



Reverse quantitative structure–activity relationship for modelling the sorption of esfenvalerate to dissolved organic matter A multivariate approach

Marianne Thomsen ^{a,*}, Shima Dobel ^b, Pia Lassen ^c, Lars Carlsen ^d,
Betty Bügel Mogensen ^e, Poul Erik Hansen ^f

^a Department of Policy Analysis and Department of Environmental Science, National Environmental Research Institute,
P.O. Box 358, DK-4000 Roskilde, Denmark

^b Office of Chemicals, Danish Environmental Protection Agency, Strandgade 29, DK-1401 Copenhagen-K, Denmark

^c Department of Chemistry and Microbiology, National Environmental Research Institute, DK-4000 Roskilde, Denmark

^d Department of Environment, Technology and Social Studies, Roskilde University, DK-4000 Roskilde, Denmark

^e Department of Atmospheric Chemistry, National Environmental Research Institute, DK-4000 Roskilde, Denmark

^f Department of Life Science and Chemistry, Roskilde University, DK-4000 Roskilde, Denmark

Received 6 November 2001; received in revised form 9 July 2002; accepted 23 August 2002

Abstract

The sorption of the pyrethroid, esfenvalerate, to the dissolved and/or dispersed fraction of eight different natural humic compounds has been investigated. The dissolved organic matters (DOMs) included in this study originate from ground water, soil pore water, and surface waters. Sorption was modelled at DOM concentration levels where equilibrium partitioning of esfenvalerate between DOM and the aqueous bulk phase prevails. The inherent characteristics of the eight different humic materials, quantified in the preceding paper by Thomsen et al. (2002, this issue (PII: S0045-6535(02)00335-1)), have been used as explanatory variables for modelling this equilibrium partitioning. Using a reverse QSAR approach based on by projection-into-latent-structure regression (PLS-R) inherent sorbent properties determining for the sorption affinity of esfenvalerate to DOM were analysed. For all humic substances a decrease in the DOM-normalised equilibrium-partitioning coefficient, K_{DOM} , with increasing concentration of DOM was observed. Significant variations in K_{DOM} values, as function of the inherent characteristics of the individual humic substances, were found at DOM concentrations of 75 and 100 ppm, respectively. The latter is a strong indication of variations in sorption mechanisms of esfenvalerate to DOM of varying inherent properties. Groupings in the principal property space quantifying DOMs may indicate that separate models are needed for quantifying the equilibrium partitioning to different classes of DOM.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Dissolved and/or dispersed organic matter; Esfenvalerate; Reverse QSAR; Equilibrium partitioning coefficient; K_{DOM}

1. Introduction

Environmental risk assessment of pollutants have traditionally been based solely on the inherent properties of the individual chemical compounds,

* Corresponding author. Tel.: +45-46-30-1358; fax: +45-46-30-1114.

E-mail address: mth@dmu.dk (M. Thomsen).

i.e. physicochemical properties such as solubility, octanol–water partitioning, sorption and degradation rate in different media. These physicochemical properties are used as input parameters in models for calculating the fate, e.g. transport and distribution, of chemicals in environmental compartments air, water and soil (European Commission, 1996). However the effect of binding of pollutants to a mobile fraction of organic matter is not included in the current assessment of the fate and effects of pollutants within the environments (European Commission, 1996).

The natural occurrences of dissolved organic matter (DOM) in surface waters, soil and sediment pore water have been found in the range of 2–50, 10–1000 and above 100 mg C/l, respectively (Caron and Suffet, 1989; Kukkonen and Oikari, 1991; Cao et al., 1999). The impact of DOM, and thereby the possibility of a third-phase effect (Lee and Kuo, 1999), may be a significant factor in relation to the fate and effects of environmental pollutants. Sorption to DOM may significantly affect the apparent solubility and migration potential of especially hydrophobic substances (Fauser and Thomsen, 2002). Furthermore, the presence of DOM reduces the bio-availability, and thereby toxicity, of pollutants in most cases (Kukkonen and Oikari, 1991; Haitzer et al., 1998; Steinberg et al., 2000).

Experimental parameters, such as concentration of DOM, pH and ionic strength (Ghosh and Schnitzer, 1980), contribute to the generally high uncertainty or variability in measured equilibrium partitioning to DOM (Staples et al., 1997; Cousins and Mackay, 2000). Despite this, partitioning coefficients are often considered unique values independent of varying inherent properties and concentration of the organic sorbent. Thus the sorption of polycyclic aromatic hydrocarbons (PAHs) to DOM increases with an increase in aromaticity of the DOM due to van der Waals interactions between the PAHs and aromatic substructures of DOM (Nielsen et al., 1997; Lassen and Carlsen, 1999; Perminova et al., 1999). However, for structurally more complicated molecules such as the pesticides, including the pyrethroids, no such clear relationship can be quantified (Piccolo et al., 1996; Oesterreich et al., 1999). The apparent lack of simple relationships describing the organic matter partitioning coefficients for the pyrethroids may be explained through the structural complexity of this group of compounds.

The main goal of this study is to analyse the explanatory significance of DOM descriptors for estimating the sorption of esfenvalerate to different types of DOM.

Through the use of reverse QSARs, i.e. using the inherent properties of the humic substances for quantifying the $\log K_{\text{DOM}}$ of esfenvalerate, the predictability of PLS-regressions based on different type of DOM descriptors is evaluated.

2. Descriptors quantifying the inherent properties of DOM

Descriptors included are derived from liquid-state ^{13}C -NMR and UV–VIS spectroscopy, respectively (data given in Thomsen et al., 2002). These descriptors quantify the amount and distribution of the different functional groups within the DOMs. The aliphaticity and aromaticity descriptors contain secondary information concerning the rigidity/flexibility of the DOM macromolecules.

The NMR data comprises 18 descriptors quantifying the percentage distribution of substructural groups within the humic macromolecular structure. The descriptors are integrated peak areas quantifying the content of C atoms of quinonic and ketonic groups CO_i , carboxylic and ester C, COO_i , aromatic O-substituted C, ArO_i , un-substituted aromatic C, ArHC_i , N-alkyl and methoxy C including a major fraction of carbohydrates, AlkO_i and un-substituted aliphatic C, AlkHC_i . The same type of descriptors was calculated based on peak widths, the subscript being replaced by w , and for peak areas divided by widths, the subscript replaced by i/w . All descriptors calculated according to the chemical shift ranges (ppm) of the functional subgroups within the molecular structure of DOM.

The UV–VIS descriptors quantify information concerning the degree of aromaticity, average molecular weight (MW), size and oxygen content. The E_4/E_6 ratio is found to increase with decreasing average MW and increasing oxygen content, the E_2/E_3 ratio increases with a decrease in aromaticity, and finally the absorptivity quantifies the aromatic $\pi \rightarrow \pi^*$ transitions at 272 nm.

The selection of descriptors, included in the modelling of the sorption of esfenvalerate to DOM of different origin, is based on a preceding evaluation of the explanatory significance of different spectroscopic methods and elemental analysis data (Thomsen et al., 2002).

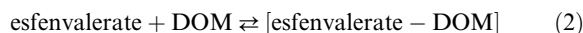
3. Equilibrium partitioning of esfenvalerate to DOM- K_{DOM}

The nominal concentration of the solute, esfenvalerate, was 2.57 $\mu\text{g/l}$ in all measurements, well below the aqueous solubility of esfenvalerate, i.e. 0.3 mg/l (Worthing and Hance, 1991). Experimental system parameters of aqueous media were kept at constant pH 8.3, ionic strength of 0.05 M NaCl and temperature at 25 °C. As such the only varying parameter is the change in DOM concentration. The equilibrium constant for the process of complex formation between esfenvalerate and DOM in an aqueous bulk phase were determined DOM concentrations of 10, 20, 30, 40, 50, 75 and 100 mg/l. The

so-called DOM-normalised partitioning coefficient calculated as

$$K_{\text{DOM}} = \frac{C_{\text{esfen}}^{\text{DOM}}}{C_{\text{esfen}}^{\text{aq}} C_{\text{DOM}}} \quad (1)$$

where $C_{\text{esfen}}^{\text{DOM}}$ and $C_{\text{esfen}}^{\text{aq}}$ are the concentration of esfenvalerate bound to DOM and the concentration of freely dissolved esfenvalerate, respectively. The concentration of esfenvalerate bound to DOM, $C_{\text{esfen}}^{\text{DOM}}$, is given in $\mu\text{g/l}$ solution, the concentration of DOM, C_{DOM} , is given in kg/l solution and the concentration of freely dissolved esfenvalerate, $C_{\text{esfen}}^{\text{aq}}$, is given in $\mu\text{g/l}$. The equilibrium constant, as expressed in Eq. (1), refers to the process



A system described as in Eq. (1) and equilibrium (2) is defined as being a mixture of dilute pollutant, e.g. esfenvalerate, and DOM in an aqueous bulk phase. As observed from Table 1 the measured K_{DOM} values decrease at increasing DOM concentrations. The decrease in K_{DOM} may be explained by a change in the configuration of the DOM at increasing DOM concentration. At low concentrations the configuration of DOM may be described as flexible, linear polyelectrolytes. At higher concentrations these building blocks coalesce to form rigid structures consisting of several monomer units forming a distinct two-phase system (Schnitzer, 1991). Equilibrium partitioning coefficients, independent of the DOM concentration, are expected only when a two-phase system prevails, i.e. the activity of DOM phase equals one (Thomsen, 2001). According to theory this seems to be the case at concentration levels of DOM above 60–90 mg/l , where K_{DOM} approaches a constant value (cf. Fig. 1).

The equilibrium concentration of esfenvalerate in bulk water and bound to DOM by complex formation was determined by an inverse column elution method (Kukkonen et al., 1990). The method is based on separating the freely dissolved fraction of esfenvalerate from the DOM-complexed fraction by eluting the aqueous

equilibrium mixtures through an inert support column. For hydrophobic compounds with slow desorption kinetics, the freely dissolved hydrophobic compounds will be retained on the column. The DOM molecules, and thereby the DOM-bound fraction of esfenvalerate, will show lower retention times (Landrum et al., 1984). The esfenvalerate concentration, in the continuous bulk water phase and bound to DOM, is determined by scintillation using ^{14}C -labelled esfenvalerate. Average K_{DOM} values based on triple to quintuple repeated sample measurements is given in Table 1.

The equilibrium partitioning coefficients as function of DOM concentration, for Water pond HS and Gohy-573-HS-(H^+)II, are given in Fig. 1.

As seen from Fig. 1, and Tables 1 and 2, the dependence of K_{DOM} on C_{DOM} is most significant at low concentrations of DOM. At high DOM concentrations, above approximately 90 ppm for Water pond HS and 60 ppm for Gohy-573-HS-(H^+)II, K_{DOM} is approaching constant values, i.e. the partitioning coefficient becomes independent of the concentration of DOM (Schwarzenbach et al., 1993; Hiemenz and Rajagopalan, 1997). The standard deviations on reproduced determinations of K_{DOM} , given as error bars in Fig. 1, are observed to decrease significantly at increasing DOM concentration. This pattern is similar for all humic samples, as given in Table 2.

The standard deviation on the DOM-normalised partitioning coefficients for esfenvalerate, is similar to the standard deviations found for benzo[α] pyrene, anthracene, biphenyl, *p,p'*-DDT, 2,5,2',5'-tetrachlorobiphenyl and 2,4,5,2',4',5'-hexachlorobiphenyl based on the same reversed-phase method (Landrum et al., 1984).

Analysis of homogeneity of variances (Funk et al., 1995) of measured DOM-complexed, $C_{\text{esfen}}^{\text{DOM}}$, and freely dissolved concentrations of esfenvalerate, $C_{\text{esfen}}^{\text{aq}}$, (cf. Eq. (1)) as function of the DOM concentrations was performed to evaluate the relative contributions to the standard deviation in K_{DOM} values. The results showed that variance inhomogeneity is highest for measuring of the dissolved esfenvalerate concentrations in respect to

Table 1

Average DOM normalised partitioning, K_{DOM} values for esfenvalerate^a as function of DOM concentrations

DOM (mg/l)	Aldrich HA (Na^+)	DE72	FA surface	Gohy-573- HA-(H^+)II	Gohy-573- HS-(H^+)II	Kranichsee HA	Purified Aldrich HA (Na^+)	Water pond HS
10	23 829	5316	10 089	7612	5702	20 329	28 484	13 283
20	18 167	4589	6093	5282	3125	7924	16 878	9213
30	11 494	3338	2356	3999	2215	4486	13 701	8014
40	9512	2294	2042	3024	1559	3312	9976	4942
50	8222	1903	1897	3187	1091	2357	8981	4227
75	6514	1513	820	1857	792	1776	7716	2311
100	5150	1393	377	2052	584	1251	6233	1669

^a The nominal concentration of esfenvalerate was 2.57 $\mu\text{g/l}$ in all partition experiments.

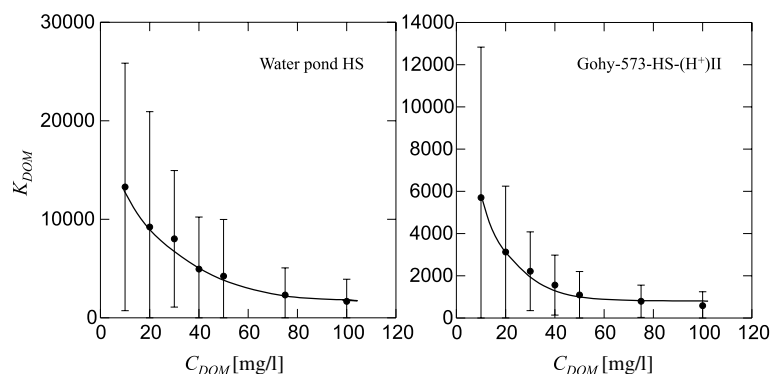


Fig. 1. Average DOM-normalised partitioning coefficients, K_{DOM} , as function of the concentration of DOM, for Water pond HS and Gohy-573-HS-(H⁺)II. A decrease in the standard deviations on repeated K_{DOM} measurements, are illustrated by error bars on each average value.

Table 2

Standard deviations on the reproducibility of K_{DOM} measurements

HS concentration (mg/l) Humic substances	^a Standard deviations on K_{DOM} measurements						
	10	20	30	40	50	75	100
Aldrich HA (Na ⁺)	4186	2070	2066	1145	1677	534	652
DE72	2476	2514	1881	977	1003	407	470
FA surface	6027	623	307	835	984	285	54
Gohy-573-HA-(H ⁺)II	2509	1441	1164	347	1079	413	617
Gohy-573-HS-(H ⁺)II	1378	747	644	177	174	99	95
Kranichsee HA	3506	5202	2038	1355	993	75	109
Purified Aldrich HA (Na ⁺)	937	2280	1383	752	1004	561	121
Water pond HS	851	2226	1285	1082	1322	516	499

^a Based on triple to quintuple measurements.

the variance in measurements on the DOM-complexed fraction (data not given).

Significance testing of the changes in the DOM-normalised partitioning coefficients in the concentration range of 10–100 mg DOM/l is relevant in relation to the need for correction factors when quantifying the partitioning to DOM. This due to natural variations in DOM concentration levels in different natural compartment systems. If the dependence of K_{DOM} on the concentration DOM concentration is not significant, then there is no need to investigate this aspect any further (cf. Fig. 1 and Table 2).

Another aspect of partitioning to DOM is the effect of variations in the inherent properties of DOM. Modelling the complex formation between esfenvalerate and DOM at the different concentration levels, require that significant differences in K_{DOM} at the individual DOM concentration levels exist.

In summary it is of utmost importance that the variation between objects, i.e. in this case K_{DOM} values for the partitioning of esfenvalerate to DOM of different origin, is significant. Therefore, before calibration of any SAR/QSAR model, the test data should be validated by

pre-processing of endpoint data (Cousins and Mackay, 2000; Thomsen, 2001), e.g. as described below.

3.1. Significance testing of K_{DOM} data

Pre-processing of data is often an overlooked step in the majority of SAR/QSAR investigations for estimating endpoints for use in environmental risk assessment at any level. Simple endpoints such as the aqueous solubility and octanol–water partition coefficients shown significant variations according to experimental standard methods used for the measuring the endpoint.

Prior to the development of QSARs, validation of the quality of endpoint data, as well as significance testing, in respect to the variation in endpoint data, is performed (Thomsen, 2001). Furthermore, the homogeneity in spanning of the descriptor space used for quantifying the endpoint is investigated (Thomsen and Carlsen, 2001).

As partition coefficients to natural organic matter in general display high variabilities (Staples et al., 1997; Cousins and Mackay, 2000), the variances in K_{DOM} for each DOM at each concentration level, σ_{within} , is compared to the variance between DOMs at each DOM

concentration level, $\sigma_{\text{between},1}$. The minimum requirement to data is that the variance of each average K_{DOM} value is significantly lower than the variance in K_{DOM} between DOMs at each concentration level. A one-side F -test (Miller and Miller, 1988) for this hypothesis is defined as

$$H_1(\sigma_{\text{between},1}^2 > \sigma_{\text{within}}^2) : T_1 > F_{1(1-0.05)} \quad (3)$$

If H_1 is fulfilled at 95% confidence level, then the variance in K_{DOM} values between DOMs is significantly higher than the variance in K_{DOM} for the individual DOMs at individual DOM concentration levels. Thus, the data quality accepted for further analysis. PLS-regression models may be developed for each concentration level, where H_1 is fulfilled.

A second hypothesis, H_2 , was performed to validate the concentration dependence of K_{DOM} . Hypothesis, H_2 , is a test to evaluate if the variance in K_{DOM} between DOM concentration levels for each specific DOM ($\sigma_{\text{between},2}^2$) is significantly higher than the variance on K_{DOM} for each DOM at each concentration level, σ_{within} , i.e.

$$H_2(\sigma_{\text{between},2}^2 > \sigma_{\text{within}}^2) : T_2 > F_{2(1-0.05)} \quad (4)$$

The rejection or acceptance of the hypothesis H_1 and H_2 are summarised in Table 3.

In general the variance of K_{DOM} values between the different DOMs, at constant concentration level of DOM, are in the same order of magnitude as found for, e.g., the low molecular weight phthalates, and lower than variances of K_{DOM} values for the high molecular weight phthalates (Staples et al., 1997). Furthermore the variances within each K_{DOM} measurement are much lower than found the pesticides between different unknown sorbents (NERI, 2001). It is observed that the frequency of rejected hypothesis, H_1 and H_2 increases at decreasing DOM concentration level. The results of the analysis of variance indicates that variations in K_{DOM} values may not only be due to the heterogeneity and complexity of natural organic matter, but also to uncertainties in the experimental methods used for measuring the partition of organic pollutants to natural organic matter. This is due to the systematic increase in standard deviation for all DOMs, as well as high contribution from uncertainties from the measuring the concentration of freely dissolved esfenvalerate.

The variation in K_{DOM} between DOMs of different origin at DOM concentrations of 10, 20, 30, 40 and 50 mg/l is insignificant, as tested by H_1 (cf. Table 3). Therefore, it is only reasonable to model the sorption of esfenvalerate to DOMs of different origin at the two highest DOM concentration levels, i.e., 75 and 100 mg/l, respectively.

4. QSARs based on partial least-squares regression

The QSAR paradigm is based on the assumption the variation in the activity of chemical compounds can be modelled through a quantification of the molecular inherent structural and electronic properties. In this study however the approach is reversed and the change in the partitioning of single compound, esfenvalerate, to DOM is modelled through a quantitative description of variations in the inherent properties natural DOM. The objective is to study the influence of changes in micro-environment surrounding a single pollutant molecule on the measured endpoint, i.e. K_{DOM} .

The absolute value of the DOM-normalised equilibrium coefficient (cf. Eq. (1)), as described in the preceding sections, depends on the concentration as well as the origin of DOM. This is a major problem for most environmental and ecotoxicological QSARs as they do not include any effects from varying environmental conditions.

The results of the four best performing PLS-models at DOM concentration levels of 75 and 100 mg/l are given in Table 4.

The correlation coefficient, R^2 , expresses the fraction of the variance in $\log K_{\text{DOM}}$ that is explained by the models, and Q^2 , the cross-validated correlation coefficient, expresses the fraction of predicted variance according to the leave-one-out method (Höskuldsson, 1996; CAMO ASA, 1998). The standard deviations on endpoint values, i.e. the K_{DOM} measurements, show significant influence on the model performance. The robustness of the models is reflected in the differences between Q^2 and R^2 , and as seen from Table 4, the robustness decreases significantly by decreasing DOM concentration. The same trend is observed for the root mean square error of calibration (RMSEC) versus the root mean square error of predictions (RMSEP) by the leave-on-out cross-validation method as seen from Table 4. The RMSEP is a measure of the average differences between predicted and measured K_{DOM} values. In this case the RMSEP is obtained through a full cross-validation with six different DOMs, and six different models based on five possible subsets of five DOMs, i.e. leaving one sample out at a time. For each of the left-out DOM samples predictions are made, and the RMSEP is calculated through the expression

$$\text{RMSEP} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{I}} \quad (5)$$

I is the number of calibration samples, i is the left out sample, \hat{y}_i is the predicted K_{DOM} by the model calibrated on the I samples, and y_i is the K_{DOM} values estimated by the reference models based on the six DOMs.

Table 3

Results of hypothesis testing based on analysis of variance on individual K_{DOM} values, as well as within and between DOM concentration levels

HS concentration (mg/l): Humic substances	10	20	30	40	50	75	100	
	$F_1^{a,c}$	$T_1 = \sigma_{\text{between},1}^2 > \sigma_{\text{within}}^2$						
Aldrich HA (Na ⁺)	19.353	4.5	7.4	4.5	8.5	3.2	25	11
DE72	19.353	13	5.1	5.4	12	9.0	43	21
FA surface	19.353	2.1	82	200	16	9.3	88	1600
Gohy-573-HA-(H ⁺)II	6.094	12	15	14	93	7.8	42	12
Gohy-573-HS-(H ⁺)II	19.353	41	57	46	360	300	730	510
Kranichsee HA	19.353	6.4	1.2	4.6	6.1	9.2	1300	390
Purified Aldrich HA	19.353	89	6.1	10	20	9.0	23	320
Water pond HS	19.353	110	6.4	12	9.6	5.2	27	19
Rejection (N) or acceptance (Y) of H ₁								
Aldrich HA (Na ⁺)	N	N	N	N	N	Y	N	
DE72	N	N	N	N	N	Y	Y	
FA surface	N	Y	Y	N	N	Y	Y	
Gohy-573-HA-(H ⁺)II	Y	Y	N	Y	Y	Y	Y	
Gohy-573-HS-(H ⁺)II	N	Y	Y	Y	Y	Y	Y	
Kranichsee HA	N	N	N	N	N	Y	Y	
Purified Aldrich HA	Y	N	N	Y	N	Y	Y	
Water pond HS	Y	N	N	N	N	Y	N	
	$F_2^{b,c}$	$T_2 = \sigma_{\text{between},2}^2 > \sigma_{\text{within}}^2$						
Aldrich HA (Na ⁺)	19.296	2.6	11	11	35	16	160	110
DE72	19.296	0.4	0.4	0.7	2.5	2.4	15	11
FA surface	19.296	0.3	31	130	17	13	150	4100
Gohy-573-HA-(H ⁺)II	6.256	0.7	2.0	3.0	34	3.5	24	11
Gohy-573-HS-(H ⁺)II	19.296	1.7	5.8	7.8	100	110	330	350
Kranichsee HA	19.296	4.4	2.0	13	30	55	8000	3800
Purified Aldrich HA	19.296	62	10	28	96	54	170	3700
Water pond HS	19.296	24	3.5	11	15	9.9	65	69
Rejection (N) or acceptance (Y) of H ₂								
Aldrich HA (Na ⁺)	N	N	N	Y	N	Y	Y	
DE72	N	N	N	N	N	N	N	
FA surface	N	Y	Y	Y	N	Y	Y	
Gohy-573-HA-(H ⁺)II	N	N	N	Y	N	Y	Y	
Gohy-573-HS-(H ⁺)II	N	N	N	Y	Y	Y	Y	
Kranichsee HA	N	N	N	Y	Y	Y	Y	
Purified Aldrich HA	Y	N	Y	Y	Y	Y	Y	
Water pond HS	Y	N	N	N	N	Y	Y	

^a The degrees of freedom for the variance between DOMs in the H₁-test is seven (N-1), as there are eight different DOMs at each concentration level.

^b The degrees of freedom for the variance for each DOM between DOM concentration levels in H₂-test is five (N-1), as there are six concentration levels of DOM.

^c The degrees of freedom for the variance within each K_{DOM} measurement is from two and four, as the K_{DOM} measurements was reproduced three to five times.

A bi-plot of second, PC2 versus the first principal component, PC1, showing the loading weights of the NMR descriptors and the K_{DOM} loading are given in Fig. 2.

The most significant principal component, PC1, explains 78% of the variation in K_{DOM} . From Fig. 2, it is

showed that the partitioning to DOM increases with increasing un-substituted or C-substituted aromaticity, quantified by the descriptors ArHC_i and $\text{ArHC}_i/\text{AlkHC}_i$. The $\log K_{\text{DOM}}$ value is inversely related to the aliphatic carbon shape descriptors, $\text{AlkHC}_{i/w}$ and carbohydrate, $\text{AlkO}_{i/w}$, descriptors. Furthermore a signifi-

Table 4

Model performance parameters of reverse QSAR models for estimating the sorption of esfenvalerate to DOM of different origin

Descriptors	Endpoint	<i>N</i>	<i>R</i> ²	<i>Q</i> ²	RMSEC	RMSEP	ExpXcal	ExpXval	ExpYcal	ExpYval
NMR	log <i>K</i> _{DOM} (100)	3	0.99	0.65	0.036	0.325	82	27	97	61
NMR/UV	log <i>K</i> _{DOM} (100)	3	0.99	0.60	0.031	0.340	74	19	98	56
NMR	log <i>K</i> _{DOM} (75)	2	0.94	0.33	0.083	0.385	28	39	20	67
NMR/UV	log <i>K</i> _{DOM} (75)	2	0.93	0.28	0.088	0.386	63	2	86	31

N is the number of principal components included in the model, *R*² the correlation coefficient, *Q*² the cross-validated correlation coefficient, RMSEC is the root mean square of error of the model, RMSEP the root mean square error of predictions. ExpXcal is the total explained X-variance used for explaining the variation in log *K*_{DOM} in the calibrated model, ExpXval the total explained X-variance used for explaining the variation in log *K*_{DOM} in the cross-validated model. ExpYcal is the total explained variance in log *K*_{DOM} and ExpYval the total explained variance in log *K*_{DOM} by cross-validation.

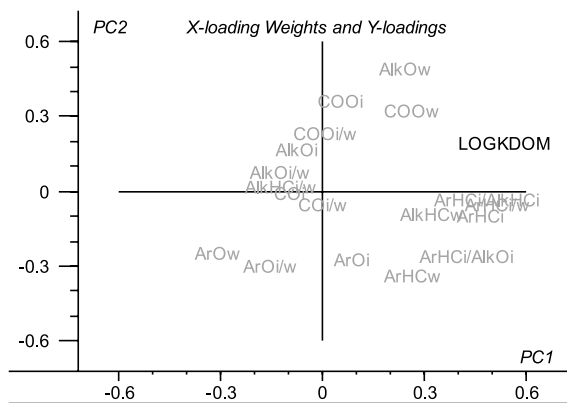


Fig. 2. Bi-plots of second versus first principal component, PC2 versus PC1, showing the loading weights, i.e. correlation patterns and importance, of individual original NMR-descriptor with respect to the *K*_{DOM}.

cant inverse relation to the width descriptors, ArO_w and ArO_{i/w} is observed.

In PC2 explaining 16% of the variation in *K*_{DOM}, the descriptors COO_i, AlkO_w and COO_w have high positive loading weights, whereas the descriptors ArO_i, ArO_{i/w} and ArHC_w have high negative loading weights. The partitioning to DOM is inversely related to the O-substituted aromaticity shape and width descriptors in the third quadrant. The NMR shape and width descriptors increase the model performance significantly in addition to models based on solely the integrated area descriptors. The effect is mainly on the robustness of the model. In the above models the width of carbohydrate peaks has high loading weight in PC2, which contribute significant to the homogeneous spanning of the DOM-property space in the PLS regression models when compared to a preceding study of the inherent properties of DOM. This is illustrated below in Fig. 3.

By use of the reverse QSAR concept it is possible to quantify varying sorbent sorption capacities. The standard errors of predicted *K*_{DOM} values are significantly lower than the standard deviation given in Table 2 in all

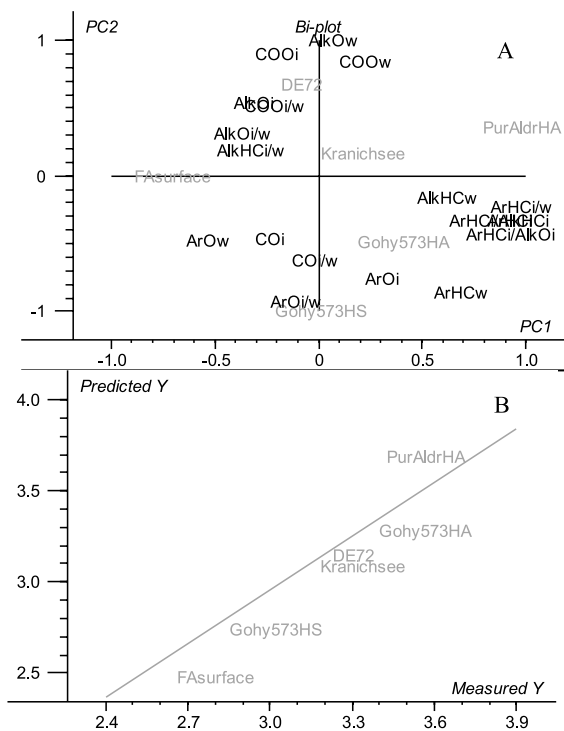


Fig. 3. The bi-plot, A, shows a homogeneous spanned X-space consisting of FA-surface, DE72, Gohy-573-HS-(H⁺)II, Kranichsee HA, Purified Aldrich HA and Gohy-573-HA-(H⁺)II. The regression line, B, illustrates the calibrated model predicted versus measured *K*_{DOM} values.

of the tested models. The sorption generally increases with increasing aromaticity, but still the aliphaticity and carbohydrate descriptors are just as significant in the investigated models for quantifying *K*_{DOM} of esfenvalerate. By fitting the X-matrix to log *K*_{DOM} the weighting of the original descriptor variables changes compared to the PCA model based exclusively on the DOM property descriptors as investigated in a preceding study (Thomsen, 2001). The shape and size descriptors generally

show high significance in explaining the variation in $\log K_{\text{DOM}}$.

5. Discussion and conclusions

There are several theories concerning the dependence of K_{DOM} on the DOM concentration. DOM may consist of different types of sorption sites, and in this case the solute will show highest affinity towards the most energetically favourable sorption sites, secondly the next highest etc., which could explain a change in K_{DOM} by a change in the DOM concentration (Schwarzenbach et al., 1993). Another explanation that may contribute to explaining the concentration dependence of K_{DOM} is the colloidal nature of DOMs, and the degree of inter- and intra-molecular associations within and between the organic macromolecules (Wershaw, 1999). At low concentration of DOM in the aqueous bulk, the system may be described as an aqueous true solution of humic monomeric macromolecules and esfenvalerate, respectively, as described by Eqs. (1) and (2). However, at increasing concentration the size of the organic macromolecular colloids increases, and the system changes from being a solution to being a two-phase system. In this case, under dilute solute conditions for esfenvalerate, the partition to DOM is expected to be independent of the DOM concentration (Schwarzenbach et al., 1993; Hiemenz and Rajagopalan, 1997).

Clearly the uncertainty level of K_{DOM} measurements is expected to be significantly higher than compared to the uncertainty on octanol–water partitioning measurements. This simply due to the complexity of DOMs, caused by the heterogeneity in structural characteristics such as shape and size as function of concentration level. Significant variations in the partitioning of esfenvalerate to DOM have been quantified solely based on a quantification of substructural functional group within the different humic materials. For an in-depth analysis of the effects of the inherent properties of DOM of aqueous origin, the inclusion of more fulvic type DOMs is needed.

The molecular structure of esfenvalerate includes an ester group, a cyano-group, a biphenylether, a chlorophenyl and an alkyl group. As observed for other pesticides (Piccolo et al., 1996; Oesterreich et al., 1999), this may increase the complexity, and the number of possible mechanisms, of sorbate–sorber interactions. For this reason the patterns in loadings weights of the inherent DOM property descriptors may very well vary for different e.g. pesticides. This aspect, as well as the variance inhomogeneity in K_{DOM} values for DOMs between concentration levels as well as inherent properties, suggests that classification of humic materials into similar inherent properties is required. The latter at least if the variation in sorption affinities to DOM of different

classes of environmental pollutants is to be quantified by conventional QSARs. Separate QSAR models for the different classes of humic substances will probably increase the robustness and predictability of QSAR models for estimating sorption to DOM provided that the measured equilibrium partitioning values are independent of the DOM concentration.

The present study has showed significant variation in sorption affinities of esfenvalerate to humic substances of different origin. The significance of this variation needs to be further investigated for other pesticides with respect to generally high variabilities in measured partitioning coefficient to sorber of varying composition. With respect to the bioavailability and mobility of environmental pollutants, more focus on the presence of a non-fixed mobile organic matter or third phase effects from a dispersed colloidal phase. The presence of DOM has shown impacts on the potential risks of environmental pollutants to a degree that is yet only sparsely elucidated.

References

- CAMO ASA, 1998. The Unscrambler 7.01, Oslo, Norway.
- Cao, J., Tao, S., Li, B.G., 1999. Leaching kinetics of water soluble organic carbon (WSOC) from upland soil. *Chemosphere* 39, 1771–1780.
- Caron, G., Suffet, I.H., 1989. Binding of nonpolar pollutants to dissolved organic carbon. Environmental fate modelling. In: Suffet, I.H., MacCarthy, P., (Eds.), *Aquatic Humic Substances. Influence on Fate and Treatment of Pollutants*. Advances in Chemistry Series, no. 219, pp. 117–130.
- Cousins, I., Mackay, D., 2000. Correlating the physical-chemical properties of phthalate esters using the “three solubility” approach. *Chemosphere* 41, 1389–1399.
- European Commission, 1996. Technical Guidance Document in support of Commission Directive 93/67/EEC on Risk Assessment for New and Notified Substances and Commission Directive (EC) no. 1488/94 on Risk Assessment for Existing Substances. 1996. Office for Official Publications of the European Communities, Luxembourg. Part III. Use of QSAR, Use Categories, Risk Assessment Format, CR-48-96-003-En-C, ISBN 92-827-8013-9.
- Fausner, P., Thomsen, M., 2002. Sensitivity analysis of calculated exposure concentrations and dissipation of DEHP in a topsoil compartment—The influence of the third phase effect and Dissolved Organic Matter (DOM). *Sci. Tot. Environ.* 296, 89–104.
- Funk, W., Dammann, V., Donnevert, G., 1995. *Quality Assurance in Analytical Chemistry*. VCH Verlagsgesellschaft, Weinheim.
- Ghosh, K., Schnitzer, M., 1980. Macromolecular structures of humic substances. *Soil Sci.* 129, 266–276.
- Haitzer, M., Höss, S., Traunsperger, W., Steinberg, C., 1998. Effects of dissolved organic matter (DOM) on the bioconcentration of organic chemicals in aquatic organisms—A review. *Chemosphere* 37, 1335–1362.

- Hiemenz, P.C., Rajagopalan, R., 1997. Principles of colloid and surface chemistry, third ed., Revised and Expanded Marcel Dekker, Inc, New York.
- Höskuldsson, A., 1996. Prediction Methods in Science and Technology. Thor Publishing, Denmark.
- Kukkonen, J., McCarthy, J.F., Oikari, A., 1990. Effects of XAD-8 fractions of dissolved organic carbon on the sorption and bioavailability of organic micropollutants. *Arch. Environ. Contam. Toxicol.* 19, 551–557.
- Kukkonen, J., Oikari, A., 1991. Bioavailability of organic pollutants in boreal waters with varying levels of dissolved organic material. *Water Res.* 25, 455–463.
- Landrum, P.L., Nihart, S.R., Eadie, B.J., Gardner, W.S., 1984. Reverse-phase separation method for determining pollutant binding to Aldrich humic acid and dissolved organic carbon of natural waters. *Environ. Sci. Technol.* 18, 187–192.
- Lassen, P., Carlsen, L., 1999. The effect of humic acids and the water solubility and water–organic carbon partitioning of fluorene and its NSO-heteroanalogues: Carbazole, dibenzofuran and dibenzothiophene. *Chemosphere* 38, 2959–2968.
- Lee, C.-L., Kuo, L.-J., 1999. Quantification of the dissolved organic matter effect on the sorption of hydrophobic organic pollutant: Application of an overall mechanistic sorption model. *Chemosphere* 38, 807–821.
- Miller, J.C., Miller, J.N., 1988. Statistics for Analytical Chemistry, second ed. Ellis Horwood, Chichester.
- Nielsen, T., Siigur, K., Helweg, C., Jørgensen, O., Hansen, P.E., Kirso, U., 1997. Sorption of polycyclic aromatic compounds to humic acid as studied by high-performance liquid chromatography. *Environ. Sci. Technol.* 31, 1102–1108.
- NERI National Environmental Research Institute, 2001. The Pesticide and Environmental Database (PATE).
- Oesterreich, T., Klaus, U., Volk, M., Neidhart, B., Spiteller, M., 1999. Environmental fate of amitrole: Influence of dissolved organic matter. *Chemosphere* 38, 379–392.
- Perminova, I.V., Grechishcheva, N.Y., Petrosyan, V.S., 1999. Relationships between structure and binding affinity of humic substances for polycyclic aromatic hydrocarbons: Relevance of molecular descriptors. *Environ. Sci. Technol.* 33, 3781–3787.
- Piccolo, A., Conte, P., Saccomandi, F., 1996. Interactions of different classes of herbicides with humic substances from European soils under monocultures and in a climate gradient. In: Del Re, A.A.M., Capri, E., Evans, S.P., Trevisan, M. (Eds.), *X Symposium Pesticide Chemistry—Basic Processes*, pp. 239–244.
- Schnitzer, M., 1991. Soil organic matter—the next 75 years. *Soil Sci.* 151, 41–58.
- Schwarzenbach, R.P., Gschwend, P.M., Imboden, D.M., 1993. *Environmental Organic Chemistry*. Wiley-Interscience, New York.
- Staples, C.A., Peterson, D.R., Parkerton, T.F., Adams, W.J., 1997. The environmental fate of phthalate esters: A literature review. *Chemosphere* 35, 667–749.
- Steinberg, C.E.W., Haitzer, M., Brüggemann, R., Perminova, I.V., Yashchenko, N.Y., Petrosyan, S., 2000. Towards a quantitative structure–activity relationship (QSAR) of dissolved humic substances as detoxifying agents in freshwaters. *Int. Rev. Hydrobiol.* 85, 253–266.
- Thomsen, M., 2001. QSARs in environmental risk assessment. Interpretation and validation of SAR/QSAR based on multivariate data analysis, Ph.D. dissertation, Department of Environmental Chemistry, National Environmental Research Institute and Department of Life Science and Chemistry, Roskilde University, DK-4000, Denmark.
- Thomsen, M., Lassen, P., Dobel, S., Hansen, P.E., Carlsen, L., Mogensen, B.B., 2002. Characterisation of humic substances of different origin: A multivariate approach for quantifying the latent properties of dissolved organic matter. *Chemosphere*, this issue (PII: S0045-6535(02)00335-1).
- Thomsen, M., Carlsen, L., 2001. Evaluation of empirical contra non-empirical descriptors. *SAR QSAR Environ. Chem.* 13, 525–540.
- Wershaw, R.L., 1999. Molecular aggregation of humic substances. *Soil Sci.* 164, 803–813.
- Worthing, C.R., Hance, R.J., 1991. *The Pesticide Manual: A World Compendium*. The British Crop Protection Council, Farnham.