Indian Journal of Chemical Technology
Vol. 23, March 2016, pp. 121-130

Prediction of AHAS inhibition by sulfonylurea herbicides using genetic algorithm and artificial neural network

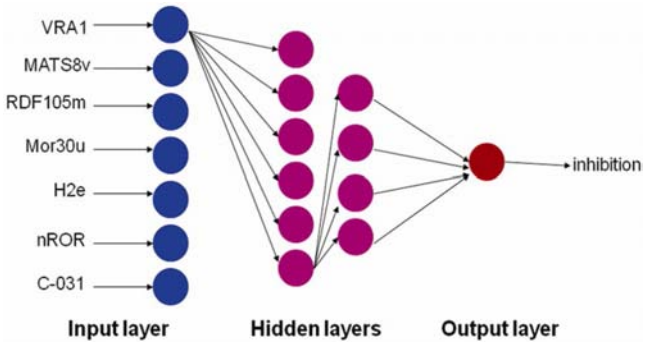
A Saidi* & M Mirzaei

Department of Chemistry, Shahid Bahonar University of Kerman, 22 Bahman Boulevard, Kerman 76169, Iran

E-mail: asiye_saidi@yahoo.com.

Received 6 February 2014; accepted 13 March 2014

Acetohydroxyacid synthase (AHAS; EC 2.2.1.6) catalyzes the first common step in branched-chain amino acid biosynthesis. The enzyme is inhibited by several chemical classes of compounds and this inhibition is the basis of action of the sulfonylurea herbicides. The negative logarithm inhibition constant (*p*Ki) of 68 sulfonylurea analogs as inhibitors of pure AHAS using quantitative structure–activity relationship (QSAR) has been calculated. Suitable set of molecular descriptors are calculated and the important descriptors are selected by genetic algorithm and stepwise multiple regression methods. These variables serve as inputs to generated neural networks. After optimization and training of the networks, they are used for the calculation of *p*Ki for the prediction set. Comparison between the results obtained, show the superiority of genetic algorithm over stepwise multiple regression method in feature-selection. For network that used the genetic algorithm for feature-selection methods there are very good agreements between calculated and experimental *p*Ki for data set. The correlation coefficient between calculated and experimental values of *p*Ki for training and prediction set are 0.988 and 0.954, respectively.



Keywords: QSAR, Genetic algorithm, Artificial neural network, Acetohydroxyacid synthase, Sulfonylurea

Since their discovery in the 1970s, sulfonylureas have emerged as a group of herbicides with many innovative natures. Their level of activity is unprecedented and may be up to 100 times that of conventional herbicides. Compared with other herbicides, sulfonylureas have much lower use range and are more rapidly degraded in soil[1,2]. They have the attributes of low application rates, environmental safety, good crop selectivity and low mammalian toxicity. Following an extensive synthetic program led by Levitt and colleagues[3] the first sulfonylurea herbicide chlorsulfuron was developed. Since that time, a large number of other sulfonylurea herbicides have been identified and are now applied widely[4]. The general features of most active compounds are an *ortho*-substituted aromatic ring attached to the sulfur atom, and a heterocyclic ring substituted in both *meta* positions and attached to the distal nitrogen atom of the sulfonylurea bridge. This heterocyclic ring is either a pyrimidine (X = CH) or triazine (X = N). The mode of action of sulfonylureas started to become clear when it was discovered that sulfometuron methyl is a potent inhibitor of bacterial acetohydroxyacid synthase[5] (AHAS; EC 2.2.1.6), the enzyme that catalyzes the first common step in branched-chain amino acid biosynthesis. Contemporary, Ray showed that chlorsulfuron inhibits plant AHAS[6]. Since then, other sulfonylurea herbicides have been shown to inhibit AHAS, and it is widely accepted that inhibition of this enzyme is the mode of action of sulfonylureas as well as several other families of herbicides[7]. An important property

of the use of such inhibitors is that there is no AHAS counterpart in humans and other animals.

Despite of the large number of papers on AHAS published in recent years on the interaction between herbicides and this enzyme there are some aspects of the inhibition that remain puzzling[8]. Thus, it is important to determine which structural features of the herbicides are responsible for the enzyme inhibition. This is essential for the design of new herbicides since its properties may be predicted prior to synthesis and consequently the design may, in this way, be guided by the results of calculations. Quantitative structure activity relationship (QSAR) modeling has shown to be very effective for this purpose. This approach provides information that is useful for molecular design and medicinal chemistry[9]. The QSAR models are mathematical equations which relate chemical structure of compounds to a wide variety of their physical, chemical, biological and technological properties. If we could elucidate in detail how these properties are determined by structure we can predict such properties simply from the molecular structure. The main task of QSAR is to obtain a reliable statistical model for the prediction of activities or properties of new chemical substances and analytical systems. Nowadays, QSAR models are rapidly developing and have been widely used by chemists for predicting different chemical and physical properties of different types of molecules. In the case of herbicides there are lots of QSAR studies[10,11]. Surprisingly, considering the similarity of the compounds applied to develop the models, they involve different number and types of descriptors, complicating the physical interpretation. On the other hand, the models are based on empirically derived descriptors which limit their application to new or developing chemicals. Duggleby *et al.* performed comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSiA) analyses of a new family of sulfonylurea herbicides[10]. Roy and Paul performed docking and 3D QSAR analysis for 45 sulfonylurea derivatives[11]. They used homogeneous compounds, relatively. Whiles, if data set are more diverse for example another sulfonylurea derivatives such as pyrazole sulfonylureas, pyridine sulfonylureas, thiazole sulfonylureas and etc, are applied, final model is more general and valid. Also they considered linear relationships but the commonly used multiple linear regression (MLR) will fail to develop an appropriate QSAR model when the nonlinear phenomenon is significant to some extent within the data investigated; therefore nonlinear modelling techniques such as artificial neural networks (ANN) were necessary to be introducing for building an accurate and reliable QSAR model. ANN has recently gained much popularity for calibrating the nonlinear relationships[12]. The major advantage of ANN lies in the inherent ability to calibrate the nonlinear relationships. Therefore, ANN has become an important modelling tool for building QSAR models[13].

In the present paper, a QSAR model for the prediction of inhibition constant of 68 diverse sulfonylurea herbicides consist of mono-substituted, bridge modification, pyrazole, pyridine, thiazole and di-substituted sulfonylureas, using artificial neural network and genetic algorithm has been presented. To the best of our knowledge, this is the first QSAR study using a hybrid method to the prediction of inhibition constant of sulfonylurea herbicides.

Experimental Section

Data set

The structures of a diverse set of 68 sulfonylurea herbicides as well as their negative logarithm inhibition constant (*p*Ki) reported in the literature [10]. The value of *p*Ki ranged between 2.64 and 8.08 for sulfonylurea and Chlorimuron ethyl herbicides, respectively. The molecules in data set randomly divided into three sets; training, test and prediction set which each of them consisting of 46, 12, 10 member, respectively. The structures of these herbicides and their experimental *p*Ki are shown in Table 1.

Structural descriptors

To obtain a QSTR model, compounds are represented by theoretical molecular descriptors. In order to compute the structural descriptors, the structures of all herbicides were drawn with HyperChem 4.0 program[14] and were optimized using the semiepirical quantum method AM1 of the HyperChem program. After geometry optimization, Hyperchem output files were used by the Dragon program as input to calculate molecular descriptors[15]. In order to reduce redundant and non-useful information, prescreening of descriptors were carried out in the following way; (1) constant or near constant descriptors were eliminated, and (2) among those descriptors whose inter-correlations exceeded 0.9 the most suitable and interpretable ones were kept while the others were deleted. The remaining

534 descriptors were used to generate the QSAR models. These parameters encoded different aspects of the molecular structure. Since the number of descriptors considered is large, a suitable feature selection method should be combined with a proper feature mapping technique. In the present work we have considered stepwise multiple regression and genetic algorithm as feature-selection tools and ANN was employed for feature mapping.

Genetic algorithm

Variable selection is always one of the most important steps in developing a QSTR model, which is especially important when one is required to deal with a large variable set. Genetic algorithm (GA) is a stochastic optimization method that has been inspired by evolutionary principles[16]. The different aspect of a GA is that it investigates many possible solution simultaneously, each of which explores distinctive regions in parameter space[17]. For the moment, one of the best available tutorial on variable selection using GA published by Leardi and Gonzalez[18]. In the present paper, GA optimization method was tried following the studies of Rogers and Hopfinger[19] and Luke[20] with a few minor modifications. In this GA an individual of the population is represented by the string of bits that encoding the selected feature. The first step in a GA is to create a gene pool of *n* individuals. Each individual (chromosome) contain some descriptors that in the first generation are chosen randomly from a common list and in a way such that no two individuals can be fined that contain exactly the same set of descriptors. The fitness of each individual in this generation is appointed by a user specified fitness function. In the next step reproduction take place, which individuals are selected probabilistically on the basis of their fitness scores and serve as parents. The selection strategy that applied in this program was random selection method. Next step is a crossover that each of parents contribute a random selection of half of its descriptors and an offspring is constructed by combining these two halves of genetic code. Therefore, the generated offspring contains characteristics from both of its parents. Finally, this offspring is subjected to a random mutation in one of its gene, i.e. one descriptor is replaced by another. This selection crossover mutation process is repeated until all of the *n* parents in the gene pool are replaced by their offspring. The fitness score of each member of this new generation is again evaluated, and the reproductive cycle is continued until a desired number of generations or target fitness score is reached. In our GA program that was written by MATLAB7 one modification is made. This is the inclusion of elitism, which protect the fittest individual in any given generation from crossover or mutation during reproduction. The genetic content of this individual simply moves on to the next generation intact. In original studies, the fitness function of the individual was determined by a function related to the residual error in the regression analysis of the training data. Here we try to use varieties of fitness functions, which are proportional to the residual error of the prediction set and the number of selected variables according to the following equation:

$$\text{Fitness} = (1 - w)\,(1/\text{MSEt}) + w\,(1/m) \qquad ...(1)$$

In this equation MSEt is mean square error of test set [21], m is the number of variables in the represented model and w is a parameter between 0 and 1 that implies the weights of m in the value of fitness. In fact the value of w determine the number of variables exist in the chromosome.

Some experiments were applied using different value for w. Results obtained showed that for small value of w the number of variables in the fittest individual was high and on the other hand if the value of w was to be high the number of variables in the best chromosome was small. Hence after some experiments the value of w was set to be 0.7. It is worth noting that the parameter of w was determined in a primary study, before the overall genetic algorithm optimization has been carried out. Here for the calculation of the fitness of each chromosome a non linear model was constructed using variables consists in each chromosome separately ANN and the value of MSEt was calculated using this model. This procedure was applied for each chromosome separately.

Artificial neural network

Artificial neural network (ANN) is generally applied as a technology offering an alternative way to simulate ambiguous and complex problems. The importance of using neural networks in process modelling is that they have learning and generalization abilities as well as nonlinearity. Numerous applications of ANN have been known in pattern recognition, materials modelling, data analysis and property prediction[22-25]. A neural network is a computational structure, consisting of a number of

highly interconnected processing units called neurons. The neurons are connected to each other by weighted links over which signals can pass. Each neuron receives multiple inputs from other neurons in proportion to their connection weights and generates a single output, which may be propagated to several other neurons[26]. Among different types of ANN models, back-propagation (BP) algorithm, an iterative gradient algorithm, is so popular that it has been used for the present work. BP neural network consists of an input layer, one or more hidden layers and output layers. In order to train the network using the BP algorithm, the differences between the ANN output and its desired value are calculated after each training iteration and the values of weights and biases modified by using these error terms. In the present work, an ANN program was written in MATLAB7 in our laboratory. This network was feed-forward fully connected that has three layers with sigmoidal transfer function. The network inputs are selected descriptors, the signal of the output node represents the *p*Ki value for sulfonylurea herbicides and the number of nodes in the hidden layer would be optimized. The Levenberg-Marquardt (LM) algorithm is one of the most efficient learning algorithms for neural networks[27]. The advantages of using LM algorithm are that specifying momentum or rate is not necessary and training processes are much more rapid. Therefore, in this work LM algorithm was used to develop the nonlinear model. The ANNs cannot be able to select important descriptors that would be used as its inputs; therefore it is necessary to apply a variable selection method. In this work, we use genetic algorithm and stepwise multiple regression feature-selection methods for these purpose. Then the optimized network was trained using training set for the adjustment of weights and biases values. It is known that a neural network can become over-fitted. An over-fitted network has usually learned the stimulus pattern it has seen perfectly, but cannot give an accurate prediction for unseen stimuli, and it is no longer able to generalize. There are several methods for overcoming this problem. One method is to use a test set to evaluate the prediction power of the network during its training. In this method after each 200 iterations, the network was used to calculate the inhibition constants of molecules included in the test set. To maintain the predictive power of the network at a desirable level, training was stopped when the value of error for the test set started to increase. Since the test error is not a good estimate of the generalization error, the prediction potential of the model was evaluated on a third set of data, named the prediction set. The compounds in the prediction set were not used during the training process and were reserved to evaluate the predictive power of the generated ANN.

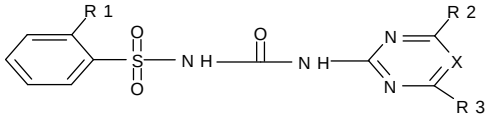
Results and Discussion

Nonlinear model

The data set and corresponding observed and predicted values of the *p*Ki of all molecules studied in this work are shown in Table 1. For the selection of the most important descriptors both genetic algorithm and stepwise multiple regression techniques were used. Then these descriptors were used as inputs for generated ANNs. In other hand, two separate ANNs were constructed that used these descriptors as inputs and their outputs are *p*Ki values of interesting molecules. These models referred as GANN and stepwise-NN, respectively. Applied GA contained a population of 100 individuals, which evolved for 300 generations, crossover probability 0.9 and mutation probability 0.01. Then by comparison between the fitness values of individuals, the best model was chosen. The process of the genetic algorithm is shown in Fig. 1 for all the generations from the beginning to the end of the process. The best fitness plot for the GA maps the gradual convergence of the best fitness values of successive generations towards the final optimum value. It indicates that for this case study, after 300 generations, the optimal results can be obtained.

Table 2 shows the names of descriptors of the ANNs models that their descriptors were chosen by stepwise multiple regression and GA methods. Although the number of descriptor in two models are identical but they are differ from each other. In order to optimize the number of nodes in hidden layer and to control over-fitting of the network, the values of mean square error of training (MSE_{Train}) and mean square error of test (MSE_{Test}) were monitored during the training procedure (Fig. 2). The minimum MSEs at the beginning of over fitting appeared in number of nodes and number of epochs equal to 7 and 6, respectively. So these values were considered as optimum. Table 3 shows the architecture and specification of the optimized network. The statistical parameters obtained by these models for the training and prediction set were shown in Table 4. These simulations demonstrated some significant differences

Table 1 — Structures of 68 herbicides and their calculated and experimental values of pK_i for this data set.

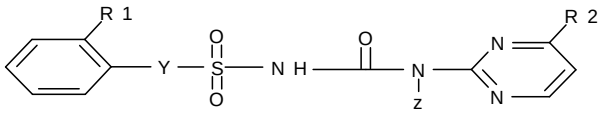


General structure for sulfonylurea

Compd	R1	R2	R3	X	pK_i(Exp)	pK_i(Cal)
1 $^{\Delta}$	NO_2	CH_3	H	CH	6.61	6.12
2	$COOC_2H_5$	CH_3	H	CH	6.57	6.34
3$^{\Delta}$	$COOCH_3$	CH_3	H	CH	6.44	6.15
*4	$COOC_2H_5$	OCH_3	H	CH	6.31	6.33
5	$COOC_2H_5$	Cl	H	CH	6.19	6.34
6$^{\Delta}$	NO_2	C_2H_5	H	CH	6.11	6.12
7$^{\Delta}$	NO_2	OC_2H_5	H	CH	5.96	5.94
8	$COOCH_2CH_2Cl$	CH_3	H	CH	5.92	5.95
*9	$COOCH(CH_3)_2$	CH_3	H	CH	5.90	4.95
*10	$COOCH(CH_3)_2$	OCH_3	H	CH	5.89	4.78
11	$COOCH_2CH_2Cl$	OCH_3	H	CH	5.85	5.68
12	NO_2	SCH_3	H	CH	5.75	5.27
13 $^{\Delta}$	$COOC_2H_5$	OC_2H_5	H	CH	5.68	5.67
14	$COOCH_2$Phenyl	OCH_3	H	CH	5.65	5.47
15	NO_2	$CH(CH_3)_2$	H	CH	5.52	5.64
16 $^{\Delta}$	$COOCH_3$	SCH_3	H	CH	5.49	5.94
17	Cl	OCH_3	H	CH	5.27	5.18
*18	$COOCH_2$Phenyl	CH_3	H	CH	5.08	4.13
19	$COOC_2H_5$	$OCH(CH_3)_2$	H	CH	4.95	4.94
20$^{\Delta}$	$COOCH_3$	OC_2H_5	H	CH	4.91	5.67
$^{\Delta}$21	$COOCH_2$Cyclohexyl	CH_3	H	CH	4.89	5.27
22	$COOCH_3$	OC_3H_7	H	CH	4.86	4.87
23	$COOCH_2$Cyclohexyl	OCH_3	H	CH	4.82	4.81
24	$COOC_2H_4OC_2H_5$	CH_3	H	CH	4.80	4.52
25	$COOC_2H_4OC_2H_5$	OCH_3	H	CH	4.60	4.48
26	$COOCH_3$	OCH_2CH_2F	H	CH	4.52	4.52
*27	NO_2	H	H	CH	4.49	4.01
28	$COOCH_3$	H	H	CH	4.48	4.44
29	Cl	OC_2H_5	H	CH	4.43	4.07
30	$NHCOCF_3$	CH_3	H	CH	4.29	4.70
31	NO_2	SC_2H_5	H	CH	4.29	4.41
*32	$NHCOCF_3$	OCH_3	H	CH	4.25	4.69
33	NO_2	SC_2H_5	H	CH	4.04	4.49
34	$NHCOCF_3$	OC_2H_5	H	CH	4.00	3.84
35	$COOC_2H_4OCH_3$	OCH_3	H	CH	3.71	4.22
36	$COOC_2H_5$	$NHCH_3$	H	CH	3.64	3.51
37	$COOC_2H_4OCH_3$	CH_3	H	CH	3.43	3.41
38	$COOC_2H_5$	$N(CH_3)_2$	H	CH	3.43	3.41
39	NO_2	$NHCH_3$	H	CH	3.22	3.83
40	$COOC_2H_5$	$OC_2H_4OCH_3$	H	CH	2.80	2.89
41	$COOC_2H_5$	$OC_2H_4OC_2H_5$	H	CH	2.64	2.59

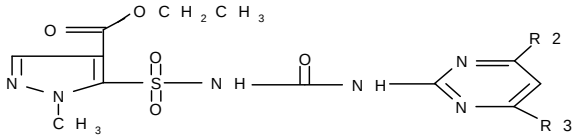
(*Contd.*)

Table 1 — Structures of 68 herbicides and their calculated and experimental values of pK_i for this data set. (*Contd*.)

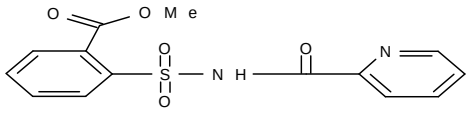
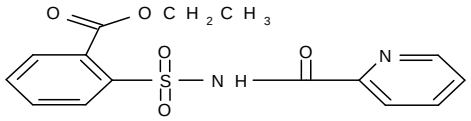
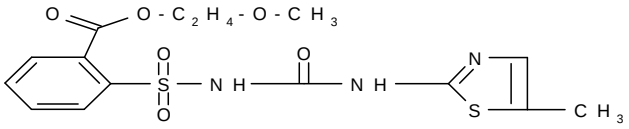
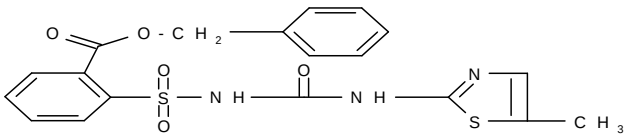


General structure for monosubstituted sulfonylurea

Compd	R1	R2	Y	Z	pK_i (Exp)	pK_i (Cal)
*42	$COOCH_3$	CH_3	CH_2	H	4.07	3.16
43	$COOCH_3$	$NHCH_3$	CH_2	H	4.03	4.11
44	$COOCH_3$	OCH_2CH_3	CH_2	H	3.84	3.80
45	$COOCH_3$	CH_3	O	H	3.88	3.74
*46	$COOCH_3$	C_2H_5	O	H	3.72	3.98
47 [Δ]	OCH_3	CH_3	O	H	3.73	4.27
48	NO_2	CH_3	-	C_4H_9	3.57	3.39
49	NO_2	CH_3	-	$(C_2H_5)CHCH_3$	3.43	3.62
50	NO_2	CH_3	-	CH_3	3.22	2.99
*51	$COOC_2H_5$	CH_3	-	CH_3	3.05	4.07

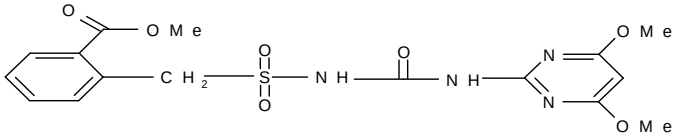
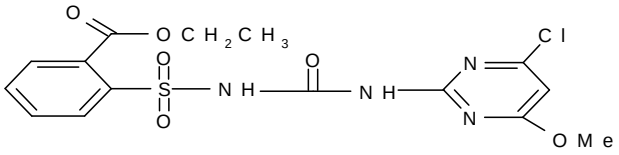
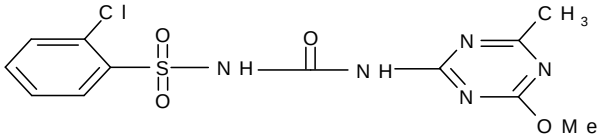
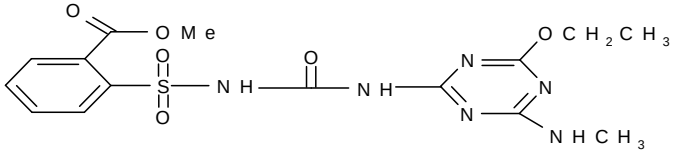
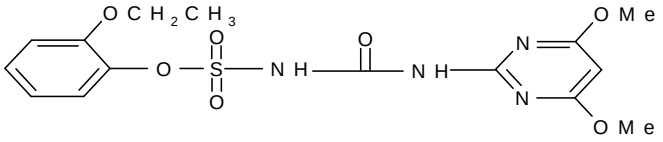
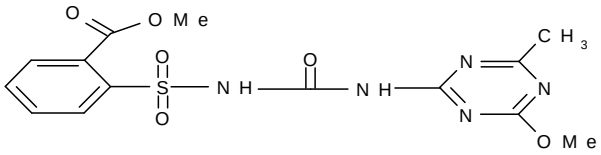
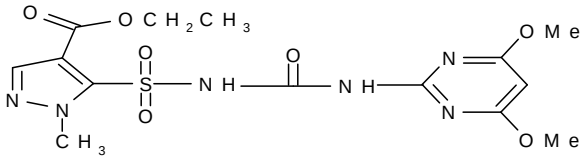
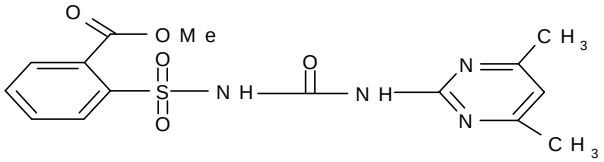


General structure for pyrazole sulfonylurea

Compd	R2	R3	pK_i(Exp)	pK_i(Cal)
52	CH_3	H	4.65	4.95
53	OC_2H_5	H	4.50	4.65
*54	SCH_3	H	3.74	3.12
55			3.70	3.70
	a typical pyridine			
56			3.22	3.47
	a typical pyridine			
57			3.53	3.44
	a typical thiazole			
58			2.92	2.97
	a typical thiazole			

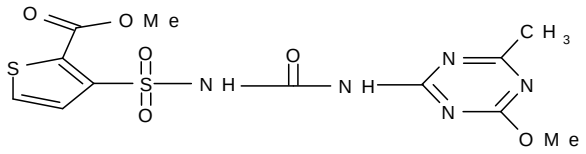
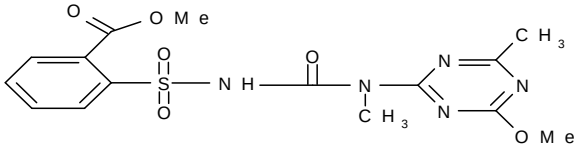
(*Contd*.)

Table 1 — Structures of 68 herbicides and their calculated and experimental values of pK_i for this data set. (*Contd*.)

59		7.79	7.77
	bensulfuron methyl		
*60		8.08	9.70
	chlorimuron ethyl		
61		7.84	7.87
	chlorsulfuron		
62		7.28	7.42
	ethametsulfuron methyl		
63		8.05	8.09
	ethoxysulfuron		
*64		7.97	8.81
	metsulfuron methyl		
65		7.00	6.87
	pyrazosulfuron ethyl		
$^{\Delta}$66		7.40	7.03
	sulfometuron methyl		

(*Contd*.)

Table 1 — Structures of 68 herbicides and their calculated and experimental values of pK_i for this data set. (*Contd*.)

67	 thifensulfuron methyl	7.22	7.18
68	 tribenuron methyl	6.50	6.45

[*] Test set
[Δ]Prediction set

Table 2—Definitions and notations of descriptors for stepwise-NN and GANN.

Stepwise-NN		GANN	
Descriptor	Notation[a]	Descriptor	Notation[a]
Highest eigenvalue	BEHm8	Randic-type eigenvector	VRA1
Superpendentic index	SPI	Moran autocorrelation	MATS8v
Average eigenvector coefficient	VEZ2	Radial distribution function	RDF105m
Complementary information content	CIC4	Molecular representation of structure	Mor30u
Structural information content	SIC2	Autocorrelation	H2e
Second Mohar index	TI2	Number of ethers	nROR
Salvation connectivity index	X5sol	No. of X—CR—X	C-031

[a]The notations are based on Dragon software.

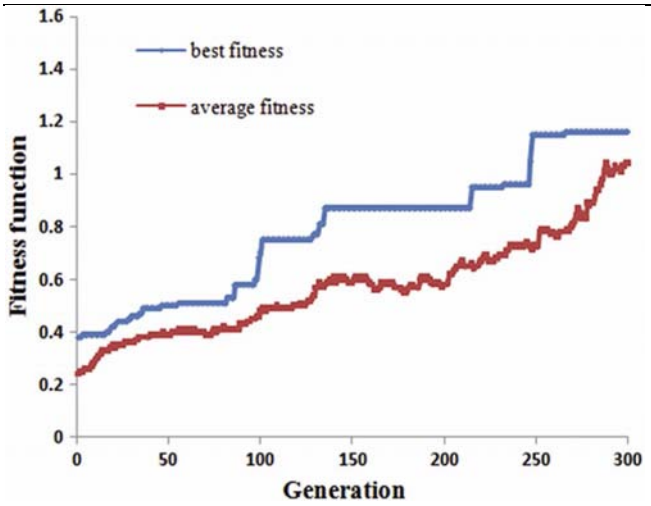


Fig. 1— The results of genetic algorithm for 300 generations.

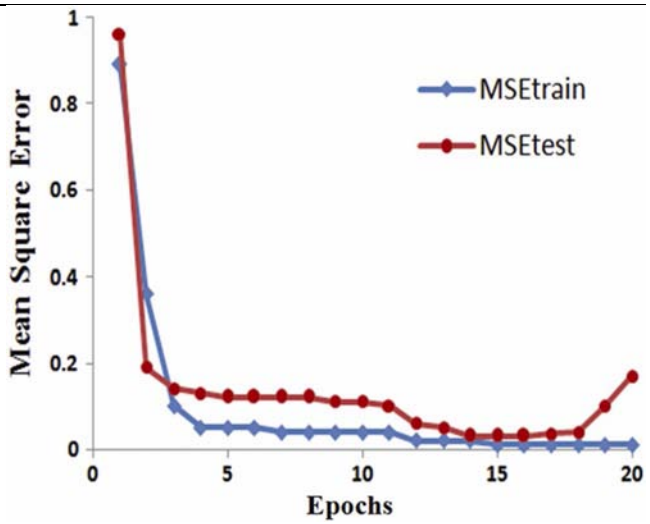


Fig. 2— Learning curve for the ANN.

between two networks. It can be seen from this table that statistical results of the GANN are better than other method. Also these results reveal that the GA is superior method for feature-selection in this QSAR study. The predicted values of the *p*Ki using GANN model for data set were shown in Table 1. Fig. 3 shows a plot of the GAANN calculated versus the experimental values of *p*Ki for the data set molecules.

Correlation coefficient of 0.957 for this plot confirms the suitability of the ANN model to predict of permeability coefficient. Results obtained reveals that there are some nonlinear relation between the inhibition constant of sulfonylurea herbicides and the selected structural molecular descriptors.

Table 3—Architectures of the optimized ANN.

Number of nodes in the input layer	7
Number of nodes in the hidden layer	5
Number of nodes in the output layer	1
Number of epoch in the beginning of over-fitting	16
Momentum	0.001
Training function	trainlm

Table 4—Statistical parameters obtained using stepwise-NN and GANN models.

Model	MSEt	MSEp	Rt	Rp	Ft	Fp
GANN	0.084	0.325	0.988	0.954	1823	80
stepwise-NN	0.337	0.445	0.917	0.862	263	17

[t]training set
[p]prediction set

Model validation

In spite of good accuracy and apparent mechanistic appeal, QSPR models should pass rigorous validation tests to be useful as reliable screening tools. Y-randomization test is a tool used in validation of QSTR models, whereby the performance of the original model in data description is compared to that of models built for permuted (randomly shuffled) response, based on the original descriptor pool and the original model building procedure. The Yscrambling procedure was performed to ensure that there is not any chance correlation in data matrix[28]. The mean value of R after 20 times Y-scrambling was 0.346, which disapproved the chance correlation probability. The real usefulness of QSTR models is not just their ability to reproduce known data, verified by their fitting power (R), but is mainly their possibility of predictive application. For this reason internal validation, leave one out cross-validation (LOO) and leave 8 out (L8O) were applied on GANN model which resulted in square cross validated correlation coefficient R_{cv}^2 and Q^2 are (0.886) and (0.946) respectively, which confirmed good predictive ability of this model.

Interpretation of descriptors

In this work, quantitative relationships between inhibition constants of sulfonylurea herbicides and their structural descriptors were investigated by using non-linear approach.

Seven descriptors appeared in the GANN model; Randic-type eigenvector-based index from adjacency matrix (VRA1), Moran autocorrelation (MATS8v), Radial distribution function (RDF105m), 3D Molecule representation of structure based on electron

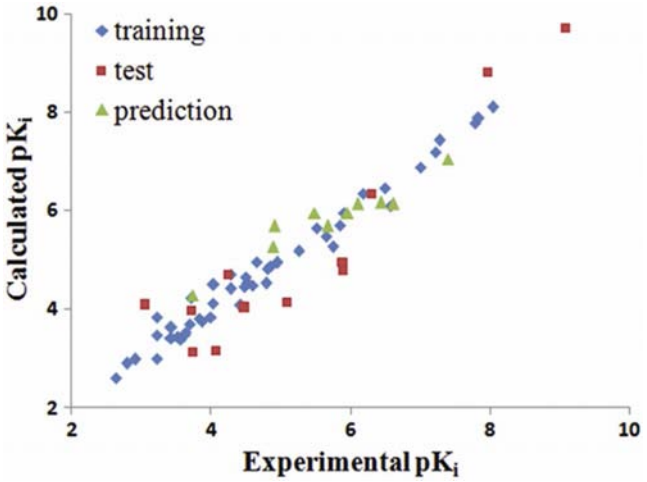


Fig. 3—Plot of the GANN calculated values of pK_i against the experimental ones.

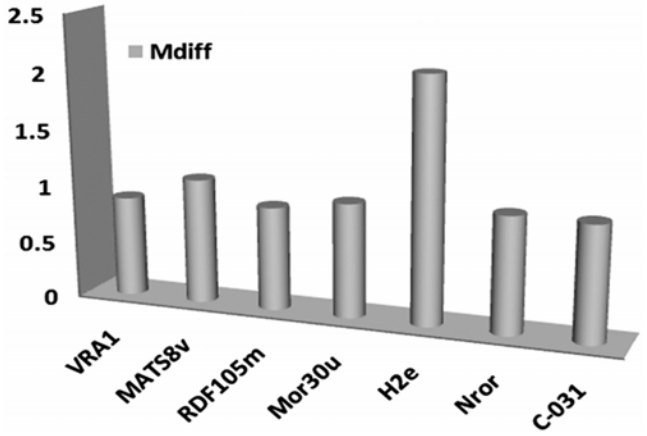


Fig. 4—Sensitivity analysis results.

diffraction (Mor30u), H autocorrelation (H2e), Number of ethers (aliphatic) (nROR), X—CR—X Atom-centred fragments (C-031)[29].

To determine the order of importance of descriptors in GANN model, the sensitivity analysis was performed. According to this method, the differences between the mean-square error (MSE) of the complete data set and the MSE were obtained when the ith variable is excluded from the trained network (MSE_i), and were shown as $Mdiff_i$ (2);

$$Mdiff_i = MSE_i - MSE \quad \ldots (2)$$

It is obvious that the most important variable is the one that leads to the highest value of Mdiffi. The values of Mdiffi for GANN model were calculated and plotted in Fig. 4. As it can be seen in this figure, the orders of importance of selected molecular descriptors are; H2e > MATS8v > nROR > C-031 >

Mor30u > RDF105m > VRA1. According to the sensitivity analysis results, among these 7 descriptors the GANN model has the least and most sensitivity to VRA1 and H2e descriptors respectively. This means that H2e is the most effective parameter in inhibition of AHAS by the sulfonylurea herbicides.

Conclusion

In the present study, both stepwise-NN and GANN approaches were used to develop the QSAR model for prediction of inhibition constant for sulfonylurea herbicides. The statistical results showed that the best model was GANN that combines genetic algorithm as variable selection technique and artificial neural network as feature mapping method. The superiority of this model accomplishes two messages. First, the evolutionary programming of genetic algorithm is very effective in the selection of the best descriptors, second, the strength of neural network in their ability to allow for flexible mapping of the selected features by manipulating their functional dependence implicitly, unlike regressions analysis. Finally descriptors appearing in these QSAR models provide related to different molecular properties, which can participate in the inhibition of AHAS by the sulfonylurea herbicides.

Acknowledgement

The authors gratefully acknowledge Shahid Bahonar University of Kerman for financial support

References

1 Penmetsa K V, Leidy R S & Shea D, *J Chromatogr A*, 766 (1997) 225.
2 Corcia A D, Crescenzi C, Samperi R & Scappaticcio L, *Anal Chem*, 69 (1997) 2819.
3 Sauers R F & Levitt G, in *Sulfonylureas: New high potency herbicides, Pesticide Synthesis through Rational Approaches*, edited by P S Magee, G K Kohn, & J J Menn, (American Chemical Society, Washington DC) 1984, p. 21.
4 Beyer E M, Duffy M J, Hay J V & Schlueter D D, in *Sulfonylureas, Herbicides: Chemistry Degradation and Mode of Action*, edited by P C Kearney & D D Kaufman, (Marcel Dekker, New York) 1987 p. 117.
5 LaRossa R A & Schloss J V, *J Biol Chem*, 259 (1984) 8753.
6 Ray T B, *Plant Physiol*, 75 (1984) 827.
7 Duggleby R G & Pang S S, *J Biochem Mol Biol*, 33 (2000)1.
8 Chipman D M, Duggleby R G & Tittmann K, *Curr Opin Chem Biol*, 9 (2005) 475.
9 Schultz T W, Cronin M T D & Netzev T I, *J Mol Struct Theochem*, 622 (2003) 23.
10 Wang J G, Li Z M, Ma N, Wang B L, Jiang L, Pang S S, Lee Y T, Guddat L W & Duggleby R G, *J Comput Aided Mol Des*, 19 (2005) 801.
11 Roy K & Paul S, *J Mol Model*, 16 (2010) 951.
12 Stefan P N, *J Mol Struct Theochem*, 622 (2003) 71.
13 Fatemi M H & Goudarzi N, *Electrophoresis*, 26 (2005) 2968.
14 HYPERCHEM. 6.03 (Hypercube), http://www.hyper.com/
15 DRAGON. 5.0 Evaluation Version, http://www.disat.unimib.it/chm.
16 Holland J, *Adaptation in natural and artificial systems*, (The University of Michigan press, Ann Arbor, MI) 1975.
17 Cartwright H M, *Applications of artificial intelligence in chemistry*, (Oxford University Press, Oxford) 1993.
18 Leardi R & Gonzalez, A L, *Chemom Intell Lab Sys*, 41(1998) 195.
19 Rogers D R & Hopfinger A, *J Chem Inf Comput Sci*, 34 (1994) 854.
20 Luke B T, *J Chem Inf Comput Sci*, 34 (1994) 1279.
21 Tan F, Fu X, Zhang Y & Bourgeois A G, *Soft Comput*, 12 (2008) 111.
22 Swarup K S, *Neurocomputing*, 71 (2008) 983.
23 Cai A H, Xiong X, Liu Y, An W K, Tan J Y & Luo Y, *Comput Mater Sci*, 48 (2010) 109.
24 Tagliaferri R, Longo G, D'Argenio B & Incoronato A, *Neural Networks*, 16 (2003) 295.
25 Singh S K, Mahesh K & Gupta A K, *Mater Design*, 31 (2010) 2288.
26 Swingler K, *Applying neural networks: a practical guide*, (Morgan Kaufman Publishers, Inc. San Francisco, California) 1996.
27 Kanzow C, Yamashita N & Fukushima M, *J Comput A Math*, 172 (2004) 375.
28 Tropsha A, Gramatica P & Gombar V K, *QSAR Comb Sci*, 22 (2003) 69.
29 Soltzberg L J & Wilkins C L, *J Am Chem Soc*, 99 (1977) 439.