Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 316 (2006) 84–97
Journal of Hydrology
www.elsevier.com/locate/jhydrol

A quantitative structure–activity relationships (QSAR) analysis of triarylmethane dye tracers

Jarai Mon, Markus Flury*, James B. Harsh

Department of Crop and Soil Sciences, Center for Multiphase Environmental Research, Washington State University, Pullman, WA 99164, USA

Received 1 December 2004; revised 7 April 2005; accepted 29 April 2005

Abstract

Dyes are important hydrological tracers. Many different dyes have been proposed as optimal tracers, but none of these dyes can be considered an ideal water tracer. Some dyes are toxic and most sorb to subsurface materials. The objective of this study was to find the molecular structure of an optimal water tracer. We used QSAR to screen a large number of hypothetical molecules, belonging to the class of triarylmethane dyes, in regard to their sorption characteristics to a sandy soil. The QSAR model was based on experimental sorption data obtained from four triarylmethane dyes: C.I. Food Blue 2 (C.I. 42090; Brilliant Blue FCF), C.I. Food Green 3 (C.I. 42053; FD&C Green No. 3), C.I. Acid Blue 7 (C.I. 42080; ORCOacid Blue A 150%), and C.I. Acid Green 9 (C.I. 42100; ORCOacid Fast Green B). Sorption characteristics of the dyes to the sandy soil were expressed with the Langmuir isotherm. Our premise was that dye sorption can be reduced by attachment of sulfonic acid (SO_3) groups to the triarylmethane template. About 70 hypothetical dyes were created and QSAR were used to estimate sorption characteristics. The results indicated that both the position and the number of SO_3 groups affected dye sorption. Sorption decreased with increasing number of SO_3 groups attached to the molecule. Increasing the number of sulfonic acid groups also decreases the toxicity of the compounds. An optimal triarylmethane water tracer contains 4 to 6 SO_3 groups.

Keywords: Dyes; Hydrological tracers; Water tracing; QSAR; Sorption

1. Introduction

Dye tracers are frequently used in hydrology to measure groundwater flow velocity, identify flow directions, hydraulic connections, and the pattern of water movement (Drew, 1968; Smart and Laidlaw, 1977; Davis et al., 1980; McLaughlin, 1982; Flury and Wai, 2003). An ideal dye to trace the movement of water should move conservatively, i.e. without interacting with the solid phase and without decaying during the time period of the tracer test. By and large, however, dyes react with subsurface media, and the fate and behavior of dyes in the subsurface depend on the environmental conditions, i.e. property of the media and solution chemistry (Smart and Laidlaw, 1977; Flury and Flühler, 1995; German-Heins and Flury, 2000). A good dye tracer under one set of environmental conditions may not perform

* Corresponding author. Tel.: +1 509 335 1719; fax: +1 509 335 8674.
E-mail address: flury@mail.wsu.edu (M. Flury).

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2005.04.027

as well in a different environment. Therefore, dye tracers should be evaluated for their suitability in specific applications.

Thousands of dyes are commercially available (Colour Index, 2001). Because experimental screening of a large number of dyes to find a good tracer for specific environmental conditions is not feasible, alternative methods for screening are needed. Quantitative structure–activity relationships (QSAR) offer the possibility for screening a large number of chemicals in a short time and with low cost. The QSAR establish a statistical relationship between biological activity or environmental behavior of the chemicals of interest and their structural properties (Sabljić, 1989; Hansch and Fujita, 1995). Using QSAR, we can obtain an estimate of the activity of a chemical from its molecular structure only.

QSAR have been successfully applied to predict soil sorption coefficients of non-polar and non-ionizable organic compounds including many pesticides (Gerstl and Helling, 1987; Pussemier et al., 1989). Sorption of organic chemicals in soils or sediments is usually described by sorption coefficients. QSAR models with the first-order molecular connectivity index have been used to estimate soil sorption coefficients of non-polar organic chemicals (Sabljić, 1987, 1989, 2001; Meylan et al., 1992). However, the first-order molecular connectivity model was insufficient for predicting sorption coefficients of polar organic compounds (Sabljić, 1989; Sekusak and Sabljić, 1992). Meylan et al. (1992) suggested that sorption coefficients of some polar organic compounds could be successfully predicted with a model that includes molecular connectivity indices (MCIs) and polarity correction factors. Sekusak and Sabljić (1992) reported that a QSAR model consisting of the second-order MCI and a descriptor representing the polar part of a molecule was able to describe the log K_{oc} of amides, whereas the first-order MCI alone failed. High-order MCIs are possible descriptors to identify pesticides susceptible to leaching through soils (Worrall, 2001).

Sorption coefficients normalized by organic carbon content (log K_{oc} or log K_{om}) have been used in several QSAR models (Pussemier et al., 1989; Sabljić, 1987, 1989, 2001; Meylan et al., 1992). Soil sorption of ionizable, hydrophobic organic chemicals could be estimated using a QSAR model which included the octanol/water partition coefficient, organic carbon content, and correction factors for acids and bases (Bintein and Devillers, 1994). The K_{oc}, however, is not an appropriate parameter for describing sorption of ionizable, polar molecules, such as organic dye tracers. Sorption of ionizable organic compounds in soils or sediments depends on the concentration of chemicals, solution pH, ionic strength, and medium properties (Schwarzenbach et al., 2003). Obtaining a soil sorption QSAR model that accounts for differences in solution chemistry and media properties may be impossible. Moreover, sorption models derived from a specific chemical class under specific environmental conditions would not be valid for general use. QSAR models are specific to chemicals and their interactions with the targeted environment (Nendza, 1998).

Dyes are grouped into classes based on their chemical structures or their methods of application. A number of hydrological tracers are from the class of triarylmethane dyes (Flury and Wai, 2003). This class includes hundreds of dyes that are mostly red, violet, blue, or green in color and is one of the largest dye groups (Colour Index, 1971). The class includes C.I. Food Blue 2 (Brilliant Blue FCF), which is a well-known vadose zone tracer, but may also contain dyes that are better tracers than C.I. Food Blue 2. In addition, new and better dye tracers could be designed by structural modification of existing tracers.

Dyes that contain more sulfonic acid groups are generally less sorbed to soils than those with fewer SO_3 groups (Corey, 1968). Testing the adsorption of dyes and selected intermediates to activated carbon, Reife and Freeman (1996) found that chemicals that contain more sulfonic acid groups in their molecular structures were less sorbed. These results suggest that as more sulfonic acid groups are attached to the molecular template, dyes become more water-soluble and more readily move with the water. Not only the number but also the positions of SO_3 groups in the structures may influence the sorption of dyes. At neutral pH, orthanilic acid (one sulfonic acid group attached at the *ortho* position on aniline) sorbed much more strongly to activated carbon than that of metanilic acid (one sulfonic acid group attached at the *meta* position) (Reife and Freeman, 1996). Similarly, the *para* isomer of Rhodamine WT exhibits less sorption to aquifer materials than does its *meta* isomer (Vasudevan et al., 2001).

The goal this study was to identify the structure of a dye best suited for tracing water flow in porous media. A best-suited dye tracer will sorb as little as possible to the porous medium to accurately trace the flow pathways. We based our selection of dyes on the class of triarylmethane dyes. The specific objective was to examine the sorption properties of dyes consisting of the same molecular template as C.I. Food Blue 2 but with different numbers and positions of sulfonic acid groups. Four commercially available triarylmethane dyes were used to develop a QSAR model that allowed us to predict the sorption characteristics of hypothetical dyes with similar molecular structure.

2. Theory

QSAR development requires three basic elements: (1) an activity or property data set, measured experimentally, (2) molecular descriptors, which are the quantitative descriptions of structural properties, and (3) statistical techniques to establish the relationship between molecular descriptors and activities. QSAR analysis consists of statistical modeling and, consequently, QSAR results are associated with some uncertainty and the predictive power of a QSAR model is related not only to the quality of the input data but also the power of the statistics (Hall et al., 2002b).

The first step in a QSAR study is the selection of a group of compounds for which experimental data are available. This group of compounds, the training set, should cover the whole range of compounds whose activities/properties are to be modeled (Hall et al., 2002b). The QSAR analysis is further based on the calculation of appropriate molecular descriptors. Many different descriptors are available, but only few of them might be used as independent variables in a QSAR model. Appropriate descriptors are selected from previous knowledge of interaction processes or by statistical techniques, i.e. by choosing descriptors that are highly correlated to the experimental parameters of interest.

One group of molecular descriptors, the MCIs, have been successfully used in QSAR to describe sorption of some hydrophobic chemicals in soils (Sabljić, 2001; Sabljić and Protic, 1982). The indices quantify the structure of a molecule by coding the electrons present in the whole molecule (Kier and Hall, 1999) and carry much structural information including intermolecular accessibility (Kier and Hall, 2000). There are two basic types of MCIs: simple MCIs and valence MCIs, which can be calculated as (Hall and Kier, 2001; Kier and Hall, 1986):

$$^1\chi = \sum(\delta_i\delta_j)^{-0.5} \tag{1}$$

and

$$^1\chi^v = \sum(\delta_i^v\delta_j^v)^{-0.5} \tag{2}$$

where $^1\chi$ is the first-order simple MCI; δ_i, δ_j are the counts of sigma electrons in adjacent non-hydrogen atoms i, j; $^1\chi^v$ is the first-order valence MCI; and $\delta_i^v\delta_j^v$ are the counts of all valence electrons in adjacent non-hydrogen atoms i, j. The lower MCIs (zeroth- to second-order) encode global or bulk properties, such as molecular size, and the higher-order MCIs (> second-order) encode local structural properties, e.g., *para* substitutions (Hall and Kier, 2001; Sabljić, 2001).

3. Materials and methods

3.1. QSAR model development

The class of triarylmethanes includes a number of dyes that are often used as hydrological tracers (Flury and Wai, 2003). Four members of this class, C.I. Food Blue 2 (Brilliant Blue FCF, C.I. 42090; CAS Reg. Nr. 3844-45-9; $C_{37}H_{34}N_2O_9S_3Na_2$), C.I. Food Green 3 (FD&C Green No. 3, C.I. 42053; CAS Reg. Nr. 2353-45-9; $C_{37}H_{34}N_2O_{10}S_3Na_2$), C.I. Acid Blue 7 (ORCOacid Blue A 150%, C.I. 42080; CAS Reg. Nr. 3486-30-4; $C_{37}H_{35}N_2O_6S_2Na$), and C.I. Acid Green 9 (ORCOacid Fast Green B, C.I. 42100; CAS Reg. Nr. 4857-81-2; $C_{37}H_{34}N_2ClO_6S_2Na$), were selected as a training set for QSAR model development. These dyes share the same molecular template but differ in the number and position of functional groups (Fig. 1).

Sorption properties of these dyes to a sandy soil were determined with laboratory batch experiments.

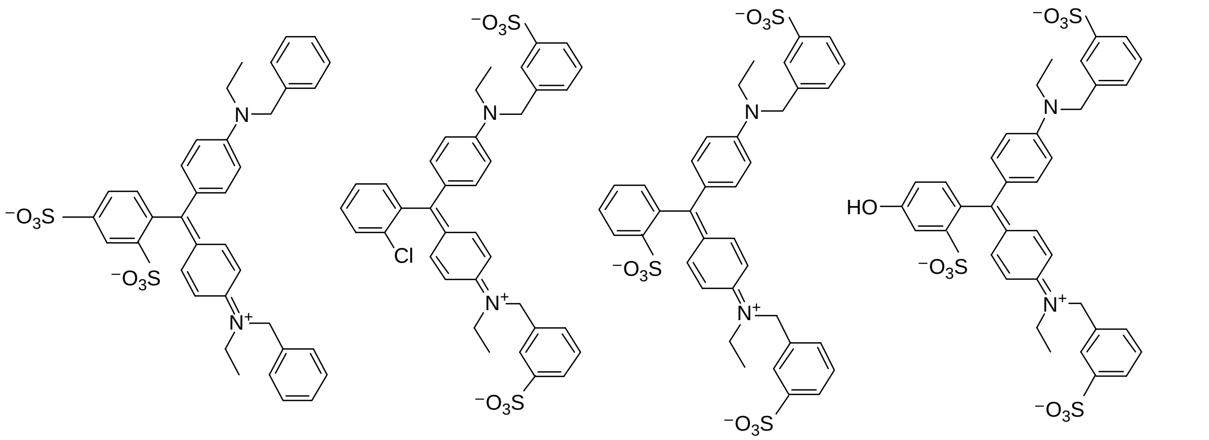

Fig. 1. Molecular structures of the four triarylmethane dyes used as the training set.

The sandy soil was a Vantage sand, and its properties were reported in German-Heins and Flury (2000): 95.7% sand, 2.5% silt, and 1.8% clay by weight; 3.51 g/kg organic carbon content; 1.68 μg/kg ammonium oxalate-extractable Fe; mineralogy dominated by quartz, plagioclase, and alkali feldspar. For the sorption experiments, we followed an experimental protocol similar to the one described in German-Heins and Flury (2000). Briefly, dye solutions with concentrations ranging from 0.0001 to 2.9 mM were equilibrated with the Vantage sand at 20–22 °C. The range of dye concentrations was chosen based on recommended dye application rates for hydrological tracing (Flury and Flühler, 1995). Equilibrium dye concentrations in solution were measured spectrophotometrically. The pH of the batch system was maintained at 8.0–8.5 using 0.01 M NaOH or HCl and the background ionic strength was 0.1 M $CaCl_2$. At pH 8, all the dyes are likely completely dissociated: the *p*Ka's for C.I. Food Blue 2 are 5.8 and 6.6 (Flury and Flühler, 1995; Flury and Wai, 2003) and it is likely that the other dyes have *p*Ka's in a similar range. The high ionic strength provides a worst-case scenario for compression of the electric double layers in soils. The sorption experiments were made in triplicate.

The experimental data were analyzed with the Langmuir sorption isotherm (Schulthess and Dey, 1996):

$$C_a = \frac{A_m K_L C_s}{1 + K_L C_s} \tag{3}$$

where C_a is the sorbed dye concentration, A_m is maximum adsorption capacity of the medium, K_L is the Langmuir coefficient, and C_s is aqueous dye concentration at equilibrium. We used the normal nonlinear least squares regression method as described in Schulthess and Dey (1996) to estimate the sorption parameters K_L and A_m. This method considers that not only dependent variables but also independent variables are subjected to errors. We chose the two sorption parameters, K_L and A_m, as experimental data (activity) for the QSAR development.

Structural properties (molecular descriptors) of the dyes to develop the QSAR model were calculated using MDL QSAR (version 2.1, 2002, MDL Information System Inc., San Leandro, CA). The MDL QSAR provides nearly 200 molecular descriptors including simple and valence MCIs up to the tenth order. The molecular structure data files were created using the ISIS/Draw program (version 2.4, 2001, MDL Information System Inc., San Leandro, CA). Default bond length, bond angle, and functional groups provided in the program template were used in creating the structure graphs. The training set contained zwitterions—molecules in which both positive and negative charges are present. As MDL QSAR does not recognize that structural type, a positive charge and a negative charge were manually assigned to the respective atoms in each molecule. This charge contributes to the electronegativity of the molecules.

The most appropriate descriptors to be included in the QSAR model were chosen by statistical considerations. Stepwise linear regression analyses were applied to determine which descriptors were well correlated to the experimental parameters (Hall et al., 2002b). Cross validation was used for validation of the model and randomization tests were performed to check the probability that the descriptors included in a model were selected by chance (Hall et al., 2002b). The MDL QSAR program carried out 100 randomizations of the activity values for each compound and then calculated multiple r^2 and the mean r^2 for all regressions. Only the model that achieved the best quality of statistics was selected for estimation of each sorption parameter. Thus, two QSAR models were established, one for estimation of the Langmuir coefficient and another for estimation of the maximum adsorption.

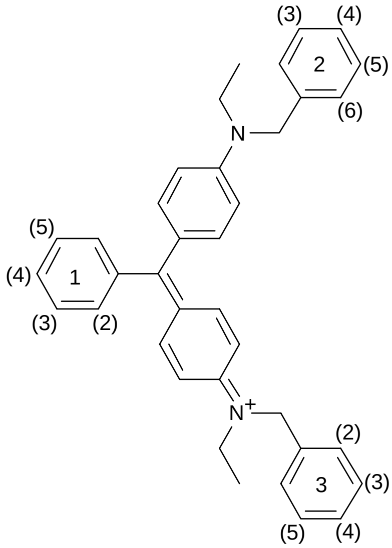

Fig. 2. Common molecular template shared by the training set and by all hypothetical molecules. Numbers identify the benzene rings and the positions (in parentheses) on each ring. These numbers are used in identifying the molecules with respect to specific positions of their functional groups.

3.2. Generation of molecular structures of potential dye tracers

Our premise was that SO_3 groups are the key components in designing an optimal water tracer. By attaching different numbers of SO_3 groups at different positions on the molecular template of the triarylmethane dyes used in the training set, we obtained many hypothetical molecules. Specifically, one or two SO_3 groups were attached to the benzene rings and the SO_3 group(s) attached on each ring were moved one position at a time. We followed the rule for the substitution of meta-directing deactivators in attaching the SO_3 on each benzene ring (McMurry, 1996). The different molecules were identified by the position of the SO_3 groups. These positions were numbered as shown in Fig. 2. Each SO_3 groups was identified by the number of the benzene ring and by the position (number in parenthesis) where the group is attached to that benzene ring. For example, a molecule is named as 1(2) if the SO_3 group is attached to ring number 1 and positions number 2 on that ring. Similarly, a molecule named 1(2)2(3) indicates that two SO_3 groups are attached to the molecule; one group is at ring number 1 and position number 2, and another at ring 2 and position 3. If two SO_3 groups are attached to the same benzene ring, the notation is abbreviated, e.g., the molecule named 1(24) contains two SO_3 groups at ring number 1 at positions 2 and 4.

We systematically varied number and positions of SO_3 groups on the hypothetical molecules. Four different sets of molecules were generated. The first set contained six molecules with different numbers of SO_3 groups: 1, 2, 3, …, 6 SO_3 groups on each molecule, respectively. The second set of molecules contained only one SO_3 group attached to the molecular template at six different positions. First, the SO_3 group was attached on benzene ring number 1 at position 2, and then the SO_3 group was moved to position 3, and position 4 in clockwise directions (Fig. 2). In doing so, three different molecules, 1(2), 1(3), and 1(4), were created. The same procedure was repeated for benzene ring 3 and thus another three molecules, 3(2), 3(3), 3(4) were obtained. Functional groups were not attached at the positions where atom overlap was indicated by the ISIS/Draw program. The same rules were applied in creating the third and the fourth set of molecules. The third set contained two

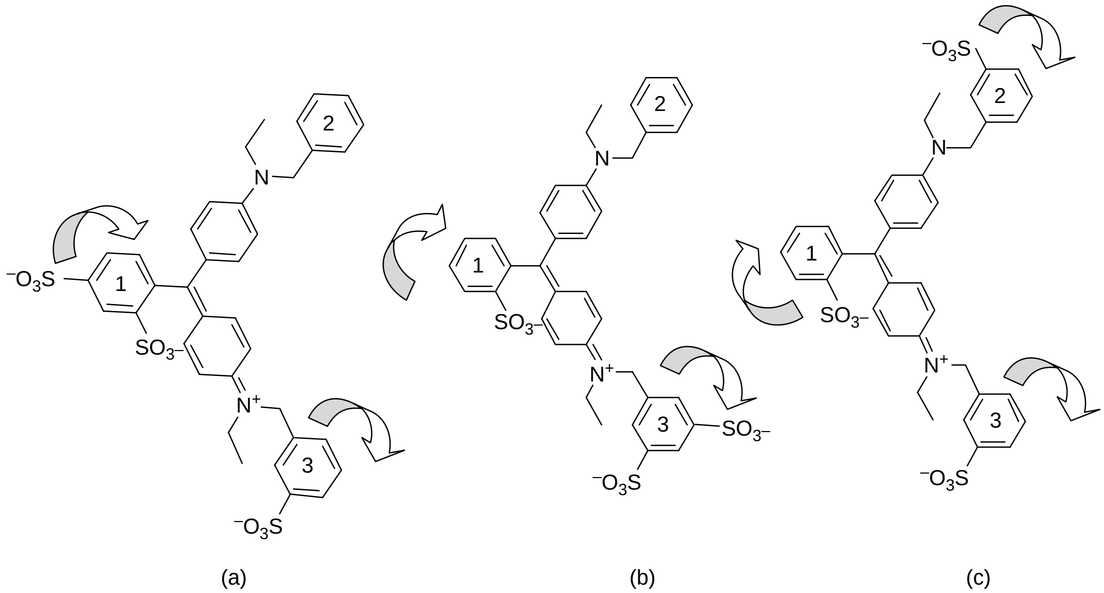

Fig. 3. Molecules with three SO_3 groups: (a) two SO_3 groups on ring 1 and one SO_3 group on ring 3, (b) one SO_3 group on ring 1 and two SO_3 groups on ring 3, (c) one SO_3 group on rings 1, 2, and 3, each. A total of 31 possible different molecules were obtained by moving each sulfonic acid group one position at a time in clockwise direction as indicated by arrows.

SO_3 groups attached to the template in all 20 possible combinations. The fourth set included molecules with three SO_3 groups attached to the molecular templates. A maximum of two SO_3 groups were attached to a benzene ring in meta-positions. By attaching three SO_3 groups in all the possible combinations, 31 molecules were obtained (Fig. 3).

4. Results and discussion

4.1. Experimental data and molecular descriptors

The adsorption of the selected dyes could be reasonably well described by a Langmuir isotherm (Fig. 4). For the low concentration range, the Langmuir isotherm did not fit well for C.I. Acid Green 9 and C.I. Acid Blue 7 (on a log scale as shown in Fig. 4). The sorption parameters estimated by the normal non-linear least squares are shown in Table 1. The deviation of the fitted isotherm from the experimental data at low concentrations did not reduce the fit's correlation coefficients because the absolute deviations in concentrations were small.

4.2. QSAR models

The regression analysis produced a number of QSAR models based on the input data and molecular descriptors. However, the results of the stepwise

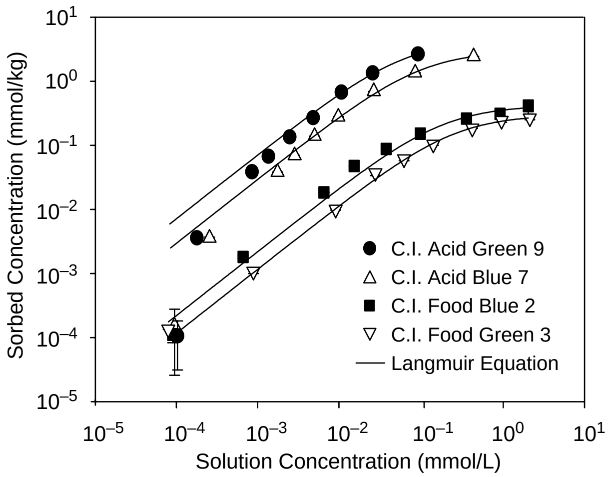

Fig. 4. Adsorption isotherms of the four test dyes. Symbols are experimental data and lines are fitted Langmuir isotherms. Error bars denote one standard deviation of three replicates. In many cases the error bars are smaller than the symbols.

Table 1
Experimental sorption parameters of the four dyes used as the training set and calculated molecular volumes and surface areas

Dyes	Number of SO_3 groups	Langmuir coefficient K_L (L/mmol)	Maximum adsorption A_m (mmol/kg)	Correlation coefficient of fit, r^2	Molecular volume[a] (Å^3)	Surface area[a] (Å^2)	SO_3 positions
C.I. Acid Blue 7	2	10.1	2.99	0.98	415	494	1(24)
C.I. Acid Green 9	2	16.5	4.40	0.99	466	546	2(3)3(5)
C.I. Food Blue 2	3	5.29	0.419	0.95	490	569	1(2)2(3)3(5)
C.I. Food Green 3	3	3.94	0.300	0.98	477	559	1(2)2(3)3(5)

[a] Calculated with MDL QSAR (version 2.1); molecular volume based on the grid method of Bodor et al. (1989).

regression suggested that K_L and A_m parameters of the four tested dyes were best correlated to the ninth-order simple path MCI ($^9\chi_p$) and the ninth-order valence path MCI ($^9\chi_p^v$), respectively. These two molecular descriptors yielded the best statistical indicators, such as correlation coefficient (r^2), standard error of estimation (s), F-statistic (F), level of significance (P), predictive correlation coefficient (Q^2), residual sum of squared error in validation (RSS), and mean correlation coefficients in randomization tests (r_k^2).

The optimal model obtained for the Langmuir coefficient was:

$$K_L = -54.47(^9\chi_p) + 183.75 \quad (4)$$

where K_L is given in L/mmol, $^9\chi_p$ is the ninth-order simple path MCI, and the statistical parameters are $r^2=0.999$, $s=0.114$, $F=7384$, $P=0.0001$, $Q^2=0.998$, RSS$=0.105$, $r_k^2=0.311$. For the maximum adsorption, the optimal model was:

$$A_m = -45.72(^9\chi_p^v) + 35.88 \quad (5)$$

where A_m is given in mmol/kg, $^9\chi_p^v$ is the ninth-order valence path MCI, and the statistical parameters are $r^2=0.996$, $s=0.125$, $F=779$, $P=0.001$, $Q^2=0.991$, RSS$=0.114$, $r_k^2=0.357$.

The QSAR models showed excellent agreement between estimated and experimentally measured sorption parameters for the four test dyes. The Q^2 values resulting from the cross-validations were similar to the r^2, which suggests that the models are relatively stable (Sabljić, 2001; Hall et al., 2002b). The Q^2 is usually considered as the estimate predictability of the models (Sabljić, 2001). The mean r_k^2 from the randomization test were relatively small, so that the probability of having a chance correlation between dependent and independent variables is small (Hall et al., 2002b). A few models with other descriptors also gave high correlation coefficients (r^2), but in these cases the predictive coefficient (Q^2) and the standard error of cross-validation were not as good as in case of our final model. We also tried to use the same descriptors to model the two sorption coefficients but did not obtain adequate predictive coefficients (Q^2) and standard errors of cross validation. The models finally chosen had the best overall statistics.

Soil sorption QSAR reported in the literature were mostly constructed with lower-order MCIs (Sabljić, 1987; Sekusak and Sabljić, 1992; Sabljić et al., 1995; Hong et al., 1997). Most of those reported models were based on relatively small, non-polar organic molecules. The sixth-order chain MCIs ($^6\chi_{ch}$) were included in a QSAR that described the log K_{oc} of some polar and non-polar organic compounds (Tao and Lu, 2000). Worrall (2001) suggested the sixth-order valence path MCI ($^6\chi_p^v$) as a discriminator between herbicides that leach and those that do not leach to groundwater. Higher-order (>second-order) MCIs presumably carry more detailed structural information than the lower-order MCIs (Hall et al., 2002b).

Our QSAR model suggests that the sorption coefficient K_L is negatively correlated with $^9\chi_p$. Higher-order simple MCIs, such as $^9\chi_p$, contain information about size, branching patterns, and positions of substituents (Hall et al., 2002b); however, the simple MCIs do not account for types of elements or types of bonds included in a molecule. According to these considerations, the QSAR model offers a generic interpretation that K_L values of the training set were dominated by the molecular size, branching pattern, and positions of substituents. Among the molecules with the same number of SO_3 groups, there seems to be a correlation between the molecular volume of the molecules and their K_L values (Table 1) but among all the molecules there were no relationships between molecular volume or surface area and K_L values (Table 1). The number of SO_3 groups, which also represents the amount of negative charge present in the molecules seems to have a major influence on the K_L of the dyes tested.

In our QSAR model, the maximum adsorption A_m was negatively correlated with ${}^9\chi_p^v$.Valence MCIs, such as ${}^9\chi_p^v$, account for all valence electrons in σ, π, and lone pair orbitals of non-hydrogen atoms. The π and lone pair electrons are more active or have a higher potential for intermolecular interactions than σ electrons (Hall et al., 2002a). Hence, A_m values might be more related to the interactions between the molecules and the soil medium surfaces than to the size or the shape of the molecules. Our QSAR analysis was based on only four test chemicals and only one experimental test condition (sandy soil with low organic matter content, on pH, and one ionic strength). Our QSAR models have to be regarded as qualitative rather than quantitative tools to investigate dye sorption behavior.

4.3. Estimation of the activity of potential dye tracers

The QSAR models developed were used to estimate the sorption parameters (K_L and A_m) of the hypothetical dyes. We first discuss the effect of the number of SO_3 groups on the sorption parameters and then the effect of position.

The more SO_3 groups attached to the molecule, the smaller were both the K_L and A_m values (Fig. 5). The figure shows a selection of dyes with a certain position of the SO_3 groups. If the position of the groups is changed, the K_L and A_m values will change as well, but the general trend observed in Fig. 5 still remains. For both K_L and A_m, the model estimated negative values for some cases. Those negative K_L and A_m suggest negative sorption, i.e., ion exclusion. The estimated parameters should be considered as relative, rather than absolute, measures for comparing the sorption of the chemicals. The trend observed in our study agrees with previous results that showed that dyes containing more SO_3 groups are less sorbed to solid media than dyes containing less SO_3 groups (Corey, 1968; Reife and Freeman, 1996).

The position of the SO_3 groups affected the predicted sorption parameters. Fig. 6 shows molecules containing one SO_3 group attached at different positions at the benzene rings. The smallest K_L and A_m were obtained for the molecule that contains the SO_3 group attached at the *ortho* position of the benzene ring 1, namely molecule 1(2). The K_L and A_m values vary not only with the position of the SO_3 group on each ring, but also with the ring to which SO_3 group is attached. Molecules with the SO_3 group attached to ring 1 appear to have smaller K_L and A_m than those with the SO_3 group attached to ring 3 (Fig. 6). The relative change of the sorption parameters as function of position of the SO_3 group was less than the change due to the number of SO_3 groups. This suggests, not surprisingly, that the number of SO_3 groups is the more dominant factor in determining the sorption characteristics.

The sorption parameters of the 22 molecules containing two SO_3 groups at different positions are shown in Fig. 7. In this figure, the molecules are grouped according to their molecular structure: sulfonic acid groups attached to rings 1, attached to rings 1 and 3, attached to rings 2 and 3, and attached to ring 3. The K_L and A_m values of the molecules again showed differences with the position of the SO_3 groups on each ring as well as with their distribution on different rings in the molecules. Attaching the two SO_3 groups on ring 1 produced smaller K_L and A_m values than attaching them on ring 3, e.g., molecules such as 1(24) or 1(35) showed generally smaller K_L and A_m values than 3(24) or 3(35). Among the molecules whose SO_3 groups were distributed on two different rings, those that contain SO_3 groups on ring 1 and ring 3, e.g., 1(2)3(2) showed smaller K_L and A_m values than those with SO_3 groups on ring 2 and ring 3, e.g., 2(3)3(2). Interestingly, the measured K_L and A_m values of C.I. Acid Blue 7, denoted as 1(24) were among the smallest. In general K_L and A_m values of molecules with two SO_3 groups attached on ring 1 tend to be smaller, whereas K_L and A_m values of molecules with two SO_3 groups attached on ring 3 tend to be larger.

By attaching three SO_3 groups to the molecular template, 31 possible molecules were created. The QSAR modeling showed that the range of K_L and A_m

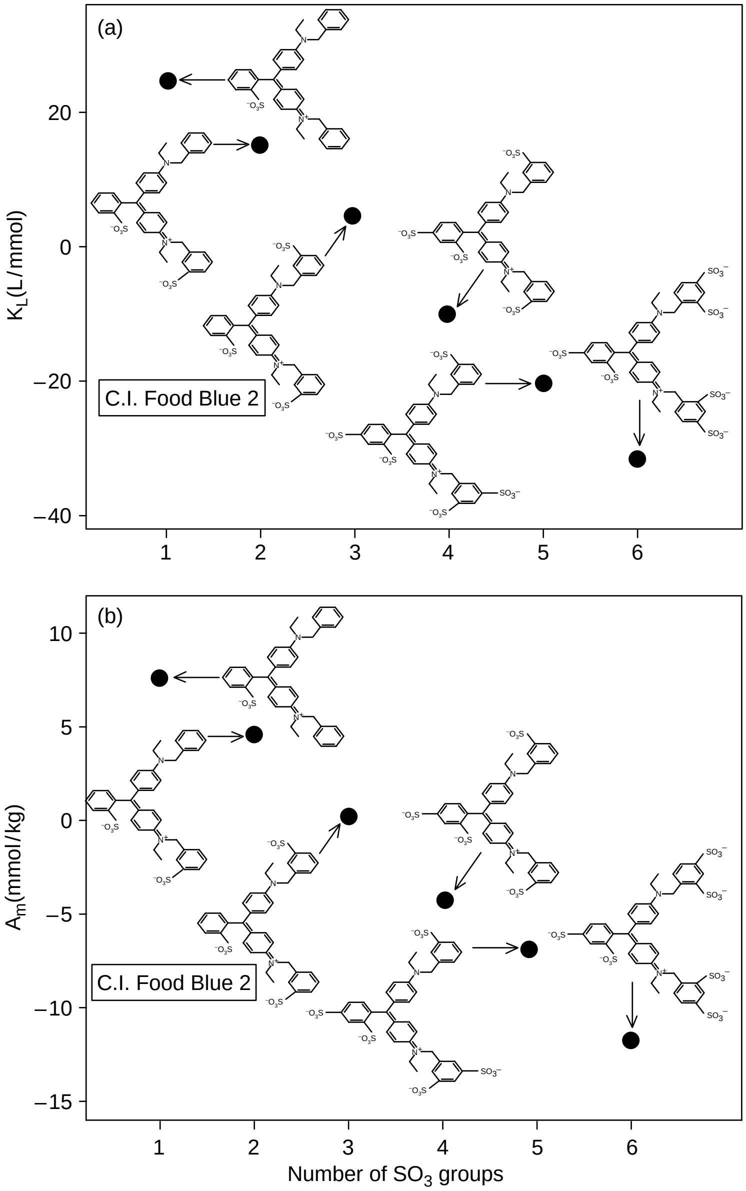

Fig. 5. Effect of the number of SO_3 groups on (a) Langmuir coefficient, K_L, and (b) adsorption maximum, A_m, of triarylmethane dyes.

of the molecules that contain two SO_3 groups attached on ring 1 and one SO_3 group attached on 2 or 3 were not significantly different than K_L and A_m of molecules that contain one SO_3 group attached on each ring (Fig. 8; molecules grouped according to molecular structure). The molecules that had the smallest K_L and A_m values contained one SO_3 group attached on ring 1 at an arbitrary position and contained the other two SO_3 groups attached at position 4 (*para* position) of rings 2 and 3 (e.g., molecules 1(2)2(4)3(4) and 1(3)2(4)3(4), Fig. 8). There seem to be many hypothetical chemicals whose K_L and A_m values are much smaller than that of C.I. Food Blue 2 and C.I. Food Green 3. C.I. Food Green 3

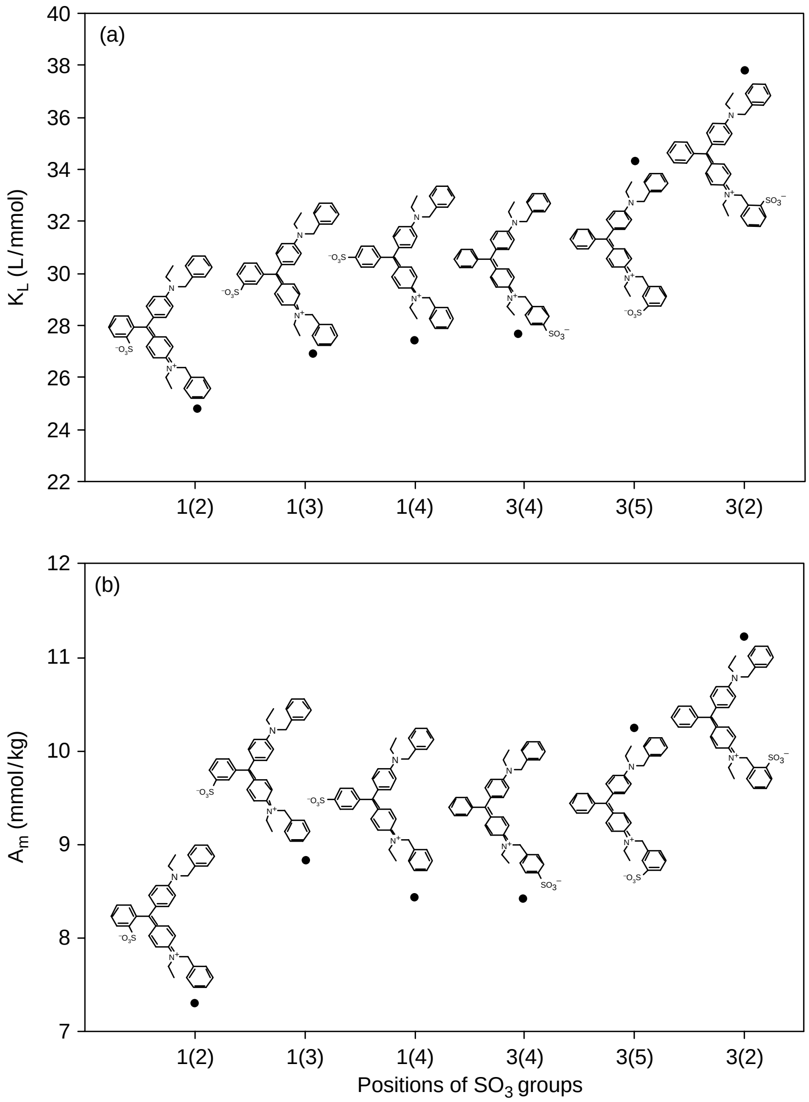

Fig. 6. Effect of the position of one SO_3 group on (a) Langmuir coefficient, K_L, and (b) adsorption maximum, A_m, of hypothetical triarylmethane dyes.

showed somewhat smaller K_L and A_m values than C.I. Food Blue 2.

In all cases investigated, we observed that the changes in K_L were correlated with the changes in A_m, i.e., molecules having large K_L values tend to have large A_m values (Figs. 5–8). When comparing the effect of position of SO_3 groups, the molecules with three SO_3 groups showed the largest variation among their K_L and A_m values. The results indicate that the effect of the position of SO_3 groups on the sorption parameters was more significant when the molecules contained more than one SO_3 group.

4.4. Recommendation for the design of an optimal dye tracer

We estimated the K_L and A_m values of nearly 70 hypothetical triarylmethane compounds using

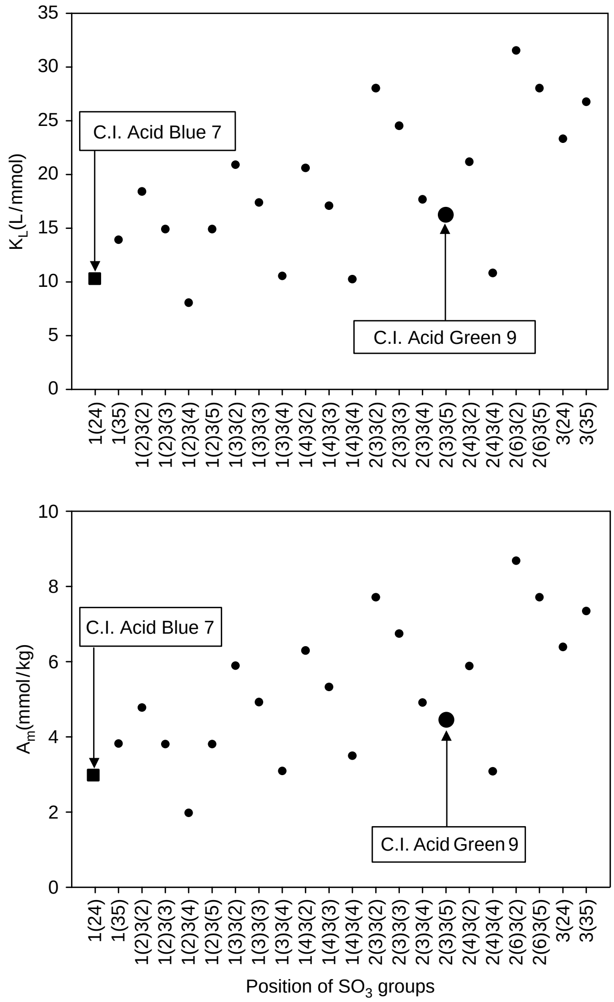

Fig. 7. Effect of the position of two SO_3 groups on (a) Langmuir coefficient, K_L, and (b) adsorption maximum, A_m, of triarylmethane dyes.

QSAR. Among those 70 compounds, there were many compounds that are possibly better tracers, in terms of their sorption under our test conditions, than the four test dyes. Their K_L and A_m values were considerably smaller than those of C.I. Food Blue 2, which is often used as a vadose zone tracer.

The number of the SO_3 groups had a dominant effect on the sorption parameters. The more SO_3 groups attached to the triarylmethane template, the smaller were the sorption parameters. The currently available triarylmethane dyes have a maximum of three SO_3 groups but hypothetical molecules with more than three SO_3 groups appear to be promising tracer candidates. As more and more SO_3 groups are attached, the molecule becomes larger and the increasing molecular size may lead to increased sorption, counteracting to some degree, the beneficial effect of the addition of SO_3 groups. The hypothetical molecules containing 4, 5, or 6 SO_3 groups, shown in Fig. 5, are likely promising hydrological tracers.

The increased number of sulfonic acid groups that are attached to the dye molecule tends to decrease the toxicity of the molecule; e.g., sulfonic acid groups tend to detoxify aromatic amines (Zollinger, 1991). Sulfonic acid groups make the dyes more water soluble and the dyes will less likely accumulate in the fat tissue of organisms (Flury and Flühler, 1994).

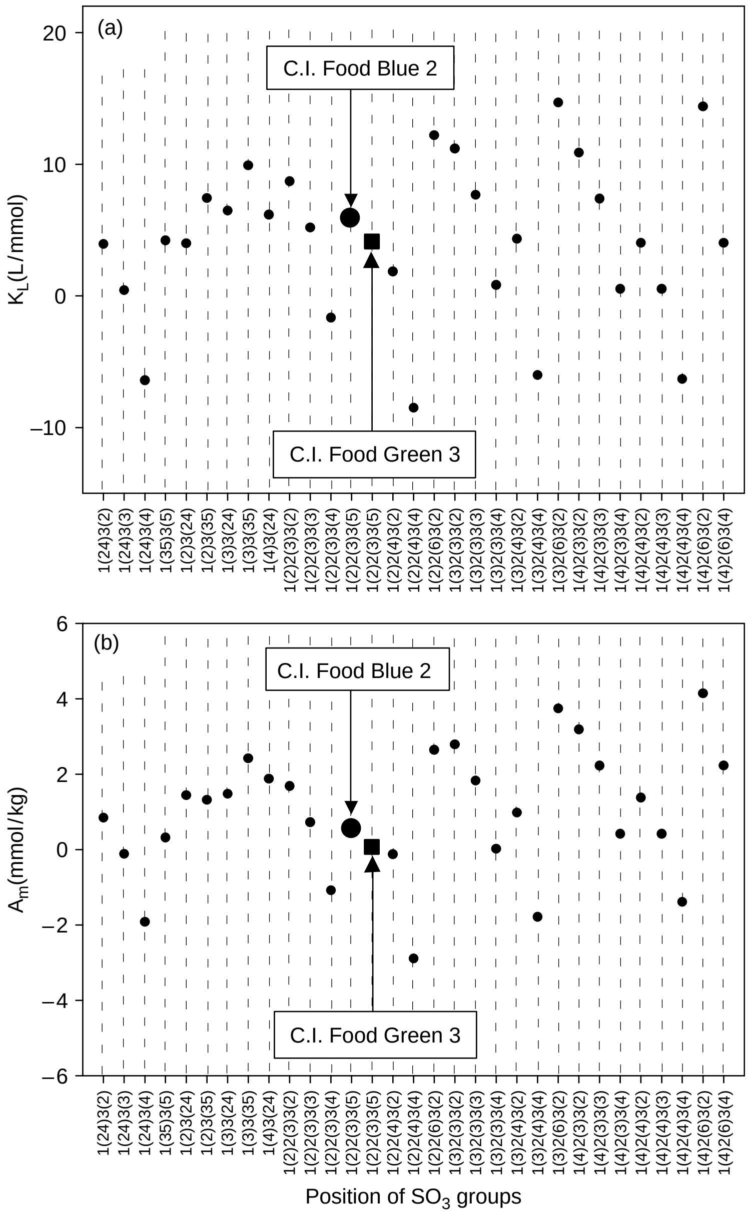

Fig. 8. Effect of the position of three SO_3 groups on (a) Langmuir coefficient, K_L, and (b) adsorption maximum, A_m, of triarylmethane dyes.

From the four dyes used in this study, the two dyes containing 3 sulfonic acid groups are certified food additives in the United States, the two dyes containing only 2 sulfonic acid groups are not.

C.I. Food Green 3, which is a commercially available food dye, sorbed less to Vantage sand than C.I. Food Blue 2. Between these two readily available dyes, C.I. Food Green 3 may be a useful alternative tracer for hydrological investigations in the vadose zone. As we only used one soil in this study, the results may not be directly transferrable to other soils; yet, the finding that increased numbers of SO_3 groups reduce sorption should be applicable to other low-organic matter soils as well.

5. Conclusions

We used QSAR modeling to screen hypothetical dyes for optimal hydrological tracer characteristics. For this study, we considered minimal sorption to soil materials as an optimal tracer characteristic. The QSAR analysis provided relative measures of soil sorption. The results showed that dyes containing more SO_3 groups are less sorbed to soil than dyes containing fewer SO_3 groups. The sorption parameters of the hypothetical compounds vary not only with the position of the SO_3 groups on each benzene ring but also depend on which benzene rings the SO_3 groups are attached to. Many hypothetical compounds seem to sorb less to soil than the currently known tracer dyes.

Acknowledgements

The Washington State Water Research Center provided financial support for this study. Warner Jenkinson Company Inc., Pylam Products Company Inc., and Organic Dyestuffs Corp. kindly provided free dye samples.

References

Bintein, S., Devillers, J., 1994. QSAR for organic chemical sorption in soils and sediments. Chemosphere 28, 1171–1188.

Bodor, N., Gabanyi, Z., Wong, C.K., 1989. A new method for the estimation of partition coefficient. J. Am. Chem. Soc. 111, 3783–3786.

Colour Index, 1971., third ed. Colour Index, vol 4. The Society of Dyers and Colourists, London.

Colour Index, 2001. Colour Index International Fourth Edition Online. Society of Dyers and Colourists and the American Association of Textile Chemists and Colorists, Bradford, England http://www.colour-index.org, accessed September 2004.

Corey, J.C., 1968. Evaluation of dyes for tracing water movement in acid soils. Soil Sci. 106, 182–187.

Davis, S.N., Thompson, G.M., Bentley, H.W., Stiles, G., 1980. Ground-water tracers—a short review. Ground Water 18, 14–23.

Drew, B.P., 1968. A review of the available methods for tracing underground waters. Proc. Br. Speleol. Assoc. 6, 1–19.

Flury, M., Flühler, H., 1994. Brilliant Blue FCF as a dye tracer for solute transport studies—a toxicological overview. J. Environ. Qual. 23, 1108–1112.

Flury, M., Flühler, H., 1995. Tracer characteristics of Brilliant Blue FCF. Soil Sci. Soc. Am. J. 59, 22–27.

Flury, M., Wai, N.N., 2003. Dyes as tracers for vadose zone hydrology. Rev. Geophys. 41, 1002. doi:10.1029/2001RG000109.

German-Heins, J., Flury, M., 2000. Sorption of Brilliant Blue FCF in soils as affected by pH and ionic strength. Geoderma 97, 87–101.

Gerstl, Z., Helling, C.S., 1987. Evaluation of molecular connectivity as a predictive method for the adsorption of pesticides by soil. J. Environ. Sci. Health, Part B 87, 55–69.

Hall, L.H., Kier, L.B., 2001. Issues in representation of molecular structure: the development of molecular connectivity. J. Mol. Graph. Model. 20, 4–18.

Hall, L.H., Kellogg, G.E., Haney, D.N., 2002a. Molconn-Z 4.00 User's Guide. Edusoft-lc, Inc, Ashland, VA online at http://www.edusoft-lc.com/molcomm/manuals/400/chaptwo.html, accessed May 2003.

Hall, L.H., Kier, L.B., Hall, L.M., 2002b. The Guide for Development of QSAR with MDL QSAR. MDL Information System, San Leandro, CA.

Hansch, C., Fujita, T., 1995. Status of QSAR at the end of the twentieth century. In: Hansch, C., Fujita, T. (Eds.), Classical and Three-Dimensional QSAR in Agro-chemistry. Vol. Symposium Series 606. American Chemical Society, Washington, DC, pp. 1–12.

Hong, H., Wang, L., Han, S., Zhang, Z., Zou, G., 1997. Prediction of soil adsorption coefficient k_{oc} for phenylthio, phenylsulfinyl and phenylsulfonyl acetates. Chemosphere 34, 827–834.

Kier, L.B., Hall, L.H., 1986. Molecular Connectivity in Structure–Activity Analysis. Wiley, New York.

Kier, L.B., Hall, L.H., 1999. Molecular Structure Description: the Electrotopological State. Academic Press, San Diego, CA.

Kier, L.B., Hall, L.H., 2000. Intermolecular accessibility: the meaning of molecular connectivity. J. Chem. Inf. Comput. Sci. 40, 792–795.

McLaughlin, M.J., 1982. A review on the use of dyes as soil water tracers. Water SA 8, 196–201.

McMurry, J., 1996. Organic Chemistry. Brooks/Cole Publishing Company, Pacific Grove, CA.

Meylan, W., Howard, P.H., Boethling, R.S., 1992. Molecular topology/fragment contribution method for predicting soil sorption coefficients. Environ. Sci. Technol. 26, 1560–1566.

Nendza, M., 1998. Structure Activity Relationships in Environmental Sciences. Chapman & Hall, London, UK.

Pussemier, L., de Borger, R., Cloos, P., van Bladel, R., 1989. Relation between the molecular structure and the adsorption of arylcarbamate, phenylurea, and anilide pesticides in soil and model organic adsorbents. Chemosphere 18, 1871–1882.

Reife, A., Freeman, H.S., 1996. Carbon adsorption of dyes and selected intermediates. In: Reife, A., Freeman, H.S. (Eds.), Environmental Chemistry of Dyes and Pigments. Wiley, New York, pp. 3–31.

Sabljić, A., 1987. On the prediction of soil sorption coefficients of organic pollutants from molecular structure: application of molecular topology model. Environ. Sci. Technol. 21, 358–366.

Sabljić, A., 1989. Quantitative modeling of soil sorption for xenobiotic chemicals. Environ. Health Perspect. 83, 179–190.

Sabljić, A., 2001. QSAR models for estimating properties of persistent organic pollutants required in evaluation of their environmental fate and risk. Chemosphere 43, 363–375.

Sabljić, A., Protic, M., 1982. Relationships between molecular connectivity indices and soil sorption coefficients of polycyclic aromatic hydrocarbons. Bull. Environ. Contam. Toxicol. 28, 162–165.

Sabljić, A., Güsten, H., Verhaar, H., Hermens, J., 1995. QSAR modelling of soil sorption. Improvements and systematic of log k_{oc} vs log k_{ow} correlations. Chemosphere 31, 4489–4514.

Schulthess, C.P., Dey, D.K., 1996. Estimation of Langmuir constants using linear and nonlinear least squares regression analyses. Soil Sci. Soc. Am. J. 60, 433–442.

Schwarzenbach, R.P., Gschwend, P.M., Imboden, D.M., 2003. Environmental Organic Chemistry, second ed. Wiley, New York, NY.

Sekusak, S., Sabljić, A., 1992. Soil sorption and chemical topology. J. Math. Chem. 11, 271–280.

Smart, P.L., Laidlaw, I.M.S., 1977. An evaluation of some fluorescent dyes for water tracing. Water Resour. Res. 13, 15–33.

Tao, S., Lu, X., 2000. Estimation of organic carbon normalized sorption coefficient (k_{oc}) for soil by topological indices and polarity factors. Chemosphere 39, 2019–2034.

Vasudevan, D., Fimmen, R.L., Francisco, A.B., 2001. Tracer-grade Rhodamine WT: structure of constituent isomers and their sorption behavior. Environ. Sci. Technol. 35, 4089–4096.

Worrall, F., 2001. A molecular topology approach to predicting pesticide pollution of ground-water. Environ. Sci. Technol. 35, 2282–2287.

Zollinger, H., 1991. Color Chemistry. Syntheses, Properties and Applications of Organic Dyes and Pigments, second, revised ed. VCH Verlagsgesellschaft, Weinheim, Germany.