

# Comparative Molecular Field Analysis of Heterocyclic Monoazo Dye–Fiber Affinities

Simona Timofei\*<sup>†</sup> and Walter M. F. Fabian<sup>‡</sup>

Institute of Chemistry, Romanian Academy, Bulevardue Mihai Viteazul 24, 1900 Timisoara, Romania, and  
Institut für Organische Chemie, Karl-Franzens Universität Graz, Heinrichstrasse 28, A-8010 Graz, Austria

Received November 26, 1997

This paper presents quantitative structure–activity relationship (QSAR) models for a series of 21 heterocyclic monoazo dyes applied on cellulose fiber by comparative molecular field analysis. The electronic and structural properties of these dyes were calculated by the semiempirical PM3 method; the results indicate the predominance of electrostatic interactions in dye–cellulose binding. The dominant contribution of the LUMO orbital molecular energy, used as descriptor, can be explained by the contribution to the solvation of the dye molecules.

## INTRODUCTION

The application of the QSAR techniques to dye–textile fiber interactions yield new insights into the mechanisms of dye binding to fibers. Classical quantitative structure–affinity relationships<sup>1–5</sup> and, also, three-dimensional QSAR studies<sup>6–9</sup> have been reported in the literature for the dye adsorption into the cellulose fiber. Electrostatic interactions have been found to dominate the steric ones. The polar effects have been expressed by dipole interactions or hydrogen bonding.

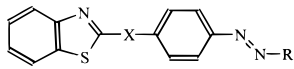
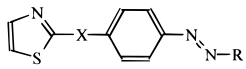
The application of QSAR techniques to dye–cellulose binding is based on the specific dye–fiber interactions. In a previous comparative molecular field analysis (CoMFA) study applied to disperse azo dyes, the dye binding to cellulose seemed to be less specific in terms of pharmacophoric constraints.<sup>9</sup> Some recent results published by Woodcock point to the specific adsorption of the anionic azo dye Congo Red on the crystalline cellulose surfaces.<sup>10</sup>

This paper presents the CoMFA results of the effects of structural modifications of a series of heterocyclic azo dyes (Table 1) on their ability to bind to cellulose fibers. Additional semiempirical molecular orbital (PM3) calculations have been performed to develop the low-energy conformations and QSAR descriptors.

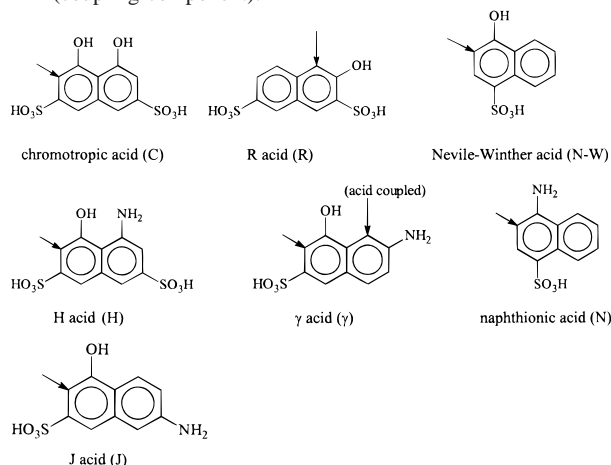
## COMPUTATIONAL DETAILS

Starting structures were generated using the SYBYL package,<sup>11</sup> with the TRIPOS force field,<sup>12</sup> and quantum mechanics have been performed for the molecular lowest energy calculations. For the lowest energy conformations, semiempirical molecular orbital calculations were done. The PM3 method<sup>13,14</sup> was employed by the MOPAC 6.0 package.<sup>15,16</sup> Geometries were completely optimized by the eigenvector following routine.<sup>17</sup> For each compound the conformation of lowest energy was used in CoMFA calculations. The alignment of the molecules was done by the RMS\_FIT option within SYBYL (with compound 15 as template). In the alignment of the dye molecules the nitrogen atoms of the azo group, the aryl carbon atoms attached to the azo group, and the phenyl carbon atom placed in the

**Table 1.** Experimental (*A*) and Calculated (*Ĥ*) Dye Affinities (for Model II, See Table 2) and HOMO (*E*<sub>HOMO</sub>) and LUMO (*E*<sub>LUMO</sub>) Orbital Molecular Energies of Heterocyclic Azo Dyes I and II

						
			I		II	
no.	X	R <sup>a</sup>	<i>A</i> (kJ/mol)	<i>Ĥ</i> (kJ/mol)	<i>E</i> <sub>HOMO</sub> (eV)	<i>E</i> <sub>LUMO</sub> (eV)
I.1	NH	C	6.78	6.84	−8.74	−1.83
I.2	NH	R	9.20	10.21	−8.78	−1.83
I.3	NH	H	12.60	9.78	−8.60	−1.72
I.4	NH	γ	15.30	12.55	−8.62	−1.41
I.5	O	C	3.26	1.47	−9.03	−1.85
I.6	O	R	5.27	5.09	−9.21	−1.85
I.7	O	H	7.61	8.70	−8.70	−1.74
I.8	O	γ	10.30	11.71	−8.59	−1.31
I.9	O	J	10.20	11.12	−8.57	−1.29
I.10	S	C	1.26	1.77	−8.86	−1.90
I.11	S	R	3.56	4.47	−8.91	−1.90
I.12	S	H	5.02	4.15	−8.69	−1.79
I.13	S	γ	8.45	8.46	−8.58	−1.37
I.14	S	J	8.12	8.51	−8.53	−1.36
II.15	NH	γ	15.33	14.55	−8.47	−1.29
II.16	NH	H	12.60	11.38	−8.60	−1.72
II.17	NH	R	9.24	10.15	−8.79	−1.83
II.18	NH	C	6.80	7.84	−8.74	−1.82
I.19	S	N–W	5.86	6.33	−8.79	−1.49
I.20	S	N	10.33	11.57	−8.59	−1.36
I.21	S	γ (acid coupled)	9.75	10.18	−8.50	−1.58

<sup>a</sup> R (coupling component):



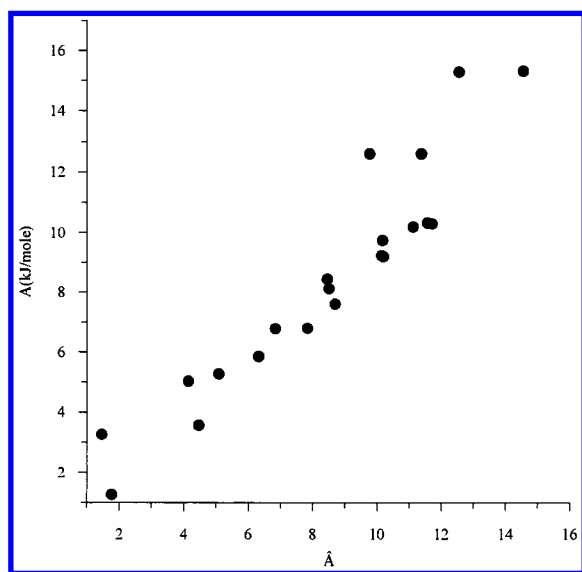
<sup>†</sup> Romanian Academy.

<sup>‡</sup> Karl-Franzens Universität Graz.

**Table 2.** CoMFA Statistical Results ("Cross-Validated  $r^2$ " ( $r_{CV}^2$ ) and Standard Errors of Predictions (SDEP) for the Optimum Number of Principal Components, Conventional  $r^2$ ,  $F$ -test, Standard Errors of Estimates (SEE), and Contributions (Normalized Coefficients and Fractions) of the Respective QSAR Descriptors

model	type <sup>a</sup>	$r_{CV}^2$	SDEP	$r^2$ <sup>b</sup>	$F$	SEE	descriptor	normalized coeff	fraction
I	CoMFA ( $S$ , $E$ )	0.624	2.645	0.970 (5)	95.88	0.751	$S$	1.589	0.395
							$E$	2.435	0.605
II	CoMFA ( $S$ , $E$ ), $E_{HOMO}$ , $E_{LUMO}$	0.766	1.960	0.886 (3)	44.10	1.367	$S$	0.722	0.263
							$E$	0.920	0.335
							$E_{HOMO}$	0.492	0.179
							$E_{LUMO}$	0.611	0.223
III	CoMFA ( $S$ , $E$ ), $E_{HOMO}$	0.618	2.581	0.899 (4)	35.70	1.326	$S$	0.864	0.298
							$E$	1.376	0.474
							$E_{HOMO}$	0.662	0.228
IV	CoMFA ( $S$ , $E$ ), $E_{LUMO}$	0.795	1.891	0.969 (4)	125.05	0.735	$S$	1.150	0.299
							$E$	1.586	0.413
							$E_{LUMO}$	1.108	0.288
V	CoMFA ( $E$ )	0.615	2.591	0.948 (4)	72.54	0.955	$E$	3.561	
VI	CoMFA ( $S$ )	0.389	3.372	0.955 (5)	63.29	0.918	$S$	4.361	
VII	CoMFA ( $S$ ), $E_{LUMO}$	0.818	1.839	0.977 (5)	126.20	0.675	$S$	2.623	0.741
							$E_{LUMO}$	0.916	0.259
VIII	CoMFA ( $S$ ), $E_{HOMO}$ , $E_{LUMO}$	0.756	2.130	0.961 (5)	74.60	0.848	$S$	2.145	0.678
							$E_{HOMO}$	0.397	0.125
							$E_{LUMO}$	0.621	0.196
IX	CoMFA ( $S$ ), $E_{HOMO}$	0.590	2.674	0.898 (4)	35.30	1.332	$S$	1.925	0.716
							$E_{HOMO}$	0.764	0.284

<sup>a</sup>  $S$  = steric field;  $E$  = electrostatic field. <sup>b</sup> The optimum principal component number used in the non-cross-validated PLS analysis is given in parentheses.

**Figure 1.** Plot of experimental ( $A$ ) versus calculated ( $\hat{A}$ ) dye affinities for model II.

para position relative to the azo group were employed. The optimized structures were taken from MOPAC calculations. Sterical fields were computed using  $sp^3$  carbon as probe atom with a charge probe of +1 for electrostatic fields. A box with dimensions  $29 \times 22 \times 19$  Å and a grid spacing of 2 Å were used. Statistical analysis was carried out by the PLS method<sup>18</sup> with the leave-one-out cross-validation procedure<sup>19</sup> to determine the optimal number of components to be used in the final PLS analysis (without cross-validation). In this procedure, the PLS components were varied from 2 to 10. All the cross-validation calculations were performed with a minimal  $\sigma$  (column filter) value of 2.00 kcal/mol. "CoMFA STD" scaling of independent variables for the steric and electrostatic fields was used.

## RESULTS AND DISCUSSION

Experimental dye affinity data ( $A$ ), taken for compounds **I.1–I.14** from ref 20, for compounds **II.15–II.18** from ref 21, and for those of **I.19–I.21** from ref 22, and, also, the results of PM3 calculations are collected in Table 1.

Using descriptors derived from quantum chemical calculations, correlations with the dye affinity have been performed. The following equations have been obtained:

$$\hat{A} = (131.87 \pm 29.24) + (14.18 \pm 3.36)E_{HOMO}$$

$$r = 0.696 \text{ SEE} = 2.75 \text{ } F = 17.83 \text{ } r_{CV(LOO)}^2 = 0.293 \quad (1)$$

$$\hat{A} = (24.42 \pm 4.84) + (9.81 \pm 2.94)E_{LUMO}$$

$$r = 0.608 \text{ SEE} = 3.04 \text{ } F = 11.12 \text{ } r_{CV(LOO)}^2 = 0.283 \quad (2)$$

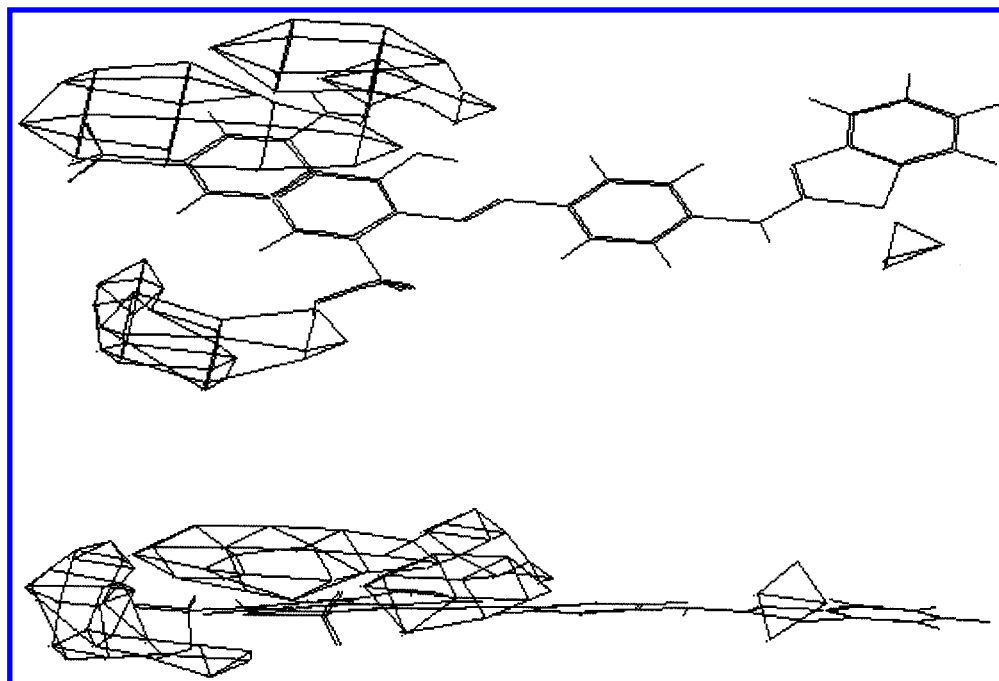
$$\hat{A} = (108.89 \pm 37.18) + (10.83 \pm 4.74)E_{HOMO} +$$

$$(3.76 \pm 3.75)E_{LUMO}$$

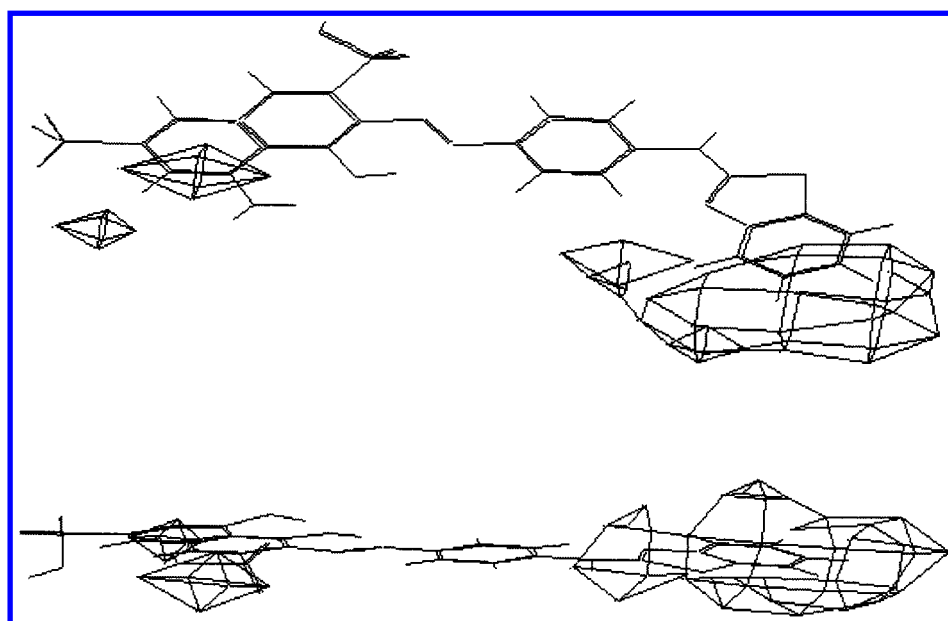
$$r = 0.715 \text{ SEE} = 2.75 \text{ } F = 9.42 \text{ } r_{CV(LOO)}^2 = 0.269 \quad (3)$$

and they show the same dependence for the HOMO ( $E_{HOMO}$ ) and LUMO ( $E_{LUMO}$ ) molecular orbital energies. In the case of the HOMO energy, the correlation has a physical meaning if the dye has the role of the electron donor. In general, affinity is associated with a low-lying LUMO molecular orbital, assuming a charge or electron transfer from the receptor (fiber in this case) to the compound. This type of correlation might be explained by the contribution of the solvation of the compounds (the higher the LUMO energy, the weaker the solvation).

The MLR (multiple linear regression) analysis has been performed by the SYSTAT program, from SYSTAT, Inc., Evanston, IL. The standard errors of the regression coefficients were not multiplied by the  $t$  (Student) test values. The statistical results of eqs 1–3 indicate poor correlations with the dye affinity and predictability. Attempts to correlate



**Figure 2.** Top and side views of regions with sterically favorable interactions (contouring levels from 0.02 to 1000).



**Figure 3.** Top and side views of regions with sterically unfavorable interactions (contouring levels from -1000 to -0.03).

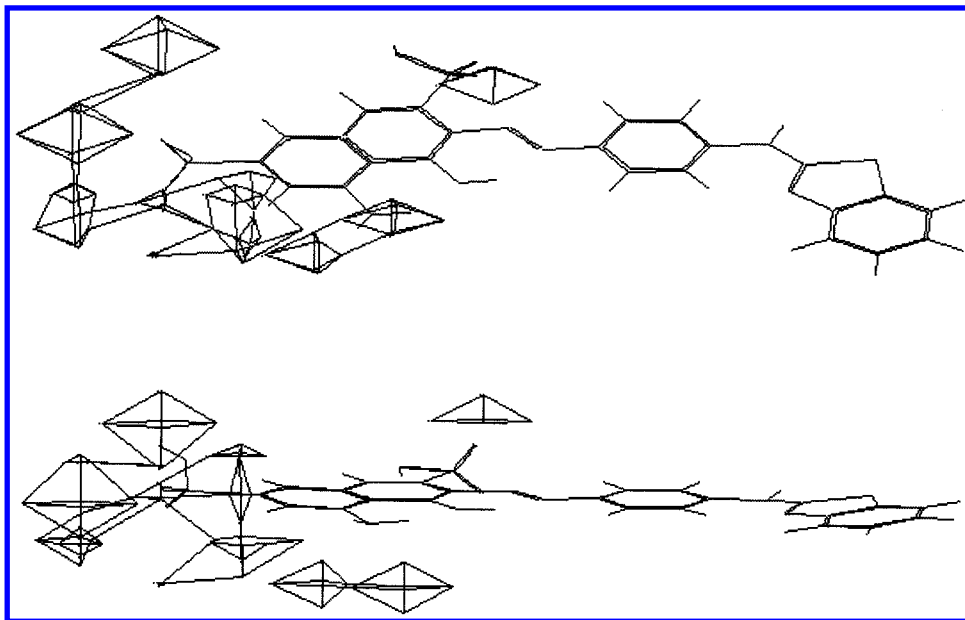
the adsorption strength with the ClogP hydrophobicity parameter<sup>23</sup> showed worse statistical results ( $r = 0.089$ ,  $r_{CV(LOO)}^2 = -0.167$ ), indicating the absence of hydrophobic dye–cellulose interactions. A molecular length descriptor, similar to that from ref 9, was calculated. Nonsatisfactory statistical results have been noticed ( $r = 0.558$ ,  $r_{CV(LOO)}^2 = +0.157$ ), in opposite to the previous<sup>9</sup> correlations with this parameter.

Besides the lowest energy conformation, for each compound other conformations have been taken into account in the MOPAC calculations, with respect to the orientation of the azo group and the position of the heterocyclic X atom, in the range of 3 kcal/mol above the lowest calculated molecular energy of one compound. The residuals of the CoMFA predicted affinity values for these conformations

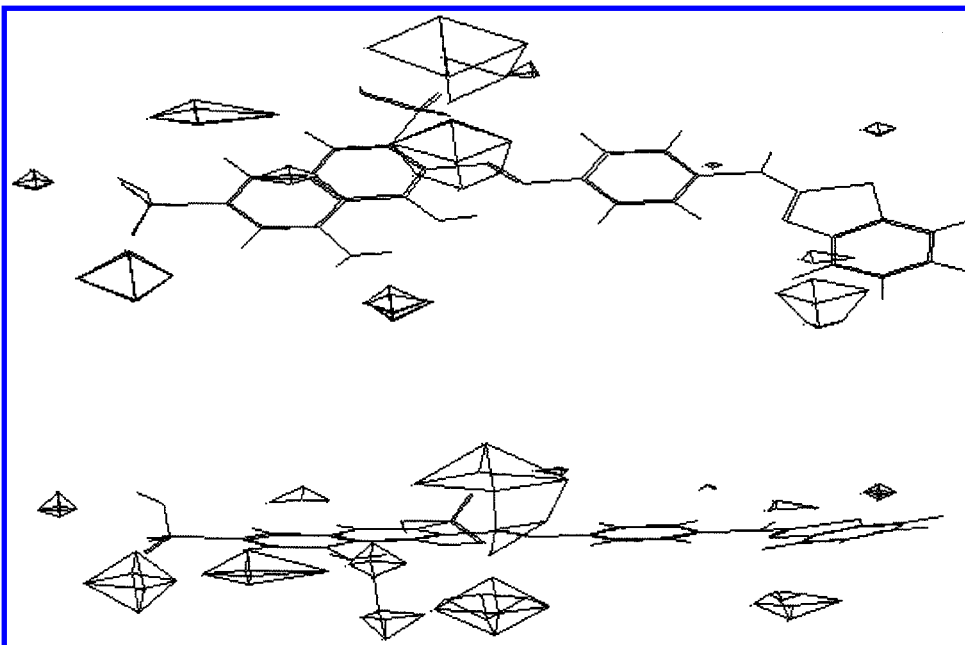
were in the range of the standard deviations of the first proposed model (steric and electrostatic field).

A summary of the statistical results of the CoMFA analysis is listed in Table 2. In model I (see Table 2) only CoMFA (steric and electrostatic) fields were used. The statistical results indicate the predominance of electrostatic effects in dye–cellulose fiber binding, as stated previously.<sup>7,8</sup>

Good correlation with the dye affinity and predictions of compounds with higher affinities have been noticed for model I. The predominance of the electrostatic field indicates polar attractions between the dye molecules and the fiber. The introduction of the HOMO and, respectively, LUMO molecular orbital energies as independent variables, together with the CoMFA steric and electrostatic fields, increased significantly the quality of model I. The LUMO molecular orbital energy has a bigger contribution to the dye



**Figure 4.** Top and side views of regions where an increase of negative charges should lead to higher affinities (contouring levels from  $-1000$  to  $-0.03$ ).



**Figure 5.** Top and side views of regions where an increase of positive charges should lead to higher affinities (contouring levels from  $0.02$  to  $1000$ ).

affinity in comparison to the HOMO molecular orbital energy. The electrostatic interactions are in this case dominant too. The actual versus calculated dye affinity plot is presented in Figure 1.

With application of the parsimony principle, model II with only three latent variables appears to be the best one. However, this model has the highest complexity toward interpretation; therefore simpler models will, also, be discussed.

Similar to model II, improvements in the statistical results have been noticed for model IV, with only LUMO molecular orbital energy used as independent variable. This type of correlation can be explained by the contribution of the solvation of the compounds. The frontier orbital calculations indicated the association of the LUMO molecular orbital energy with the nitrogen atoms of the azo groups in the dye

molecules. The introduction of HOMO molecular orbital energy as independent variable besides the CoMFA electrostatic and steric fields yielded statistical results similar to those from model I.

Similar statistical results with model I have been noticed for model V with only electrostatic field. Good correlation with the dye affinity is observed for model VI with only steric field, but poorer predictability in comparison to model V. Because of the low predictability of this model, HOMO and LUMO molecular orbital energies have been added as independent variables.

The cross-validated correlation coefficient got a high value ( $0.818$ ) and the standard error of estimates a low value ( $0.675$ ) when steric interactions and LUMO energies were used as independent variables, indicating the importance of

the LUMO energies in the dye–fiber interactions (see Table 2).

The coefficient contour maps for the steric field of model II are presented in Figures 2 and 3. They indicate as favorable for the adsorption on cellulose the presence of substituents at certain regions of the coupling components. Increased regions in the aromatic moieties of the coupling component, by the introduction of bulky substituents at the edge of the 1- and 2-naphthyl moieties (like the H and chromotropic acids) increase the dye affinity.

The presence of the heterocyclic aromatic moiety decreases the dye affinity. Also, the presence of some bulky substituents of the coupling components placed in the ortho position related to the azo group (like the sulfonic group) is detrimental for the dye–cellulose binding.

The predominance of the electrostatic field indicates polar attractions between the dye molecules and the fiber. The electrostatic coefficient contour maps indicate the predominance of the positive over the negative charges. Increased positive charged substituents in the 2-naphthyl moiety and in the regions corresponding to the substitution to the phenylic moiety of the benzothiazole group could lead to higher affinity toward the fiber. Negative charged substituents in the regions of the sulfonic groups are favorable for the dye affinity for cellulose. The electrostatic field contour maps for model II are presented in Figures 4 and 5. The plots of electrostatic fields lead to the conclusion of the beneficial contribution to the affinities for dyes having as the coupling component the H acid and increased positive charges in the substitution positions of the coupling components near the azo group.

### CONCLUSION

Comparative molecular field analysis has been applied to derive quantitative structure–activity relationships for heterocyclic monoazo dye binding on cellulose. Electrostatic contributions are dominant, as stated previously. The introduction of LUMO molecular orbital energy as descriptor may indicate the contribution of the dye molecule solvation in the cellulose dyeing. The plots of electrostatic and steric contributions are useful guides in the design of new dyes. Certainly there are considerably less specific interactions in dye–fiber adsorption than in the binding of a drug to its receptor. Nevertheless, as shown in the present paper, the pharmacophore model proves to be a valuable tool for the description of such interactions.

### ACKNOWLEDGMENT

S.T. acknowledges with thanks the financial support provided by Karl-Franzens University of Graz (Austria).

### REFERENCES AND NOTES

- (1) Timofei, S.; Schmidt, W.; Kurunczi, L.; Simon, Z.; Sallo, A. A QSAR Study of the Adsorption by Cellulose Fibre of Anthraquinone Vat Dyes. *Dyes Pigm.* **1994**, *24*, 267–279.
- (2) Timofei, S.; Kurunczi, L.; Schmidt, W.; Fabian, W. M. F.; Simon, Z. Structure-Affinity Binding Relationships by Principal-Component-Regression Analysis of Anthraquinone Dyes. *Quant. Struct.-Act. Relat.* **1995**, *14*, 444–449.
- (3) Timofei, S.; Kurunczi, L.; Schmidt, W.; Simon, Z. Structure-Affinity Binding Relationships of Some 4-Aminoazobenzene Derivatives for Cellulose Fibre. *Dyes Pigm.* **1995**, *29*, 251–258.
- (4) Timofei, S.; Kurunczi, L.; Schmidt, W.; Simon, Z. Lipophilicity in Dye-Cellulose Fibre Binding. *Dyes Pigm.* **1996**, *32*, 25–42.
- (5) Timofei, S.; Kurunczi, L.; Schmidt, W.; Simon, Z. Dye Structure-Affinity Relationships by the MTD Method. *Rev. Roum. Chim.*, in press.
- (6) Timofei, S.; Kurunczi, L. In *Quantitative Relationships between Chemical Structure and Biological Activity (QSAR). The MTD Method*; Chiriac, A., Ciubotariu, D., Simon, Z., Eds.; Mirton Publishing House: Timisoara, Romania, 1996; pp 205–220.
- (7) Fabian, W. M. F.; Timofei, S.; Kurunczi, L. Comparative Molecular Field Analysis (CoMFA), Semiempirical (AM1) Molecular Orbital and Multiconformational Minimal Steric Difference (MTD) Calculations of Anthraquinone Dye-Fibre Affinities. *J. Mol. Struct. (THEOCHEM)* **1995**, *340*, 73–81.
- (8) Fabian, W. M. F.; Timofei, S. Comparative molecular field analysis (CoMFA) of dye-fibre affinities Part 2. Symmetrical bisazo dyes. *J. Mol. Struct. (THEOCHEM)* **1996**, *362*, 155–162.
- (9) Oprea, T. I.; Kurunczi, L.; Timofei, S. QSAR Studies of Disperse Azo Dyes. Toward the Negation of The Pharmacophore Theory of Dye-Fiber Interaction? *Dyes Pigm.* **1997**, *33*, 41–64.
- (10) Woodcock, S.; Henrissat, B.; Sugiyama, J. Docking of Congo Red to the Surface of Crystalline Cellulose Using Molecular Mechanics. *Biopolymers* **1995**, *36*, 201–210.
- (11) SYBYL 6.0, Tripos Associates, St. Louis, MO.
- (12) Clark, M.; Cramer, R. D., III; van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (13) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.
- (14) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods II. Applications. *J. Comput. Chem.* **1989**, *10*, 221–264.
- (15) MOPAC 6.0, QCPE Program No. 455.
- (16) Stewart, J. J. P. MOPAC: A semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1–105.
- (17) Baker, J. Algorithm for the Location of Transition States. *J. Comput. Chem.* **1986**, *7*, 385–395.
- (18) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *Siam J. Sci. Stat. Comput.* **1984**, *5*, 735–743.
- (19) Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank E. I. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.
- (20) Alberti, G.; Seu, G. Dyeing Substantivity of Anionic Dyes with Angular Molecule. *Ann. Chim. (Rome)* **1983**, *73*, 737–740.
- (21) Alberti, G.; Cerniani, A.; De Giorgi, M. R.; Seu, G. Dyeing of Cellulose with Direct Dyes. *Ann. Chim. (Rome)* **1983**, *73*, 265–272.
- (22) Alberti, G.; Cerniani, A.; De Giorgi, M. R.; Seu, G. *Thermodynamic data of dyeings on rayon and cellulose of some azo dyes having a benzothiazole ring. Tinctoria* **1980**, *5*, 141–145.
- (23) ClogP, version 1.0.0, BioByte Corp., Claremont, CA.

CI9704367