ELSEVIER

Dyes and Pigments, Vol. 34, No. 3, pp. 181–193, 1997

PII: S0143-7208(96)00081-2 0143-7208/97 $17.00+0.00

Multiple Linear Regression (MLR) and Neural Network (NN) Calculations of some Disazo Dye Adsorption on Cellulose[1]

Simona Timofei,[a]* Ludovic Kurunczi,[a] Takahiro Suzuki,[b] Walter M. F. Fabian[c] & Sorel Mureşan[d]

[a]Institute of Chemistry, Romanian Academy, Bul. Mihai Viteazul 24, 1900, Timisoara, Romania
[b]Research Laboratory of Resources Utilization, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, 226 Japan
[c]Institut für Organische Chemie, Karl-Franzens-Universität Graz, Heinrichstrasse 28, A-8010, Graz, Austria
[d]Politehnica University of Timisoara, Faculty of Industrial Chemistry, Piaţa Victoriei Nr. 2, 1900, Timisoara, Romania

(Received 15 July 1996; accepted 16 August 1996)

ABSTRACT

Multiple Linear Regression (MLR) analysis and Neural Network (NN) calculations are applied to a series of 21 disazo anionic dyes. Three-dimensional QSAR parameters were derived from the Cartesian coordinates of the dye molecules. Low energy conformations were obtained by molecular mechanics and quantum chemical calculations. Electronic and steric effects in the dye–cellulose binding are present. The proposed MLR models are rough approximations of nonlinear models. Good correlation with the dye affinity from the MLR calculations and a significantly improved fitting of the NN over the MLR models are observed. The model validity was checked for two proposed models derived from different sets of structural parameters by the leave-one-out cross-validation procedure. For the first model, a better validity ('cross-validated r^2*' value, of 0.622) of the NN model is noticed by leaving out one compound (found as outlier) from the training set, in comparison to that of the MLR model obtained for the same set of compounds (*$q^2 = 0.434$*). The* q^2 *value of a second MLR proposed model is better than that of the corresponding NN model.* © *1997 Elsevier Science Ltd*

Keywords: Multiple Linear Regression (MLR) analysis, Neural Networks (NNs), dye adsorption, cellulose fibre.

[1]Presented in part at the Symposium on Computational Chemistry, held on 16–17 May 1996 in Tokyo, Japan

*Corresponding author.

INTRODUCTION

The dyeing of textile fibres led to many theories, which tried to explain the mechanism of dye uptake. The known models describe generally limiting cases, which reflect partially the practical processes [1].

In many studies reported in the literature, dye molecular fragments were related qualitatively to the adsorption on fibres. Intermolecular dye–cellulose interactions similar to some extent to the ligand–biological receptor ones, are also reported.

Classical and three-dimensional QSAR techniques have been applied to model quantitatively dye–structure affinity relationships for many types of dyes for cellulose fibre [2–5]. In a previous CoMFA study [6] an analysis and prediction of the effect of structural modifications of some disazo direct dyes have been performed.

In this paper statistical results obtained by a classical QSAR and neural network study are compared for a series of disazo direct dyes [7] to find possible effects in the dye–cellulose fibre binding.

METHODS

Definition of parameters used

The dye affinity for cellulose fibre (A) is used similarly to the biological activity from QSAR calculations, as dependent variable.

The length of the conjugated chain of the dye molecule (n) is considered to determine a possible physical attraction between the π electronic system of the dye molecule and the fibre [8].

The van der Waals volumes (VW) are calculated by the additivity of van der Waals volume increments [9]. For the sulphonic and azo groups, increment values are taken from the literature [10].

Quantum chemical parameters, i.e. the dipole moment (μ), the HOMO molecular orbital energy (E_{HOMO}) and the LUMO molecular orbital energy (E_{LUMO}) were obtained by MOPAC [11]. Calculations were performed with the semiempirical PM3 method [12]. A molecular correction for the amide group was applied (keyword MMOK).

As steric parameter, a molecular shape 'metameter' (λ_f, $f = 1, 2, 3$), derived from Cartesian coordinates of the dye molecule, is used [13,14]. The product of these 'lambda values' represents the 'internal scattering' of the molecule.

Other steric parameters derived from Cartesian coordinates are the ellipsoidal parameters (a, b, c) [15a,b]. These parameters are synthetic indicators of shape and represent the semiaxes of triaxial ellipsoid that embeds the

molecule in the 'hard spheres' approximation. The dye molecules were aligned by the MULTIFIT option from the SYBYL package [6], along the azo groups. In the present calculations the 'a' ellipsoidal parameter corresponds to the longest molecule axis. These parameters were calculated by a program developed by Ciubotariu [15a].

Hydrophobicity effects are generally expressed by the octanol/water partition coefficients (logP). Peters [16] suggests that a dye molecule with a large hydrophobic area could be adsorbed on the hydrophobic crystalline cellulose regions, even if the cellulose fibre is considered with many hydrophilic regions. LogP values were calculated by the CHEMICALC-2 program [17]. Experimental dye affinity values and the structural parameters of disazo dyes are collected in Table 1.

Multiple linear regression (MLR) analysis

In this method one experimental variable y_k or dependent variable (in our case the dye affinity) is correlated with one or several structural variables x_i, by the equation [18]:

$$y_k = b_0 + \sum ib_i \cdot x_{ik} + e_k \tag{1}$$

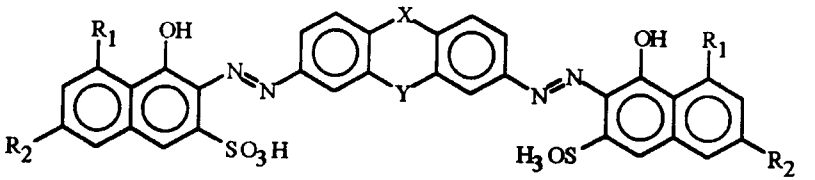
where b represents partial regression coefficients and e_k the deviations and residuals. The MLR calculations were performed by the MASCA package [19].

The neural network (NN) approach

The neural networks (NNs) have an inherent ability to provide non-linear and cross product terms for QSAR modelling. The NNs may be especially useful when a rigid theoretical basis and/or mathematical relationship to describe a phenomenon to be modeled is not available in advance.

From many NN approaches, both different in architecture and in learning algorithms, the three-layer NNs with the back-propagation of errors were employed in this study. Since the theory and practical application of the NN are popular, an explanation of the methodology can be delegated to the literature [20]. The most commonly used logsigmoid transfer function and the delta rule for the error correction formula were used in the networks. The NN calculations were carried out by the program written in the BASIC language [21].

TABLE 1
Dye Affinities (A) and Structural Parameters of Disazo Direct Dyes



No.	R_1	R_2	X	Y	A (*kJ/mole*)	λ_1	λ_2	λ_3	n	V_W (*cm³/mole*)	E_{HOMO} (*eV*)	*logP*	*a* (*Å*)	*b* (*Å*)	*c* (*Å*)
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
1	NH_2	SO_3H	—	none	24.60	1.88	1.02	0.10	23	382.12	−8.81	0.29	14.61	9.52	3.76
2	NH_2	SO_3H	—	CONH	24.10	1.93	0.97	0.11	23	373.88	−8.79	1.42	14.63	9.82	4.21
3	NH_2	SO_3H	—	CO	23.70	1.86	0.98	0.16	22	366.62	−8.85	0.73	15.17	8.67	4.06
4	NH_2	SO_3H	NHCONH	none	23.70	2.53	0.41	0.06	26	409.98	−8.70	0.60	10.00	13.87	6.65
5	NH_2	SO_3H	—	NH	23.60	2.37	0.44	0.19	22	363.00	−8.78	0.99	12.94	11.45	4.45
6	NH_2	SO_3H	CONH	none	21.00	2.64	0.29	0.07	25	401.90	−8.73	0.61	14.94	10.89	5.95
7	NH_2	SO_3H	NH	—	19.00	1.66	0.95	0.32	21	363.00	−8.66	0.66	12.15	9.01	3.90
8	NH_2	SO_3H	NH	none	18.90	2.44	0.46	0.09	24	390.2	−8.67	1.00	13.76	7.10	8.15
9	NH_2	SO_3H	CO	none	18.30	1.81	1.09	0.10	24	393.82	−8.83	0.99	14.42	6.40	6.99
10	NH_2	SO_3H	none	CONH	17.60	1.61	0.83	0.56	23	401.90	−8.68	−0.11	9.26	10.42	4.82
11	NH_2	SO_3H	none	NHCONH	16.40	2.20	0.62	0.18	24	409.98	−8.81	0.95	8.14	9.33	9.83
12	NH_2	SO_3H	none	CO	13.90	2.21	0.57	0.22	22	393.82	−8.83	0.99	8.31	10.79	8.16
13	H	NH_2	—	none	30.40	1.93	0.95	0.13	25	382.96	−8.63	2.25	13.46	9.23	3.15
14	H	NH_2	—	CONH	28.90	1.86	0.91	0.24	25	340.50	−8.61	1.71	13.65	9.27	3.97
15	H	NH_2	NHCONH	none	27.40	2.53	0.41	0.06	28	356.82	−8.52	1.24	9.23	12.67	6.21
16	H	NH_2	—	CO	27.10	1.74	1.11	0.15	24	332.42	−8.66	0.75	14.22	8.34	3.65
17	H	NH_2	CONH	none	27.10	2.66	0.28	0.07	27	348.74	−8.53	0.90	14.07	9.53	6.17
18	H	NH_2	none	NHCONH	22.90	1.56	1.00	0.44	26	356.82	−8.66	1.24	7.95	8.65	9.22
19	H	NH_2	none	CO	20.40	2.41	0.50	0.08	24	340.66	−8.69	1.28	10.48	11.68	5.54
20	H	NH_2	none	CONH	22.40	1.61	0.87	0.52	25	348.74	−8.51	0.90	8.72	9.99	4.78
21	H	NH_2	CO	none	21.00	2.43	0.49	0.08	26	340.66	−8.68	1.15	9.03	12.03	5.69

RESULTS

Regression analysis results

Molecular mechanics calculations were performed for the 21 disazo anionic dyes by the SYBYL package [6,22]. Many isomers were obtained; their geometries were optimized by MOPAC analysis [11]. In the present calculations only conformers of lowest energy were considered.

For each dye molecule, the matrix of (N,3) Cartesian coordinates (where N represents the number of atoms and 3 corresponds to the 3 spatial Cartesian coordinates x, y, z) was employed. Each molecule is considered a cluster of N atoms in the 3D Euclidean space. Eigenvalues are estimated from the (3,3) correlation matrix. As molecular surface 'metameter', the vector-valued observation $\lambda = (\lambda_1\ \lambda_2\ \lambda_3)$T is used [13]. PCA (Principal Component Analysis) calculations were performed by the MASCA package [19].

In a first attempt, correlations with dye affinity were performed for compounds having the same coupling component. The following starting variables were used: V_W, μ, E_{HOMO}, E_{LUMO} for compounds **1–12** (see Table 1). By a variable selection procedure based on the T-square test used as local criterion and the largest root criterion as global criterion, the following variables were chosen: μ, E_{HOMO} and E_{LUMO}. In accordance to the test statistics, which was performed at a significance level of 1%, the proposed model was not robust.

MLR analysis was applied to the series of 21 disazo dyes, using the following structural variables: λ_1, λ_2, λ_3, n, V_W, μ, EHOMO, ELUMO and logP. By variable selection based on the T-square criterion [14], used as test statistics (TS) at a significance level of 1% or less as local criterion and the largest root criterion (at a significance level of 5% or less) as global criterion, the following equation was obtained:

$$\hat{A} = 287.24 + 8.21(\pm 7.88)\lambda_2 - 17.59(\pm 14.87)\lambda_3 + 30.75(\pm 20.42)E_{HOMO}$$
$$N = 21\ r = 0.768\ s = 2.99\ TS_1 = 3.02\ TS_2 = 3.43\ TS_3 = 4.36\ r^2_{CV} = 0.397 \quad (2)$$

The outliers were tested by the externally Studentized residuals; high-leverage points were examined by the main diagonal of the so-called hat matrix and the influential points by the likelihood function distance criterion [19]. Neither outliers, high leverage points nor influential points were observed. Equation (2) expresses a robust model, without collinearity or multicollinearity. The validity of the proposed model was checked by the

'leave-one-out' (LOO) cross-validation procedure [23], to have an idea of the predictive ability of the equation.

Another model was developed from the following starting variables: λ_1, λ_2, λ_3, n, V_W, μ, E_{HOMO}, E_{LUMO}, logP, a, b and c. By a variable selection procedure based on the largest root criterion [14], used as test statistics (TS) at a significance level of 1% or less as local criterion and the largest root criterion (at a significance level of 5% or less) as global criterion, the following equation was obtained:

$$\hat{A} = 231.44 + 25.22(\pm 18.55)E_{HOMO} + 0.87(\pm 0.72)a$$
$$r = 0.721\ s = 3.13\ TS_1 = 3.1\ TS_2 = 3.25 \tag{3}$$

The 'leave-one-out' cross-validation procedure gave $q^2 = 0.383$. Neither outliers, high leverage points nor influential points were observed.

Neural Network results

The architecture of the feed-forward model used in this study is shown in Fig. 1 for the model derived from equation (2). The network consists of three layers: an input layer, a hidden layer and an output layer. Each layer is comprised of individual processing units called neurons and represented here as circles. As described above, in the MLR analysis three descriptors in the first model [see equation (2)], respectively, two descriptors in the second model [see equation (3)] were found to be effective for the given set of 21 disazo dyes. The same descriptors were implemented as inputs for the NN to compare the results with those obtained by MLR analysis. Since the number of hidden neurons strongly influences the predictive quality of the derived model [24], the number of hidden neurons were set to be variable. Both input and hidden layers have an additional neuron, termed a bias neuron, respectively.

The criteria used for the comparison with the MLR analysis are the correlation coefficient (r), the 'cross-validated r^2' (q^2) value and the root mean-square error, RMSE, defined by:

$$\mathrm{RMSE} = \sum (A_k - \hat{A}_k)^2 / N \tag{4}$$

where A_k is the kth training or experimental value of the dye affinity, $\hat{A}_k$ is the kth output or calculated value of the dye affinity, and N is the number of compounds, respectively.

The normalization of the descriptor values (input scaling) was performed according to the following equation:

$$x_i' = 0.9\frac{x_i - x_{min}}{x_{max} - x_{min}} + 0.05 \quad (5)$$

where x_i is the ith input value, x_{min} and x_{max} are its minimum and maximum data over the data set. The same transformation was applied to the training values (output scaling).

The range of initial values of connection weights was set to random values between 0.5 and +0.5. Besides the number of hidden neurons (NH), the performance of NNs depends on other model parameters like the learning rate η, the momentum factor μ', the starting configuration, and the number of iteration cycles (epochs) [25–27]. The NH, the initial values of η and μ', were varied from 1 to 3, 0.1 to 0.8, and 0.2 to 0.95, respectively, in both NN calculations. After 1000 epochs, both the values of η and μ' were gradually decreased. Training continued up to 10 000 epochs. After extensive tests, the best values for model parameters, NH, initial values of η and μ', were found to be 2, 0.7 and 0.9, respectively.

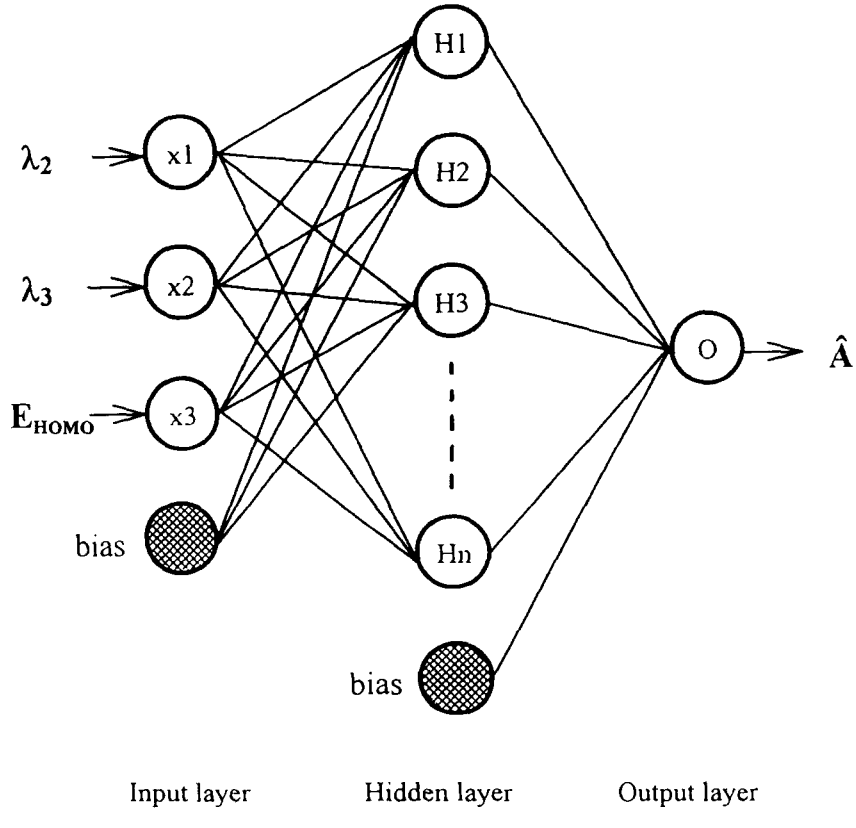


Fig. 1. Architecture of the three-layer neural network used in this study.

DISCUSSION

The model proposed by equation (2) reflects a satisfactory correlation with the dye affinity. The LOO cross-validation procedure gave the q^2 value of 0.397. Even if equation (2) is robust from a statistical point of view, the physical meaning of the three-dimensional vector-valued observation (derived from Cartesian coordinates of the dye molecules), reduced by the omisssion of the first eigenvalue, can be questionable. An insight into the problem of the 'vector-valued observation' (λ) could be the topic of further investigations. From equation (3) steric and electronic effects could be expected in dye–fibre interactions. The 'a' ellipsoidal molecule axis from the final model expresses more appropiate the dye–fibre binding along the longest dye molecule axis, as stated previously [2,3,28], in comparison to the length of the conjugated chain of the dye molecule. Disazo dyes would be expected to act as donors in the dye–cellulose interactions. Equations (2) and (3) have similar statistical results, but only the last one can give an interpretation of dye–fibre interactions. The calculated octanol/water coefficients indicate low lipophilicity of the dye molecules. The correlation coefficient between the dye affinity and logP values (of 0.460) and, also, the absence of the hydrophobicity parameter in the final MLR proposed models suggest the absence of hydrophobic effects in dye–cellulose binding.

The results obtained for the training set of 21 compounds by the NN (NN1, NN2) are shown in Table 2 together with those results fitted by the MLR. Underlined value indicates the analysis giving the more accurate calculation. The NN1 model was calculated more accurate for 14 of the 21 data. The quality of the fitting, RMSE value of 2.12 and r^2 of 0.743 with the NN model, was found to be better than that obtained with the previous first MLR model, of 2.70 and 0.590, respectively. However, from the results of 'leave-one-out' cross-validation, there was significant difference between the goodness of fit with the training set and the prediction power in a cross-validation set. The 'cross-validated r^2' score was much less than that of MLR. The performance is probably due to overfitting or inclusion of outliers.

When the data set is fairly consistent and the model is not biased by any particular data point, the 'jackknifed r' values do not indicate any unduly high variation. As can be seen from the values in Table 2, only compound **3** appears to be an outlier causing a considerably higher 'jackknifed r' of 0.940, while all the others are reasonably consistent (around 0.86). In order to avoid masking or swamping by the outlier, compound **3** was omitted from the training data set for NN. The fitted affinities by the resulting NN model (NN2 model) are shown in Table 2. The RMSE value decreased to 1.46 in this new NN model. The 'jackknifed r' values and the predicted dye affinities by the LOO cross-validation procedure are shown in Table 2. The NN2

TABLE 2
Calculated Dye Affinities by Multiple Linear Regression (MLR) and Neural Networks (NN1 and NN2)

No.	$\hat{A}_1$			$\hat{A}_2$		*Cross validation*	
	MLR	*NN1*	*jackknifed* r^2	*NN2*	*jackknifed* r^3	$\hat{A}_3$	$\hat{A}_4$
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1	22.95	21.80	0.877	24.06	0.938	21.72	23.26
2	22.98	21.83	0.872	24.47	0.939	22.11	24.88
3	20.33	20.83	**0.940**	—	—	—	—
4	22.03	21.46	0.871	21.40	0.948	21.79	21.10
5	17.53	20.43	0.903	20.90	0.949	15.86	20.22
6	19.94	21.48	0.862	21.39	0.939	19.69	21.38
7	23.12	22.10	0.866	21.66	0.949	23.44	23.59
8	22.83	21.69	0.868	21.60	0.947	24.01	21.95
9	22.91	21.95	0.896	18.34	0.934	23.02	20.16
10	17.29	17.62	0.851	17.63	0.935	16.98	21.05
11	18.26	20.17	0.882	17.01	0.937	18.16	19.74
12	16.53	14.07	0.816	13.03	0.924	16.84	13.03
13	27.38	28.70	0.836	28.72	0.933	25.57	28.12
14	25.73	27.49	0.846	27.93	0.932	25.26	25.36
15	27.56	28.64	0.854	28.66	0.938	28.21	29.11
16	27.42	28.73	0.858	28.73	0.940	27.14	29.41
17	26.01	26.98	0.851	26.71	0.938	26.25	22.00
18	21.42	21.32	0.866	21.23	0.943	28.87	20.90
19	22.72	21.50	0.862	21.61	0.941	22.92	21.78
20	23.55	22.28	0.864	21.33	0.941	24.79	21.17
21	22.95	21.57	0.862	21.71	0.940	23.14	21.81
RMSE	2.70	2.12		1.46		3.22	2.63
r^2	0.590	0.743		**0.884**			
q^2						0.434	0.622

model gives the 'cross-validated r^2' value of 0.622, which is better than that of the reduced MLR model (without compound **3**): $q^2 = 0.434$. Figure 2 illustrates the relationship between the r^2 value and the number of training epochs for the data set of the 20 compounds. Both r^2 values reached their maximum at 10000 epochs. The ratio of training examples to adjustable connection weights is 1.82 and this value is in the recommended range of 1.8–2.2 [29]. Thus the weights at this point were saved as the best values for this NN model. To make it possible to use the NN model developed in this paper, Table 3 represents the parameters for the NN2 model.

The relative dependence of each descriptor in affecting the dye affinities in the NN model was assessed. The variations of predicted affinities of the NN model were monitored by changing the value of one descriptor while the other two descriptors are constant at 0.25 of their normalized scales. The results are presented in Fig. 3. The plots show that the dye affinity has a strong nonlinear dependence on the three descriptors. However, it is

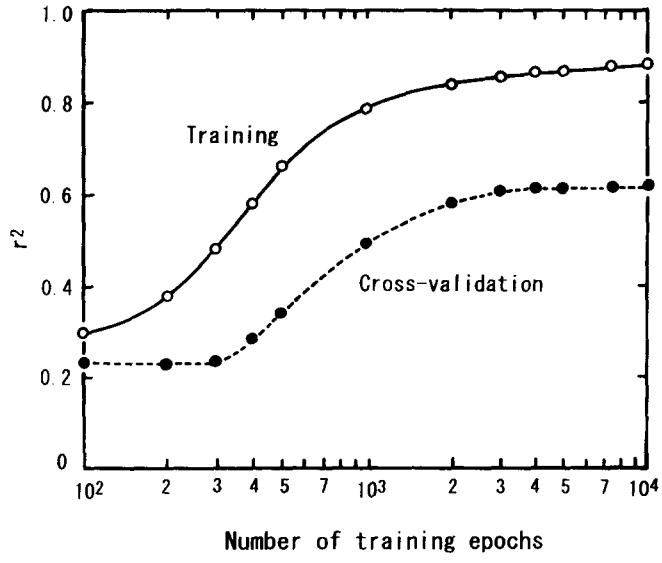


Fig. 2. Effects of the number of training epochs on the performance of the neutral network model (NN2) used for 20 compounds (compound **3** was left out).

TABLE 3
Parameters of the Neural Network Model 2 (NN2)[a] for Dye Affinities

	Connection weights between the inputs and hidden layers			
	x_1	x_2	x_3	*Bias*
H_1	−18.148	24.165	−18.610	15.380
H_2	2.139	13.632	−29.108	−2.094
	Connection weights between the hidden and input layers			
	H_1	H_2	Bias	
O	−1.990	−9.993	1.809	

[a]Compound 3 was removed from the training data set.

interesting to note that the positive slope of the λ_2 and E_{HOMO}, and the negative slope of λ_3 in equation (2) are consistent with this NN model. Therefore, it can be concluded that the former MLR model is a rough approximation of a non-linear model. The calculation steps based on this model are shown in the Appendix.

Similar results were obtained for the NN3 model where the descriptors from equation (3) were implemented as inputs. The NN3 model presents a better fitting in comparison to the MLR model described by equation (3) (r^2 value of 0.52, respectively, 0.71). The q^2 and RMSE values derived from the LOO cross-validation procedure of the last MLR model (0.383, respectively, 3.43) were better than those of the corresponding NN3 model

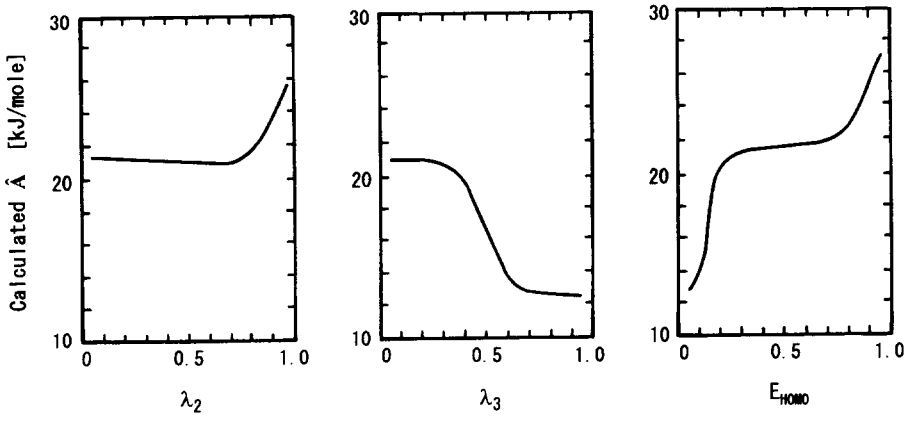


Fig. 3. Neural network calculated dye affinities as a function of individual descriptors λ_2, λ_3 and E_{HOMO}. In each graph other normalized values of descriptors were held constant at 0.25.

($q^2 = 0.13$; RMSE = 2.27). The relative dependence of each descriptor in affecting the dye affinities showed a strong non-linear dependence on the two descriptors.

From these results, one can see that the improved fitting of the NN model over the MLR model is significant. This improvement is probably due to the NNs ability to use interactions among descriptors as well as non-linearities.

CONCLUSION

Correlations with dye affinity were performed by MLR analysis for a series of 21 disazo anionic dyes. The final model proposed by equation (3) suggests the presence of electronic and steric effects in the dye binding to cellulose and the donor ability of dye molecules.

The chosen descriptors from MLR calculations were implemented as inputs in the three-layer NNs with the back-propagation of errors. A better fitting of the NN over the MLR model was noticed. The NN1 model validity was improved by leaving out one compound from the training set (found as outlier). The q^2 value for the NN2 model was better than that of the MLR model proposed for the same set of dyes. The q^2 values of the second MLR model was better than that of the NN3 model. The former MLR proposed models are rough approximations of nonlinear models.

The MLR approach leads to a better interpretation of the contribution of individual terms, but NNs can extract more 'information' from the data than statistical methods, especially where non-linear relationships are involved.

ACKNOWLEDGEMENTS

The authors are indebted to Dr Peter P. Mager from the University of Leipzig (Germany), for the permission of access to the MASCA package and beneficial suggestions and to Professor Dr Z. Simon from the University of Medicine and Pharmacy of Timisoara (Romania) for useful discussions.

REFERENCES

1. Flath, H. J., *Melliand Textilberichte*, **2** (1991) 132.
2. Timofei, S., Schmidt, W., Kurunczi, L., Simon, Z. and Salló, A., *Dyes and Pigments*, **24** (1994) 267.
3. Fabian, W. M. F., Timofei, S. and Kurunczi, L., *Journal of Molecular Structure (Theochem.)*, **340** (1995) 73.
4. Timofei, S., Kurunczi, L., Schmidt, W. and Simon, Z., *Quantitative Structure–Activity Relationships*, **14** (1995) 444.
5. Timofei, S., Kurunczi, L., Schmidt, W. and Simon, Z., *Dyes and Pigments*, **29** (1995) 251.
6. Fabian, W. M. F. and Timofei, S., *Journal of Molecular Structures (Theochem.)*, **362** (1996) 155.
7. Szadowski, J., *Dyes and Pigments*, **14** (1990) 217.
8. Giles, C. H. and Hassan, A. S. A., *Journal of the Society of Dyers and Colourists*, **74** (1958) 846.
9. Bondi, A., *Journal of Physics and Chemistry*, **68** (1964) 441.
10. Shibusawa, T., *Sen'i Gakkaishi*, **43** (1987) 401.
11. *MOPAC 6.0*: J. J. Stewart, QCPE program no. 455
12. Stewart, J. J. P., *Journal of Computational Chemistry*, **10** (1989) 209–221.
13. Mager, P. P., *European Journal of Medicinal Chemistry*, **29** (1994) 369.
14. Mager, P. P., In *QSAR in Design of Bioactive Compounds*. Proc. 2nd Telesymp. Med. Chem., M. Kuchar (ed.). J. R. Prous, Barcelona, 1992, pp. 446–469.
15. a. Ciubotariu, D., In *Modeling of Cancer Genesis and Prevention*, Voiculetz, N., Balaban, A. T., Niculescu-Duvăz, I and Simon, Z. (eds). CRC Press, Boca Raton, 1991, pp. 176–188; b. Ciubotariu, D., Deretey, E., Medeleanu, M., Mureçan, S., Bologa, C., Novac, A. and Simon, Z., In *Trends in QSAR and Molecular Modelling* '92, Wermuth, G. (ed.). ESCOM, Leiden, The Netherlands, 1993, pp. 358–359.
16. Peters, R. H., In *Textile chemistry. The Physical Chemistry of Dyeing*. Elsevier, Amsterdam, 1975.
17. *CHEMICALC-2*, Version 1.0, T. Suzuki, QCPE program no. 608
18. Wold, S. and Dunn III, W. J., *Journal of Chemical Information and Computer Sciences*, **23** (1983) 6.
19. Mager, P. P., Rothe, H., Mager H. and Werner, H., In *QSAR in Design of Bioactive Compounds*, J. R. Prous, South Africa, 1992, pp. 131–182.
20. Zupan, J. and Gasteiger, J., *Neural Networks for Chemists: An Introduction*. VCH, Weinheim, 1993.
21. Suzuki, T. and Ishida, M., *Fire and Materials*, **19** (1995) 179.
22. SYBYL 6.0, *Tripos Associates*, St. Louis, MO, USA

23. Franke, R., *Theoretical Drug Design Methods*. Akademie, Berlin, 1984.
24. Livingstone, D. J. and Salt, D. W., *Bioorganic and Medicinal Chemistry Letters*, **3** (1992) 213.
25. Bodor, N., Harget, A. and Huang, M-J., *Journal of the American Chemical Society*, **113** (1991) 9480.
26. Egolf, L. M. and Jurs, P. C., *Journal of Chemical Information and Computer Sciences*, **33** (1993) 616.
27. Schüürmann, G. and Müller, E., *Environmental Toxicological and Chemistry*, **13** (1994) 743.
28. Timofei, S., Ph.D. thesis, Western University of Timisoara, Romania, 1995.
29. Andrea, T. A. and Kalayeh, H., *Journal of Medicinal Chemistry*, **34** (1991) 2824.

APPENDIX

Calculation of the dye affinity of compound 1

1. Normalization of input data: by using equation (5) the above three descriptors are rescaled to the following values: $x1 = 0.8524$, $x2 = 0.1220$ and $x3 = 0.1559$
2. Calculation of outputs from the hidden layers:

$$\begin{aligned} H_1 &= 1/\left[1 + e^{-(x_1 \cdot w_{11} + x_2 \cdot w_{21} + x_3 \cdot w_{31} + w_{41})}\right] \\ &= 1/\left[1 + e^{-\{(0.8524)(-18.148)+(0.1220)(24.165)+(0.1559)(-18.610)+15.380\}}\right] \\ &= 0.4894 \quad H_2 = 0.0413 \end{aligned}$$

3. Calculation of outputs from the output layer:

$$\begin{aligned} 0 &= 1/\left[1 + e^{-(H_1 \cdot \nu_1 + H_2 \cdot \nu_2 + \nu_3)}\right] \\ &= 1/\left[1 + e^{-\{(0.9424)(-1.990)+(0.0002)(-9.993)+1.809\}}\right] \\ &= 0.604 \end{aligned}$$

4. By unscaling the above value, the dye affinity is calculated to be 24.06 kJ/mole.