CROATICA CHEMICA ACTA
CCACAA **79** (2) 227–236 (2006)
ISSN-0011-1643
CCA-3086
Original Scientific Paper

Quantitative Structure-Retention Relationships (QSRR) for Chromatographic Separation of Disazo and Trisazo 4,4'-Diaminobenzanilide-based Dyes

Simona Funar-Timofei,[a] Walter M. F. Fabian,[b] Georgeta M. Simu,[a] and Takahiro Suzuki[c,*]

[a]*'Coriolan Dragulescu' Institute of Chemistry, Romanian Academy, Bul. Mihai Viteazul 24, 300223 Timisoara, Romania*

[b]*Institut für Chemie (IfC), Karl-Franzens Universität Graz, Heinrichstr. 28, A-8010 Graz, Austria*

[c]*Natural Science Laboratory, Toyo University, 2-11-10 Hakusan, Bunkyo-ku, Tokyo112-8606, Japan*

RECEIVED MARCH 23, 2005; REVISED OCTOBER 14, 2005; ACCEPTED OCTOBER 17, 2005

For a series of 23 disazo and trisazo 4,4'-diaminobenzanilide-based direct dye molecules, the chromatographic mobilities, extrapolated to modifier-free conditions (R_{M0} values), were determined from reverse-phase thin-layer chromatography (RP-TLC) experiments. Traditional and rational QSAR/QSPR modelling techniques have been applied to find a quantitative structure-retention relationship (QSRR) for the dyes. Molecular dye structures were energy minimized by both molecular mechanics and quantum chemical calculations. A variety of 1D to 3D molecular descriptors taking into account size, shape, symmetry, electronic structure, atom or group distribution, and hydrophobicity of the dyes was derived from the optimized three-dimensional geometries. Multiple linear regression (MLR) and artificial neural network (ANN) modelling revealed that the R_{M0} values can be successfully expressed by a combination of hydrophobic and polarity dye structural parameters. Additional comparative molecular field (CoMFA) and similarity index (CoMSIA) analyses suggested almost equal contribution of both steric and electrostatic fields to the chromatographic mobility and the major role of hydrophobic and electrostatic interactions with the chromatographic environment.

Key words
• direct dyes
• reverse-phase thin-layer chromatography
• quantitative structure-retention relationships
• artificial neural network
• comparative molecular field analysis
• comparative molecular similarity index analysis

INTRODUCTION

The importance of the concept of lipophilicity in the research of pharmacological and toxicological potency of many drugs has been recognized since a century.[1] It can affect the affinity of dyestuffs for the textile substrates,[2,3] dye-dye aggregation[4–6] and dye-surfactant complexation.[7,8] Furthermore, toxicity can correlate with lipophilicity in many cases.[9] The most popular quantitative scale to measure the lipophilicity of compounds is the logarithm of the partition coefficient (called the log *P* parameter) between 1-octanol and water, introduced by Hansch and Leo.[10] For dye molecules, the experimental difficulties in the reproducibility of log *P*, on one hand, and the observations of linear relationship between log *P* and chromatographic retention parameters, on the other hand, suggested that the former could be substituted by the available chromatographic data. Sangster[11] pointed out the difficulties in estimating the partition coefficients of dye molecules by fragmental methods.

* Author to whom correspondence should be addressed. (E-mail: suzuki@toyonet.toyo.ac.jp)

The chromatographic R_M values were successfully related to log *P* by Pliška *et. al.*[12] Boyce and Millborow[13] suggested also the use of the chromatographic R_M value in order to avoid the practical difficulties that often arise in the direct determination of the partition coefficient (because of the low solubility of the solute in the 1-octanol/water system and the presence of impurities). Therefore, reverse-phase high performance liquid chromatography (HPLC) and reverse-phase thin-layer chromatography (RP-TLC) can be considered as important alternatives to the 1-octanol/water partition.[12,14] Chromatographic mobility derived from RP-TLC has been extensively used as a measure of the lipophilic character, though to a lesser extent for dyestuffs or in textile chemistry.[15]

The determination of lipophilicity by means of the RP-TLC technique is mainly based on a simple linear regression model between the chromatographic R_M values and the organic solvent concentrations in the mobile phase.[16] The R_M value is described by the following linear equation:

$$R_M = R_{M0} + bC \quad (1)$$

where *C* represents the organic phase concentration *v*/*v* and the value of *b*, the slope of the regression line, characterizes the molecular lipophilicity. The value of *b* has been regarded as a characteristic of the specific hydrophobic surface area of the compounds.[17] The value of R_{M0} (*i.e.*, that extrapolated to modifier-free conditions), the intercept of the line, can be obtained from the extrapolation for 0 % organic solvent. The R_{M0} value better approximates the experimental conditions of lipid-phase/water partitioning systems in comparison to the R_M values.[12]

Over the past decade, the quantitative structure-activity/property relationships (QSAR/QSPR) have become a powerful theoretical tool for the description and prediction of molecular systems in chromatographic research.[18–21] Numerical techniques[22] were applied for rational classification and selection of TLC systems of some flavonoids and phenolic acids. Principal component analysis (PCA), principal component regression analysis (PCRA) and multiple linear regression (MLR), have been used to chromatographic R_M values of some arylamides of *o*-hydroxycarboxylic acid of the Naphthol AS type[23] and R_F chromatographic data of non-benzidinic disazo dyes[24] obtained by reverse-phase thin-layer chromatography (RP-TLC).

The aim of this study was to explore the chromatographic mobilities of a series of disazo and trisazo direct dyes with 4,4'-diaminobenzanilide as diazo component experimentally and the QSAR/QSPR modelling techniques were applied to the data to identify the significant molecular properties contributing to the retention of the dyes.

EXPERIMENTAL

Synthesis of Dyes

The investigated dyestuffs were disazo and trisazo direct dyes containing 4,4'-diaminobenzanilide as diazo component, of general formula (I) (see Table I). All dyes were synthesized and purified as described previously (dyes no. **2**, **14**, **15** in Ref. 25; dye no. **5** in Ref. 26; dye no. **6** in Ref. 27; dye no. **9** in Ref. 28; dye no. **11** in Ref. 29; dye no. **12** in Ref. 30).

Measurement of R_F *values*

The measurement of the chromatographic R_F values was performed by RP-TLC analysis. All reagents used for the PR-TLC analysis were of analytical grade and were supplied by Reactivul Bucuresti (Bucharest, Romania). The dyes were dissolved in methanol to form 0.1 % solutions. The dye solutions were spotted onto 20 cm x 20 cm Merck (Darmstadt, Germany) silica gel TLC plates (of DC Fertigplatten Kieselgel 60 type, with 0.25 mm layer thickness). The plates were pre-developed during 24 hours in a 95:5 (*v*:*v*) hexane : paraffin oil mixture. The elution was carried out in a 6 cm × 23 cm × 23 cm tank (of Camag type). The eluting solvent, which is imposed to be a binary system, was isopropanol : NH_3 25 %, at 60–85 % ratios (*v*/*v*), which varied with an increment of 5% of organic phase. The R_F values were calculated as the ratio of the migration distance of the dye to that of the solvent front (which was of 11 cm from the origin).

To ascertain the reproducibility of the experimental chromatographic mobilities, the dyes were each spotted ten times onto the RP-TLC plates. Average R_F values were obtained by successive experiments. Chromatographic R_M (respectively R_{M0}) values were calculated from the experimental R_F values.[12] The resulting chromatographic R_{M0} data are listed in the forth column in Table I, and the data were used as dependent variable for QSPR modelling.

COMPUTATIONAL METHODS

Molecular Descriptors

Molecular dye structures were constructed by the Sybyl molecular modelling package[34] and energy minimized using the Tripos force field[35] with Gasteiger-Hückel charges.[36] Conformational analysis was performed by a grid search of all the individual rotatable bonds (step size = 30 degrees) using the Tripos force field with Gasteiger–Hückel charges. Each conformation found in this way for the respective dye molecule was then used for energy minimization by the semiempirical AM1 Hamiltonian[37] (using AMPAC software, as implemented in Sybyl).

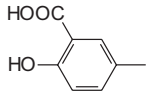
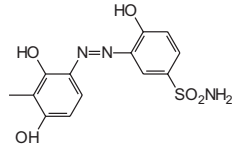
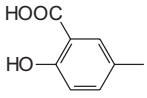
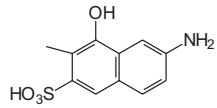
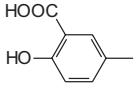
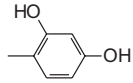
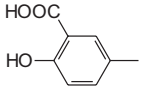
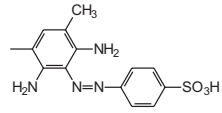
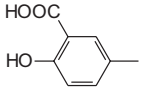
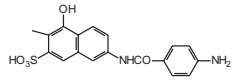
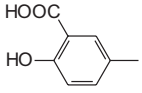
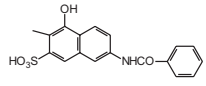
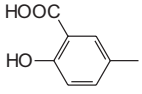
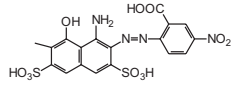
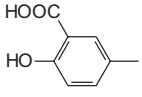
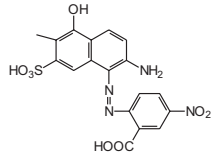
Several structural descriptors for the investigated dyes were calculated from these optimized geometries. They are presented in Table II.

Multiple Linear Regression (MLR)

Multiple linear regression (MLR)[45] calculations were performed by the STATISTICA package.[46] The leave-one-out[47] cross validation procedure was applied in order to verify the reliability of our results.

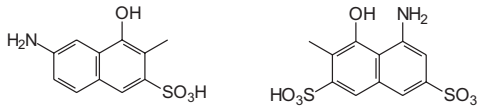
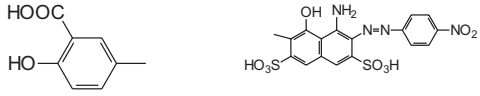
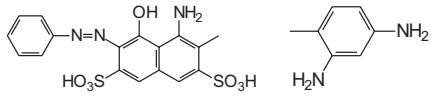
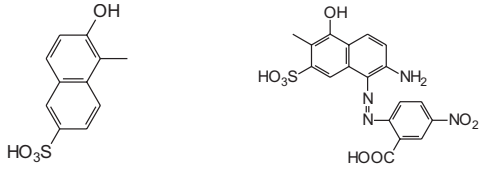
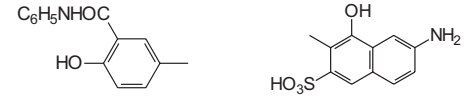
TABLE I. Chromatographic R_{M0} values and structural descriptors of the direct azo dyes (**I**) included in the final MLR and ANN models[a]

I

No	C_1	C_2	R_{M0}	log P	*n*COOHPh	Hy
1			–5.61	–1.21	1	1.42
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)ccc(N=Nc5cc(ccc5O)S(=O)(=O)N)c4O[b]					
2			–5.33	1.71	1	1.45
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5cc(N)ccc5cc4S(=O)(=O)O[b]					
3			–5.37	1.18	1	0.84
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4ccc(O)cc4O[b]					
4			–4.75	1.77	1	2.09
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4cc(C)c(N)c(N=Nc5ccc(cc5)S(=O)(=O)O)c4N[b]					
5			–5.31	2.30	1	1.35
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5ccc(NC(=O)c6ccc(N)cc6)cc5cc4S(=O)(=O)O[b]					
6			–5.45	3.59	1	0.15
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5ccc(NC(=O)c6ccccc6)cc5cc4S(=O)(=O)O[b]					
7			–6.63	0.95	2	1.36
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5c(cc4S(=O)(=O)O)cc(c(N=Nc6ccc(cc6C(=O) O)N(=O)=O)c5N)S(=O)(=O)O[b]					
8			–5.29	1.54	2	1.35
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5ccc(N)c(N=Nc6ccc(cc6C(=O)O)N(=O)=O)c5c c4S(=O)(=O)(O) [b]					

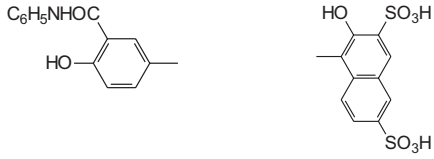
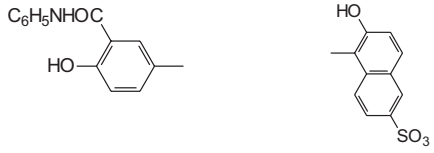
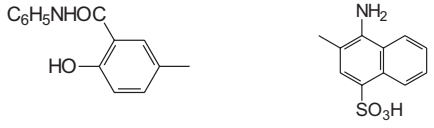
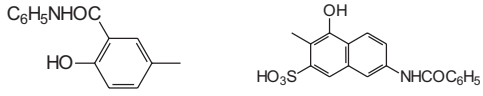
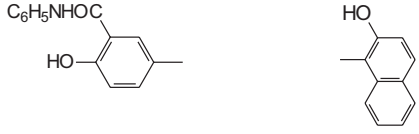
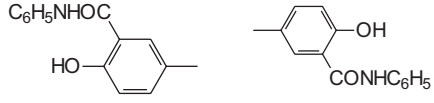
continued

TABLE I. (continued)

9		–4.05	2.60	2	0.22
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4ccc(O)c(c4)C(=O)O[b]				
10		–6.60	0.64	0	2.80
	c1(O)c2cc(N)ccc2cc(c1N=Nc3ccc(cc3)C(=O)Nc4ccc(cc4)N =Nc5c(O)c6c(N)cc(cc6cc5S(=O)(=O)O)S(=O)(=O)O)S(=O)(=O)O[b]				
11		–4.01	1.37	1	1.37
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5c(cc4S(=O)(=O)O)cc(c(N=Nc6ccc(cc6)N(=O) =O)c5N)S(=O)(=O)(O) [b]				
12		–4.24	0.06	0	3.51
	c1(N)c(N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N=Nc4ccc(N)cc4N) c(cc5c1c(O)c(N=Nc6ccccc6)c(c5)S(=O)(=O)O)S(=O)(=O)O[b]				
13		–4.19	1.93	1	1.31
	c1cc(O)c(N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5ccc(N)c(N=Nc6ccc(cc6C(=O)O)N(=O)=O)c5cc4S (=O)(=O)O)c7c1cc(cc7)S(=O)(=O)O[b]				
14		–3.35	1.71	1	1.45
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5ccc(N)cc5cc4S(=O)(=O)O[b]				
15		–4.30	3.59	1	0.14
	c1(O)ccc(cc1C(=O)O)N=Nc2ccc(cc2)C(=O)Nc3ccc(cc3)N =Nc4c(O)c5cc(ccc5cc4S(=O)(=O)O)NC(=O)c6ccccc6[b]				
16		–2.69	2.69	0	1.36
	c1(O)ccc(cc1C(=O)Nc2ccccc2)N=Nc3ccc(cc3)C(=O)Nc4ccc (cc4)N=Nc5c(O)c6c(N)cc(cc6cc5S(=O)(=O)O)S(=O)(=O)O[b]				
17		–2.53	2.88	0	1.34
	c1(O)ccc(cc1C(=O)Nc2ccccc2)N=Nc3ccc(cc3)C(=O) Nc4ccc(cc4)N=Nc5c(O)c6cc(N)ccc6cc5S(=O)(=O)O[b]				

continued

TABLE I. (continued)

18		–2.49	3.99	0	0.17
	c1(O)ccc(cc1C(=O)Nc2ccccc2)N=Nc3ccc(cc3)C(=O)Nc4ccc(cc4)N=Nc5c(O)c(cc6cc(ccc56)S(=O)(=O)O)S(=O)(=O)O[(b)]				
19		–2.06	4.17	0	0.13
	c1(O)ccc(cc1C(=O)Nc2ccccc2)N=Nc3ccc(cc3)C(=O)Nc4ccc(cc4)N=Nc5c(O)ccc6cc(ccc56)S(=O)(=O)O[(b)]				
20		–2.00	4.09	0	0.70
	c1(O)ccc(cc1C(=O)Nc2ccccc2)N=Nc3ccc(cc3)C(=O)Nc4ccc(cc4)N=Nc5cc(c6ccccc6c5N)S(=O)(=O)O[(b)]				
21		–2.14	4.76	0	0.09
	c1(O)ccc(cc1C(=O)Nc2ccccc2)N=Nc3ccc(cc3)C(=O)Nc4ccc(cc4)N=Nc5c(O)c6ccc(NC(=O)c7ccccc7)cc6cc5S(=O)(=O)O[(b)]				
22		–2.85	4.35	0	0.09
	c1(O)ccc(cc1C(=O)Nc2ccccc2)N=Nc3ccc(cc3)C(=O)Nc4ccc(cc4)N=Nc5c(O)ccc6ccccc56[(b)]				
23		–1.66	4.94	0	0.08
	c1(O)ccc(cc1C(=O)Nc2ccccc2)N=Nc3ccc(cc3)C(=O)Nc4ccc(cc4)N=Nc5ccc(O)c(c5)C(=O)Nc6ccccc6[(b)]				

[(a)] log *P* – logarithm of the 1-octanol/water coefficient, calculated by the Chemicalc-2 software;[31] *n*COOHPh – number of aromatic carboxylic acids; Hy-hydrophilic factor.[32]

[(b)] SMILES notation.[33]

Nonlinear Modelling by Artificial Neural Networks (ANNs)

The theory and general practice of artificial neural networks (ANNs) and their applications in chemistry and drug design have been reviewed in depth.[48] A fully connected three-layer neural networks was used to relate the R_{M0} values to a selected set of molecular descriptors used for MLR (vide infra). The hidden layer contained variable nodes, and the input and hidden variables each had a bias neuron. A sigmoid transfer function was used for each neuron, and connection weights were adjusted iteratively by back-propagation using the generalized delta rule to minimize mean square errors between desired and actual outputs.

Input and output data were normalized by dividing the twice of their maximum values, and models were evaluated on the basis of square of correlation coefficient r^2, cross-validated (leave-one-out) q^2(LOO) and root-mean-square error (RMS). The calculation was performed using a commercial program (Sanuk Co., ltd., Tokyo, Japan) on a microcomputer running Windows XP as operating system.

TABLE II. Structural descriptors of direct azo dyes

Descriptor	Property description	Reference
SAS	Connolly Solvent Accessible Surface Area	38
MS	Connolly Molecular Surface Area	39
SEV	Connolly Solvent-Excluded Volume	39
(*Q*MAX*Q*MIN)	maximum and minimum atom charge in molecule	40, 41
Q+	atomic formal charge on the most positive hydrogen atom	
MAXMIN	the difference between the maximum and minimum atomic charge in a molecule	40, 41
E_{HOMO}, E_{LUMO}	HOMO and LUMO molecular orbital energies, respectively	
μ	dipole moment	
log *P*	logarithm of 1-octanol/water partition coefficient	31, 42, 43, 44
log *S*	logarithm of aqueous solubility	31, 43
1D to 3D Dragon descriptors		32
Hy	hydrophilic factor	32
*n*COOHPh	number of aromatic carboxylic acids	32
MW	molecular weight	
MR	molar refraction	39

Comparative Molecular Field (CoMFA) and Similarity Index (CoMSIA) Analysis

For the comparative molecular field (CoMFA) and similarity index (CoMSIA) analysis,[49–52] besides the conformations of lowest energy (series **A**), three additional series (**B**–**D**) of the dye structures were used, because for the individual dye molecules, the minimum energy conformations considerably differed. These additional series were chosen on the basis of an energy criterion (only those within ± 1.5 kcal mol^{-1} of the lowest energy structure) and optimal superposition of the common azo substructure **I** (Table I), taking into account the same orientation of the azo groups.

Alignment of the individual molecules was accomplished with the aid of the "database alignment" tool provided by Sybyl using the azo moiety **I** (Table I) as a common substructure. In each of the four series three different template structures, including those with the largest and smallest, respectively, R_{M0} – values (**10**, R_{M0} = – 6.60; **23**, R_{M0} = – 1.66, and **1**, R_{M0} = – 5.61), were used. In CoMFA both steric and electrostatic field (based on the Mulliken charges provided by the semiempirical AM1 calculations) were included; in CoMSIA additionally the hydrophobic field was considered. The standard settings implemented in Sybyl were used throughout. Statistical analysis was done using partial least squares[53] with the leave-one-out crossvalidation.[47] The number of components used in the PLS analysis was varied in the range 2–5, and additionally 10 components were also tested. According to the parsimony principle, if q^2 did not change significantly, less than the proposed optimum number of components were used in the final PLS analyses without crossvalidation.[52]

Model Calibration and Validation

The calibration quality of MLR, NN and PLS models have been characterized using the squared correlation coefficient

$$r^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{SS}} \tag{2}$$

and standard error

$$\mathrm{SE} = \sqrt{\frac{\mathrm{RSS}}{n-m-1}} \tag{3}$$

with SS (sum of squares) and RSS (residual sum of squares) being defined as

$$\mathrm{SS} = \sum_{i=1}^{n} (y_i - y_0)^2 \tag{4}$$

and

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - y_i^{\mathrm{calc}})^2 \tag{5}$$

where n is the number of compounds and m is the number of optimized (independent in MLR) variables, y_i the i-th experimental value (in our case: chromatographic mobility – R_{M0}), y_0 the respective mean and y_i^{calc} the i-th calculated value using an MLR or PLS model calibrated with the total set of compounds.

As a rough measure of the prediction capability, the predictive squared correlation coefficient

$$q^2 = 1 - \frac{\mathrm{PRESS}}{\mathrm{SS}} \tag{6}$$

with

$$\mathrm{PRESS} = \sum_{i=1}^{n} (y_i - y_i^{\mathrm{calc}})^2 \tag{7}$$

(predictive residual sum of squares) based on the leave-one-out scheme were used. y_i^{calc} denotes the i-th predicted value of the i-th sub-model calibrated without the i-th experimental value. Whilst the leave-one-out approach is not conservative but tends to increasingly overestimate the prediction power of regression models with increasing number of compounds,[53] it is still a reasonable method for smaller data sets.

TABLE III. The one- and two-descriptor MLR models selected from 113 descriptors[a]

No.	Equation	n	r^2	SE	F	q^2(LOO)	Outlier
(8)	R_{M0} = –5.76(±0.36) + 0.75(±0.13) log P	22	0.638	0.935	35.538	0.556	No. **6**
(9)	R_{M0} = –4.22(±0.39) + 0.46(±0.10) log P – 1.05(±0.23)*n*COOHPh	21	0.823	0.639	41.59	0.761	Nos. **6**, **10**
(10)	R_{M0} = –2.28(±0.31) + 0.51(±0.21) Hy – 1.55(±0.25)*n*COOHPh	22	0.714	0.811	23.73	0.617	No. **10**

[a] n – number of compounds, r – correlation coefficient, SE – standard error, F-test, q^2(LOO) – crossvalidated leave-one-out correlation coefficients.

RESULTS AND DISCUSSION

Modelling by MLR and ANN

Starting from the entire data set of 23 compounds, several subsets of descriptors were considered based on the intercorrelations between them. Variable selection by a stepwise regression procedure based on the Fischer test was performed for each subset. All statistical tests were performed at a significance level of 5 % or less. Outliers were detected by estimating the standard residuals, as implemented in the Statistica software.[46]

From a variety of potential MLR models with various combinations of descriptors, the statistically significant MLR models obtained by the leave-one-out procedure containing one or two descriptors were generated. Details about the three final MLR models and the values of their statistical parameters (correlation coefficient r, standard error SE, F-test, and crossvalidated correlation coefficients q^2(LOO)) are given in Table III.

Best statistical results were obtained by Eq. (9) with 2D molecular descriptors, log P and *n*COOHPh. The plot of the calculated versus experimental R_{M0} of the 21 dye molecules using the two-descriptor model is shown in Figure 1. None of the 3D descriptors were significant. Two dyes, no. **6** and **10**, were identified as outliers. The statistical test for outliers was based on standard residual and actually the MLR models including these compounds were not stable statistically. Although the reason is not known clearly at the present stage, we will have to consider the reliability of the predicted log P values. Deviations from linearity might be expected with the MLR models, indicating that there might be room for improvement through inclusion of quadratic terms or of other molecular descriptors. Thus, improved statistical results were obtained in the equation employing log P^2 as quadratic term (n = 22 – compound **6** found as outlier was omitted – r^2 = 0.657, SE = 0.913, F = 38.24, q^2(LOO) = 0.598) and in equation including the log P value together with the cross-product between log P and *n*COOHPh (n = 22 – compound **10** found as outlier was omitted – r^2 = 0.734, SE = 0.784, F = 26.17, q^2(LOO) = 0.611).

Since there is a suggestive colinearity between log P and Hy (r = –0.76), Eq. (10) can be considered as an alternative model for Eq. (9), indicating the hydrophilic factor has an important contribution to the dye lipophilicity. The dye lipophilicity can be expressed by structural parameters indicative of both the hydrophobic nature and the bulkiness of the dye molecules. The number of aromatic carboxylic acids also influences the dye lipophilicity.

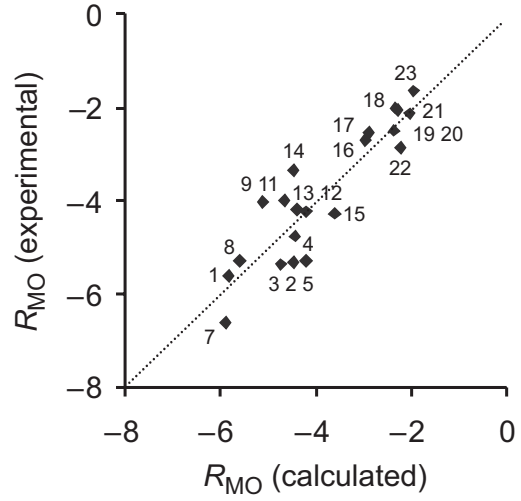


Figure 1. Scatter plot of the experimental *versus* calculated R_{M0} values for 21 dyes by the MLR model described by Eq. (9). The numbers show dye molecules in Table I.

Since there might be a strong nonlinear component in the relationship between R_{M0} and the two descriptors in Eq. (9), QSRR modelling was also performed using ANNs to get the ANN counterpart to Eq. (9). The architecture of the alternative ANN model was (2 inputs + 1 bias) : (3 hidden-layer nodes + 1 bias): (1 output). The relevant statistical parameters of the best ANN model were as follows: n = 21 (outliers, no. **6** and **10**), r^2 = 0.829, q^2(LOO) = 0.625, RMS = 0.579. This nonlinear model had a slightly-improved value of r^2 compared to Eq. (9), but the q^2 score, was just comparable to the model achieved using linear regression.

Modelling by CoMFA and CoMSIA

CoMFA and CoMSIA were performed for the entire set of compounds with improved statistical results, in terms of crossvalidated q^2, conventional r^2, standard errors SE, and F-test. They are completely independent on the tem-

Table IV. CoMFA and CoMSIA results

CoMSIA	Comp	q^2 (LOO)	SE	r^2	F
Steric	4	0.472	0.374	0.953	68.875
Electrostatic	5	0.387	0.450	0.932	46.677
Both	5	0.453	0.370	0.954	70.751
Hydrophobic	4	0.522	0.472	0.921	52.347
Hydrophobic + steric	4	0.545	0.406	0.941	72.416
Hydrophobic + electrostatic	7 (5)	0.552 (0.506)	0.250 (0.349)	0.981 (0.959)	113.245 (79.666)
Hydrophobic + steric + electrostatic	7 (5)	0.541 (0.509)	0.244 (0.344)	0.982 (0.960)	118.947 (82.337)
CoMFA					
Steric	4	0.523	0.305	0.967	132.030
Electrostatic	5	0.423	0.232	0.982	184.584
Both	4	0.450	0.292	0.970	144.417

plate (**1**, **10** or **23**) used for alignment. However, they significantly differ for the four series of conformations of the individual compounds: for series **A**, *i.e.*, using the minimum energy conformation for each dye **1**–**23**, both CoMFA and CoMSIA gave essentially no correlation at all with chromatographic R_{M0} – values. In view of the rather different minimum energy structures of the individual compounds, this result is not unexpected. Similarly, series **D** also gave rather unsatisfactory results, whereas both series **B** and **C** resulted in statistically quite significant models. CoMFA and CoMSIA results for series **C** (crossvalidated correlation coefficient resulting from the leave-one-out procedure q^2(LOO), conventional correlation coefficient r^2, standard errors SE, F-test) obtained using steric, electrostatic, and in case of CoMSIA, also hydrophobic fields and combinations thereof, are summarized in Table IV (dye **1** as template).

TABLE V. Contribution (normalized coefficient and fraction) of the individual CoMSIA and CoMFA fields

CoMSIA		Normalized coefficient	Fraction
Steric + electrostatic	steric	1.065	0.306
	electrostatic	2.414	0.694
Steric + hydrophobic	steric	1.045	0.363
	hydrophobic	1.832	0.637
Electrostatic + hydrophobic[(a)]	electrostatic	1.788	0.555
	hydrophobic	1.436	0.445
Steric + electrostatic + hydrophobic[(a)]	steric	0.588	0.181
	electrostatic	1.499	0.462
	hydrophobic	1.158	0.357
CoMFA			
Steric + electrostatic	steric	1.692	0.494
	electrostatic	1.735	0.506

[(a)] number of components = 5

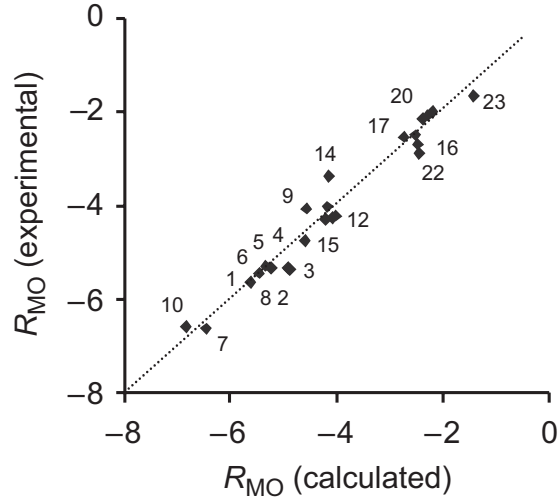


Figure 2. Plot of experimental chromatographic *versus* calculated R_{M0} mobilities (CoMSIA, steric + electrostatic + hydrophobic fields, 5 components) for 23 dye molecules. The numbers show dye molecules in Table I.

Contributions (normalized coefficients and fractions) of the individual fields (steric, electrostatic, and in case of CoMSIA, also hydrophobic) are provided in Table V. In the framework of the CoMFA procedure, both steric and electrostatic fields contribute almost equally to the chromatographic mobility; the CoMSIA approach indicates greater importance of the electrostatic interactions. Moreover, with CoMSIA, which also allows the description of hydrophobic interactions, their importance clearly is evident (Table V). A plot of calculated *vs*. experimental R_{M0} – values obtained by the CoMSIA 3D-QSAR method including steric, electrostatic, and hydrophobic fields, is shown in Figure 2.

CONCLUSIONS

A quantitative structure-retention relationship (QSRR) study of chromatographic mobilities of a series of disazo and trisazo direct dyes with 4,4'-diaminobenzanilide as diazo component has been presented. Experimental chromatographic R_F values were measured by reverse-phase thin-layer chromatography (RP-TLC). Chromatographic R_M (respectively R_{M0}) values were calculated from the R_F values and were used in further analysis.

Multiple linear regression (MLR) and nonlinear modelling using artificial neural networks (ANN) was applied to model the chromatographic R_{M0} mobilities. The results indicated that the dye lipophilicity can be explained by hydrophobic and polarity dye structural parameters. Good statistical results for calibrated models with the experimental R_{M0} values were obtained by comparative molecular field (CoMFA) analysis and comparative molecular similarity index (CoMSIA) approach, but less robust models as compared to the MLR and NN ones, as seen from the q^2(LOO) cross-validation values. While in CoMFA almost equal contribution of both steric and electrostatic fields to the chromatographic mobility was observed, the CoMSIA approach indicates greater importance of the electrostatic and hydrophobic interactions.

REFERENCES

1. A. Nasal, D. Siluk, and R. Kaliszan, *Cur. Med. Chem.* **10** (2003) 381–426.
2. W. Biedermann, *Rev. Prog. Col.* **10** (1979) 1–10.
3. K. Greider, *J. Soc. Dyers Colour.* **92** (1976) 8–13.
4. D. G. Duff, D. J. Kirkwood, and D. M. Stevenson, *J. Soc. Dyers Colour.* **93** (1977) 303–306.
5. H. Zollinger, *J. Soc. Dyers Colour.* **81** (1965) 345–350.
6. T. Imae, C. Mori, and A. Ikeda, *J. Chem. Soc., Faraday Trans. 2* **78** (1982) 1359–1368.
7. W. Biedermann and A. Datyner, *J. Colloid Interface Sci.* **82** (1981) 276–285.
8. D. M. Stevenson, D. G. Duff, and D. J. Kirkwood, *J. Soc. Dyers Colour.* **97** (1981) 13–17.
9. H. van de Waterbeemd and B. Testa, *The Parametrization of Lipophilicity and other Structural Properties in Drug Design*, in: B. Testa (Ed.), *Advances in Drug Research* Vol. 16, Academic, London, 1987, pp. 85–225.
10. C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley-Interscience, New York, 1979.
11. J. Sangster, *Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry, Wiley Series in Solution Chemistry*, Vol. 2, John Wiley & Sons, Chichester, 1997, p.133.
12. V. Pliška, B. Testa, and H. van De Waterbeemd (Eds.), *Lipophilicity in Drug Action and Toxicology*, VCH, Weinheim, 1996.
13. C. B. C. Boyce and B. V. Millborow, *Nature* **208** (1965) 537–539.
14. P. Mukerjee and A. K. Ghosh, *J. Am. Chem. Soc.* **92** (1970) 6403–6407.
15. E. Tomlinson, *J. Chromatogr.* **113** (1975) 1–45.
16. G. L. Biagi, A. M. Barbaro, A. Sapone, and M. Recanatini, *J. Chromatogr. A.* **662** (1994) 341–361.
17. C. Horvath, W. Melander, and I. Molnar, *J. Chromatogr.* **125** (1976) 129–156.
18. W. Guo, Y. Lu, and X. M. Zheng, *Talanta* **51** (2000) 479–488.
19. R. Kaliszan, *J. Chromatogr. B* **715** (1998) 229–244.
20. R. Kaliszan, M. A. van Straten, M. Markuszewski, C. A. Cramers, and H. A. Claessens, *J. Chromatogr. A* **855** (1999) 455–486.
21. D. G. Duff, R. W. Horobin, and G. B. Proctor, *Dyes Pigm.* **6** (1985) 61–73.
22. M. Medić-Šarić, I. Jasprica, A. Smolčić-Bubalo, and A. Mornar, *Croat. Chem. Acta* **77** (2004) 361–366.
23. S. Funar-Timofei, E. Sarandan, A. Salló, F. Elenes, and E. Crasmareanu, *Rev. Chim.-Bucharest* **10** (2003) 802–806.
24. G. M. Simu, S. Funar-Timofei, and E. Sarandan, *Rev. Chim.-Bucharest* **7** (2003) 603–607.
25. G. Simu, S. Hora, M. Grad, and E. Sisu, *Rev. Chim.-Bucharest* **11** (2004) 873–349.
26. G. Simu, S. Funar-Timofei, L. Kurunczi, and W. Schmidt, *Rev. Roum. Chim.* **49** (2004) 345–349.
27. G. Simu, S. L. Funar-Timofei, S. G. Hora, W. E. Schmidt, L. Kurunczi, E. N. V. Sisu, and N. Morin, *Rev. Chim.-Bucharest* **53** (2002) 826–829.
28. G. M. Simu, S. A. Chicu, N. Morin, W. Schmidt, and E. Sisu, *Turk. J. Chem.* **28** (2004) 579–585.
29. G. Simu, S. Funar-Timofei, S. Hora, and L. Kurunczi, *Mol. Cryst. Liq. Cryst.* **416** (2004) 97–104.
30. G. Simu, S. Funar-Timofei, L. Kurunczi, S. Hora, W. Schmidt, and M. Grad, *Cellulose Chem. Technol.* **38** (2004) 409–415.
31. CHEMICALC-2, version 1.0., T. Suzuki, QCPE program No. 608.
32. DRAGON v1.1: Chemometrics and QSAR Research Group, Dipartimento di Scienze dell'Ambiente e del Territorio, Universita' degli Studi di Milano-Bicocca, Italy; available from http://www.disat.unimib.it/chum/Dragon.html.
33. D. Weininger, *J. Chem. Inf. Comput. Sci*. **28** (1988) 31–36.
34. Sybyl 6.9, Tripos Inc., St. Louis, MO, USA.
35. M. Clark, R. D. Cramer III, and N. van Opdenbosch, *J. Comput. Chem.* **10** (1989) 982–1012.
36. J. Gasteiger and M. Marsili, *Tetrahedron* **36** (1980) 3219–3222.
37. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* **107** (1985) 3902–3909.
38. M. L. Connolly, *J. Mol. Graphics* **11** (1993) 139–141.
39. ChemOffice 6.0, CambridgeSoft. Com, Cambridge, MA, U.S.A.
40. M. Turowski, R. Kaliszan, C. Lüllmann, G. Genieser, and B. Jastorff, *J. Chromatogr. A* **728** (1996) 201–211.
41. K. Osmialowski, J. Halkiewicz, A. Radecki, and R. Kaliszan, *J. Chromatog*. **346** (1985) 53–60.
42. W. M. Meylan and P. H. Howard, *J. Pharm. Sci.* **84** (1995) 83–92.
43. I. V. Tetko and V. Yu. Tanchuk, *J. Chem. Inf. Comput. Sci.* **42** (2002) 1136–1145, available from: http://146.107.217.178/lab/alogps/start.html.

44. ClogP v.4.0, available from: http://146.107.217.178/lab/alogps/index.html.
45. S. Wold and W. J. Dunn III, *J. Chem. Inf. Comput. Sci.* **23** (1983) 6–13.
46. STATISTICA package 6.0, StatSoft Inc., Tulsa, OK, U.S.A.
47. A. R. Leach and V. J. Gillet, *An Introduction to Chemoinformatics*, Kluwer Academic Publishers, Dordrecht, Boston, London, 2003, p. 83.
48. J. Zupan and J. Gasteiger, *Neural Networks in Chemistry and Drug Design: An Introduction*, 2nd Edition, Wiley-VCH, Weinheim, 1999.
49. R. D. Cramer III, D. E. Patterson, and J. D. Bunce, *J. Am. Chem. Soc.* **110** (1988) 5959–5967.
50. G. Klebe, U. Abraham, and Th. Mietzner, *J. Med. Chem.* **37** (1994) 4130–4146.
51. K. H. Kim, G. Greco, and E. Novellino, *A Critical Review of Recent CoMFA Applications*, in: H. Kubinyi, G. Folkers and Y. C. Martin (Eds.), *3D QSAR in Drug Design: Recent Advances*, Vol. 3, KLUWER/ESCOM, Leiden, The Netherlands, 1998, pp. 257–315.
52. H. Kubinyi, *Comparative Molecular Field Analysis*, in: P. Von Rague Schleyer (Ed.), *Encyclopedia of Computational Chemistry*, John Wiley & Sons, Ltd., Chichester 1998, pp. 448–460.
53. S. Wold, A. Ruhe, H. Wold, and W. J. Dunn, *Siam J. Sci. Stat. Comput.* **5** (1984) 735–743.

SAŽETAK

Kvantitativni odnosi strukture i retencije (QSRR) za kromatografsku separaciju boja temeljenih na diazo i triazo 4,4'-diaminobenzanilidima

Simona Funar-Timofei, Walter M. F. Fabian, Georgeta M. Simu i Takahiro Suzuki

Za niz od 23 molekule boja iz skupine diazo i triazo 4,4'-diaminobenzanilidina eksperimentalno je određena kromatografska pokretljivost, ekstrapolirana na uvjete u kojima nema modifikatora (vrijednost R_{MO}), pomoću tankoslojne kromatografije obrnutih faza (RP-TLC). Tradicionalne i kvantativne QSAR/QSPR metode modeliranja primijenjene su kako bi se pronašao kvantitativni odnos između strukture i retencije boja (QSRR). Molekularne strukture boja optimirane su minimalizacijom energije pomoću molekularno mehaničkih i kvantno kemijskih proračuna. Iz optimiranih trodimenzionalnih geometrija izračunani su različiti jednodimenzionalni i trodimenzionalni molekularni deskriptori (opisivači) koji sadrže informaciju o veličini, obliku, simetriji, elektronskoj strukturi, rasporedu atoma i kemijskih skupina, te hidrofobnosti molekula boja. Modeliranjem pomoću višestruke linearne regresije (MLR) i metode umjetnih neuronskih mreža (ANN) pokazalo se da se vrijednosti R_{MO} mogu uspješno iskazati kombinacijom parametara koji opisuju hidrofobna i polarna svojstva strukture boja. Dodatna analiza usporedbom molekularnih polja (CoMFA) i analiza indeksa sličnosti (CoMSIA) ukazuje na gotovo jednaki doprinos steričkog i elektrostatskog polja kromatografskoj pokretljivosti, te na glavnu ulogu hidrofobnih i elektrostatskih međudjelovanja molekula boja s kromatografskom okolinom.