Revue Roumaine de Chimie, **1998**, *43* (4), 347–354

ARTIFICIAL NEURAL NETWORKS APPLICATIONS. PART 7[1]

ESTIMATION OF BIOCONCENTRATION FACTORS IN FISH USING SOLVATOCHROMIC PARAMETERS

Ovidiu IVANCIUC

"Politehnica" University, Faculty of Industrial Chemistry, Department of Organic Chemistry of Bucharest, Splaiul Independenței 313, 77206 Bucharest, Roumania
E-mail: o_ivanciuc@chim.upb.ro

Received February 28, 1996

The bioconcentration factors in fish for 51 organic nonelectrolytes are estimated by a neural network model, using as structural descriptors three solvatochromic parameters: V_I, the intrinsic solute molecular volume, β, the hydrogen bond acceptor basicity, and α, the hydrogen bond donor acidity parameters. The neural network quantitative structure-activity relationship gives better estimations than a multiple linear regression model, both for model calibration and prediction. The predictions of the neural model are affected by large errors for patterns whose input and/or output values are outside the range of the corresponding values in the training set.

INTRODUCTION

Neural networks are algorithmic systems derived from a simplified concept of the brain.[2–8] In a neural network, a number of units, called neurons, are interconnected into a net-like structure. A Multi-Layer Feed-forward (MLF) Artificial Neural Network (ANN) is constructed with three or more layers of neurons: input neurons, output neurons, and one or more layers of intermediate elements called hidden neurons. Neural networks have learning features that allow them to be trained to recognize patterns in data of high dimensionality. The neural networks do not require any formulation of rules about the investigated phenomenon to make decisions because they form an internal model by extracting information directly from the properly selected examples belonging to the so-called training set.

Recently there has been growing interest in the application of neural networks in the field of Quantitative Structure-Activity Relationships (QSAR),[9–16] and it has been demonstrated that this new technique is often superior to the traditional multilinear regression analysis. The key strength of the neural networks is that with the presence of hidden layers, they are able to perform nonlinear mapping of the physicochemical parameters to the corresponding biological activity. ANNs offer some advantage over standard statistical methods of modeling data since they can recognize complex relationships in the data without these having to be explicitly included in the analysis.

The goal of the present paper is to compare the performance of the Multi-Linear Regression (MLR) model and of MLF neural model in estimating the bioconcentration factors in fish for 51 organic nonelectrolytes.

RESULTS AND DISCUSSION

The BioConcentration Factor (BCF) is the concentration of a chemical compound in an organism divided by its concentration in water, and represents an important indicator for the fate of chemicals in the environment. An MLR model relating the BCF data to molecular parameters was recently reported.[17] The BCF data for 51 organic compounds were taken from the paper of Sabljić,[18] who developed a QSAR model using topological indices as independent variables. The MLR model used as molecular descriptors three solvatochromic parameters from the Linear Solvation Energy Relationship (LSER) model:[19–21] V_I, the intrinsic solute molecular volume, β, the hydrogen bond acceptor basicity parameter, and α the hydrogen bond donor acidity parameter. The structure of the 51 organic compounds, the solvatochromic parameters, and the experimental data, lg BCF, are presented in Table 1 and were used in ref. 17 to develop an MLR QSAR:

$$\lg \mathrm{BCF} = -0.894(\pm 0.271) + 4.606(\pm 1.395)V_I/100 - 3.716(\pm 1.125)\beta + 0.766(\pm 0.232)\alpha \qquad (1)$$

$$n = 51 \quad r = 0.944 \quad s = 0.439 \quad mres = 0.350$$

where n is the number of compounds used in the correlation, r is the correlation coefficient, s is the standard deviation, and *mres* is the mean residual computed with the following equation: $mres = \Sigma\,|\lg \mathrm{BCF}_{calc} - \lg \mathrm{BCF}_{exp}|\,/n$. The standard error of estimation of each coefficient at the 95% confidence level is given in parentheses. The partial correlation coefficients are $r(V_I/100) = 0.890$, $r(\beta) = -0.276$, and $r(\alpha) = -0.225$, leading to the conclusion that no single parameter can estimate with high accuracy the lg BCF values. The largest error in estimating the lg BCF values was obtained for the compound 43, 2,2′,4,4′,5,5′-PCB, with a computed lg BCF value of 5.83, and a residual equal to –0.85, which is in the range of $\pm 2s$ from the experimental value.

The collinearity of the three variables is low, as indicated by the intercorrelation matrix:

	VI/100	β	α
VI/100	1.000	0.025	–0.234
β		1.000	0.339

In order to investigate the predictive character of the MLR model, we have used a Leave-20%-Out (L20%O) cross-validation method. Cross-validation was applied by deleting 20% of the patterns from the data set, then computing the MLR model with the remaining 80% data, and finally predicting the lg BCF values for the deleted patterns. This procedure is repeated five times, until all points are deleated once and only once. The five sets of cross-validation patterns are presented in Table 2; while the first set consisted of 11 patterns, the remaining four sets of data contained each 10 different patterns. The lg BCF values predicted by the L20%O method were very close to the experimental values, as indicated by the regression equation:

$$\lg \mathrm{BCF}_{exp} = 0.245 + 0.917 \lg \mathrm{BCF}_{MLR\,cv} \qquad (2)$$

$$n = 51 \quad r = 0.920 \quad s = 0.510 \quad mres = 0.419$$

The statistical results for the MLR models developed on the basis of the five cross-validation sets and the corresponding modeling sets are presented in Table 3. While the MLR models for the sets 1, 4, and 5 give better statistics than the whole set of data, the sets 2 and 3 give lower statistical indices, but the differences are not high, and can be explained by the inhomogeneous distribution of the patterns in the five sets of data. The statistical results for the cross-validation sets are lower than that of the MLR models for sets 1, 4, and 5, while for sets 2 and 3 the correlations show better statistics.

Table 1

Structure, solvatochromic parameters, and experimental and calculated lg BCF for the set of 51 organic compound used in the QSAR models

No.	Compound	$V_i/100$	β	α	lg BCF		
					exp	NN1	residual
1	toluene	0.592	0.11	0.00	0.92	0.88	0.04
2	ethylbenzene	0.687	0.12	0.00	1.19	1.49	–0.30
3	1,2-dimethylbenzene	0.671	0.12	0.00	1.15	1.34	–0.19
4	1,3-dimethylbenzene	0.671	0.12	00.0	1.17	1.34	–0.17
5	1,4-dimethylbenzene	0.671	0.12	0.00	1.17	1.34	–0.17
6	isopropylbenzene	0.775	0.12	0.00	1.55	2.20	–0.65
7	1,2,40trimethylbenzene	0.769	0.13	0.00	2.12	2.14	–0.02
8	naphthalene	0.753	0.15	0.00	2.20	1.91	0.29
9	2-methylnaphthalene	0.851	0.16	0.00	2.61	2.43	0.18
10	phenanthrene	1.015	0.20	0.00	3.42	3.19	0.23
11	2-methylphenanthrene	1.113	0.21	0.00	3.48	3.55	–0.07
12	anthracene	1.015	0.20	0.00	3.13	3.19	–0.06
13	9-methylanhracene	1.113	0.21	0.00	3.66	3.55	0.11
14	benzo[a]anthracene	1.277	0.25	0.00	4.00	3.73	0.27
15	acenaphthene	0.896	0.17	0.00	2.60	2.60	0.00
16	pyrene	1.156	0.25	0.00	3.43	3.63	–0.20
17	benzo[a]pyrene	1.418	0.30	0.00	3.70	3.74	–0.04
18	fluorene	0.960	0.21	0.00	3.11	2.88	0.23
19	chlorobenzene	0.581	0.07	0.00	1.08	1.30	–0.22
20	1,2-dichlorobenzene	0.671	0.03	0.00	1.95	2.01	–0.06
21	1,3-dichlorobenzene	0.671	0.03	0.00	1.82	2.01	–0.19
22	1,4-dichlorobenzene	0.671	0.03	0.00	2.10	2.01	0.09
23	1,2,3-trichlorobenzene	0.761	0.00	0.00	2.69	3.03	–0.34
24	1,2,4-trichlorobenzene	0.761	0.00	0.00	3.23	3.03	0.20
25	1,3,5-trichlorobenzene	0.761	0.00	0.00	3.24	3.03	0.21
26	1,2,3,5-tetrachlorobenzene	0.851	0.00	0.00	3.50	3.53	–0.03
27	1,2,3,4-tetrachlorobenzene	0.851	0.00	0.00	3.58	3.53	0.05
28	1,2,4,5-tetrachlorobenzene	0.851	0.00	0.00	3.65	3.53	0.12
29	pentachlorobenzene	0.941	0.00	0.00	3.74	3.81	–0.07
30	hexachlorobenzene	1.031	0.00	0.00	4.23	4.19	0.04
31	2-chlorophenanthrene	1.105	0.16	0.00	3.63	3.74	–0.11
32	1,2-dichloroethane	0.442	0.10	0.00	0.30	0.49	–0.19
33	1,1,2-trichloroethylene	0.492	0.05	0.00	1.20	1.08	0.12
34	tetrachloroethylene	0.578	0.05	0.00	1.70	1.45	0.25
35	biphenyl	0.920	0.20	0.00	2.42	2.65	–0.23
36	4-chlorobiphenyl	1.010	0.17	0.00	2.77	3.16	–0.39
37	4,4'-dichlorobiphenyl	1.100	0.14	0.00	4.10	4.05	0.05
38	2,4,4'-trichlorobiphenyl	1.190	0.10	0.00	4.66	4.80	–0.14
39	2,2',5-trichlorobiphenyl	1.190	0.10	0.00	4.69	4.80	–0.11
40	2,2'4,4'-tetrachlorobiphenyl	1.280	0.06	0.00	4.85	4.88	–0.03
41	2,2'5,5'-tetrachlorobiphenyl	1.280	0.06	0.00	4.86	4.88	–0.02
42	2,2'4,5,5'-pentachlorobiphenyl	1.370	0.03	0.00	4.83	4.91	–0.08
43	2,2',4,4',5,5'-hexachlorobiphenyl	1.460	0.00	0.00	4.98	4.92	0.06
44	3-chlorophenol	0.626	0.23	0.69	1.25	1.26	–0.01
45	4-bromophenol	0.669	0.23	0.69	1.56	1.58	–0.02
46	4-nitrophenol	0.676	0.32	0.82	2.10	2.08	0.02
47	aniline	0.562	0.50	0.16	0.78	0.90	–0.12
48	N,N-diethylaniline	0.948	0.44	0.00	2.08	2.13	–0.05
49	nitrobenzene	0.631	0.30	0.00	1.18	0.80	0.38
50	1-chloro-4-nitrobenzene	0.721	0.26	0.00	1.00	1.18	–0.18
51	cyclohexane	0.598	0.00	0.00	2.22	2.15	0.07

Table 2

The composition of the five cross-validation sets used in the leave - 20% - out method. The numbers correspond to the compounds in Table 1

Set	Compound										
1	5	7	11	13	27	30	33	34	47	48	49
2	2	3	9	21	22	35	37	38	39	40	
3	8	10	15	19	23	29	31	32	36	44	
4	4	14	16	17	20	24	26	42	50	51	
5	1	6	12	18	25	28	41	43	45	46	

Table 3

Cross-validation results for the MLR model obtained for the five sets of data used in the leave - 20% - out method

Data Set	MLR Model			Cross-Validation		
	r	*s*	*mres*	*r*	*s*	*mres*
1	0.947	0.435	0.349	0.939	0.445	0.435
2	0.939	0.445	0.357	0.984	0.282	0.383
3	0.941	0.463	0.371	0.975	0.276	0.292
4	0.952	0.412	0.320	0.905	0.567	0.529
5	0.950	0.413	0.317	0.922	0.567	0.455

Also, the performance of the MLR model was tested with a set of 8 compounds not used in developing eq. (1). The chemical compounds used in the prediction set, together with their solvatochromic parameters, are presented in Table 4. The comparison of predicted and experimental values for lg BCF gives the following equation:

$$\lg BCF_{exp} = 0.185 + 0.845 \lg BCF_{MLR\ p} \tag{3}$$

$$n = 8 \quad r = 0.981 \quad s = 0.218 \quad mres = 0.221$$

Table 4

Structure, solvatochromic parameters, and experimental and calculated lg BCF for the set of 8 organic compounds used in the prediction set

Compound	$V_I/100$	β	α	lg BCG		
				exp	NN1	residual
acenaphthene	0.896	0.17	0.00	2.59	2.60	-0.01
benzene	0.491	0.10	0.00	1.10	0.56	0.54
perylene	1.415	0.30	0.00	3.86	3.74	0.12
tetrachloromethane	0.514	0.10	0.00	1.36	0.62	0.74
chloroform	0.427	0.10	0.35	0.78	0.46	0.32
pentachloroethane	0.700	0.10	0.00	1.83	1.70	0.13
1,1,2,2-tetrachloroethane	0.617	0.10	0.00	1.25	1.17	0.08
1,1,1-trichloroethane	0.519	0.10	0.00	0.95	0.64	0.31

The ANNs used in the present study are three-layer MLF networks, with three input neurons representing the three solvatochromic parameters used to develop the MLR model (V_I, β, and α), and one output unit, representing lg BCF. The training was performed with the backpropagation algorithm,[2,3] and we have used three activation functions: the hyperbolic tangent (tanh), the linear function, and a bell-shape activation function with the formula $Act(z) = 1/(1 + z^2)$.[1,22] Other important specifications for the ANNs used in simulations are: input scaling between –0.9 and 0.9, initial weights scaling

between –0.1 and 0.1, and the momentum equal to 0.8. The training continues until the correlation coefficient between experimental and computed lg BCF improves with less than 10^{-4} in 100 epochs.

To optimize the ANNs with respect to the QSAR model quality, network architecture, training regime, we have used several statistical indices: correlation coefficient r, the standard deviation s, and the mean residual *mres*, of the linear correlation between lg BCF and lg BCF of the type: $\lg BCF_{exp} = a + b \lg BCF_{ANN}$.

The number of neurons in the hidden layer was selected on the basis of empirical trials, in which ANNs with different number of hidden neurons are trained to predict the $\lg BCF_{exp}$ values. The training was done by randomly presenting to the network the patterns for the set of 51 compounds, using a hyperbolic tangent as activation function, a learning rate equal to 0.01, and output scaling between –0.9 and 0.9. For a size of the hidden layer between two and ten neurons, the correlation coefficient between experimental and estimated lg BCF was between 0.983 and 0.985, values higher than that of the MLR model.

In QSAR studies it is very important to take into consideration that feedforward neural networks are universal approximators: they are capable of arbitrarily accurate approximation to arbitrary mappings, provided sufficiently many hidden units are available.[23–25] The potential of chance correlations in QSAR models was investigated,[26–28] and it was proposed to characterize a network by a structural parameter ρ, which is the ratio of the number of patterns in the training set to the number of connections. The optimum ρ values for the MLF networks lay in the range $1.8 < \rho < 2.2$, and it was reported that networks having ρ values in excess of 2.2 failed to extract relevant features and gave poor predictions. In this paper we will use networks with four hidden neurons, whose ρ parameter, equal to 2.43, is close to the optimum value.

In order to establish the role of randomness in developing ANNs models, we have simulated a learning set of data by retaining the values of the input patterns assigned to the corresponding compounds, and randomly assigning a real lg BCF value to a compound not necessarily possessing this activity. A neural network with four hidden neurons and tanh activation functions is trained for 2000 epochs using this simulated training set. If the fit obtained with the real data is consistently better than that with the randomly assigned output data, the correlation obtained with the real data can confidently be said not to be due to chance factors. The procedure of generating simulated training sets was repeated 10 times, and the mean correlation coefficient between experimental and computed lg BCF is 0.379, while the highest correlation coefficient is 0.585. Because the statistical indices in the runs using randomly reassigned activity data are considerably lower than those obtained when actual activity data are used, we are confident that a valid relationship has indeed been established between the structural descriptors and the observed biological properties.

Because in certain cases a linear output function provides a better QSAR model, we have investigated an ANN model with tanh hidden activation function, and linear output activation function. The performance of a neural network with a linear output activation function is strongly dependent on the range of the output scaling, and an output scaling between –5 and 5 provided the best results. This network, denoted by NN1, with an output learning rate equal to 0.005, was trained for 2900 epochs, and provided a good estimation of the experimental lg BCF values:

$$\lg BCF_{exp} = -0.048 + 1.007 \lg BCF_{NN1} \qquad (4)$$

$$n = 51 \quad r = 0.990 \quad s = 0.192 \quad mres = 0.147$$

If we compare the predictions of the MLR model represented by eq. (1) with the predictions of the ANN model in eq. (3), it is clear that the neural network outperforms regression analysis and provides superior mapping of solvatochromic parameters to the BCF data. The lg BCF values estimated by the NN1 network and the residuals are presented in Table 1.

The L20%O cross-validation was performed with the same sets of data used in the MLR model. The statistical indices for the estimated and predicted lg BCF values in the five sets of data are reported in Table 5. The estimations in the five runs have correlation coefficients higher than 0.990,

Table 5

Cross-validation results for the neural network model obtained for the five sets of data used in the leave – 20% – out method

Data Set	MLR		Model Cross-Validation			
	r	*s*	*mres*	*r*	*s*	*mres*
1	0.992	0.165	0.120	0.920	0.506	0.314
2	0.990	0.180	0.142	0.987	0.253	0.225
3	0.994	0.145	0.115	0.922	0.478	0.414
4	0.991	0.179	0.134	0.963	0.360	0.257
5	0.994	0.147	0.113	0.917	0.584	0.341

while the cross-validation r, r_{cv}, is lower, mainly in the case of sets 1, 3, and 5. When we have examined the predicted values, we discovered that an excessively low r_{cv} was usually associated with a small number of points with very large residuals, and that these occurred in points in which one or more input and/or output patterns were extremal (with values outside the range of the corresponding values in the training set). In the first set, the compound 47 is extremal in β, and has a very poor predicted value, 2.02, and the largest residual in this cross-validation set, –1.24. In the third set, the compound 32 is extremal both in V_I and lg BCF, and as expected, has a predicted value equal to 1.14 and the largest residual in its prediction set, –0.84. Finally, in the fifth set compound 46 is extremal in α, with a predicted value of 0.50 and a residual equal to 1.60, again the largest in this set. Extremal points are absent from the two cross-validation sets with a high r_{cv}, the second and the fourth. Thus, although the only technique for assessing the quality of the fit is cross-validation, it leads to very pessimistic values for r_{cv} when the data in the cross-validation set are extremal in one or more variables, because it involves extrapolation of the scaling. The large residuals for extremal points, which have been consistently observed in our computations, are a warning that the ANN data should not be extrapolated, and also that the r_{cv} is unduly pessimistic in this cases.

The L20%O cross-valiation method gives the following correlation between predicted and experimental lg BCF values:

$$\lg BCF_{exp} = 0.078 + 0.961 \lg BCF_{NN\,cv} \tag{5}$$

$$n = 51 \quad r = 0.938 \quad s = 0.450 \quad mres = 0.310$$

The statistical indices of eq. (5) are higher than of the corresponding MLR cross-validation equation, but much lower than of the estimations of the neural model. As explained above, this situation appears because the cross-validation sets contain extremal data.

The second cross-validation method used was the leave-one-out (LOO) technique. In this approach, an untrained network is first created. Then one pattern is taken out of the training set of patterns and the network is trained with the remaining patterns. When the learning process is finished, the network predicts the output value for the pattern which was eliminated from the learning set. The pattern is then put back in the set and the next one is taken out to repeat the process, starting with the untrained network. In this manner, each pattern serves as an unknown once and as a training pattern all the other times. The predictions of the ANN model using the LOO method are well correlated with the experimental values:

$$\lg BCF_{exp} = 0.034 + 0.967 \lg BCF_{NN\,LOO} \tag{6}$$

$$n = 51 \quad r = 0.954 \quad s = 0.390 \quad mres = 0.280$$

While the predictions in the LOO method are better than that of the L20%O method, one must take into account that the LOO method requires the computation of 51 networks with a training set consisting of 50 patterns. Because the LOO method is computationally more complex than the L20%O method, one can apply it only on small sets of data. Also, the largest prediction errors in the LOO method are provided by the extremal patterns: compound 8, with lg BCF_{pred} = 1.38 and a residual of 0.82; compound 32, with lg BCF_{pred} = 1.10 and a residual of –0.80;compound 47, with lg BCF_{pred} = 2.03 and a residual of –1.25; and compound 48, with lg BCF_{pred} = 3.08 and a residual of –1.00.

The predictive performance of the NN1 network was tested with the same set of 8 compounds from Table 4 used to evaluate the MLR model. The ANN prediction of lg BCF values is of equal statistical quality with that of the MLR model:

$$\lg BCF_{exp} = 0.476 + 0.862 \lg BCF_{NN\,p} \tag{7}$$

$$n = 8 \quad r = 0.983 \quad s = 0.209 \quad mres = 0.280$$

The predicted lg BCF values and the residuals for the compounds in the prediction set are presented in Table 4.

Because there is no theoretical way to establish the optimal topology of a neural network, we have investigated a large number of networks with different characteristics. Some selected results are presented, in order to show that similar results can be obtained with different ANN topology. A neural network, denoted by NN2, with tanh activation function for both hidden and output layers, a learning rate equal to 0.01, and an output scaling between –0.9 and 0.9, was trained for 3500 epochs and provided the following estimation of lg BCF:

$$\lg BCF_{exp} = -0.020 + 1.003 \lg BCF_{NN2} \tag{8}$$

$$n = 51 \quad r = 0.987 \quad s = 0.212 \quad mres = 0.163$$

Because in cases where the dependence between inputs and outputs is highly nonlinear a bell-shape activation function for the hidden layer gives better performances, we have trained the network NN3, with the following specifications: hidden activation function $Act(z) = 1/(1 + z^2)$; tanh output function; output scaling between –0.9 and 0.9. After 7000 epochs, the estimations were of the same quality as those obtained with the network NN2:

$$\lg BCF_{exp} = -0.030 + 1.003 \lg BCF_{NN3} \tag{9}$$

$$n = 51 \quad r = 0.987 \quad s = 0.208 \quad mres = 0.164$$

In some cases, using a gain factor γ improves the performances of the network. We have used a hidden layer activation function $Act(z) = \tanh(\gamma z)$, with gain values different for the four hidden neurons: 2, 1, 0.5, and 0.1. The convergence was obtained after 5100 epochs, with no improvement in the estimation of the lg BCF values:

$$\lg BCF_{exp} = -0.018 + 1.004 \lg BCF_{NN4} \tag{10}$$

$$n = 51 \quad r = 0.986 \quad s = 0.214 \quad mres = 0.164$$

The last experiment reported here used a generalized neural network, with a tanh output function, a hidden neuron with a bell-shape activation function, and three hidden neurons with tanh func-

tions and γ values equal to 1, 0.5, and 0.1. The learning phase was finished after 3400 epochs, providing a good estimation of the experimental data, but with no advantage over the other networks:

$$\lg BCF_{exp} = -0.019 + 1.006 \lg BCF_{NN5} \quad (11)$$

$$n = 51 \quad r = 0.985 \quad s = 0.223 \quad mres = 0.179$$

CONCLUSIONS

The neural network approach gives better estimations than the usual MLR model for the computation of the bioconcentration factors in fish, using as structural descriptors solvatochromic parameters. The predictions of the two models are of the same statistical quality.

Also, it was demonstrated that the usual cross-validation procedure gives too pessimistic values for r_{cv}, because the calculation of the output values for patterns which are extremal in one or more variables involves extrapolation, and neural networks give poor results in these cases. The leave-one-out cross-validation method gives also large residuals for extremal patterns. The use of leave-one-out analysis with large data sets cannot be considered routine, however, as it is extremely demanding of computer time.

The random reassignment of the dependent variable in the training set is a method which can demonstrate the relevance of the neural model to the problem under study, and its use is recommended at least with small samples.

ACKNOWLEDGEMENT. Financial support was obtained from the Ministry of Research and Technology under Grant 310 TA10 and from the Ministry of the National Education under Grant 7001 T34.

REFERENCES

1. Part 6: O. Ivanciuc, *Rev. Roum. Chim.* **1995**, *40*, 1093–1101.
2. D. E. Rumelhart and J. L. McClelland, "Parallel Distributed Processing", MIT Press, Cambridge, MA, 1986.
3. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Nature*, **1986**, *323*, 533–536.
4. P. D. Wasserman, "Neural Computing", Van Nostrand Reinhold, New York, 1989.
5. J. Zupan and J. Gasteiger, *Anal. Chim. Acta*, **1991**, *248*, 1–30.
6. B. J. Wythoff, *Chemom. Intell. Lab. Syst.*, **1993**, *18*, 115–155.
7. J. Gasteiger and J. Zupan, *Angew. Chem. Int. Ed. Engl.*, **1993**, *32*, 503–527.
8. J. R. M. Smits, W. J. Melssen, L. M. C. Buydens, and G. Kateman, *Chemom. Intell. Lab. Syst.*, **1994**, *22*, 165–189.
9. V. S. Rose, I. F. Croall and H. J. H. MacFie, *Quant. Struct.-Act. Relat.*, **1991**, *10*, 6–15.
10. N. Ghoshal, S. N. Mukhopadhyay, T. K. Ghoshal and B. Achari, *BioMed. Chem. Lett.*, **1993**, *3*, 329–332.
11. D. Zakarya, L. Farhaoui and S. Fkih-Tetouani, *Tetrahedron Lett.*, **1994**, *35*, 4985–4988.
12. M. A. Sofan, A. E.-S. Abdel-Megied, M. B. Pedersen, E. B. Pedersen and C. Nielsen, *Synthesis*, **1994**, 516–520.
13. M. Wiese and K.-J. Schaper, *SAR QSAR Environ. Res.*, **1993**, *1*, 137–152.
14. J. Devillers, *SAR QSAR Environ. Res.*, **1993**, *1*, 161-167.
15. M. Brinn, P. T. Walsh, M. P. Payne and B. Bott, *SAR QSAR Environ. Res.*, **1993**, *1*, 169–210.
16. O. Ivanciuc, *Rev. Roum. Chim.*, **1996**, *41*, 645–652.
17. J. H. Park and E.H. Cho, *Bull. Korean Chem. Soc.*, **1993**, *14*, 457–461.
18. A. Sabljić, *Z. Gesamte Hyg.*, **1987**, *33*, 493–496.
19. M. J. Kamlet, R. M. Doherty, J.-L. M. Abboud, M. H. Abraham, and R. W. Taft, *CHEMTECH*, **1986**, *16*, 566–576.
20. M. J. Kamlet, R. M. Doherty, M. H. Abraham, Y. Marcus and R. W. Taft, *J. Phys. Chem.*, **1988**, *92*, 5244–5255.
21. M. H. Abraham, *Chem. Soc. Rev.*, **1993**, 73–83.
22. O. Ivanciuc, J.-P. Rabine, D. Cabrol-Bass, A. Panaye and J. P. Doucet, *J. Chem. Inf. Comput Sci.*, **1996**, *36*.
23. G. Cybenko, *Math. Control Signals Syst.*, **1989**, *2*, 303–314.
24. K. Funahashi, *Neural Networks*, **1989**, *2*, 183–192.
25. K. Hornik, M. Stinchcombe and H. White, *Neural Networks*, **1989**, *2*, 359–366.
26. T.A. Andrea and H. Kalayeh, *J. Med. Chem.*, **1991**, *34*, 2824–2836.
27. D. T. Manallack and D. J. Livingstone, *Med. Chem. Res.*, **1992**, *2*, 181–190.
28. D. J. Livingstone and D. T. Manallack, *J. Med. Chem.*, **1993**, *36*, 1295–1297.