has achieved so far. The first step of the algorithm is to randomly initialize the position and velocity of each particle in the swarm, dispersing

them uniformly across the search space. The ith particle is represented as $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})$. Velocity, the rate of the position change for particle i is represented as $v_i = (v_{i1}, v_{i2}, \ldots, v_{id})$. PSO posits that particles should move toward some combination of their personal best position and the global best position. The personal best position $p_i = (p_{i1}, p_{i2}, \ldots, p_{id})$ is the best previous position of the ith particle that gives the best fitness value. The global best position $p = (p_{g1}, p_{g2}, \ldots, p_{gd})$ is the best particle among all the particles in the population. Therefore, all of the data required to update the particle positions for each iteration can stored in four $M \times N$ matrices, where M is the number of particles in the simulation and N is the number of dimensions of the problem. The position and velocity of each particle is updated at discrete intervals according to the following equation:

$$v_{id} = v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \quad (1)$$

where c_1 and c_2 are two positive constants named as learning factors, r_1 and r_2 are uniform random variables in the range of (0, 1). The three terms of this expression are the initial velocity of the particle at the beginning of the iteration, and the distance of the particle from the personal and global best positions. Hence, the particle velocities are likely to be large in the early stages of the optimization and will become substantially smaller as the swarm converges to the optimum. Finally, for each dimension, the particles move in the direction specified by the velocity matrix according to a simple relationship given by

$$x_{id} = x_{id} + v_{id} \quad (2)$$

Such an adjustment of the particle's movement through the space causes it to search around the two best positions. If the minimum error criterion is attained or the number of cycles reaches a user-defined limit, the algorithm is terminated.

In past several years, PSO has been successfully applied in many research and application areas, such as analysis of human tremor [16], controlling reactive power and voltage [17], distribution state estimation [18] and so on. It is demonstrated that PSO gets better results in a faster, cheaper way compared with other methods.

2.2. Modified particle swarm optimization

In PSO, only best positions give out the information to others; while in GAs, chromosomes share information with each other. As information sharing mechanism in PSO is significantly different from GAs, all the particles tend to converge to the best solution quickly. But the disadvantage of PSO is easily getting into local optima. According to information sharing mechanism of PSO, a modified PSO for variable selection is proposed as follows. The velocity v_{id} of each individual is a random number in the range of (0,1). The resulting change in position then is defined by the following rule:

$$\text{If } 0 < v_{id} \leq a, \quad \text{then } x_{id\,(\text{new})} = x_{id\,(\text{old})} \quad (3)$$

$$\text{If } a < v_{id} \leq \tfrac{1}{2}(1+a), \quad \text{then } x_{id} = p_{id} \quad (4)$$

$$\text{If } \tfrac{1}{2}(1+a) < v_{id} \leq 1, \quad \text{then } x_{id} = p_{gd} \quad (5)$$

where a is a random value in the range of (0,1) named static probability. The initial value of a is 0.5.

Though the velocity in the modified PSO is different from that in continuous version of PSO, information sharing mechanism and updating model of particle by following the two best positions is the same in two PSO versions. Some elements in x_i are kept fixed according to Eq. (3) which is similar to the first term of the right side of Eq. (1). Without this part, the "flying" particle are only determined by their best positions in history, and all particles would tend to move toward the same position resembling a local search. In this sense, Eq. (3) really provides the particles a tendency to expand the search space. Eqs. (4) and (5) are similar to the second and third terms of the right side of Eq. (1). Without these two parts, the particles would keep on "flying" randomly and PSO would not be able to find a meaningful solution. Even so, preliminary experiments indicate that such a PSO version still tends to converge to local optima. To circumvent this drawback and improve the ability of the modified PSO algorithm to overleap local optima, ten percent of particles are forced to fly randomly not following the two best particles.

Using decreasing static probability and defined percent of randomly flying particle to overleap local optima, the modified PSO remains having satisfactory converging characteristics.

2.3. Fitness function

In the modified PSO, the performance of each particle is measured according to a pre-defined fitness function. The modified C_p statistic as an objective function is applied to variable selection in the modified PSO. The modified C_p in MLR is expressed as follows

$$C_{p_1}(p) = \frac{\mathrm{RSS}_p}{\hat{\sigma}^2_{\mathrm{PLS}}} - (n - 2p) \quad (6)$$

Here, n is the number of dependent variables, and p is the number of independent variables. RSS_p is the residual sum of the squares of p-variable model, $\hat{\sigma}^2_{\mathrm{PLS}}$is defined as the value of RSS corresponding to the minimum number of principal components when further increase of the number of principal components does not cause a significant reduction in RSS. The details of modified C_p have been described elsewhere [19].

2.4. Data sets

Forty-two 1,2-diarylimidazole derivatives with the corresponding inhibitory activities were used as a data set for variable selection and QSAR analysis. The chemical structures and inhibitory activity of them are shown in Table 1. The activity is expressed as IC_{50}, the molar concentration of the compound causing 50% inhibition of enzyme. We randomly divided the data taken from the study by Khanna et al. [13] 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 consists 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 ($r \geq 0.950$) encoded similar information, and one of them should be removed. Descriptors that contain a high percentage ($\geq$90%) of identical values are also discarded. Thus, only 85 of total descriptors were remained which were listed in Table 2.

All these molecular descriptors were generated using Cerius^{2}3.5 software on Silicon Graphics R3000 workstation. The modified PSO, GAs and MLR algorithms were written in Matlab 5.3 and run on a personal computer.

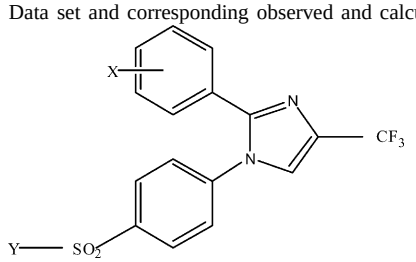
3. Results and discussions

3.1. Variable selection and QSAR analysis on COX-2 inhibitory activity

For the selection of the most important descriptors the modified PSO was used, which contained a population of 100 individuals and evolved for 100 generations. Then the selected descriptors were taken as independent variables to build QSAR models with MLR method (PSO–MLR). The names of selected descriptors by modified PSO and statistical parameters obtained by these models for the training and prediction set are shown in Table 3. The best model with the lowest C_p (fitness) value includes four descriptors and the second best one includes three descriptors. Comparing to the modified PSO, GAs were used for variable selection and the number of selected descriptors was four and three, respectively. The selected descriptors were also used for constructing QSAR models (GA–MLR) and we obtained the statistical results (Table 3). It can be seen from this table that although the number of descriptors in these models are identical but they differ from each other. The correlation coefficient of training set obtained by PSO–MLR is lower than that obtained by GA–MLR, but the correlation coefficient of prediction set is increased. Also, these results reveal that in this QSAR study PSO–MLR is superior to GA–MLR in predictive ability. The calculated activity values using PSO–MLR method are listed in Table 1. In the case of PSO–MLR, the relative standard deviation between the calculated and experimental values for training and predicting set are 26.44 and 32.44%, respectively. Fig. 1 shows the relationship between the experimental and calculated values using PSO–MLR.

The descriptors selected by modified PSO are Jurs-PNSA-2, Jurs-RPCG, shadow-*X*length and $F_{\mathrm{H_2O}}$. These descriptors encode different aspects of the molecular structure. Jurs-PNSA-2 is the total charge weighted negative surface area: partial negative solvent-accessible surface area multiplied by the total negative charge. Jurs-RPCG is the relative positive charge: charge of the most positive atom divided by

Table 1
Data set and corresponding observed and calculated properties of compounds



Compound	X	Y	ln(1/IC50) (COX-2)		Selectivity[a] (COX-1/COX-2)	
			Experimental	Calculated	Experimental	Calculated
1	4Cl	Me	2.2073	2.7205	5.3471	6.2112
2	4F	Me	2.3026	2.1018	5.8861	6.7927
3	H	Me	2.1203	1.2760	8.7037	7.5328
4[b]	4Me	Me	1.8326	0.8649	5.0752	5.3631
5	4Ome	Me	0.5621	0.8694	1.6094	3.8499
6	4NHMe	Me	−0.3853	0.4273	3.5835	2.9795
7	$4NMe_2$	Me	0.3567	0.0676	1.9459	1.3844
8[b]	4Sme	Me	1.8326	1.7820	2.5649	3.4144
9	4Cl	NH_2	4.6052	4.6400	5.0752	4.7916
10	4F	NH_2	4.6052	3.8932	5.2470	5.5540
11	H	NH_2	3.2189	3.0949	6.1779	5.6658
12	4Me	NH_2	3.2189	2.5339	4.7449	3.5683
13	3Cl	Me	2.8134	2.9992	8.6995	7.7633
14[b]	3Me	Me	2.8134	0.8434	7.3132	7.1684
15	$3NMe_2$	Me	−1.1632	0.0461	2.5649	3.9105
16	3Cl	NH_2	4.8283	4.6616	6.6529	7.0941
17	3F	NH_2	3.5066	3.8136	7.7187	7.0709
18[b]	3Br	NH_2	4.9618	4.8319	6.4615	7.0596
19[b]	3Me	NH_2	3.5066	2.6348	4.6634	4.7095
20	2Me	Me	0.2231	0.8574	6.5793	6.7143
21	2F	NH_2	2.3026	3.6323	5.4072	5.7699
22	2Me	NH_2	1.6094	2.6065	3.7136	5.2308
23	3F-4Ome	Me	1.8971	1.6641	5.7900	5.6203
24	3Cl-4OMe	Me	2.0402	1.7800	7.7319	5.0407
25[b]	$3Cl\text{-}4NMe_2$	Me	1.1394	1.2258	1.6094	3.9972
26	$3F\text{-}4Nme_2$	Me	1.1087	0.4564	3.9512	4.2758
27	3Cl-4Me	Me	3.5066	2.5603	5.9915	6.4278
28	3Me-4F	Me	1.7720	1.8522	4.9558	6.0488
29	3Me-4Cl	Me	2.4079	2.4817	4.4659	5.3624
30[b]	$3,4\text{-}OCH_2O–$	Me	1.7720	1.4121	4.2627	4.3785
31	$3,4\text{-}Me_2$	Me	1.1087	0.6218	4.5109	3.8987
32	3F-4Ome	NH_2	3.5066	3.1142	4.8442	5.5771
33	3Cl-4Ome	NH_2	3.9120	3.6392	5.6699	5.5685
34	3Br-4Ome	NH_2	3.5066	4.0746	4.5109	5.2185
35[b]	3Cl-4Sme	NH_2	4.6052	4.3285	3.5553	4.1895
36	3Cl-4Me	NH_2	5.8091	4.2298	5.2470	5.3642
37	3Ome-4Cl	NH_2	3.9120	4.0156	6.0638	5.8041
38	$3,4\text{-}F_2$	NH_2	3.5066	3.9583	6.9007	6.8380
39	3Me-5F	NH_2	3.5066	2.8032	7.9121	6.3701
40	$3,5\text{-}Me_2\text{-}4Ome$	Me	0.3285	−0.0926	4.8363	3.8500
41	$3,5\text{-}F_2\text{-}4Ome$	NH_2	3.5066	4.3655	7.0622	7.1132
42	4Cl	Me	2.2073	2.7097	5.3428	5.1819

[a] Selectivity is expressed as natural logarithm of IC_{50}(COX-1)/IC_{50}(COX-2).
[b] The compounds used for prediction.

Table 2
The descriptors used for selection in this study

Functional families of descriptors	Descriptors
Conformational descriptors	Energy
Electronic descriptors	Apo1 (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	Rad of gyration (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	$A\log P$ (log of the partition coefficient), F_{H_2O} (desolvation free energy for water), F_{oct} (desolvation free energy for octanol), mol ref (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

the total positive charge. It can be seen from Table 3 that Jurs descriptors play the most important roles in COX-2 inhibitory activity. Jurs descriptors (charged partial surface area descriptors) encode features responsible for polar interactions between molecules [20]. This set of descriptors, combining 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. Shadow-Xlength belongs to shadow indices descriptors. This set of descriptors helps to characterize the shape of the molecules [21]. They are calculated by projecting the molecular surface on three mutually perpendicular planes, *XY*, *YZ*, and *XZ*. They depend not only on conformation but also on the orientation of the molecule. Among them shadow-*X*length is the length of molecule in the *X* dimension. Its negative coefficient accounts for why the length of *X* substituents at the second position of phenyl ring is detrimental to activity. The above three descriptors all belong to structural descriptors. F_{H_2O}, the aqueous desolvation free energy, is physiochemical properties associated with linear free energy models

Table 3
Selected descriptors used for MLR analysis and the statistical results in COX-2 inhibitory activity QSAR study

	Descriptors	Coefficients[a]	R_t[b]	R_p[b]
PSO–MLR	Jurs-PNSA-2	−0.0064	0.9209	0.8815
	Jurs-RPCG	67.8020		
	Shadow-*X*length	−1.7280		
	F_{H_2O}	−0.0055		
GA–MLR	Jurs-DPSA-3	0.1855	0.9569	0.8395
	Jurs-FNSA-1	−28.4062		
	Jurs-RPCG	120.7818		
	S_sNH$_2$	0.3924		
PSO–MLR	Jurs-FPSA-3	−89.2982	0.9108	0.9000
	Jurs-RASA	−10.7658		
	S_sNH$_2$	0.2822		
GA–MLR	Jurs-FPSA-3	251.8937	0.9380	0.8505
	Jurs-WNSA-3	−0.2200		
	Jurs-RPCG	127.7156		

[a] The corresponding coefficient of descriptor in the regression equation.

[b] *R*, correlation coefficient, t is for training set, p is for prediction set.

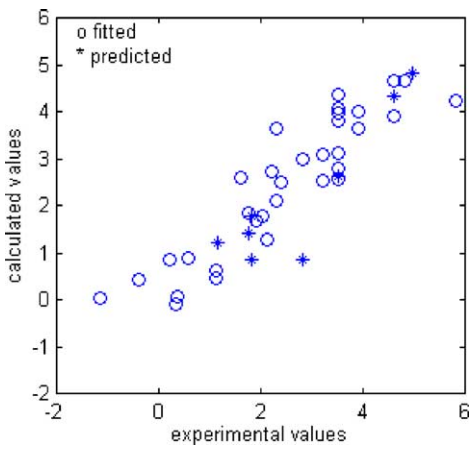


Fig. 1. Calculated vs. experimental $\ln 1/IC_{50}$ values using PSO–MLR method.

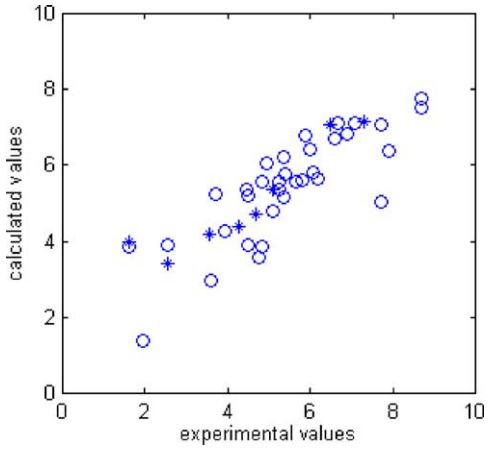


Fig. 2. Calculated vs. experimental selectivity values using PSO–MLR method.

of a molecule. QSAR calculates F_{H_2O} for each molecule by searching the molecule for recognizable substituent groups and their bonding patterns, and summing the substituent constants contributions for each group that is present in the molecule. The negative coefficient shows that the smaller the value of F_{H_2O}, the higher the activity. It is apparent that the substituent of amido is benefit while methyl is detrimental to activity. This is because the value of F_{H_2O} of amido is negative while that of methyl is positive.

3.2. Variable selection and QSAR analysis on the selectivity to COX-2

The relationship between the COX-2 inhibitory activity and the molecule structures (properties) has been studied. But only considering this point is not sufficient because at the same time the COX-1 inhibitory activity is perhaps also enhanced, which will lead to a series of unpleasant effects. Hence, the relationship between selectivity and the molecular structures (properties) should be built so as to design more reasonable drugs. The variables were selected separately using PSO–MLR and GA–MLR. With the selected variables, QSAR models have been constructed and some results are shown in Table 4. The calculated values of the best model are listed in Table 1. The relative standard deviation of training set and predictive set is 17.67 and 23.01%, respectively. Fig. 2 shows

Table 4
Selected descriptors used for MLR analysis and the statistical results in COX-2 selectivity QSAR study

	Descriptors	Coefficient[a]	R_t[b]	R_p[b]
PSO–MLR	Jurs-FPSA-1	102.0326	0.8231	0.9263
	Jurs-FNSA-1	−24.7637		
	Jurs-RPCG	221.8227		
	Jurs-RNCG	118.1285		
	Shadow-*XZ*frac	−29.0363		
	PMI-mag	−0.0041		
PSO–MLR	Dipole-Y	0.2097	0.8297	0.8287
	Jurs-PNSA-1	−0.0567		
	Jurs-FNSA-2	−11.4841		
	Jurs-RPCG	64.7096		
	Jurs-RNCG	148.3189		
	Shadow-*XZ*frac	−23.8165		
GA–MLR	Radius of gyration	11.8726	0.8587	0.8040
	Jurs-PNSA-2	−0.0499		
	Jurs-WNSA-1	−0.3468		
	Jurs-RPCG	328.2977		
	Shadow-*XZ*frac	−33.9385		
	Shadow-nu	−0.8533		

[a] The corresponding coefficient of descriptor in the regression equation.

[b] *R*, correlation coefficient, t is for training set, p is for prediction set.

the relationship of the calculated and experimental values.

The selected variables of the best model are Jurs-FPSA-1, Jurs-FNSA-1, Jurs-RPCG, Jurs-RNCG, shadow-*XZ*frac, and PMI-mag. These descriptors all belong to spatial descriptors, so the spatial descriptors are the most important factors in the selectivity of inhibition. Jurs-FPSA-1 and Jurs-FNSA-1 are fractional charged partial surface areas. Jurs-FPSA-1 is obtained by dividing sum of the solvent-accessible surface areas of all positively-charged atoms by the total molecular solvent-accessible surface area. Jurs-FNSA-1 is obtained by dividing total charge weighted negative surface area by the total molecular solvent-accessible surface area. Total charge weighted negative surface area is partial negative solvent-accessible surface area multiplied by the total negative charge. Jurs-RNCG is relative negative charge: charge of most negative atom divided by the total negative charge. In terms of their coefficients in Table 4, a conclusion can be deduced that the selectivity increased with the increased relative charge. The charge distribution within the

molecules acts as the driving force for intermolecular interactions and the more the relative charge the larger the interactions. Shadow-*XZ*frac is fraction of area of molecular shadow in the *XZ* plane over area of enclosing rectangle. Its coefficient being negative shows that the selectivity increased with the decreased value of shadow-*XZ*frac. That is to say the smaller area of molecular shadow in the enclosing rectangle will benefit the selectivity. PMI-mag is to calculate the principal moments of inertia about the principal axes of a molecule. Through other several models we found the spatial descriptors indeed play the key roles in the compounds' activity and selectivity to COX-2, especially Jurs descriptors. All models include several Jurs descriptors, which indicate that polar interactions between molecules are the principal interactions. The strength of these interactions was considered to be a function of contacting area between molecules. Also, the charge distribution within the molecules acts as the driving force for these interactions.

4. Conclusions

The modified PSO algorithm has been employed in variable selection and sastifactory results have been obtained. In the selected descriptors, Jurs descriptors are the most important descriptors both in predicting COX-2 inhibitory activity and in COX-2 selectivity. Therefore, one can infer that polar interactions are the principal binding strength between compounds and COX-2 enzyme. In addition, the F_{H_2O} value of substituent will affect the COX-2 inhibitory activity, while the charge distribution can affect the selectivity to COX-2.

QSAR models were constructed with selected variables to predict new compounds' activity. The results suggest that a small number of chemically meaningful descriptors will provide the most predictive QSAR model. The modified PSO has been testified to be an effective method for variable selection comparable to GA.

Acknowledgements

The work was supported by grants from the Natural Science Foundations of China (grants no. 20105007, 20205005 and 20375012).

References

[1] R.M. Smith, D.L. DeWitt, R.M. Garavito, Annu. Rev. Biochem. 69 (2000) 145–182.
[2] H. Herschman, Adv. Exp. Med. Biol. 469 (1999) 3–8.
[3] (a) W. Xie, J.G. Chipman, D.L. Robertson, R.L. Erikson, D.L. Simmons, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 2692;
(b) D. Kujubu, B.S. Fletcher, B.C. Varnum, R.W. Lim, H.R. Herschman, J. Biol. Chem. 266 (1991) 12866;
(c) M.K. O'Banion, H.B. Sadowski, V. Winn, D.A. Young, J. Biol. Chem. 266 (1991) 23261.
[4] K. Seibert, Y. Zhang, K. Leahy, S. Hauser, J. Masferrer, W. Perkins, L. Lee, P. Isakson, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 12013–12017.
[5] G.P. O'Neill, A.W. Ford-Hutchinson, FEBS Lett. 330 (1993) 156–160.
[6] S. Kargman, S. Charleson, M. Cartwright, J. Frank, D. Riendeau, J. Mancini, J. Evans, G. O'Neill, Gastroenterology 111 (1996) 445–454.
[7] (a) M.C. Allison, A.G. Howatson, C.J. Torrence, F.D. Lee, R.I. Russel, N. Engl. J. Med. 327 (1992) 749–754;
(b) S.J. Sontag, Drugs 32 (1986) 445–457;
(c) R.T. Schoen, R.J. Vender, Am. J. Med. 86 (1989) 449–458.
[8] (a) C. Patrono, M.J. Dunn, Kidney Int. 32 (1987) 1–12;
(b) D.M. Clive, J.S. Stoff, N. Engl. J. Med. 310 (1984) 563–572.
[9] (a) J.L. Masferrer, B.S. Zweifel, P.T. Manning, S.D. Hauser, K.M. Leahey, W.G. Smith, P.C. Isakson, K. Seibert, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 3228–3232;
(b) K. Seibert, J.L. Masferrer, Receptor 4 (1994) 17–23.
[10] (a) E.A. Meade, W.L. Smith, D.L. DeWitt, J. Biol. Chem. 268 (1993) 6610–6614;
(b) J.A. Mitchell, P. Akarasereemont, C. Thiemerman, R.J. Flower, J.R. Vane, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 11693–11697;
(c) J.R. Vane, J.A. Mitchell, I. Appleton, A. Tomlinson, D. Bishop-Bailey, J. Croxtail, D.A. Willoughby, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 2046–2050.
[11] Y. Harada, K. Hatanaka, M. Saito, M. Majima, M. Ogina, M. Kawamura, T. Ohno, Q. Yang, M. Katori, S. Yamamoto, Biomed. Res. 15 (1994) 127–130.
[12] (a) P. Emery, Scand. J. Rheumatol. 25 (Suppl. 102) (1996) 23–28.;
(b) A. Carabaza, F. Cabre, E. Rotllan, M. Gomez, M. Gutierrez, L. Garcia, D. Mauleon, J. Clin. Pharmacol. 36 (1996) 505–512;
(c) J.K. Gierse, J.J. McDonald, S.D. Hauser, S.H. Rangwala, C.M. Koboldt, K.A. Seibert, J. Biol. Chem. 271 (1996) 15810–15814.
[13] I.K. Khanna, R.M. Weier, Y. Yu, X.D. Xu, F.J. Koszyk, P.W. Collins, C.M. Koboldt, A.W. Veenhuizen, W.E. Perkins, J.J. Casler, J.L. Masferrer, Y.Y. Zhang, S.A. Gregory, K. Seibert, P.C. Isakson, J. Med. Chem. 40 (1997) 1634–1647.
[14] Q. Shen, J.H. Jiang, G.L. Shen, R.Q. Yu, Eur. J. Pharm. Sci., in press.

[15] J. Kennedy, R. Eberhart, Particle Swarm Optimization, in: Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia, 1995, pp. 1942–1948.
[16] R.C. Eberhart, X. Hu, in: Proceedings of the Congress on Evolutionary Computation, Washington DC, USA, 1999, pp. 1927–1930.
[17] H. Yoshida, K. Kawata, Y. Fukuyama, S. Takayama, Y. Nakanishi, IEEE Trans.Power Syst. 15 (2000) 1232–1239.
[18] S. Naka, T. Genji, T. Yura, Y. Fukuyama, IEEE Trans.Power Syst. 18 (2003) 60–68.
[19] Q. Shen, J.H. Jiang, G.L. Shen, R.Q. Yu, Anal. Bioanal. Chem. 375 (2003) 248–254.
[20] D.T. Stanton, P.C. Jurs, Anal. Chem. 62 (1990) 2323–2329.
[21] R.H. Roxburgh, P.C. Jurs, Anal. Chim. Acta 987 (1991) 99–109.