Comparative Molecular Surface Analysis (CoMSA) for Modeling Dye−Fiber Affinities of the Azo and Anthraquinone Dyes

Jaroslaw Polanski,*,† Rafal Gieleciak,† and Miroslaw Wyszomirski‡

Department of Organic Chemistry, Institute of Chemistry, University of Silesia, PL-40-006 Katowice, Poland, and University of Bielsko-Biala, PL-43-310 Bielsko-Biala, Poland

Received April 17, 2003

Despite recent investigations aimed at modeling 3D QSAR for dye molecules a controversy still exists: can a pharamacophore hypothesis be used for such purposes. In the present publication we reported on the application of the CoMSA method for modeling 3D QSAR of azo and anthraquinone dyes. We obtained very predictive models, which significantly outperform those reported in the previous CoMFA studies, especially for the azo dyes. Our results proved the previous conclusion that steric requirements are far less pronounced for the cellulose cavities than for the classical drug receptor. Moreover, our results indicate that all molecular surface segments are important for dye−fiber interactions, which also makes an important difference in relation to the classical drug pharmacophore. On the other hand, high predictivity of the CoMSA models indicates that a pharmacophore concept is suitable for the description of the dye−fiber interactions. However, this pharmacophore must substantially differ from the drug pharmacophore used for the illustration of the drug-receptor interactions. From a theoretical point of view dye−cellulose interactions can be an interesting case in which shape decides the activity rules not by the steric repulsion but as a cofactor determining the electrostatic potential distribution.

INTRODUCTION

Interaction between dye molecule and cellulose is a complicated phenomenon, which can be described by the Langmuir isotherm.[1,2] This does not, however, provide a molecular description of the process. Moreover, we cannot use such an approach for the optimization of the dye molecule structure. The influence of the electrostatic, van der Waals, or hydrogen bonding as well as hydrophobic forces on such processes has been investigated. On the other hand, the arrangement of the dye molecules on cellulose surface suggested that specific binding sites could exist which are patterned by the supramolecular cellulose structure.[3] Although it is not quite clear if we can treat it similarly to the contacts taking place during targeting a receptor by a drug molecule, several QSAR studies have been published recently[4−11] that make use of the pharmacophore concept in investigations of cellulose dying. Both 2D and 3D QSAR modeling have been applied for this purpose including the Hansch, MTD, and CoMFA methods that appeared to provide satisfactory models for anthraquinone vat dyes,[8] symmetrical bisazo dyes,[9] heterocyclic monoazo dyes,[12] and disperse azo dyes.[10] In particular, the results of CoMFA method indicated that it is the electrostatic field that dominantly contributes to the dying affinity. On the other hand, binding affinity seemed to be less specific than biological responses characterizing classical drug pharmacophores.[2]

Usually, a physicochemical meaning underlying the notion of pharmacophore emphasizes the fact that what we are doing is the attempting to identify a certain subset of the atoms that forms a common or similar pattern for all active molecules. In fact, we do not have any information on what the relation between a real receptor and this subset is. Pharmacophore mapping is a broad strategy which includes a variety of experimental and computational approaches. Quantitative Structure Activity Relationship (QSAR) and all its variants, i.e., QSPR, QSRR, and finally three-dimensional (3D) approaches[13−16] are possible strategies that can also be followed in such cases. In previous papers we have described the alternative method of the Comparative Molecular Surface Analysis (CoMSA) that provides a more flexible scheme for the comparison of the molecules, namely, molecular surfaces and modeling 3D QSARs.[17−20]

The aim of this study is the systematic analysis of a few series of dyes for which QSAR modeling has been performed previously. Moreover, we would like to verify the applicability of the pharmacophore concept in the dye chemistry by the application of the Comparative Molecular Surface Analysis. Thus, in this paper we discuss the application of the CoMSA to the series of azo and anthraquinone dyes.

THEORETICAL BACKGROUND

Self-organizing neural network (SOM) is a technique designed to reduce the dimensionality of the data while preserving topology. Bioinformatics is one of the recent implementations that illustrate the importance of this technique.[21−23] The method has also been applied in chemistry,[24,25] in particular, for two-dimensional mapping of the electrostatic potential on the three-dimensional molecular surfaces[25,26] or partial atomic charges for the atomic molecular representation.[27] The ability to compress the size of the data and to reconstruct the 3D object from the 2D representation makes this procedure an interesting tool for molecular design.[24,25,28] Such maps were used for the visualization of

* Corresponding author e-mail: Polanski@us.edu.pl.
† University of Silesia.
‡ University of Bielsko-Biala.

10.1021/ci0340761 CCC: $25.00
Published on Web 08/30/2003

the interactions of individual molecules with biological receptors or drug design.[24,25]

In our previous publications we described the use of the Kohonen neural network in QSAR investigations, in particular we designed a scheme of 3D QSAR method by a coupled neural network and PLS system which we called the Comparative Molecular Surface Analysis (CoMSA).[17−20,29] Some further improvements to the CoMSA have also been published by Hasegawa et al.[30]

EXPERIMENTAL SECTION

Model Building. All the experimental data, i.e., the dye affinity $\Delta\mu^\circ$ of anionic azo dyes and dye affinity of anthraquinone vat dyes, are extracted from the refs 12 and 9, respectively, and are given in Tables 1 and 2. We used Gesteiger's software package for modeling purposes. The 3D coordinates of all molecules were obtained by the 3D structure generator CORINA[31−33] Partial atomic charges were calculated by the PEOE method,[34,35] and the SURFACE program was used for the calculation of the Coulomb electrostatic potential on the molecular surface.

Data Analysis. Kohonen Mapping. The competitive Kohonen strategy[36] was used to construct a two-dimensional topographic map obtaining the signals from the points sampled randomly at the molecular surface. As molecular surfaces are continuous the plane of projection was also selected to be a continuous surface. Thus we used a torus for this purpose, which was cut along two perpendicular lines and then spread into a plane. Each neuron, *j*, was then defined by three weights, w_{ji}. The competitive training of the network was based on the rule that each point, *s*, of the molecular surface was projected into that neuron, *c*, that has weights, w_{ci}, that come closest to the Cartesian coordinates, x_{si}, of this point, *s* (eq 1)

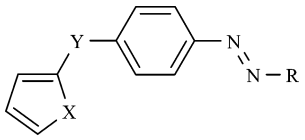
$$\text{out}_c \rightarrow \min\left[\sum_{i=1}^{m}(x_{si} - w_{ji})^2\right] \qquad (1)$$

where *m* is a number of weights per neuron.

A projection of the electrostatic potential value (MEP) from the surface points, *s*, into such a two-dimensional arrangement of neurons, after calculating the average MEP value within this particular neuron and scaling this values into the respective colors results in the so-called feature map.

Comparative Kohonen Mapping. In fact, such a map illustrates the property (MEP) of a single molecule. As, the weights of the Kohonen network contain the shape of the certain molecular surface, it can be used to compare the geometries of molecular surfaces of other molecules. In such a method, the trained Kohonen network processes the signals coming from the surface of other molecule(s), i.e., the electrostatic potential of each input vector was projected through the network to obtain a series of comparative maps both for the template molecule and each analyzed molecule. This allows us to compare these parts of the molecule surfaces that can be superimposed. If the surfaces cannot be superimposed on the reference molecule (template), then the respective output neurons get no signal from the molecules processed. This results in the appearance of the so-called empty neurons that are coded by zeros. Before further processing we prepared three different potential matrix types,

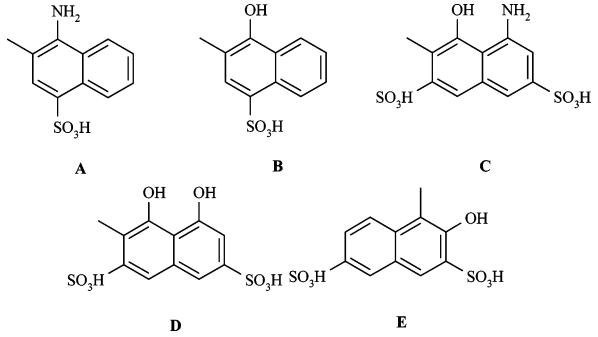
Table 1. Experimental Dye Affinity ($-\Delta\mu^\circ$)[11] of Azo Dyes (I) and This Predicted by the CoMSA Method



I

no.	X	Y	R	$-\Delta\mu^\circ$ (kJ/mol)[a]	$-\Delta\mu^\circ$ [b]	$-\mu^\circ$ [c]
1.	-S-	-CH=CH-	A	15.80	14.52	*d*
2.	-CH=CH-	-CH=CH-	A	14.25	13.77	14.71
3.	-S-	-CONH-	A	13.08	12.26	*d*
4.	-CH=CH-	-CONH-	A	12.00	12.55	13.10
5.	-S-	-CH=CH-	B	9.66	10.80	*d*
6.	-S-	-CH=CH-	C	9.45	9.91	9.39
7.	-CH=CH-	-CH=CH-	B	9.20	8.74	*d*
8.	-S-	-CONH-	C	9.03	8.78	8.44
9.	-S-	-CO-	A	8.78	9.07	*d*
10.	-CH=CH-	-CH=CH-	C	8.40	8.29	7.77
11.	-CH=CH-	-CONH-	C	8.28	7.70	*d*
12.	-S-	-CONH-	B	7.15	8.04	8.12
13.	-S-	-CH=CH-	D	7.06	6.83	*d*
14.	-CH=CH-	-CO-	A	7.02	9.26	10.02
15.	-CH=CH-	-CONH-	B	6.52	7.28	*d*
16.	-S-	-CH=CH-	E	6.27	6.03	6.05
17.	-S-	-CONH-	D	6.23	6.49	*d*
18.	-CH=CH-	-CH=CH-	D	6.02	6.58	6.55
19.	-CH=CH-	-CH=CH-	E	5.81	6.09	*d*
20.	-CH=CH-	-CONH-	D	5.18	5.90	6.21
21.	-S-	-CONH-	E	5.10	5.78	*d*
22	-S-	-CO-	C	4.64	5.99	4.83
23.	-CH=CH-	-CONH-	E	4.26	3.72	*d*
24.	-S-	-CO-	B	4.22	4.11	4.63
25	-CH=CH-	-CO-	C	4.10	4.04	*d*
26.	-CH=CH-	-CO-	B	4.05	4.52	5.36
27.	-S-	-CO-	D	3.85	2.80	*d*
28.	-CH=CH-	-CO-	D	3.43	1.45	2.91
29.	-S-	-CO-	E	3.22	1.78	*d*
30.	-CH=CH-	-CO-	E	2.84	2.73	2.84

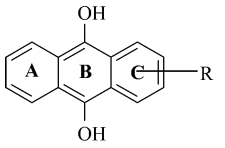
[a] Experimental affinities acc. to ref 11. [b] Cross-validated (LOO) values, details in text. [c] External predictions, details in text. [d] Training series, details in text.



as shown in Figure 1. The matrices that contain all elements including empty neurons form the first group. We named such matrices en-matrices (en) (Figure 1a). Then we eliminated empty neurons within all comparative patterns to obtain nonempty neurons matrices (nen) (Figure 1b). This means that these elements that are represented in any of the matrix within the series by 0 are eliminated in each matrix of the series. Inversely, the third types of the electrostatic potential matrices, non-nen-matrices (non-nen) (Figure 1c), are those that include only such elements which obey the rule of having in any matrix of the series at least one empty neuron.

All the molecules were superimposed before the calculation of the molecular surfaces. Covering each atom of the

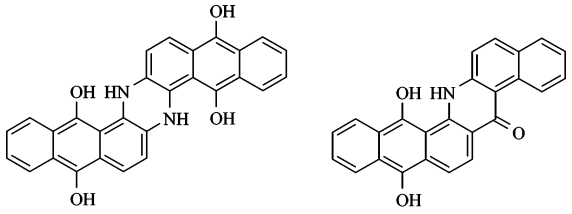
Table 2. Experimental Dye Affinity ($-\Delta\mu^0$)[8] of Anthraquinone Dyes (II) and This Predicted by the CoMSA Method



II

no.	R	$-\Delta\mu^{0\,a}$	$-\mu^{0\,b}$	$-\Delta\mu^{0\,c}$	no.	R	$-\Delta\mu^{0\,a}$	$-\Delta\mu^{0\,b}$	$-\Delta\mu^{0\,c}$
1.	H	0	0.38	*d*	26.	1-[4-$OCH_3C_6H_4CONH$-]	2.90	3.13	3.17
2.	1-Cl	0	0.31	0.53	27.	1-NH_2; 4-$NHCOC_6H_5$	2.94	3.12	*d*
3.	1-CH_3	0	0.30	*d*	28.	1-[4-ClC_6H_4CONH-]	3.02	3.23	3.28
4.	1-OCH3	0	−0.04	0.17	29.	1-$NHCOC_6H_5$; 5-$NHCOC_6H_5$	3.57	4.38	*d*
5.	1-NH_2	0	0.98	*d*	30.	1-$NHCOC_6H_5$; 5-OCH_3	3.59	3.84	3.87
6.	2-NH_2	0	0.31	0.34	31.	1-$NHCOC_{10}H_7$−2′	3.61	2.73	*d*
7.	1-N(CH_3)($COCH_3$)	0	0.29	*d*	32.	1-$NHCOC_6H_5$; 4-OCH_3	3.59	1.81	2.22
8.	1-$NHCH_2C_6H_5$	0	0.70	0.72	33.	1-[4-$CH_3C_6H_4CONH$-]; 5-[4-$CH_3C_6H_4CONH$-]	3.73	3.83	*d*
9.	1-N(CH_3)(COC_6H_5); 4-N(CH_3)-(COC_6H_5)	0	0.03	*d*	34.	1-[3-$CH_3C_6H_4CONH$-]; 5-[3-$CH_3C_6H_4CONH$-]	3.77	3.10	3.04
10.	1-NH_2; 4-NH_2	1.46	1.79	1.65	35.	1-[3-$OCH_3C_6H_4CONH$-]; 5-[3-$OCH_3C_6H_4CONH$-]	3.86	4.07	*d*
11.	1-$NHCH_3$	1.46	0.70	*d*	36.	1-[4-$OCH_3C_6H_4CONH$-]; 5-[4-$OCH_3C_6H_4CONH$-]	4.11	4.05	3.83
12.	1-$NHCOCH_3$	1.46	2.10	1.87	37.	1-[4-ClC_6H_4CONH-]; 5-OCH_3	4.06	3.65	*d*
13.	1-NH_2; 8-NH_2	1.49	0.46	*d*	38.	1-[3-ClC_6H_4CONH-]; 5-[3-ClC_6H_4CONH-]	4.17	4.63	3.88
14.	1-$NHCOCH_3$	1.52	1.40	1.46	39.	1-$NHCOC_6H_5$; 4-$NHCOC_6H_5$;	4.29	4.15	*d*
15.	2-$NHCOCH_3$	1.53	1.39	*d*	40.	1-[4-ClC_6H_4CONH-]; 5-[4-ClC_6H_4CONH-]	4.37	4.46	3.92
16.	1-$NHCH_3$; 4-$NHCH_3$	1.55	1.70	1.59	41.	1-[4-$OCH_3C_6H_4CONH$-]; 5-[4-$OCH_3C_6H_4CONH$-]	4.44	4.61	*d*
17.	(1,2)−$(CH_2)_4$	1.91	0.90	*d*	42.	1-[3-$CH_3C_6H_4CONH$-]; 4-[3-$CH_3C_6H_4CONH$-]	4.56	4.29	4.21
18.	2-$NHCOC_6H_5$	2.05	1.90	1.95	43.	1-$NHCOC_6H_5$; 4-$NHCOC_6H_5$; 5-$NHCOC_6H_5$;	4.56	4.17	*d*
19.	1-NH_2; 5-NH_2	2.31	2.37	*d*	44.	1-[4-$CH_3C_6H_4CONH$-]; 4-[4-$CH_3C_6H_4CONH$-]	4.57	4.08	4.05
20.	1-$NHCOC_6H_5$	2.37	2.85	2.91	45.	1-[3-ClC_6H_4CONH-]; 4-[3-ClC_6H_4CONH-]	4.98	4.84	*d*
21.	1-NH_2; 5-$NHCOC_6H_5$	2.39	2.24	*d*	46.	1-[4-ClC_6H_4CONH-]; 4-[4-ClC_6H_4CONH-]	5.20	4.45	4.44
22.	1-$NHCOC_6H_5$; 8-$NHCOC_6H_5$	2.60	3.21	3.57	47.	A[e]	6.80		
23.	1-[4-$CH_3C_6H_4CONH$-]	2.68	3.11	*d*	48.	B	4.70	4.49	*d*
24.	1-[3-ClC_6H_4CONH-]	2.73	3.46	3.46	49.	1-$NHCOC_6H_5$; 4-OH; 5-$NHCOC_6H_5$; 8-OH	4.49	2.88	2.74
25.	1-[3-$CH_3C_6H_4CONH$-]	2.74	3.07						

[a] Experimental affinities acc. to ref 8. [b] CoMSA (size of Kohonen maps: *15* × 15; MD = 0.2; non-nen). [c] External predictions, details in text. [d] Training series, details in text. [e] The compound was excluded from the CoMSA model.



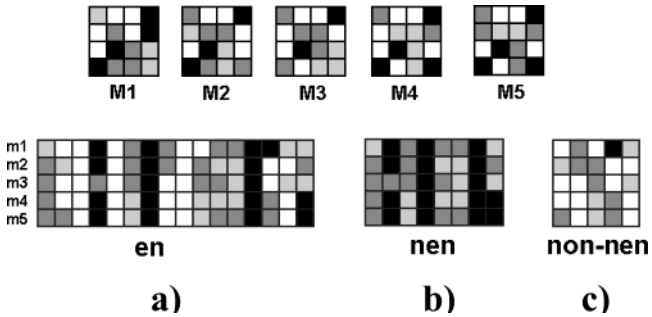


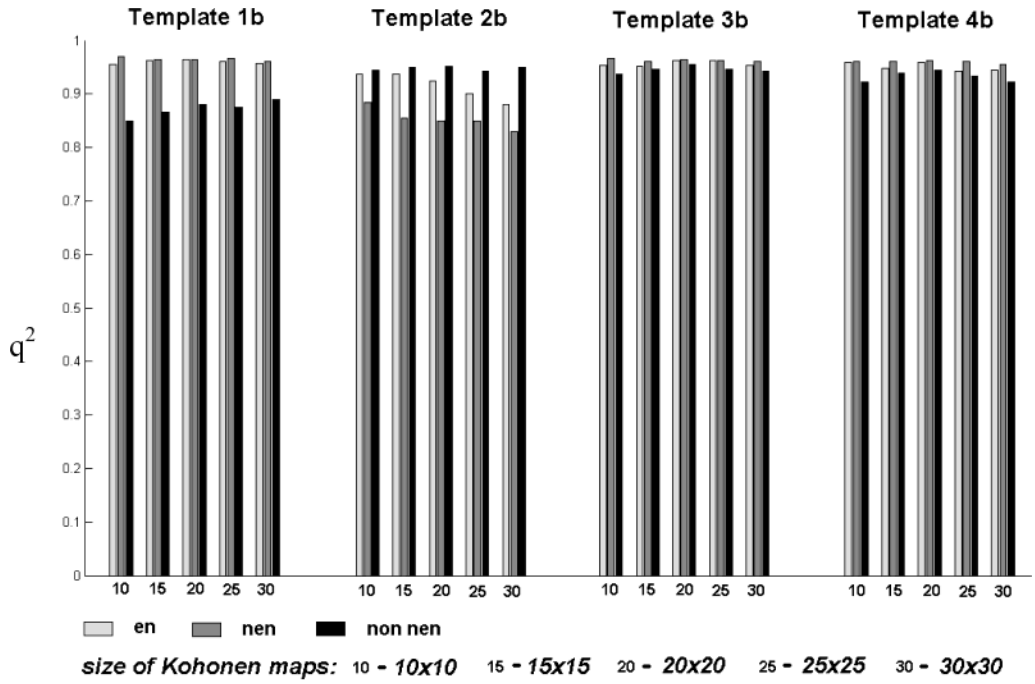
Figure 1. A schematic illustration of preprocessing the electrostatical potential matrices (M1−M5) resulted from the comparative Kohonen mapping. (a) The matrices that contain all elements including empty neurons (white boxes). (b) The matrices after eliminating columns with empty neurons. (c) The matrices included only such columns which have at least one empty neuron in any of the (M1−M5) matrices.

template molecule with the respective atoms of the analyzed molecule is performed to achieve molecular superimposition. In practice, we used Match3D program[37] to carry out this operation. The KMAP 3.0 program[37] was used for the simulation of the Kohonen networks. The size of these networks was varied from *10* × *10* to *30* × *30* neurons with step 5. The output from this program was used for the calculation of the mean electrostatic potential values within each neuron, and respective feature maps were transformed to respective N^2 element vectors, where N is the number of neurons forming the Kohonen map.

PLS Analysis. Vectors obtained were processed by the PLS analysis with a leave-one-out cross-validation procedure. The PLS procedures were programmed within the MATLAB environment (MATLAB). A PLS model was constructed for the centered data, and its complexity was estimated based on the leave-one-out cross-validation procedure (CV). The date was recentered (but not rescaled) for each cross-validation run. In the leave-one-out CV one repeats the calibration m times, each time treating the ith left-out object as the prediction object. The dependent variable for each left-out object is calculated based on the model with one, two, three, etc. factors. The Root Mean Square Error of CV for the model with j factors is defined as

$$\mathrm{RMSECV}_j = \sqrt{\frac{\sum(\mathrm{obs}_i - pred_{i,j})^2}{m}} \qquad (2)$$

where obs denotes the assayed value; pred is the predicted



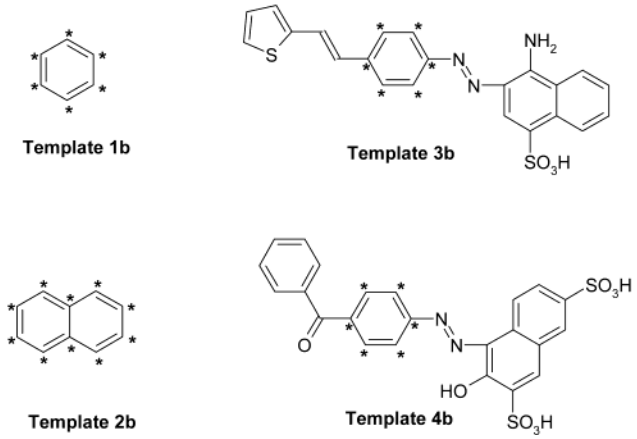


Figure 2. Comparison of the CoMSA models obtained for a series of azo dyes for four different templates. The numbers indicate the size of Kohonen maps, and the asterisks show the atoms that are covered during molecular superimposition.

value of dependent variable and *i* refers to the object index, which ranges from 1 to *m*. Model with *k* factors, for which RMSECV reaches a minimum, is considered as an optimal one.

We used the performance metrics that are accepted and widely used in CoMFA analyses, i.e., cross-validated q^2_{CV}

$$q_{CV}^{\ 2} = 1 - \frac{\sum(\text{obs}_i - \text{pred}_i)^2}{\sum(\text{obs}_i - \text{mean(obs)})^2} \qquad (3)$$

where obs is the the assayed value, pred is the predicted value, mean is the mean value of obs, and *i* refers to the object index, which ranges from 1 to *m*; and cross-validated standard error *s*

$$s = \sqrt{\frac{\sum(\text{obs}_i - \text{pred}_i)^2}{m - k - 1}} \qquad (4)$$

where *m* is the number of objects and *k* is the number of PLS factors in the model.

The quality of external predictions was measured by the SDEP parameter

$$\text{SDEP} = \sqrt{\frac{\sum(\text{pred}_i - \text{obs}_i)^2}{n}} \qquad (5)$$

where pred is the predicted value, obs is the observed value, and *n* is the number of compounds included in the test series.

RESULTS AND DISCUSSION

Comparative Molecular Surface Analysis of Azo and Anthraquinone Dyes. Figure 2 compares the CoMSA models obtained for a series of azo dyes. These models are much better than the CoMFA ones reported by Timofei et al.[11,12] Thus, the q^2 performance for the CoMFA models ranges from 0.444 to 0.731, while for our best models this parameter take a values of $q^2 = 0.829 - 0.970$. Practically, the CoMSA q^2 performances do not depend on the template indicated. This fact makes an important difference to the

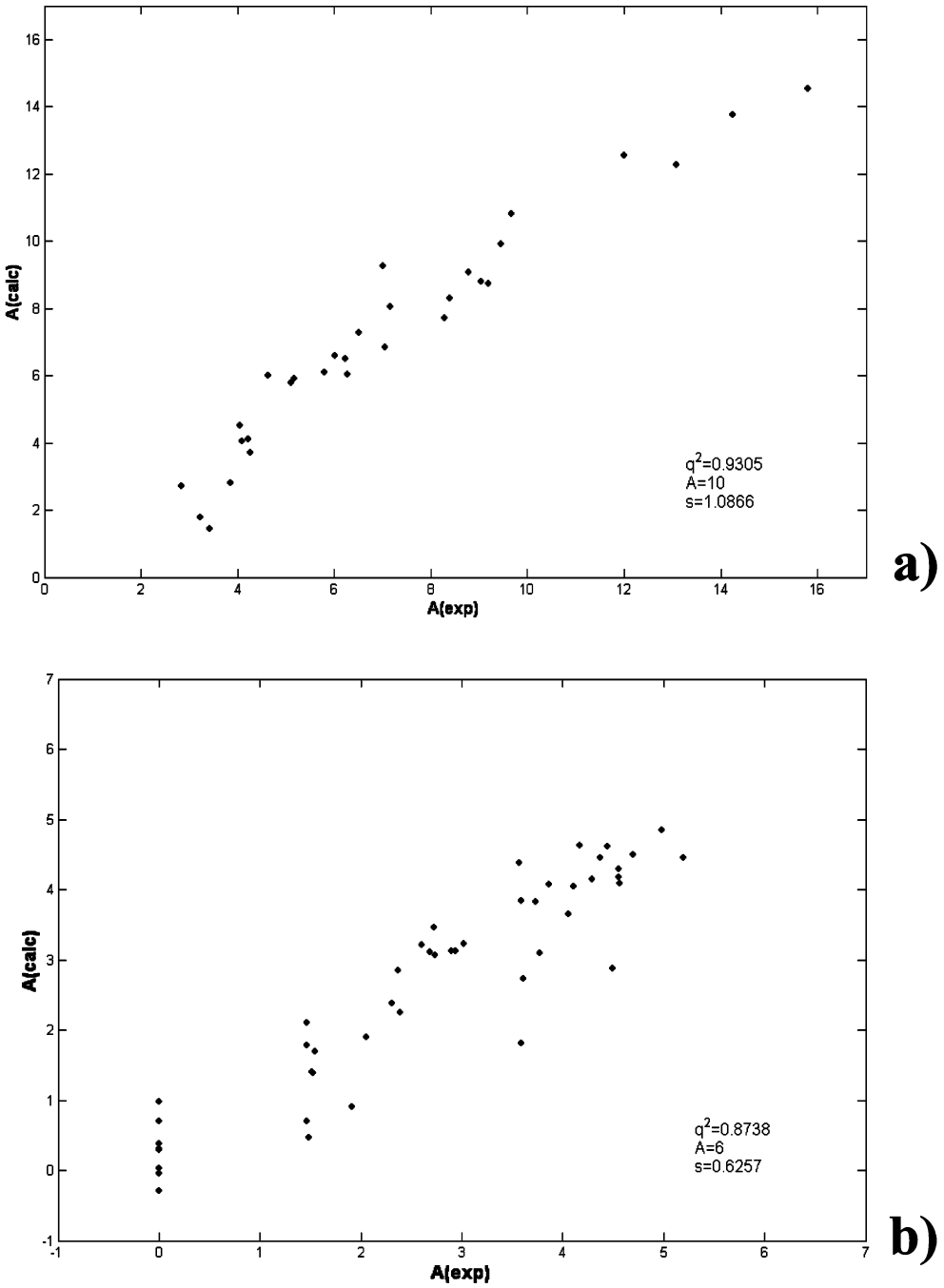


Figure 3. The CoMSA models of dye affinity to cellulose. Plot of experimental versus calculated dye affinities for (a) azo and (b) anthraquinone dyes, respectively.

results that are usually obtained for drug molecules, for which the selection of the template generally influences the result.[20] PLS analysis of electrostatic potential matrices with all variables provides slightly worse models than the PLS analysis of the matrices containing only nonzeros (nen-matrices) elements (compare Figure 2). This indicates that electrostatic potential better accounts for the fiber−dye affinity than shape and electrostatic properties together. This effect can be observed for almost all the templates tested. Such results seem to correspond well with the CoMFA conclusions which indicate that electrostatic field dominates the steric one.[8,9,12] We decided, however, to further verify this fact by analyzing the electrostatic potential matrices that would include only elements that obey the rule of having at least one zero value in any one of the comparative pattern (non-nen). Inversely to the nen-matrices, such matrices should emphasize shape effects. The black boxes in Figure 2 record the CoMSA q^2 performance for the non-nen protocol. Quite surprisingly these performances reach almost the same values as those obtained for the *en* and *nen* protocols (compare in Figure 2), which is especially true for the templates that comprise a larger part of the molecules. This indicates that shape factors are also quite important for resolving dying affinity. Figure 3a shows the relationship between the experimental affinity and the predicted from the *nen* model CoMSA.

Figure 4 compares the CoMSA models obtained for a series of anthraquinone dyes. The q^2 performances of the models range from 0.68 to 0.88, depending upon the template used for comparative mapping. This compares nicely with

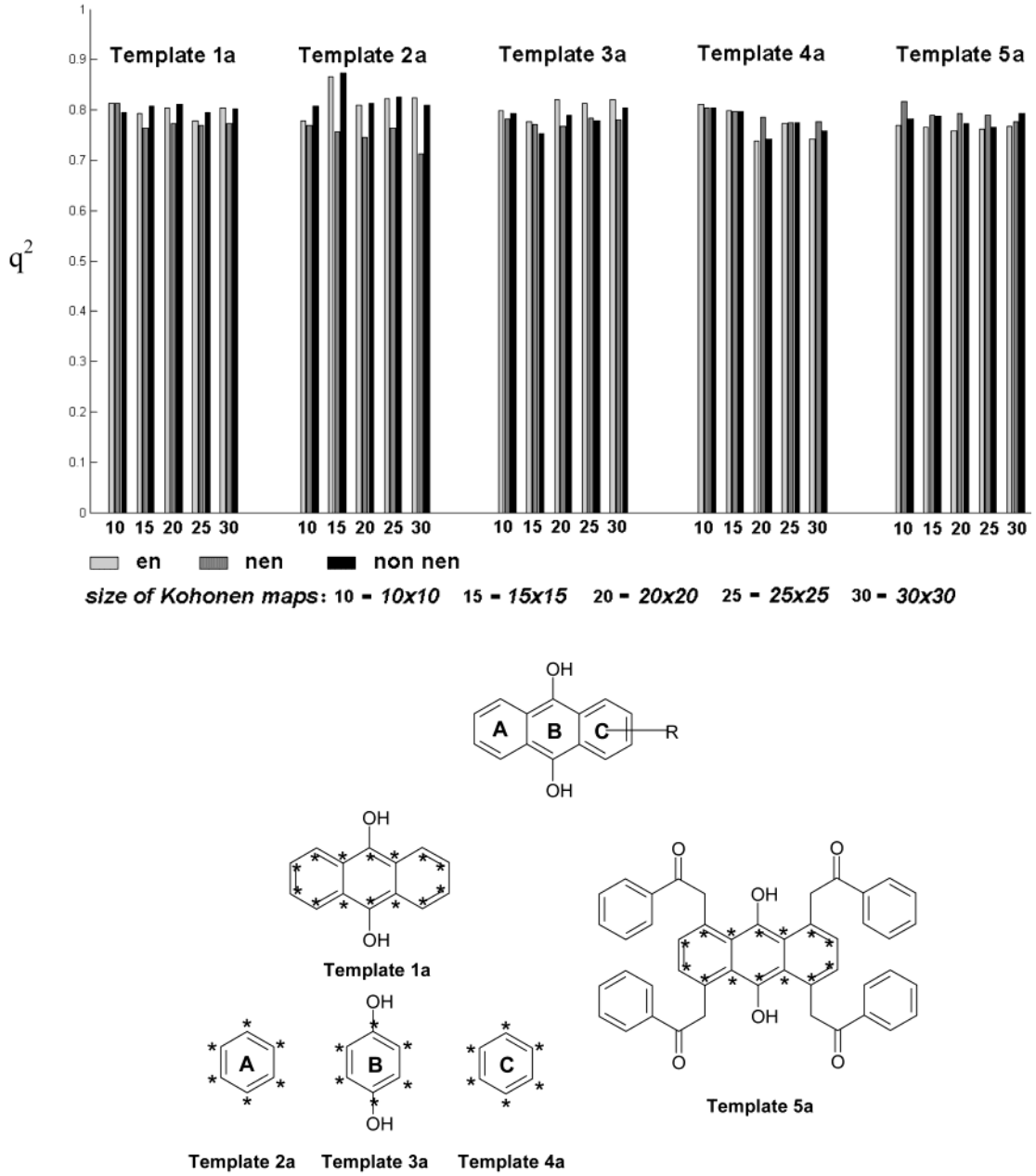


Figure 4. Comparison of the CoMSA models obtained for a series of anthraquinone dyes for five different templates. The numbers indicate the size of Kohonen maps, and the asterisks show the atoms that are covered during molecular superimposition.

the performance of the CoMFA reported by Fabian et al.[9] ($q^2 = 0.860$). Similarly to the azo dyes these values only slightly depend on the template selected, and CoMSA modeling ability only slightly depends on the type of the electrostatic potential matrix (*en, nen, non-nen*) analyzed. A plot of experimental vs calculated affinity for the best model *non-nen* CoMSA is shown in Figure 3b. Also similarly, to the azo dyes both *en* and *non-nen* modeling, i.e., differentiating the relative influence of steric and electrostatic factors, does not change the q^2 statistics.

As the current knowledge on 3D QSAR modeling requires further verification of the model performance by external modeling. In the current work we used the standard deviation of error of prediction (SDEP) for the estimation the model predictivity. Thus we divided both azo and anthraquinone dyes into two groups: training and test set (each consisting 50% of all compounds). Then we used the training group to predict the affinity for the test group. The accurate prediction values for azo and anthraquinone dyes are shown in Table 1 (column 7) and Table 2 (columns 5 and 10), respectively. The SDEP parameter values amount to 0.644 and 1.024 for the CoMSA models of azo and anthraquinone dyes, respectively. In addition, we used the Galbraikh−Tropsha criterion[38] to verify the predictivity of the best models, which is shown in Figure 5. This clearly proves that CoMSA provides reliable high predictive models for both series.

CoMSA for Verification of a Pharmacophore Concept for the Description of the Dye−Cellulose Interactions. Results from the previous CoMFA analyses of the azo dye series on the applicability of the pharmacophore concept were controversial. It was observed[10] that modeling performance was insensitive to the way in which molecule atoms had been covered. This together with the fact that a two-dimensional

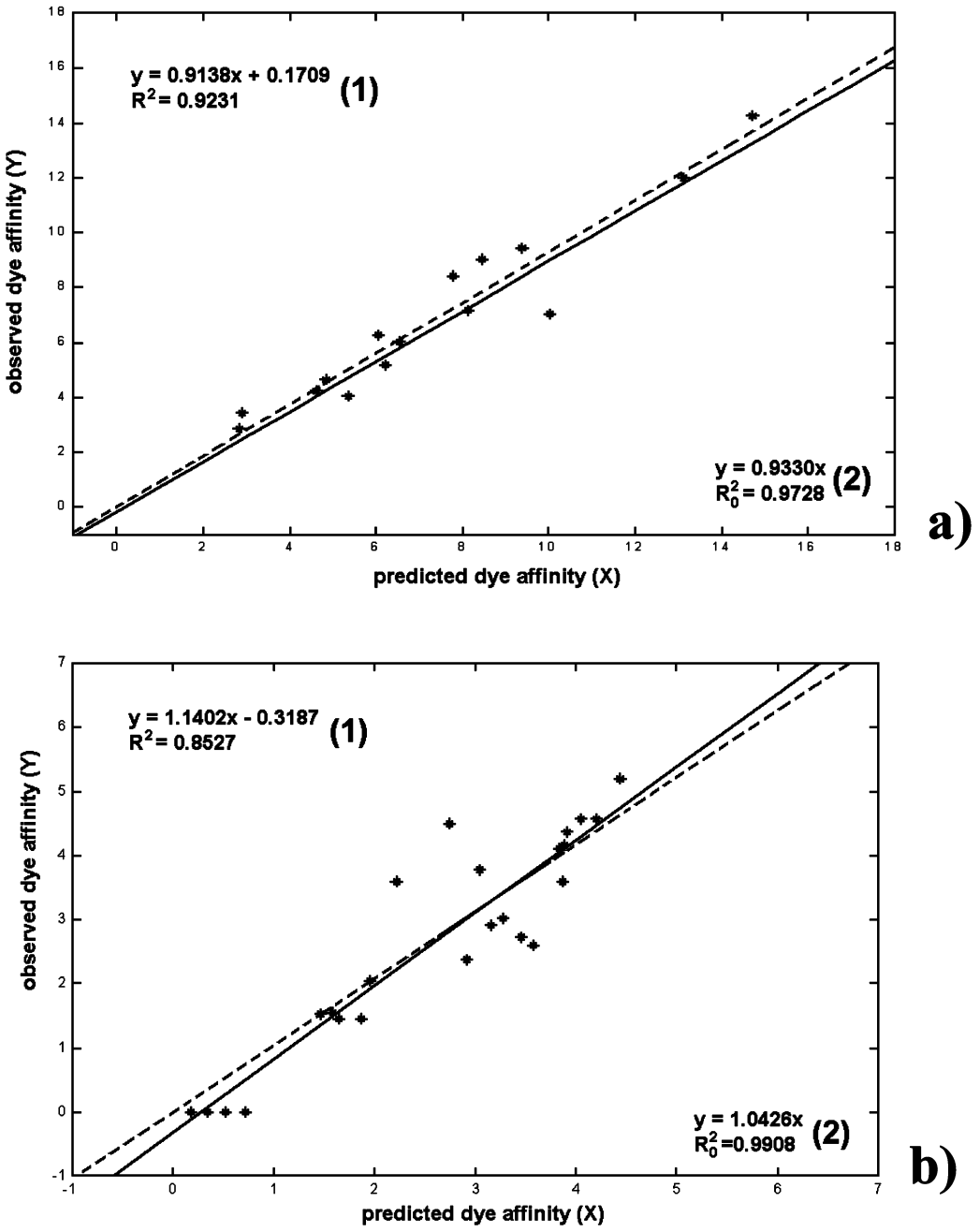


Figure 5. Validation of the best model by the Golbraikh−Tropsha criteria for azo (a) and anthraquinone (b) dyes. We tried to keep the style of the presentation of the authors.[38] The regression between observed (Y) and predicted (X) activity values for the test set. The solid line shows the regressional equation given by (1). The dotted line illustrates the regression without the bias (2). The closer are these linear plots, the better is the model predictivity. Calculations after

$$\mathrm{pred}_i^{\,0} = k^*\mathrm{pred}_i$$

$$k = \frac{\sum \mathrm{obs}_i \mathrm{pred}_i}{\sum \mathrm{pred}_i^{\,2}}$$

$$R_0^{\,2} = 1 - \frac{\sum (\mathrm{pred}_i - \mathrm{pred}_i^{\,0})^2}{\sum (\mathrm{pred}_i - \mathrm{mean(pred)})^2}$$

where upper index 0 relates to regression observed (Y) vs predicted (X); k is the slope of the regression through the origin (2); and R_0^2 is the correlation coefficient for the regression of observed (Y) vs predicted (X) without bias. $[(R^2 - R_0^2)/R^2] < 0.1$ and $0.85 \le k \le 1.15$ as recommended by Golbraikh and Tropsha.[38]

descriptor is capable of a better description of the dye−fiber affinity lead the authors to questioning the pharmacophore theory of dye adsorption.[11] However, in their more recent study they concluded that a *specific affinity of binding, opposite to previous findings, even if it is less specific in comparison with ligand-biological interactions* exists.[11]

Accordingly, the description of the dye affinity by a 3D model as well as interpretation of this model is not an easy

task. The same goes for the CoMSA models obtained in our current work. It is very surprising that we obtained very predictive models, for all the templates tested. The q^2 values and predictivity depends only to a minor extent upon the template selected. Also, the affinity can be modeled both by the inclusion of the steric and electrostatic interactions, irrespective of if they are separated or brought together. We have never observed such a situation for the drug pharmacophores.[20] We propose the following explanation of these results. It seems that electrostatic property is a decisive factor in determining binding affinity and the importance of shape could be clarified only if we realized that a molecule shape is decisive for the distribution of the surface electrostatic potential. The example of the dye molecules interacting with the relatively large supramolecular cellulose receptor cavity clearly illustrates the fact that two basic types of shape effects can operate during receptor−ligand recognition. First are the effects of steric hindrance. Steric effects of this type are very restrictive and even very minute differences in the molecular configuration can result in large activity changes. The smaller the receptor cavity size, the larger is the resulted steric hindrance. As it is the main case for the biopolymer drug receptors such a situation is typical during drug pharmacophore investigations. Probably, it is also this effect in such cases that makes CoMFA and related method extremely sensitive to the way in which molecular superposition is being performed. On the other hand, molecular shape is also important for the determination of the electrostatic potential in the molecular environment. It can be speculated that it is the case for the dye−cellulose interactions where a relatively large cellulose cavity does not create steric hindrance to the dye molecules. Thus, we do not necessarily need to explain the results that are obtained in previous studies by hypothesizing that molecule shape is not the activity-determining factor. It is rather a completely different type of shape effects that should be taken into account in this case.

An interesting remark concluded during the evaluation of this publication comes from an anonymous reviewer. Thus, *cellulose in contrast to the active site of an enzyme does not represent a unique and more or less singular area of binding but contains crystalline, polymorphs as well as amorphous regions.* This can suggest that *different dyes can probably bind to different parts of cellulose material*. On the other hand, the interactions of the binding dye molecule and such a material significantly differ from those appearing during drug-receptor recognition. However, these interactions must be quite specific, because the PLS analysis of the electrostatic potential on the molecular surface enables us to obtain highly predictive models of binding affinity. High CoMSA performances, independent of the template used, seems to indicate that all molecular surface areas are important during dye−fiber interaction. Moreover, high performances indicate that the shape of the dye molecules meets the requirements of the cellulose cavities.

Consequently, it seems that for the dye series analyzed, a pharmacophore significantly differs from the classical drug pharmacophore mostly in that the interactions taking place are transmitted through the whole molecular environment. This is probably the most important difference to the drug pharmacophore that can be represented efficiently by a few point model.

CONCLUSIONS

Despite recent investigations aimed at modeling 3D QSAR for dye molecules a controversy still exists: can a pharamacophore hypothesis be used for such purposes. In the present publication we reported on the application of the CoMSA method for modeling 3D QSAR of azo and anthraquinone dyes. We obtained very predictive models which significantly outperform those reported in the previous CoMFA studies, especially for the azo dyes. Our results proved previous conclusion that steric requirements are far less pronounced for the cellulose cavities than for the classical drug receptor. Moreover, our results indicate that all molecular surface segments are important for dye−fiber interactions, which also makes an important difference in relation to the classical drug pharmacophore. On the other hand, high predictivity of the CoMSA models indicates that a pharmacophore concept is suitable for the description of the dye−fiber interactions. However, this pharmacophore must substantially differ from the drug pharmacophore used for the illustration of the drug-receptor interactions. From a theoretical point of view dye−cellulose interactions can be an interesting case in which shape decides the activity rules not by the steric repulsion but as a cofactor determining the electrostatic potential distribution. We hope that in a near future our studies will enable us to design new effective dyes.

ACKNOWLEDGMENT

The authors thank Professor Johann Gasteiger of the University of Erlangen-Nürnberg, BRD both for his valuable discussion and for facilitating access to the programs of CORINA, PETRA, SURFACE, and KMAP. The financial support of the KBN Warsaw, grants nos. T08E02820, PBZ KBN - 040 P04/08, and 4 T09A 034 24, is gratefully acknowledged.

REFERENCES AND NOTES

(1) Peters, R. H. *Textile chemistry. The physical chemistry of dyeing*; Elsevier: Amsterdam, 1975; Vol. III.

(2) Timofei, S.; Schmidt, W.; Kurunczi, L.; Simon, Z. A Review of QSAR for Dye Affinity for Cellulose Fibres. *Dyes Pigm.* **2000**, *47*, 5−16.

(3) French, A. D.; Battista, O. A.; Cuculo, J. A.; Gray, D. G. In *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th ed.; Wiley: New York, 1993; Vol. 5, p 476.

(4) Timofei, S.; Schmidt, W.; Kurunczi, L.; Simmon, Z.; Sallo, A. A QSAR Study of the Adsorption by Cellulose Fibre of Anthraquinone Vat Dyes. *Dyes Pigm.* **1994**, *24*, 267−279.

(5) Timofei, S.; Kurunczi, L.; Schmidt, W.; Fabian, W. M. F.; Simon, Z. Structure-Affinity Binding Relationships by Principal Component Regresion Analysis of Anthraquinone Dyes. *Quant. Struct. Act. Relat.* **1995**, *14*, 444−449.

(6) Timofei, S.; Kurunczi, L.; Schmidt, W.; Simon, Z. Structure-Affinity Binding Relationships of Some 4-Aminobenzene Derivatives for Cellulose Fibre. *Dyes Pigm.* **1995**, *29*, 251−258.

(7) Timofei, S.; Kurunczi, L.; Schmidt, W.; Simon, Z. Lipophilicity in Dye-Cellulose Fibre Binding. *Dyes Pigm.* **1996**, *32*, 25−42.

(8) Fabian, W. M. F.; Timofei, S.; Kurunczi, L. Comparative Molecular Field Analysis (CoMFA), Semiempirical (AM1) Molecular Orbital and Multiconformational Minimal Steric Difference (MTD) Calculation of Anthraquinone Dye-Fibre Affinities. *J. Mol. Struct. THEOCHEM* **1995**, *340*, 73−81.

(9) Fabian, W. M. F.; Timofei, S. Comparative Molecular Field Analysis (CoMFA) of Dye-Fibre Affinities II: Symmetrical Bisazo Dyes. *J. Mol. Struct. THEOCHEM*. **1996**, *362*, 155−162.

(10) Oprea, T. I.; Kurunczi, L.; Timofei, S. QSAR Studies of Disperse Azo Dyes. Towards the Negation of the Pharmacophore Theory of Dye − Fibre Interaction? *Dyes Pigm.* **1997**, *33*, 44−64.

(11) Funar-Timofei, S.; Schrüümann, G. Comparative Molecular Field Analysis (CoMFA) of Anionic Azo Dye-Fiber Affinities I: Gas-Phase Molecular Orbital Descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 788−795.
(12) Timofei, S.; Fabian, W. M. F. Comparative Molecular Field Analysis (CoMFA) of Heterocyclic Monoazo Dye-Fibre Affinities. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1218−1222.
(13) Kubinyi, H. QSAR and 3D QSAR in Drug Design. Part 1: Methodology. *Drug Discovery Today* **1997**, *2*, 457−467.
(14) Kubinyi, H. QSAR and 3D QSAR in Drug Design. Part 2: Applications and Problems. *Drug Discovery Today* **1997**, *2*, 538−546.
(15) Kubinyi, H. QSAR: Hansch Analysis and Related approaches. In *Methods and Principles in Medicinal Chemistry*; Mannhold, R., Krokgsgaard- Larsen, P., Timmerman, H., Eds.; VCH: Weinheim, 1993.
(16) Katrizky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M. Structurally Diverse Quantitative Structure − Property Relationship Correlations of Technologically Relevant Physical Properties. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1−18.
(17) Polanski, J. The Mapping of the Molecular Surfaces by Means of Self-organizing Neural Networks within MATLAB 5.2 for WINDOWS-95. *Acta Pol. Pharm.* **1999**, *56*, 80−84.
(18) Polanski, J.; Walczak, B. The Comparative Molecular Surface Analysis (CoMSA): A Novel Tool for Molecular Design. *Comput. Chem.* **2000**, *24*, 615−625.
(19) Polanski, J.; Gieleciak, R.; Bąk, A. The Comparative Molecular Surface Analysis (CoMSA) − a Nongrid 3D QSAR Method by a Coupled Neural Network and PLS System: Predicting pK_a Values of Benzoic and Alkanoic Acids. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 184−191.
(20) Polanski, J.; Gieleciak, R. The Comparative Molecular Surface Analysis (CoMSA) with Modified Uninformative Variable Elimination-PLS (UVE-PLS) Method: Application to the Steroids Binding the Aromatase Enzymol. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 656−666.
(21) Gerstein, M.; Greenbaum, D.; Luscombe, M. N. What is Bioinformatics? A Proposed Definition and Overview of the Field. *Method Inform. Med.* **2001**, *40*, 346−358.
(22) Brazma, A.; Vilo, J. Gene Expression Data Analysis. *FEBS Lett.* **2000**, *480*, 17−24.
(23) Toronen, P.; Kolehmainen, M.; Wong, G.; Castren, E. Analysis of Gene Expression Data Using Self-Organizing Maps. *FEBS Lett.* **1999**, *451*, 142−146.
(24) Anzali, S.; Gasteiger, J.; Holzgrabe, U.; Polanski, J.; Teckentrup, A.; Wagener M. The Use of Self-Organizing Neural Networks in Drug Design. *Perspect. Drug Discov. Design* **1998**, *9/10/11*, 273−299.
(25) Zupan, J.; Gasteiger, J. *Neural Networks and Drug Design for Chemists*, 2nd ed.; VCH: Weinheim, 1999.
(26) Gasteiger, J.; Li, X.; Rudolph, Ch.; Sadowski J.; Zupan, J. The Representation of Molecular Electrostatic Potentials by Topological Feature Maps. *J. Am. Chem. Soc.* **1994**, *116*, 4608−4620.
(27) Polanski, J. The Receptor-Like Neural Network for Modeling Corticosteroid and Testosterone Binding Globulins. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 478−484.
(28) Anzali, S.; Barnickel, G.; Krug, M.; Sadowski, J.; Wagener, M.; Gasteiger, J.; Polanski, J. The Comparison of Geometric and Electronic Properties of Molecular Surfaces by Neural Networks: Application to the Analysis of Corticosteroid Globulin Activity of Steroids. *J. Comput.-Aided Mol. Design* **1996**, *10*, 521−540.
(29) Polanski, J.; Gasteiger, J.; Jarzembek, K. Self-Organizing Neural Networks for Screening and Development of Novel Artificial Sweetener Candidates. *Combin. Chem. High Throughput Screen.* **2000**, *3*, 481−495.
(30) Hasegawa, K.; Matsuoka, S.; Arakawa, M.; Funatsu, K. New molecular surface-based 3D-QSAR method using Kohonen neural network and 3-Way PLS. *Comput. Chem*. **2002**, *26*, 583−589.
(31) Gasteiger, J. CORINA for the information, see: http://www.mol-net.de.
(32) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(33) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(34) Gastaiger, J.; Saller, H. Calculation of the Charge Distribution in Conjugated Systems by a Quantification of the Resonance Concept. *Angew. Chem.* **1985**, *97*, 699−701.
(35) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity − a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(36) Kohonen, T. *Self-Organization and Associative Memory*, 3rd ed.; Springer: Berlin, 1989.
(37) Gasteiger, J. Match3D; KMAP for the information, see: http://www2.ccc.uni-erlangen.de.
(38) Golbraikh, A.; Thropsha, A. Beware of q^2! *J. Mol. Graph. Mod.* **2002**, *20*, 269−276.

CI0340761