Indian Journal of Chemistry
Vol. 46B, March 2007, pp. 478-487

Prediction of acidity constant for substituted acetic acids in water using artificial neural networks

Aziz Habibi-Yangjeh* & Mohammad Danandeh-Jenagharad

Department of Chemistry, Faculty of Science, University of Mohaghegh Ardebili, P.O. Box 179, Ardebil, Iran

E-mail: habibiyangjeh@yahoo.com

Received 11 November 2005; accepted (revised) 4 September 2006

Linear and non-linear quantitative structure-activity relationships have been successfully developed for the modelling and prediction of acidity constant (pK_a) of 87 substituted acetic acids with diverse chemical structures. The descriptors appearing in the multi-parameter linear regression (MLR) model are considered as inputs for developing the back-propagation artificial neural network (BP-ANN). ANN model is constructed using two molecular descriptors; the most positive charge of acidic hydrogen atom (q^+) and most negative charge of the carboxylic oxygen atom (q^-) as inputs and its output is pK_a. It has been found that properly selected and trained neural network with 53 substituted acetic acids could fairly represent dependence of the acidity constant on molecular descriptors. For evaluation of the predictive power of the generated ANN, an optimized network has been applied for prediction pK_a values of 17 compounds in the prediction set. Mean percentage deviation (MPD) for prediction set using the MLR and ANN models are 9.135 and 1.362, respectively. These improvements are due to the fact that the pK_a of substituted acetic acids demonstrates non-linear correlations with the molecular descriptors.

Keywords: Quantitative Structure-Activity Relationship, artificial neural network, acidity constant, theoretical descriptors, substituted acetic acids

IPC: Int.Cl.[8] C07C

Quantitative Structure-Property/Activity relationships (QSPRs/QSARs) now correlate chemical structure to a wide variety of physical, chemical, biological (including biomedical, toxicological, ecotoxicological) and technological properties[1-6]. QSPR/QSAR models are essentially calibration models in which the independent variables are molecular descriptors that describe the structure of molecules and the dependent variable is the property/activity of interest[7-10]. The development of a QSPR/QSAR models dependent upon the availability of a set of compounds (the training or calibration set) for each of which the value of the property/activity of interest is known and the necessary molecular descriptors can be calculated. Since these theoretical descriptors are determined solely from computational methods, *a priori* predictions of the properties/activities of compounds are possible, no laboratory measurements are needed thus saving time, space, materials, equipment and alleviating safety (toxicity) and disposal concerns. To obtain a significant correlation, it is crucial that appropriate descriptors be employed[11,12].

Many different techniques for constructing QSPR/QSAR models have been used including multi-parameter linear regression (MLR), principal component analysis (PCA) and partial least-squares regression (PLS)[13-15]. In addition, artificial neural networks (ANNs) have become popular due to their success where complex non-linear relationships exist amongst data[16-18]. ANNs are biologically inspired computer programs designed to simulate the way in which the human brain processes information. ANNs gather their knowledge by detecting the patterns and relationships in data and learned (or trained) through experience, not from programming. The behaviour of a neural network is determined by transfer functions of its neurons, by learning rule, and by the architecture itself. An ANN is formed from artificial neuron or processing elements (PE), connected with coefficients (weights), which constitute the neural structure and are organized in layers. The layers of neurons between the input and output layers are called hidden layers. The wide applicability of ANNs stems from their flexibility and ability to model non-linear

systems without prior knowledge of an empirical model. Neural networks do not need explicit formulation of the mathematical or physical relationships of the handled problem. These give ANNs an advantage over traditional fitting methods for some chemical applications. For these reasons in recent years, ANNs have been applied to a wide variety of chemical problems such as simulation of mass spectra, ion interaction chromatography, aqueous solubility and partition coefficient, simulation of nuclear magnetic resonance spectra, prediction of bioconcentration factor, solvent effects on reaction rate, prediction of normalized polarity parameter in mixed solvent systems and acidity constant of phenols and benzoic acid[19-25].

The interpretation and prediction of pK_a values for chemical compounds are of general importance and usefulness for chemists[26]. The acidity of molecules relates to molecular structure in a complex way. Although in the last few years several theoretical studies have been performed for correlation of acidity constant of acids, but in these studies linear equations have been used[27-30].

Very recently, QSAR models have been developed for correlation of acidity constant of phenols and benzoic acids in water[31,32]. In order to develop the idea, in this work the method has been applied for acidity constant of substituted acetic acids in water. In the first step, a MLR model was constructed. Then for inspection of non-linear interactions/relation between different molecular descriptors of the acids, an ANN model was generated for prediction of the pK_a values and the results were compared with the experimental and calculated values using the MLR model.

Results and Discussion

Multi-parameter linear correlation of pK_a values for 53 substituted acetic acids *versus* the molecular descriptors in the training set gives Eqn. 1.

$$pK_a = 26.897(\pm3.210) - 127.808(\pm11.889)q^+ + 26.580(\pm3.060)q^- \quad \ldots (1)$$

$n = 53$; $R^2 = 0.816$; MPD = 11.094; RMSE = 0.4307; $\beta_{q^+} = -0.656$; $\beta_{q^-} = 0.530$

It is clear that the pK_a of the acids correlates with most positive charge of acidic hydrogen atom (q^+) and most negative charge of carboxylic oxygen atom (q-) descriptors. As can be seen, acidity of acetic acids increases with increasing q^+ and decreases with q^-. With increasing q^+, interactions of water with acidic hydrogen of acetic acids increases, then it can be easily removed from the compounds. Acidity constant of the compounds decreases with increasing q-descriptor, because basicity of carboxylic oxygen atom increases with increasing this descriptor. Effects of q^+ on the acidity are higher than that of q^-, because standardized coefficients of q^+ is higher than that of the other descriptor. The calculated values of pK_a for the compounds in training, validation and prediction sets using the MLR model have been plotted *versus* the experimental values of it (**Figure 1**).

In the MLR model it is assumed that all the molecular descriptors are independent of each other and truly additive as well as relevant to the property under study. ANNs are particularly well-suited for QSAR/QSPR models because of their ability to extract nonlinear information present in the descriptors. For this reason the next step in this work was generation of the ANN model. There are no rigorous theoretical principles for choosing the proper network topology; so different structures were tested in order to obtain the optimal hidden neurons and training cycles[22]. Before training the network, the number of nodes in the hidden layer was optimized. In order to optimize the number of nodes in the hidden layer, several training sessions were conducted with different numbers of hidden nodes (from one to seventeen). The root mean squared error of training (RMSET) and validation (RMSEV) sets were obtained at various iterations for different number of neurons at the hidden layer and the minimum value of RMSEV was recorded as the optimum value. Plot of RMSET and RMSEV *versus* the number of nodes in the hidden layer has been shown in **Figure 2**. It is clear that the fifteen nodes in hidden layer is optimum value.

This network consists of two inputs (including q^+ and q^- descriptors), the same descriptors in the MLR model, and one output for pK_a. Then an ANN with architecture 2-15-1 was generated. It is noteworthy that training of the network was stopped when the RMSEV started to increase *i.e.* when overtraining begins. The overtraining causes the ANN to loose its prediction power[25]. Therefore, during training of the networks, it is desirable that iterations are stopped when overtraining begins. To control the overtraining of the network during the training procedure, the values of RMSET and RMSEV were calculated and recorded to monitor the extent of the learning in various iterations. Results showed that overfitting does not seen in the optimum architecture (**Figure 3**).

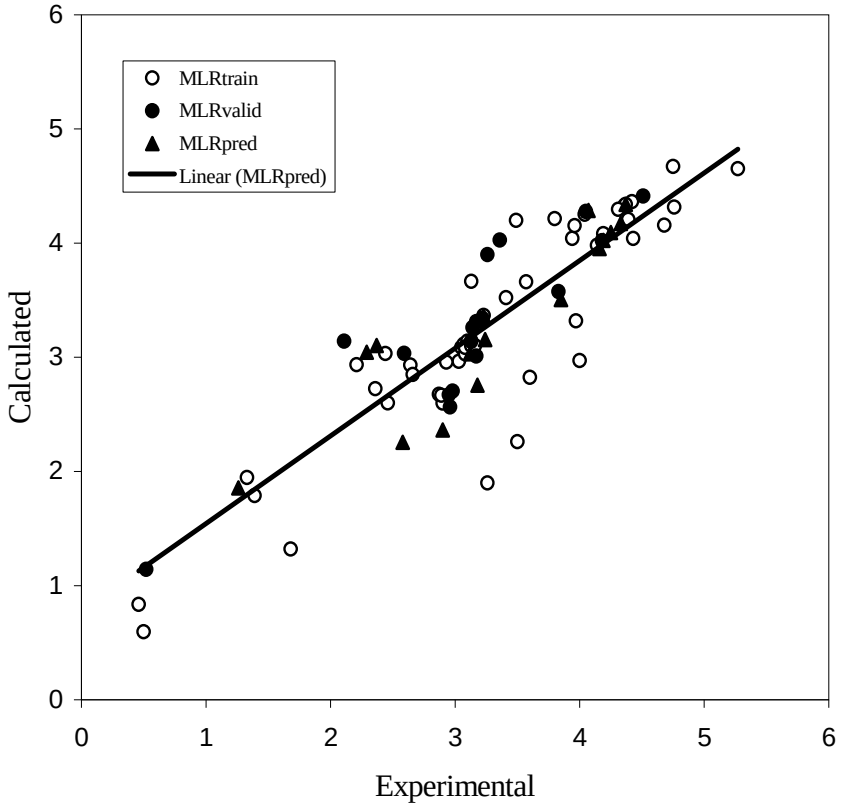

Figure 1 — Plot of the calculated values of pK_a from the MLR model *versus* the experimental values of it for training, validation and prediction sets

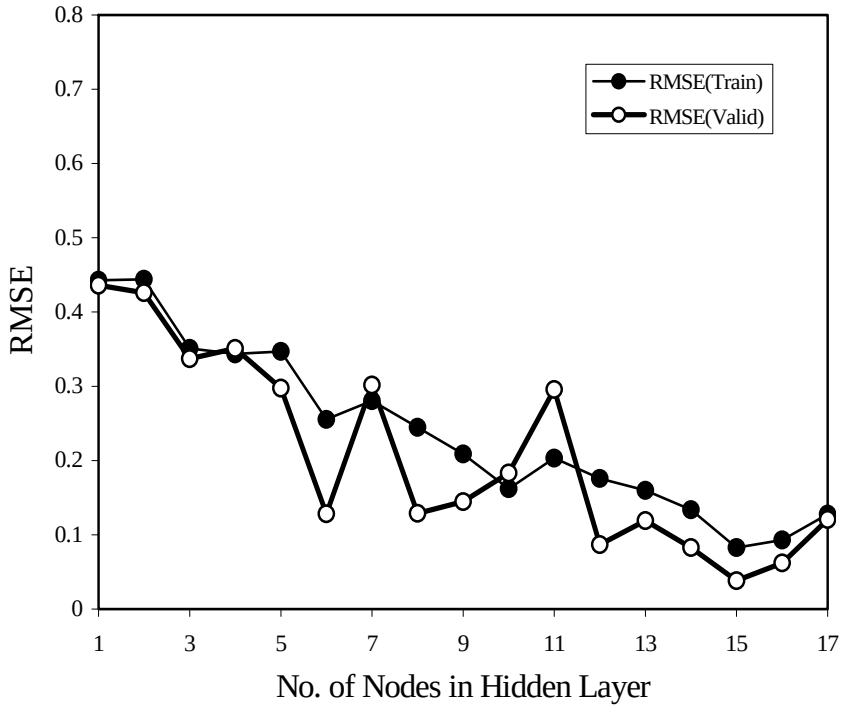

Figure 2 — Plot of RMSE for training and validation sets *versus* the number of nodes in hidden layer

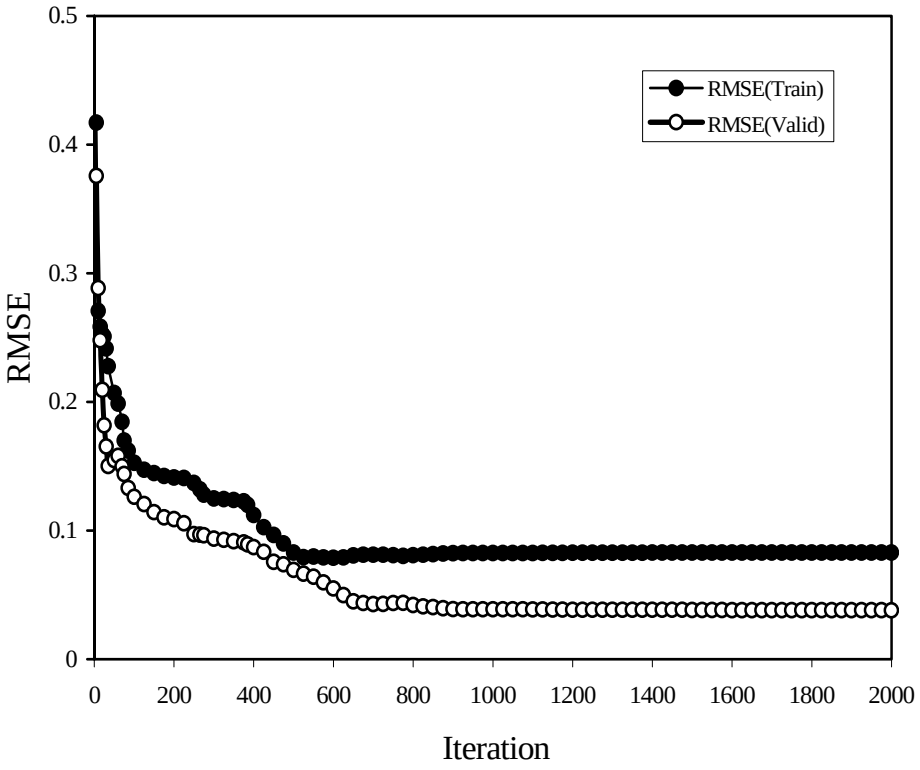

Figure 3 — Plot of RMSE for training and validation sets *versus* the number of iterations

The generated ANN was then trained using the training and validation sets for optimization of the weights and biases. For the evaluation of the predictive power of the generated ANN, an optimized network was applied for prediction the pK_a values of various acetic acids in the prediction set, which were not used in the modelling procedure (**Table I**). The calculated values of pK_a for the compounds in training, validation and prediction sets using the ANN model have been plotted *versus* the experimental values of it in **Figure 4**.

As expected, the calculated values of pK_a are in very good agreement with those of the experimental values. The correlation equation for all of the calculated values of pK_a from the ANN model and the experimental values is as follows:

$$pK_a\ (\text{cal}) = 0.9939\ pK_a\ (\text{exp}) + 0.0199 \quad \ldots (2)$$

(R^2 = 0.993; MPD = 1.307; RMSE = 0.0770; F = 12815.68)

Similarly, the correlation of pK_a (cal) *versus* pK_a (exp) values in the prediction set gives equation (3):

$$pK_a\ (\text{cal}) = 0.9722\ pK_a\ (\text{exp}) + 0.0751 \quad \ldots (3)$$

(R^2 = 0.991; MPD = 1.362; RMSE = 0.0857; F = 1569.37)

Plot of IPD for pK_a values in the training, validation and prediction sets *versus* the experimental values of it has been illustrated in **Figure 5**. The propagation of errors in both sides of zero is random and the slope (0.9939) and intercept (0.0199) of the linear regression of the calculated values *versus* the experimental values are very close to the ideal behaviour (ideal slope 1 and the ideal intercept 0).

Table II compares the results obtained using the MLR and ANN models. The squared correlation coefficient (R^2) and RMSE of the models for total, training, validation and prediction sets demonstrate potential of the ANN model for prediction of pK_a values of various substituted acetic acids in water.

As a result, it was found that properly selected and trained neural network could fairly represent dependence of the acidity constant of substituted acetic acids in water on the molecular descriptors. Then the optimized neural network could simulate the complicated nonlinear relationship between pK_a values and the molecular descriptors. The squared correlation coefficients (R^2) and RMSE are 0.991 and 0.0857 for the prediction set by the MLR model should be compared with the values of 0.809 and 0.3738, respectively, for the ANN model. It can be seen from **Table II** that although the parameters

Table I — Experimental and calculated values of pK_a for various substituted acetic acids in water at 25ºC for training, validation and prediction sets by multi-parameter linear regression (MLR) and artificial neural network (ANN) models along with individual percent deviation (IPD)[a] **—** *Contd*

No.	**Compd**	Exp	MLR	IPD_{MLR}	ANN	IPD_{ANN}
	Training set					
1	Acetic acid	4.76	4.315	9.357	4.782	-0.462
2	Allylacetic acid	4.68	4.157	11.171	4.373	6.556
3	Bromoacetic acid	2.90	2.597	10.442	2.916	-0.562
4	2-(3'-Bromophenoxy)acetic acid	3.09	3.097	-0.224	3.138	-1.566
5	2-(4'-Bromophenoxy)acetic acid	3.13	3.101	0.930	3.099	1.000
6	2-Bromo-2-phenylacetic acid	2.21	2.935	-32.783	2.172	1.715
7	4-*tert*-Butylphenylacetic acid	4.42	4.363	1.287	4.386	0.765
8	Chloroacetic acid	2.87	2.675	6.788	2.847	0.791
9	Chlorodifluoroacetic acid	0.46	0.836	-81.733	0.428	6.935
10	2-Chlorophenoxyacetic acid	3.05	3.086	-1.196	3.138	-2.879
11	3-Chlorophenoxyacetic acid	3.07	3.113	-1.398	3.121	-1.664
12	4-Chlorophenoxyacetic acid	3.10	3.141	-1.335	3.069	0.994
13	3-Chlorophenylacetic acid	4.14	3.980	3.874	4.293	-3.696
14	4-Chlorophenylacetic acid	4.19	4.082	2.575	4.182	0.193
15	Cyanoacetic acid	2.46	2.600	-5.673	2.456	0.150
16	*m*-Cyanophenoxyacetic acid	3.03	2.964	2.165	3.011	0.637
17	*p*-Cyanophenoxyacetic acid	2.93	2.957	-0.935	2.949	-0.631
18	1,1-Cyclohexanediacetic acid	3.49	4.199	-20.302	3.751	-7.476
19	1,1-Cyclopentyldiacetic acid	3.80	4.215	-10.926	3.855	-1.455
20	*trans*-Cyclopentane-1,2-diacetic acid	4.43	4.040	8.798	4.504	-1.675
21	Dibromoacetic acid	1.39	1.788	-28.658	1.392	-0.151
22	2,4-Dichlorophenoxyacetic acid	2.64	2.932	-11.060	2.685	-1.697
23	4,6-Dichlorophenoxy-2-methylacetic acid	3.13	3.664	-17.069	3.359	-7.307
24	Difluoroacetic acid	1.33	1.947	-46.386	1.330	-0.008
25	Dimethylphenylsilylacetic acid	5.27	4.651	11.740	5.270	0.000
26	2,4-Dinitrophenylacetic acid	3.50	2.261	35.407	3.500	0.009
27	Diphenylacetic acid	3.94	4.041	-2.568	3.987	-1.193
28	2-Fluorophenoxyacetic acid	3.08	3.030	1.624	3.136	-1.802
29	3-Fluorophenoxyacetic acid	3.08	3.086	-0.192	3.143	-2.042
30	4-Fluorophenoxyacetic acid	3.13	3.144	-0.444	3.062	2.188
31	Hydroxy-iodo-phenylacetic acid	3.26	1.899	41.739	3.260	0.003
32	Hydroxy-phenyl-acetic acid	3.41	3.522	-3.279	3.223	5.475
33	Indole-3-acetic acid	4.75	4.670	1.685	4.756	-0.133
34	3-Iodophenoxyacetic acid	3.13	3.099	0.977	3.132	-0.061
35	4-Iodophenoxyacetic acid	3.16	3.101	1.870	3.099	1.940
36	2-Iodophenylacetic acid	4.04	4.252	-5.236	3.905	3.332
37	4-Isopropylphenylacetic acid	4.39	4.206	4.187	4.311	1.802
38	Mercaptoacetic acid	3.60	2.824	21.564	3.605	-0.142

— *Contd*

Table I — Experimental and calculated values of pK_a for various substituted acetic acids in water at 25ºC for training, validation and prediction sets by multi-parameter linear regression (MLR) and artificial neural network (ANN) models along with individual percent deviation (IPD)[a] **—** *Contd*

No.	Compd	Exp	MLR	IPD_{MLR}	ANN	IPD_{ANN}
39	Methoxyacetic acid	3.57	3.660	-2.532	3.571	-0.025
40	(4'-Methoxy)phenoxyacetic acid	3.21	3.320	-3.439	3.201	0.283
41	4'-Methoxyphenylacetic acid	4.36	4.328	0.743	4.381	-0.479
42	(2-Methylphenoxy)acetic acid	3.23	3.366	-4.222	3.224	0.186
43	(4-Methylphenyl)acetic acid	4.37	4.339	0.719	4.391	-0.490
44	Methylsulfonylacetic acid	2.36	2.724	-15.440	2.360	0.000
45	Nitroacetic acid	1.68	1.320	21.417	1.680	-0.012
46	(4-Nitrophenoxy)acetic acid	2.89	2.668	7.685	2.892	-0.073
47	2-Nitrophenylacetic acid	4.00	2.971	25.730	4.003	-0.077
48	3-Nitrophenylacetic acid	3.97	3.318	16.418	3.969	0.033
49	Phenylacetic acid	4.31	4.296	0.334	4.444	-3.118
50	Phenylsulfenylacetic acid	2.66	2.850	-7.145	2.662	-0.064
51	Phenylsulfonylacetic acid	2.44	3.033	-24.304	2.440	0.000
52	Trifluoroacetic acid	0.50	0.597	-19.308	0.524	-4.800
53	Triphenylacetic acid	3.96	4.152	-4.845	3.961	-0.013
	Validation set					
54	(3-Bromophenyl)hydroxyacetic acid	3.13	3.138	-0.258	3.129	0.019
55	2-(Bromophenyl)acetic acid	4.05	4.278	-5.618	4.063	-0.331
56	(4-Chloro-3-nitrophenoxy)acetic acid	2.96	2.565	13.330	2.959	0.027
57	4-Chlorophenoxy-2-methylacetic acid	3.26	3.899	-19.590	3.278	-0.537
58	o-Cyanophenoxyacetic acid	2.98	2.703	9.295	2.958	0.745
59	Cyclohexylacetic acid	4.51	4.412	2.170	4.494	0.359
60	Dichloroacetylacetic acid	2.11	3.141	-48.877	2.109	0.047
61	2-6-Dimethylphenoxyacetic acid	3.36	4.025	-19.806	3.362	-0.048
62	Fluoroacetic acid	2.59	3.034	-17.130	2.590	-0.012
63	Hydroxyacetic acid	3.83	3.575	6.660	3.805	0.648
64	2-Iodophenoxyacetic acid	3.17	3.011	5.015	3.032	4.360
65	4-Iodophenylacetic acid	4.18	4.021	3.813	4.239	-1.414
66	(3'-Methoxy)phenoxyacetic acid	3.14	3.260	-3.819	3.145	-0.172
67	(4-Methylphenoxy)acetic acid	3.22	3.339	-3.693	3.227	-0.217
68	(3-Nitrophenoxy)acetic acid	2.95	2.672	9.426	2.961	-0.373
69	Phenoxyacetic acid	3.17	3.312	-4.476	3.179	-0.274
70	Trichloroacetic acid	0.52	1.143	-119.790	0.521	-0.231
	Prediction set					
71	2-(2'-Bromophenoxy)acetic acid	3.12	3.030	2.900	3.074	1.484
72	4-(Bromophenyl)acetic acid	4.19	4.023	3.983	4.162	0.680
73	(3-Chlorophenyl)hydroxyacetic acid	3.24	3.154	2.652	3.241	-0.019
74	2-Chlorophenylacetic acid	4.07	4.286	-5.308	4.011	1.459
75	2-Cyano-2-methyl-2-phenylacetic acid	2.29	3.044	-32.923	2.269	0.926
76	Cyclohexylcyanoacetic acid	2.37	3.102	-30.905	2.442	-3.017

— *Contd*

Table I — Experimental and calculated values of pK_a for various substituted acetic acids in water at 25ºC for training, validation and prediction sets by multi-parameter linear regression (MLR) and artificial neural network (ANN) models along with individual percent deviation (IPD)[a] **—** *Contd*

No.	Compd	Exp	MLR	IPD_{MLR}	ANN	IPD_{ANN}
77	Dichloroacetic acid	1.26	1.856	-47.331	1.259	0.095
78	(3,4-Dimethoxy)phenylacetic acid	4.33	4.173	3.632	4.455	-2.889
79	4-Ethylphenylacetic acid	4.37	4.339	0.719	4.391	-0.490
80	4-Fluorophenylacetic acid	4.25	4.092	3.715	3.978	6.398
81	Iodoacetic acid	3.18	2.755	13.379	3.190	-0.299
82	3-Iodophenylacetic acid	4.16	3.953	4.985	4.010	3.603
83	(2'-Methoxy)phenoxyacetic acid	3.23	3.360	-4.010	3.249	-0.594
84	(3-Methylphenoxy)acetic acid	3.20	3.341	-4.418	3.228	-0.872
85	(2-Nitrophenoxy)acetic acid	2.90	2.363	18.526	2.908	-0.272
86	4-Nitrophenylacetic acid	3.85	3.505	8.958	3.849	0.023
87	Thiocyanatoacetic acid	2.58	2.255	12.601	2.581	-0.031

(a) Exp refers to the experimental values of pK_a, MLR and ANN refer to multi-parameter linear regression and artificial neural network calculated values of pK_a, respectively.

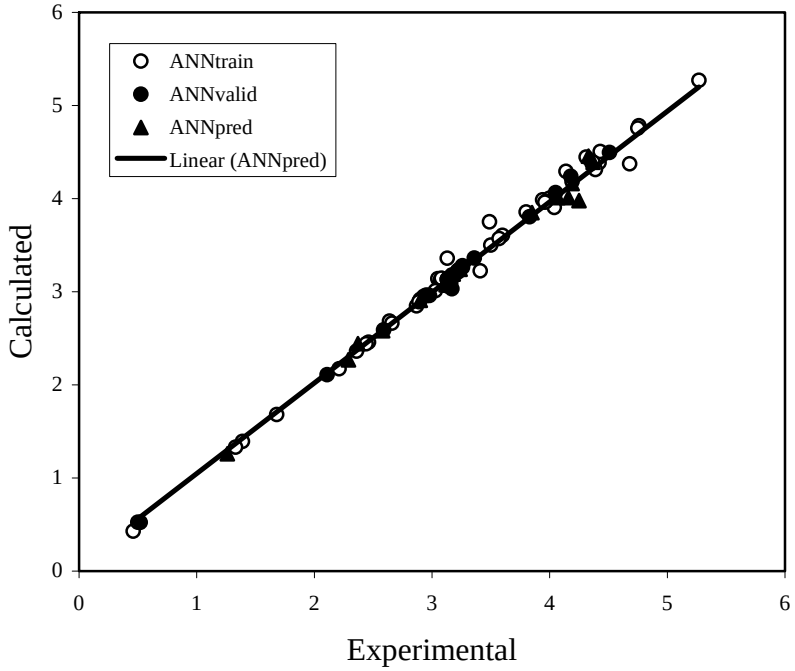

Figure 4 — Plot of the calculated values of pK_a from the ANN model *versus* the experimental values of it for training, validation and prediction sets

appearing in the MLR model are used as inputs for the generated ANN, the statistics has shown a large improvement. These improvements are due to the fact that pK_a values of acetic acids demonstrate non-linear correlations with the molecular descriptors.

Experimental Section

Descriptor generation. In order to calculate the theoretical descriptors, the z-matrices (molecular models) were constructed with the aid of HyperChem 7.0 and molecular structures were optimized using AM1 algorithm[33]. In order to calculate some of electronic theoretical descriptors, the molecular geometries of molecules were further optimized with the same algorithm in MOPAC program version 6.0. The other molecular electronic descriptors were calculated by *Dragon* package version 2.1 (Ref. 34). For this propose the output of the HyperChem

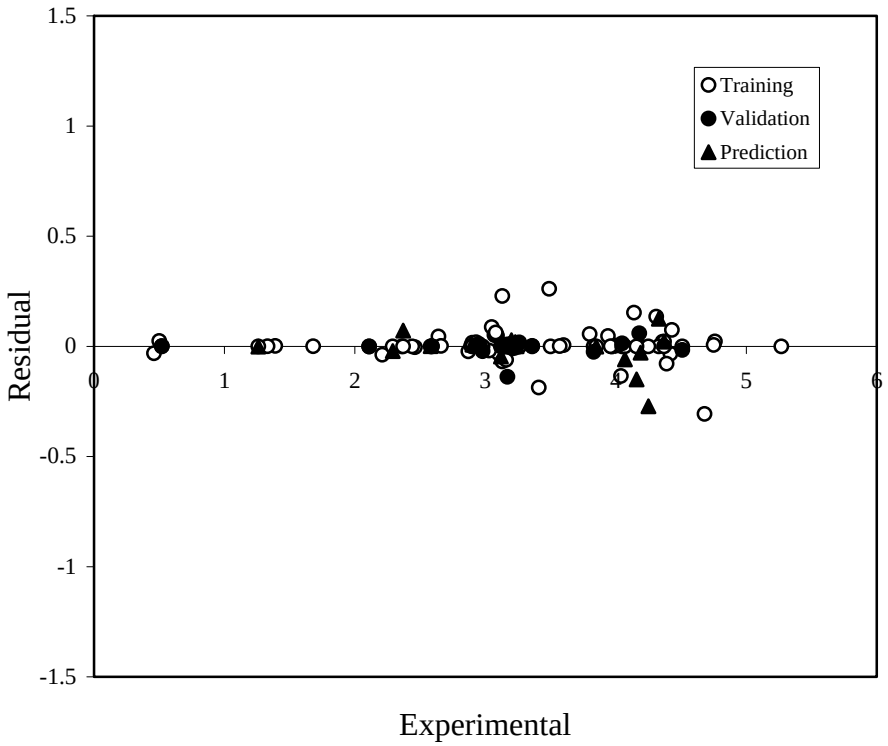

Figure 5 — Plot of the residual for calculated values of pK_a from the ANN model *versus* the experimental values of it for training, validation and prediction set

Table II — Comparsion of statistical parameters obtained by the MLR and ANN models for correlation of acidity constant of substituted acetic acids with molecular descriptors[a]

Model	R^2_{tot}	R^2_{train}	R^2_{valid}	R^2_{pred}	$RMSE_{tot}$	$RMSE_{train}$	$RMSE_{valid}$	$RMSE_{pred}$
MLR	0.805	0.816	0.787	0.809	0.4186	0.4307	0.4225	0.3738
ANN	0.993	0.993	0.998	0.991	0.0770	0.0830	0.0381	0.0857

(a) Subscript train is referring to the training set, valid is referring to the validation set and the pred is referring to the prediction set, tot is referring to the total data set, R is the correlation coefficient.

software for each compound fed into the *Dragon* program and the descriptors were calculated. As a result, a total of 18 theoretical descriptors were calculated for each compound in the data sets (87 substituted acetic acids).

Linear correlations. Acidity constant of acetic acids are literature values at 25ºC[35]. MLR model was developed for prediction pK_a values by molecular descriptors. The method of stepwise multi-parameter linear regression was used to select the most important descriptors to calculate the coefficients relating the pK_a to the descriptors. The best MLR model is one that has high correlation coefficient and F-value and low standard error. The MLR models were generated using spss/pc software package release 9.

Neural network generation. The specification of a typical neural network model requires the choice of the type of inputs, the number of hidden layers, the number of neurons in each hidden layer and the connection structure between the inputs and output layers. The number of input nodes in the ANNs was equal to the number of molecular descriptors in the MLR model. A three-layer network with a sigmoidal transfer function was designed. The initial weights were randomly selected between 0 and 1. Before training, the input and output values were normalized between 0.1 and 0.9. The optimization of the weights

and biases was carried out according to Levenberg-Marquardt algorithms for back-propagation of error[36]. The data set was randomly divided into three groups: a training set, a validation set and a prediction set consisting of 53, 17 and 17 molecules, respectively. The training set was used for training of the ANN and the validation (monitoring) set was used for determination the extent of training. The generalization ability of the model was checked using the prediction set[37].

The performances of training, validation and prediction of ANNs are evaluated by the mean percentage deviation (MPD) and root-mean squared error (RMSE), which are defined as follows:

$$\text{MPD} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{(P_i^{\exp} - P_i^{cal})}{P_i^{\exp}}\right| \quad \ldots (4)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N}\frac{(P_i^{\exp} - P_i^{cal})^2}{N}} \quad \ldots (5)$$

where P_i^{exp} and P_i^{cal} are experimental and calculated values of pK_a with the ANN model and N denote the number of data points.

Individual percent deviation (IPD) is defined as follows:

$$\text{IPD} = 100 \times \left(\frac{P_i^{calc} - P_i^{\exp}}{P_i^{\exp}}\right) \quad \ldots (6)$$

The processing of the data was carried using Matlab 6.5 (Ref. 38). The neural networks were implemented using Neural Network Toolbox Ver. 4.0 for Matlab.

Conclusion

A two-descriptor non-linear computational neural network model has been developed for prediction of acidity constant (pK_a) for various acetic acids in water using quantitative structure-activity relationship. Comparison of the values of RMSE for training, validation and prediction sets (and other statistical parameters in **Table II**) for the MLR and ANN models demonstrate superiority of the ANN model over the regression model. Root-mean square error of 0.3738 for the prediction set by the MLR model should be compared with the value of 0.0857 for the ANN model. Since the improvement of the results obtained using nonlinear model (ANN) is considerable, it can be concluded that the non-linear characteristics of the molecular descriptors on the pK_a values of substituted acetic acids in water is serious.

Acknowledgements

The authors wish to acknowledge the vice-presidency of research, University of Mohaghegh Ardebili, for financial support to this work.

References

1 Katritzky A R, Karelson M & Lobanov V S, *Pure Appl Chem*, 69, **1997**, 245.
2 Balaban A T, *J Chem Inf Comut Sc*, 37, **1997**, 645.
3 Benfenati E & Gini G, *Toxicology*, 119, **1997**, 213.
4 Cronce D T, Famini G R, Soto J A D & Wilson L Y, *J Chem Soc Perkin Trans 2*, **1998**, 1293.
5 Engberts J B F N, Famini G R, Perjessy A & Wilson L Y, *J Phys Org Chem*, 11, **1998**, 261.
6 Hiob R & Karelson M, *J Chem Inf Comut Sci*, 40, **2000**, 1062.
7 Habibi-Yangjeh A, *Indian J Chem*, 42B, **2003**, 1478.
8 Habibi-Yangjeh A, *Indian J Chem*, 43B, **2004**, 1504.
9 Nikolic S, Milicevic A, Trinajstic N & Juric A, *Molecules*, 9, **2004**, 1208.
10 Devillers J, *SAR and QSAR Environ Res*, 15, **2004**, 501.
11 Karelson M & Lobanov V S, *Chem Rev*, 96, **1996**, 1027.
12 Todeschini R & Consonni V, *Handbook of Molecular Descriptors* (Wiley-VCH, Weinheim, Germany), **2000**.
13 Kramer R, *Chemometric Techniques for Quantitative Analysis* (Marcel Dekker, New York), **1998**.
14 Barros A S & Rutledge D N, *Chemomet Intell Lab Syst*, 40, **1998**, 65.
15 Garkani-Nejad Z, Karlovits M, Demuth W, Stimpfl T, Vycudilik W, Jalali-Heravi M & Varmuza K, *J Chromatogr A*, 1028, **2004**, 287.
16 Patterson D W, *Artificial Neural Networks: Theory and Applications* (Simon and Schuster, New York), Part III, Ch. 6, **1996**.
17 Bose N & Liang P, *Neural Network Fundamentals* (McGraw-Hill, New York), **1996**.
18 Zupan J & Gasteiger J, *Neural Networks in Chemistry and Drug Design* (Wiley-VCH, Weinhein), **1999**.
19 Agatonovic-Kustrin S & Beresford R, *J Pharm Biomed Anal*, 22, **2000**, 717.
20 Fatemi M H, *J Chromatogr A*, 955, **2002**, 273.
21 Jalali-Heravi M, Masoum S & Shahbazikhah P, *J Magn Reson* 171, **2004**, 176.
22 Wegner J K & Zell A, *J Chem Inf Comput Sci*, 43, **2003**, 1077.
23 Valkova I, Vracko M & Basak S C, *Anal Chim Acta*, 509, **2004**, 179.
24 Habibi-Yangjeh A & Nooshyar M, *Bull Korean Chem Soc*, 26, **2005**, 139.
25 Habibi-Yangjeh A & Nooshyar M, *Physics and Chemistry of Liquids*, 43, **2005**, 239.
26 Hmmateenejad B, Sharghi H, Akhond M & Shamsipur M, *J Solution Chem*, 32, **2003**, 215.
27 Zhao Y-H, Yuan L-H & Wang L-S, *Bull Environ Contam Toxicol*, 57, **1996**, 242.

28 Citra M J, *Chemosphere*, 38, **1999**, 191.

29 Liptak M D, Gross K C, Seybold P G, Feldgus S & Shields G C, *J Am Chem Soc*, 124, **2002**, 6421.

30 Ma Y, Gross K C, Hollingsworth C A, Seybold P G & Murray J S, *J Mol Model*, 10, **2004**, 235.

31 Habibi-Yangjeh A, Danandeh-Jenagharad M & Nooshyar M, *Bull Korean Chem Soc*, 26, **2005**, 2007.

32 Habibi-Yangjeh A, Danandeh-Jenagharad M & Nooshyar M, *J Mol Model*, 12, **2006**, 338.

33 HyperChem, Release 7.0 for Windows, *Molecular Modeling System*, Hypercube Inc., **2002**.

34 Todeschini R, Consonni V & Pavan M, *Dragon Software* Version 2.1, **2002**.

35 Dean J A, *Lange's Handbook of Chemistry*, 15th Edn. (McGraw-Hill, Inc.), **1999**.

36 Demuth H & Beale M, *Neural Network Toolbox* (Mathworks, Natick, MA), **2000**.

37 Despagne F & Massart D L, *Analyst*, 123, **1998**, 157R.

38 Matlab 6.5. Mathworks, **1984-2002**.