Pergamon

PII: S0043-1354(97)00283-2

Wat. Res. Vol. 32, No. 3, pp. 872–881, 1998

Printed in Great Britain
0043-1354/98 $19.00 + 0.00

DEVELOPMENT OF A PREDICTIVE MODEL FOR CALCULATION OF MOLECULAR WEIGHT OF HUMIC SUBSTANCES

IRINA V. PERMINOVA[1]*, FRITZ H. FRIMMEL[2]®, DMITRII V. KOVALEVSKII[1], GUDRUN ABBT-BRAUN[2], ALEXEY V. KUDRYAVTSEV[1] and SEBASTIAN HESSE[2]

[1]Lomonosov Moscow State University, Department of Chemistry, 119899 Moscow, Russia
and [2]Engler-Bunte-Institute, University of Karlsruhe, Department of Water
Chemistry, Richard-Willstätter-Allee 5, D-76131 Karlsruhe, Germany

(First received January 1997; accepted in revised form June 1997)

Abstract—Size exclusion chromatography (SEC) is a powerful tool for the determination of molecular size of compounds. However, the conducted SEC analysis of 35 humic substances (HS) samples demonstrated, that it has some severe pitfalls for the molecular weight characterization of HS resulting from a lack of appropriate standards. To overcome this problem, a new approach to calculation of molecular weight from the SEC data was developed. It is based on the correction of the experimentally determined distribution coefficient (K_d) values according to the predictive equations derived either from the data on charge density and hydrophobicity of the analyte or from the values of its characteristic structural parameters. Verification of both predictive equations using model compounds has displayed much better prognostic potential of the polynomial equation developed with a use of integral structural parameters of the analyte. Application of this predictive equation to synthetic polyelectrolytes gave satisfactory estimates for all the polymers with $K_d > 0.15$. The model calculation for an averaged sample of HS allowed to estimate a difference between the ideal and experimental values of $K_d(\Delta K_d)$ by a value of 0.02. The correspondent correction of the SEC-data of 35 HS samples showed overestimation of the values of molecular weight of HS determined by the standard data handling technique. The obtained model estimations are to be verified along with the further development of the model based on the adjustment of the model assumptions to specific polyelectrolytic properties of HS. © 1998 Elsevier Science Ltd. All rights reserved

Key words—humic substances, size exclusion chromatography, molecular weight, predictive model

INTRODUCTION

Determination of the molecular weight or size of humic substances (HS) remains an unresolved problem in humic chemistry. This is a result of the complexity of chemical composition and irregularity of the structure of humic macromolecules. These factors render direct application of all the classical methods of molecular weight determination subject to artefacts. As a consequence the results are often contradictory. Particular criticism has been directed against the data provided by size exclusion chromatography (SEC) (De Haan *et al.*, 1987; De Nobili *et al.*, 1989; Swift, 1989). Nonetheless, the recent literature (Chin *et al.*, 1994; Wang & Benoit, 1996) demonstrates a distinct renewal of the interest to the usage of SEC for molecular weight determination of HS. This can be explained by the growing knowledge about the polyelectrolytic properties of HS, which permits application of more rigorous concepts of polymer chemistry to the description of the elution behaviour of HS on SEC columns (Chin *et al.*, 1994). Systematic studies on the sources of artefacts are required with respect to both experimental design and data analysis to understand and develop approaches to minimize problems associated with the application of SEC analysis to molecular weight determinations of HS.

The research objectives of the current study were: (1) to illustrate the role of artefacts in SEC of HS; (2) to elucidate the factors causing distortions of the elution behaviour of organic analytes in aqueous media by conducting a series of SEC analyses on an extensive set of the model compounds representing structural subunits of HS, (3) to develop a method for quantifying the deviation of experimental SEC parameters of the organic analyte from theory and explore applicability to both synthetic polyelectrolytes of known structure and HS.

MATERIALS AND METHODS

Preparation of humic substances solutions

SEC determinations were conducted using 35 HS samples isolated from fresh water, fresh water sediments, soil and peat. The HS preparations were stored as solutions at C-concentration of 500–1000 mg l^{-1}. The exact content of total organic carbon (TOC) in the solution was determined by organic carbon analysis (see below). For the SEC-analysis portions of the stock solutions of HS were diluted with a buffer identical in composition to the SEC mobile phase (0.028 M phosphate buffer) to a final C-concentration of 1–2 mg l^{-1}.

---

*Author to whom all correspondence should be addressed.

Preparation of model compound solutions

All the individual compounds used in this study as low molecular weight structural subunits of HS were purchased from Aldrich Chemicals or Fluka. All the reagents were of a grade "pure for analysis" or "purissimo". For the SEC analyses, the compounds were dissolved in MilliQ-water at a C-concentration of *ca.* 500 mg l^{-1}. Portions of these stock solutions were diluted in 0.028 M phosphate buffer to obtain a final C-concentration of 1–2 mg l^{-1}.

Molecular weight determination

All the SEC determinations were conducted according to the described procedure (Frimmel *et al.* 1992). For this purpose a liquid chromatography system consisted of a solvent pump (Shimadzu LC 9A, Germany), a packed column, a UV-VIS detector with variable wavelength, and a high sensitivity on-line DOC detector (Gräntzel, Germany), comprised of a cylindrical UV thin-film reactor and non-dispersive IR-detectors (Siemens Ultramat 3, Germany) for the continuous determination of organic and inorganic carbon was used. Samples were injected using an automatic injection system (Abimed, Bio 231, France) equipped with a rotary injector valve (Knauer, Germany) and 2 ml sample loop. Data acquisition, processing and presentation were processed according to Rhothron (Germany) on an Atari-compatible computer. The UV-absorbance of the analyte was quantified at a wavelength of 254 nm. The SEC column was a 25 × 200 mm BIAX (Chrom, Germany) with a column packing of Toyopearl HW-50S resin (Japan), with a nominal fractionation range (polyethylenglycol calibration by the supplier) of 100–20000 g/mol. Phosphate buffer (0.028 M) was used as an eluent at a flow rate of 1 ml $\min^{-1}$.

To characterize molecular weights (MW) of HS, the SEC column was calibrated by four sets of the standards. The first one was composed of polydextranes (PDX) (MW, g mol^{-1}: 0.83 K, 4.4 K, 9.9 K, 21.4 K, 43.5 K, 2000 K), mono- and oligosaccharides (MW, g mol^{-1}: 180, 342, 504), glycerol (MW, g mol^{-1}: 92) and methanol (MW, g mol^{-1}: 37). Blue dextrane (MW, g mol^{-1}: 2000 K) served as a void volume (V_0) probe, methanol served as a permeation volume (V_p) probe. The second calibration set was composed of sodium polyacrylates (PA) (MW, g mol^{-1}: 0.855 K, 1.25 K, 1.77 K, 4.0 K, 7.5 K, 16 K) and their saturated monomeric unit–propionic acid (MW, g mol^{-1}: 74). The third set was composed of sodium polymethacrylates (PMA) (MW, g mol^{-1}: 1.67 K, 2.5 K, 6.9 K, 9.8 K) and their saturated monomeric unit–butyric acid (MW, g mol^{-1}: 88). The fourth set was composed of sodium polystyrene sulfonates (PSS) (MW, g mol^{-1}: 1.37 K, 3.8 K, 6.71 K, 8.0 K, 8.6 K, 13.4 K, 16.9 K) and benzene sulfonic acid (MW, g mol^{-1}: 158). All the standard kits were purchased from the Polymer Standard Service (Mainz, Germany).

For construction of the calibration curves the distribution coefficient (K_d) was calculated for each standard by the following equation:

$$K_d = (V_e - V_o)/(V_p - V_o) \qquad (1)$$

where V_e is the elution volume, V_0 is the void volume, V_p is the permeation volume, (V_p-V_0) is the volume of the solvent inside the gel bed. For the SEC column under study V_0 determined by exclusion of blue dextrane was equal to 26.0 ml, V_p determined by exclusion of methanol was equal to 66.9 ml. The calibration curves were obtained by plotting the observed K_d value of the marker against the logarithm of its molecular weight (logMW). Molecular weights of HS and model compounds were determined by comparing their K_d values with the corresponding calibration curve.

Organic carbon analysis

Content of organic carbon in the stock solutions of HS was measured using a Shimadzu 5000 TOC analyser by the standard procedure described elsewhere (Sugimura & Suzuki, 1988). The carbon analyser was calibrated with potassium hydrogenphthalate standards at C-concentrations ranging from 10 to 500 mg l^{-1}:

RESULTS AND DISCUSSION

Results of SEC-analysis of HS and polymers of different partial charge

The typical UV- and DOC-elution profiles of HS observed under the chosen experimental conditions are shown for the IHSS FA in Fig. 1. All the HS under study were characterized by a broad, monomodal distribution (in some cases with shoulders and small sidepeaks). The elution volumes (V_e) of the HS samples differed greatly depending on their origin (water, sediment, soil or peat) and fractional composition (humic acids (HA) versus fulvic acids (FA)). The corresponding K_d values, averaged for each group of HS samples, are summarized in Table 1. It can be seen that the obtained K_d values ranged from 0.1 to 0.3. According to a magnitude of the corresponding K_d value, the molecular weight values of the HS samples under study could be put in the following order: peat HS > soil HA > sediment HS > soil FA > aquatic FA. The given sequence reflects natural differences in molecular weight values intrinsic to each group of humics. To quantify these differences numerically, the convertion of the K_d values into corresponding molecular weight values via calibration of the SEC-column is needed.

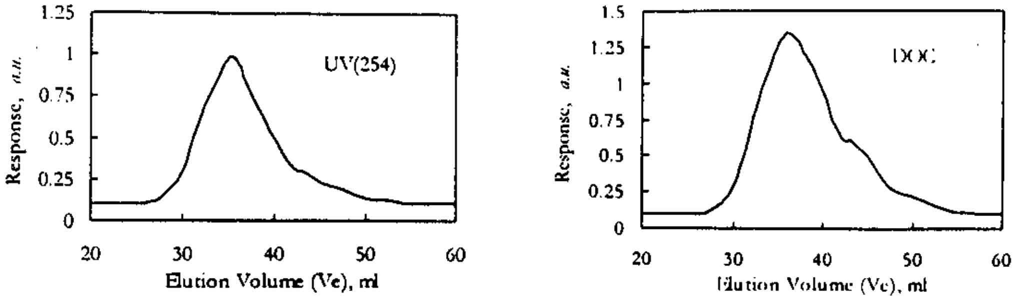

Fig. 1. SEC chromatogram of IHSS FA.

Table 1. Experimental K_d values for HS samples of different origin (confidence level 0.95, n-number of the samples is given in the table)

HS sample	K_d
Aquatic FA ($n = 9$)	0.27 ± 0.02
Fresh water sediment HS ($n = 2$)	0.16 ± 0.01
Soil FA ($n = 8$)	0.22 ± 0.02
Soil HA ($n = 8$)	0.16 ± 0.01
Peat HS ($n = 8$)	0.14 ± 0.04

According to the practice of SEC, the most important prerequisite for obtaining a true value of the molecular weight of an analyte is the right choice of the calibrants–substances of the known molecular weight and structure. In the ideal case, the calibrant and the analyte have to be authentic. This is obviously not achievable for HS, which are known to be random, acid groups containing polyelectrolytes of natural origin. That is why the usual practice is to use the standard calibrants–polydextranes and polyethylenglycols–for quantification of the SEC-results of HS, neglecting the polyelectrolytic character and structural peculiarities of HS (De Nobili *et al.*, 1989).

However, the recent investigations of Chin *et al.* (1994) introduced polystyrene sulfonates (PSS) as alternative calibrants for the SEC-analysis of HS. The authors stated that due to the polyelectrolytic nature and acidic character, the SEC-characteristics of PSS (hydrodynamic radii, viscosity etc.) are much more similar to HS than those of polydextranes or polyethyleneglycoles. Therefore, usage of these calibrants was supposed to provide a much more realistic evaluation of molecular weights of HS. This was confirmed by comparison of the obtained results with the data of other molecular weight determination techniques. Unfortunately, the authors have not conducted any comparative studies on synthetic polyanions other than PSS. From our point of view, without these studies the choice of PSS (from a broad variety of other polyanions) as the most appropriate calibrant for the SEC-analysis of HS is not convincing.

The main problem here is that under the same gel-chromatographic conditions (concentration of the analyte, pH, ionic strength and composition of the eluent), the elution behaviour of a polyanion will be strongly dependent on its charge density. We have demonstrated this by the SEC-analysis of the homological series of polyacrylates (PA), polymethacrylates (PMA), polystyrene sulfonates (PSS) and non-ionic polymers–polydextranes (PDX) (Fig. 2). The charge densities (Z) of the polymers under study were roughly estimated with a use of the following expression:

$$Z = (\text{number of ionogenic groups per mole}/\text{molecular weight}) \cdot 10^3$$

The corresponding values were 13.8 for PA, 11.6 for PMA, 5.4 for PSS and 0 for PDX. As it can be seen from Fig. 2, all the four polymers series were characterized by a different slope of the corresponding molecular weight - K_d dependence. The minimum slope was observed for a series of polydextranes as non-ionic polymers. All the series of the polyanions were characterized by the steeper slopes, the values of which increased with increasing charge density of the polyanions.

The obtained results show that the usage of the calibrants of the four polymer series differing in the charge density would have brought four different sets of the molecular weight values of HS. The smaller the charge density of the chosen calibrants, the higher the value of molecular weight will be. In our case, the highest values of molecular weights will be obtained with a PDX-calibration curve, the lowest ones–with a PA-calibration curve. To discover which of the calibrants provide the most appropriate evaluation of the molecular weights of HS, we come back to the consideration of the main factors impacting the elution behaviour of the polyelectrolytes on the SEC column.

Description of the elution behaviour of the organic analyte on the SEC column in aqueous medium as a function of its charge and hydrophobicity

To set the scene for the forces (other than diffusion and dispersion) influencing the elution behaviour of an organic analyte on the SEC column, we have to consider analyte-gel interactions occurring in the aqueous medium. We postulate that these interactions serve as a main source of artefacts in SEC of water-soluble organic compounds (Yau *et al.*, 1979; Belen'kii and Vilenchik, 1983). Fractionation is assumed to occur on the gel which is absolutely inert to the solute. In reality, it is not possible to create a hydrophilic gel which would be completely inert in aqueous medium (Dubin, 1988). In practice, a vast majority of hydrophilic gels whose spatial structure is based on ether bridging (polydextranes, polysaccharose, Bio-Rad, Toyopearl) are weakly negatively-charged due to partial hydrolysis of the ether sites in aqueous medium. In addition, the cross-linked structure of the

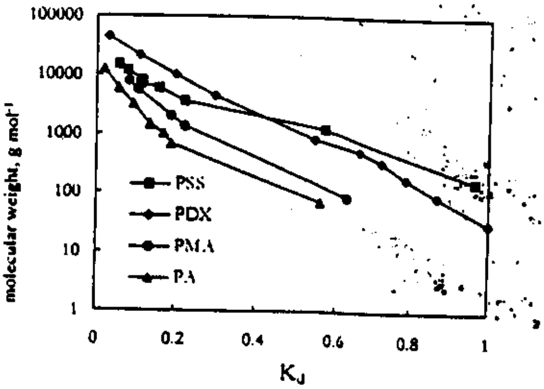

Fig. 2. Calibration curves of the Toyoperl HW-50S column by the kits of sodium polystyrene sulfonates (PSS), polydextranes (PDX), sodium polymethacrylates (PMA), sodium polyacrylates (PA).

gel creates hydrophobic sites on its surface (Dubin, 1988). According to these chemical features of the gel, it is likely that an organic analyte with functional groups will experience both electrostatic and hydrophobic interactions with the gel. In case of a negatively charged analyte, for example a deprotonated organic acid, the electrostatic interactions will be repulsive and can be accompanied with an effect of "overexclusion"–the analyte will be eluted from the column earlier than would be expected from a value of its molecular weight alone. The larger the partial charge of the analyte is, the stronger this effect will be. On the other hand, the hydrophobic interactions with the gel can cause a retardation of the analyte on the SEC column–in addition to the mere molecular weight effect. The larger the hydrophobicity of the analyte is, the stronger the retardation of the analyte by the gel will be.

With these considerations in mind, a value of K_d(exp), determined from the experimentally observed elution volume of the charged organic analyte from the SEC column can be represented as a function of the three parameters: molecular weight of the analyte (MW), its charge (Z) and hydrophobicity (H) (Eq. 2):

$$K_d = K_d f(\mathrm{MW}) + K_d f(\mathrm{Z}) + K_d f(\mathrm{H}) \qquad (2)$$

where $K_d f(\mathrm{MW})$ corresponds to the distribution coefficient of an ideal (nonionogenic and hydrophilic) analyte of molecular weight MW. This term of the equation can be designated as an ideal distribution coefficient K_d^0. To facilitate quantitative estimation of K_d^0, we have assumed, that nonionogenic, hydrophilic molecules of sugars and polydextranes satisfactory fit the given requirements to the ideal analyte. In this case for PDX the obtained plot of Molecular weight vs. K_d (Fig. 2) can be referred to as an "ideal" calibration curve of the given SEC-column. As a consequence it can be used for estimation of the "ideal" value of K_d (K_d^0) of nonideal (charged and/or hydrophobic) analyte of the known molecular weight MW.

To express $K_d f(\mathrm{Z})$ via experimental parameters, we have assumed that under conditions of the same solution pH and ionic strength, the charge density intrinsic to the molecules of low molecular weight carboxylic acids is determined only by the number of carboxyl groups per molecular weight of the acid (A_{COOH}). In this case $K_d f(\mathrm{Z}) = f(A_{\mathrm{COOH}})$. For estimation of the numerical values of A_{COOH} the following expression can be used:

$$A_{\mathrm{COOH}} = (\text{number of carboxyl groups per mole/molecular weight}) \times 10^3.$$

To express $K_d f(\mathrm{H})$ via experimental parameters, we have taken into consideration that the most developed and widespread measure of hydrophobicity of organic compound is a logarithm of its octanol-water distribution coefficient $\log K_{ow}$ (Hansch and Leo, 1979).

Thus equation 2 can be transformed to:

$$K_d(\mathrm{exp}) = K_d^0 + f(A_{\mathrm{COOH}}) + f(\log K_{ow}) \qquad (3)$$

A rearrangement of eq 3 by means of a transfer of K_d^0 into the left side of the equation and follow up change of the marks makes it possible to obtain on the left hand side the expression for deviation between the "ideal" (determined from the PDX-calibration curve for the non-ideal analyte of the known molecular weight) and experimental K_d (determined by eq 1): $K_d^0 - K_d(\mathrm{exp}) = \Delta K_d$. Thus, the numerical value of ΔK_d can be calculated for each analyte, if its elution volume and molecular weight are known. ΔK_d determined by means of the above described calculation technique will be referred to in the further parts of the paper as "experimental" ΔK_d. On the other hand, ΔK_d can be expressed as a function of charge and hydrophobicity of the analyte, as it follows from eq 3:

$$K_d^0 - K_d(\mathrm{exp}) = \Delta K_d = f(A_{\mathrm{COOH}}) + f(\log K_{ow}) \quad (4)$$

It can be easily shown that for charged, highly hydrophilic analytes ($f(\log K_{ow})$ approaches 0) a value of ΔK_d is a function of A_{COOH} only, whereas

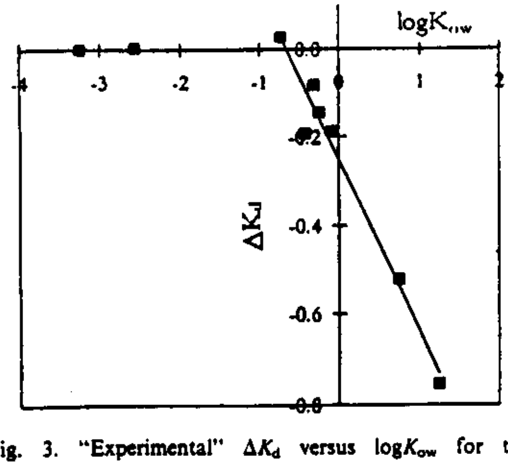

Fig. 3. "Experimental" ΔK_d versus $\log K_{ow}$ for the uncharged analytes (see Table 2).

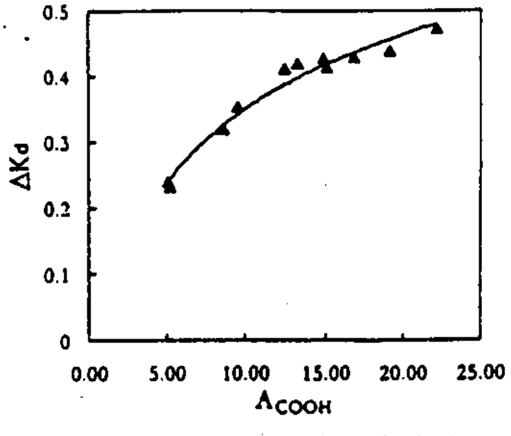

Fig. 4. "Experimental" ΔK_d *versus* charge density (A_{COOH}) of the hydrophilic analytes (see Table 3).

Table 2. List of uncharged organic compounds used for calculation of the relationship ΔK_d versus $\log K_{ow}$ from Fig. 3

Compound	$\log K_{ow}$
Glucose	−3.24
Glycerol	−2.57
Methanol	−0.74
1,4-Dioxane	−0.42
Ethanol	−0.32
Acetone	−0.24
2-Propanol	−0.1
2-Butanol	0.74
2-Pentanol	1.25

for hydrophobic, nonionogenic analytes ($f(A_{COOH})$ approaches 0) a value of ΔK_d is a function of $\log K_{ow}$ only. Thus, by plotting the values of "experimental" ΔK_d against A_{COOH} and $\log K_{ow}$, respectively, for a series of model compounds, both relationships can be determined from SEC data.

Determination of $\Delta K_d(A_{COOH}, \log K_{ow})$ from the SEC data of individual organic compounds

A set of uncharged organic compounds of known hydrophobicity (Table 2) was analysed by SEC to establish a shape of the relationship between ΔK_d and $\log K_{ow}$. The values of $\log K_{ow}$ for these and further calculations were taken from Leo *et al.* (1971) and Sangster (1989) or, as specifically noted, calculated with the use of ACDLabs Program. By applying the above described calculation scheme to the experimental data, the value of "experimental" ΔK_d was calculated for each compound. The obtained numerical values of ΔK_d were plotted against $\log K_{ow}$. The corresponding relationship is given in Fig. 3.

It can be seen that ΔK_d is directly proportional to $\log K_{ow}$ ($r^2 = 0.97$) in the range of $\log K_{ow}$ values from −0.5 up to 1.3 and equals 0 at all the values of $\log K_{ow}$ less than −0.5. The obtained data display that for the SEC system under study hydrophobic interactions with the gel can cause retardation of the analyte on the column only if it has a value of $\log K_{ow} > -0.5$. From Fig. 3 it can be seen that the retardation on the gel (negative deviation of the experimental K_d from the ideal one) is satisfactorily approximated with the following equation ($r^2 = 0.97$)

$$\Delta K_d = -0.38 \log K_{ow} - 0.26 \qquad (5)$$

The found relationship displays an existence of the specific sorption on the gel (negative slope of the plot), the effect of which is strengthened with increasing hydrophobicity of the analyte.

Table 3. List of the charged hydrophilic compounds used for calculation of the relationship ΔK_d versus A_{COOH} from Fig. 4

Compound	A_{COOH}	$\log K_{ow}$
Glucuronic acid	5.13	−1.65 ± 0.52*
Quinic acid	5.21	−2.01 ± 0.4*
Maleic acid	8.62	−0.56
Glucaric acid	9.50	−1.62 ± 0.49*
Pimelic acid	12.50	0.27 ± 0.2*
Tartaric acid	13.33	−0.76
Malic acid	14.93	−0.3
Glutaric acid	15.15	−0.47
Succinic acid	16.95	−0.59
Malonic acid	19.23	−0.91
Oxalic acid	22.22	−0.62

*The values of $\log K_{ow}$ calculated with a use of ACDLabs LogP program.

To establish the form of the relationship between ΔK_d and A_{COOH}, a set of carboxylic acids with $\log K_{ow} < -0.5$ was used (Table 3). The range of $\log K_{ow}$ values close to or less than −0.5 were chosen according to the observations described above. The obtained relationship is given in Fig. 4. It can be seen that the plot of ΔK_d versus A_{COOH} is well approximated with eq 6 ($r^2 = 0.991$):

$$\Delta K_d = 0.07 \log A_{COOH} - 0.0203 \qquad (6)$$

The positive slope of the relationship exhibits the effect of overexclusion, which increases with the charge of the analyte.

The results obtained from the two relationships make it possible to present $\Delta K_d(A_{COOH}, \log K_{ow})$ by the following equation:

$$\Delta K_d = 0.07 \log A_{COOH} - 0.3805 \log K_{ow} - 0.276 \quad (7)$$

Theoretically, eq 7 can be used for a prediction of ΔK_d of an organic analyte of known A_{COOH} and $\log K_{ow}$. This value of ΔK_d will be referred to in the further parts of the paper as "calculated" ΔK_d (to distinguish it from the earlier introduced "experimental" ΔK_d). The predicted value of "calculated" ΔK_d will be meaningful only in cases where our initial assumptions about electrostatic and hydrophobic solute-gel interactions being the only source

Table 4. List of the compounds used for verification of the predictive model $\Delta K_d = f(A_{COOH}, \log K_{ow})$

No	Compound	A_{COOH}	$\log K_{ow}$
1	Cinnamic acid	6.76	2.13
2	Benzyl alcohol	0.00	0.99
3	3,5-Dihydroxybenzoic acid	6.49	1.39
4	2-Butanol	0.00	0.71
5	3,4,5-Trihydroxybenzoic acid	5.88	−0.25
6	3,4-Dihydroxybenzoic acid	6.49	1.4
7	4-Hydroxybenzoic acid	7.25	1.58
8	Hydrocinnamic acid	6.67	1.84
9	1,4-Dioxane	0.00	−0.42
10	2-Propanol	0.00	−0.1
11	Acetone	0.00	−0.24
12	Benzoic acid	8.20	1.87
13	Ethanol	0.00	−0.32
14	Phenylacetic acid	7.35	1.41
15	Hexanoic acid	8.62	1.92
16	Valeric acid	9.80	0.99
17	Isovaleric acid	9.80	1.4
18	Methacrylic acid	11.63	0.93
19	Crotonic acid	11.63	0.72
20	Butyric acid	11.36	0.79
21	Lactic acid	11.11	−0.62
22	Maleic acid	8.62	−0.56
23	Terephthalic acid	12.05	0.3
24	Acrylic acid	13.89	0.37
25	Phthalic acid	12.05	0.26
26	Propionic acid	13.51	0.29
27	Acetic acid	16.67	−0.17
28	Adipic acid	13.70	0.08
29	Glutaric acid	15.15	−0.47
30	Malic acid	14.93	−0.3
31	Succinic acid	16.95	−0.59
32	Oxalic acid	22.22	−0.62

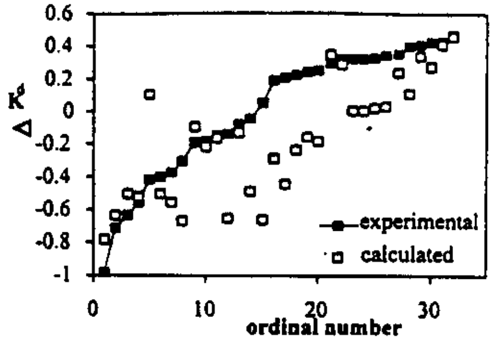

Fig. 5. Comparison of the values of "experimental" ΔK_d and ΔK_d calculated as $f(A_{COOH}, \log K_{ow})$ by eq 7 for 32 low molecular weight model compounds (see Table 4).

of artefacts in SEC analysis of organic compounds in aqueous medium are true. Therefore, the applicability of eq 7 for predictive calculations had to be verified by comparing the values of "calculated" and "experimental" ΔK_d for a large set of relatively hydrophobic carboxylic acids ($\log K_{ow} > -0.5$).

The model compounds included both aliphatic and aromatic carboxylic acids and their hydroxy-derivatives (Table 4). The chosen compounds can be considered as the simplest structural subunits of humic substances which are known to be a mixture of high molecular weight polyhydroxy-polycarboxylic acids both of aromatic and aliphatic character (Hayes *et al.*, 1989).

A comparison of the "experimental" and "calculated" values of ΔK_d is given in Fig. 5, where the values of "experimental" ΔK_d are plotted in the increasing order versus the corresponding ordinal number of the model compound in Table 4.

It can be seen that the values of "experimental" ΔK_d are positive (effect of overexclusion) for the whole set of aliphatic acids, and negative for the whole set of aromatic acids (sorption on the gel). The curve in the range of positive values allows a linear relationship and has a very small slope, whereas in the range of the negative values the slope becomes much steeper and reaches the maximum values for aromatic acids. The curve of the "calculated" ΔK_d values approximates a general trend in the change of "experimental" ΔK_d of aliphatic acids. However, the difference in the position of the "calculated" and "experimental" curves in the plot shows a general underestimation of the influence of $\log K_{ow}$ or overestimation of the impact of the acidity on the elution behaviour of the analyte in the framework of the developed predictive model.

It can be concluded that the applicability of the proposed model is very limited. Generalisation of the model requires abandoning our initial assumption that electrostatic and hydrophobic interactions are the only factors influencing elution behaviour of the organic solute on the SEC column in aqueous medium. Instead, the whole spectrum of possible analyte-gel interactions has to be taken into consideration. To achieve this goal, we decided to use the chemical structure of the analyte for the description of the vast variety of possible interactions.

Description of the elution behaviour of the organic analyte on the SEC column as a function of its characteristic structural parameters

To develop a new model which includes a link between the structure of the organic compound and the deviation from its ideal elution behaviour, we had to modify equation 4 by the following way:

$$K_d^o - K_d(\exp) = \Delta K_d = f(\mathrm{CSP}) \tag{8}$$

where $f(\mathrm{CSP})$ is the function of characteristic structural parameters which contribute the most to gel-analyte interaction under the given experimental conditions. Under the assumption that this function can be described by a linear combination of the variables of structural parameters, we can write the following equation:

$$\Delta K_d = b + AX_1 + BX_2 + CX_3 + \ldots \tag{9}$$

where b is a free member of the polynomial equation, A, B, C are linear coefficients, and X_1, X_2, X_3 are variables of structural parameters of the analyte. In this case, a search of the best fit of equation 9 in relation to the observed experimental data can be conducted by the method of linear regression. Application of this method includes generation of the different sets of the variables of structural parameters, calculation of the corresponding linear coefficients and statistical analysis of the observed results.

To generate the set of structural parameters two possible approaches were considered: (1) an exact description of the structural formula of the compound by the method of chemical topography, (2) description of the structure in terms of integral characteristics referring to atomic ratios between main constituent atoms of the analyte and their distribution between the most abundant structural fragments. Taking into consideration that a vast majority of natural polyelectrolytes, including HS, have neither stoichiometric composition nor a strictly defined structural formula, we have decided to use the second approach for generating the set of structural parameters governing the elution behaviour of the analyte on the SEC column.

The following parameters were taken into consideration: atomic ratios (C/O, C/H, (N + S)/C); distribution of oxygen between carboxyl, hydroxyl and phenolic groups (COOH/O, OH/O, PhOH/O); distribution of carbon between aliphatic and aromatic structures (C_{al}/C_{ar}, C_{ar}/C), carboxyl groups (COOH/C).

The set of the analytes was expanded up to 63 compounds, following the goal to encompass a much broader range of structural subunits of HS and to reduce the statistical uncertainty in the calculations. It is of particular importance that in this set of analytes, amino acids were included as the

Table 5. List of the compounds used for a development of the predictive model $\Delta K_d = f(\text{CSP})$

No	Brutto-formula	Compound
1	$C_7H_6O_4$	3,4-Dihydroxybenzoic acid 3,5-Dihydroxybenzoic acid
2	$C_9H_{10}O_5$	Syringic acid
3	$C_7H_6O_5$	3,4,5-Trihydroxybenzoic acid
4	$C_7H_6O_3$	3-Hydroxybenzoic acid 4-Hydroxybenzoic acid
5	$C_9H_{10}O_2$	Hydrocinnamic acid
6	$C_9H_{11}O_2N$	Phenylalanine
7	$C_4H_8O_2$	1,4-Dioxane
8	C_3H_6O	Acetone
9	$C_7H_6O_2$	Benzoic acid
10	C_2H_6O	Ethanol
11	$C_5H_{11}O_2NS$	Methionine
12	$C_8H_8O_2$	Phenylacetic acid
13	$C_{18}H_{16}O_5$	4-[4-(2-Carboxybenzoyl)-phenyl]butyric acid
14	$C_6H_{13}O_2N$	Leucine
15	$C_6H_{12}O_6$	Glucose
16	$C_3H_8O_3$	Glycerol
17	CH_4O	Methanol
18	$C_9H_8O_4$	Acetylsalicylic acid
19	$C_8H_8O_3$	2-Methoxybenzoic acid
20	$C_5H_{11}ONS$	Valine
21	$C_6H_{12}O_2$	Hexanoic acid
22	$C_6H_{12}O_4N_2S_2$	Cystine
23	$C_5H_9O_2N$	Proline
24	$C_4H_9NO_3$	Threonine
25	$C_3H_7O_3N$	Serine
26	$C_3H_7O_2N$	Alanine
27	$C_2H_5O_2N$	Glycine
28	$C_5H_{10}O_2$	Isovaleric acid Valeric acid
29	$C_7H_{12}O_6$	Quinic acid
30	$C_4H_6O_2$	Crotonic acid Methacrylic acid
31	$C_6H_{10}O_7$	Glucuronic acid
32	$C_4H_8O_2$	Butyric acid
33	$C_9H_8O_4$	2-Carboxyphenylacetic acid
34	$C_{10}H_8O_4$	2-Carboxycinnamic acid
35	$C_{10}H_{10}O_4$	Benzylmalonic acid 1,2-Phenylenediacetic acid
36	$C_3H_6O_3$	Lactic acid
37	$C_5H_9O_4N$	Glutamic acid
38	$C_4H_7O_4N$	Aspartic acid
39	$C_4H_4O_4$	Maleic acid
40	$C_8H_6O_4$	Phthalic acid Terephthalic acid
41	$C_3H_4O_2$	Acrylic acid
42	$C_2H_2O_3$	Glyoxylic acid
43	$C_3H_6O_2$	Propionic acid
44	$C_2H_4O_3$	Glycolic acid
45	$C_2H_4O_2$	Acetic acid
46	$C_6H_{10}O_8$	Glucaric acid
47	CH_2O_2	Formic acid
48	$C_6H_{10}O_4$	Adipic acid
49	$C_7H_{12}O_4$	Pimelic acid
50	$C_5H_8O_4$	Glutaric acid
51	$C_5H_6O_4$	Glutaconic acid
52	$C_4H_6O_6$	Tartaric acid
53	$C_6H_8O_7$	Citric acid
54	$C_4H_6O_5$	Malic acid
55	$C_4H_6O_4$	Succinic acid
56	$C_3H_4O_4$	Malonic acid
57	$C_2H_2O_4$	Oxalic acid

simplest models of heteroatom containing structural subunits of HS. It must be also specifically mentioned that given the chosen approach to describe the structural features of the model compounds by the linear combination of their integral structural parameters, all the data obtained for the stoichiometric isomers (identical brutto formula), needed to be averaged prior to their use in model calculations. This explains the specific layout of Table 5, where to each group of stoichiometric isomers only one ordinal number is assigned.

All the model compounds were analysed by the above described SEC procedure. The obtained data on elution volume were used for calculation of corresponding values of experimental K_d . The values of K_d^0 were calculated from molecular weight of the analyte. The obtained data were used for calculation of "experimental" ΔK_d value for each analyte. A set of the obtained values of "experimental" ΔK_d was arranged according to increasing value of ΔK_d and plotted against ordinal numbers of the compound in the sequence of Table 5. The curve of the values of "experimental" ΔK_d is given in Fig. 6.

The search of the best fit for the curve of the "experimental" ΔK_d was conducted by a method of multiple regression including all the above described structural variables. The results of this procedure for the selected structural parameters which gave the best fits among all the tested ones are given in Table 6. Application of the Fisher criterium to the obtained polynomial equations (Table 5) showed that starting with the fit describing the distribution of oxygen among the functional groups of model compounds (C/O, PhOH/O, COOH/O) and ratio of heteroatoms to carbon (N + S)/C the difference between the "experimental" and "calculated" curves of ΔK_d values was getting statistically insignificant. The addition of the further parameter (C/H) was accompanied with an increase in correlation coefficient from 0.9191 to 0.9225. The graphic representation of the polynome with the highest correlation coefficient (the last fit from Table 6) is given in Fig. 6. The polynomial equation can be written as:

$$\begin{aligned}\Delta K_d = {} & 0.21 - 1.24\text{PhOH/O} - 0.18\text{C/O} \\ & +0.35\text{COOH/O} - 0.47(\text{N}+\text{S})/\text{C} \\ & +0.07\text{C/H} \end{aligned} \tag{10}$$

Theoretically, eq 10 can be used for a prediction of ΔK_d ("calculated" ΔK_d) of an organic analyte of known oxygen distribution and elemental composition.

Recalling the starting point of our research, we explored the applicability of the above described

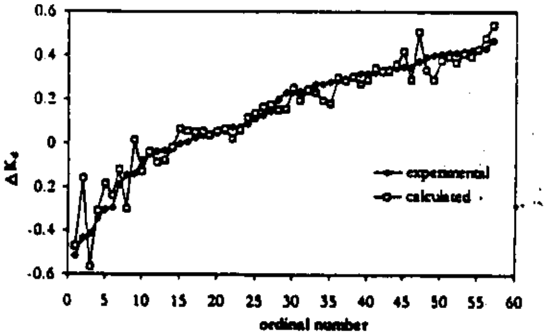

Fig. 6. Comparison of the values of "experimental" ΔK_d and ΔK_d calculated as *f*(CSP) by eq 10 for 63 low molecular weight model compounds (12 of them are stoichiometric isomers, see Table 5).

Table 6. Results of the application of multiple regression with the given structural variables for a search of the best fit for a curve of the values of "experimental" ΔK_d and their statistical estimation. (F - calculated (calc) or tabulated (tabl) value of Fisher-criterium, n - degrees of freedom, confidence level –0.95).

b	OH/O	C/O	COOH/O	(N + S)/C	C/H	F_{calc}	F_{tabl}	n_1	n_2	R^2
0.18	−1.31	–	–	–	–	–	–	–	–	0.3423
0.44	−1.32	−0.15	–	–	–	4.16	1.64	55	54	0.6775
0.23	−1.13	−0.17	0.32	–	–	4.00	1.65	54	53	0.8388
0.24	−1.17	−0.18	0.37	−0.53	–	3.97	1.65	53	52	0.9191
0.21	−1.24	−0.18	0.35	−0.47	0.07	1.09	1.66	52	51	0.9225

model to synthetic polyelectrolytes, which were used as the simplest macromolecular structural analogues of HS and to HS themselves.

Application of the developed predictive equation to synthetic polyelectrolytes and HS

The developed eq 10 was applied to calculation of ΔK_d for a set of synthetic polymers of known composition, structure and molecular weight (sodium polyacrylates, polymethacrylates and polystyrene sulfonates) which were already used in the previous experiments. The values of "calculated" ΔK_d were found to be 0.35, 0.25 and 0.09 for PA, PMA and PSS, respectively. These values were used for calculation of the corresponding values of ideal K_d (K_d^0) by the equation:

$$K_d^o = K_d(\exp) + \Delta K_d \tag{11}$$

The obtained values of K_d^0 were used for calculation of the molecular weight of the corresponding analyte by the equation of the calibration curve of polydextranes from Fig. 2:

$$\log \mathrm{MW} = -3.08 K_d + 4.66 \tag{12}$$

The calculated values were compared with the true values of the molecular weight of the analysed polymers (supplied by the manufacturer, determined by light-scattering technique) and with the determined ones from the calibration curve of the polydextranes according to the standard procedure (PDX). The obtained results are presented in Fig. 7. It can be seen, that the model provides adequate estimates

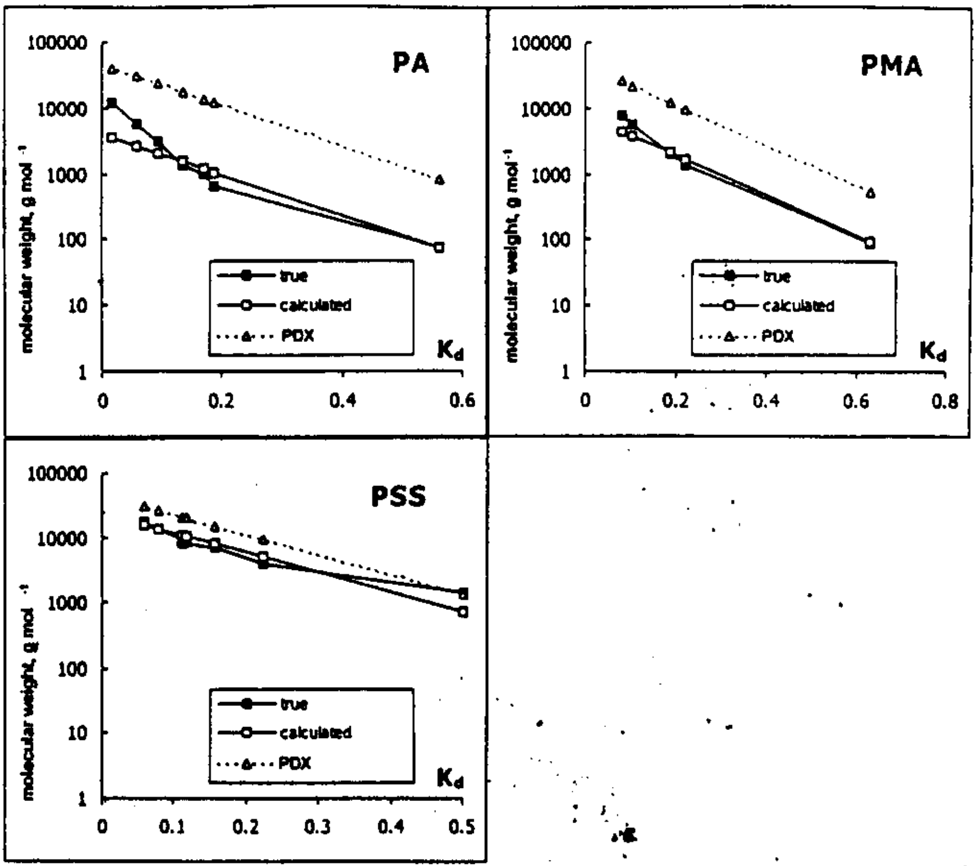

Fig. 7. Comparison of the calculated (using polynomial eq 10) and true values of molecular weight of PA, PMA and PSS.

for all the polymers under study with the values of K_d not less than 0.15. For the lower values of K_d the model estimates remain adequate for the polyelectrolyte of the least charge density–PSS–and become inaccurate with the increase of the charge density for PMA and PA, in particular. For the two last polyelectrolytes the value of the calculated molecular weight was much less than the true one. These results could be expected from the specific protolytic properties of polyelectrolytes related to an abrupt increase in a value of pK of acidic groups with an increase in dissociation degree of the polyanion (Tenford, 1965). The described phenomenon can cause an overestimation of the contribution of the structural parameter COOH/O to the deviation of the experimental values of K_d from the ideal ones. This effect will be larger, the higher the amount of the acidic groups per molecular weight of the polyelectrolyte is.

Keeping in mind the above discussed constraints, it was of interest to conduct the model calculation for HS. For this purpose eq 10 was solved with a set of values given in the literature (see below) for the structural variables under consideration for an averaged (1:1 mixture of HA and FA) sample of HS. The set was derived on the basis of the review of the published data on atomic ratios (Steelink, 1985; Rice and MacCarthy, 1991) and distribution of oxygen among the functional groups (Schnitzer and Khan, 1972; Stevenson, 1982) intrinsic to HS, and included the following structural variables and their values, respectively: C/O = 1.6, C/H = 1.0, (N + S)/C = 0.05, OH/O = 0.1, Ar = 0.5. It should be stressed that this data set is only one possibility out of a broad variety, despite the fact, that according to our knowledge and the literature cited above, it can be considered as rather representative to the general structural features of an averaged HS sample. The value of "calculated" ΔK_d for the given structural variables was found to be 0.02. This value is lower than for all the synthetic polyelectrolytes under study. It is likely that the difference results from the presence of phenolic structures in the molecules of HS. The estimated positive value of ΔK_d shows that the "calculated" molecular weight values of HS are less than those deduced from the PDX calibration curve.

To demonstrate this in practice, we applied the obtained value of "calculated" ΔK_d to calculation of the corresponding values of "ideal" K_d for the analysed set of 35 HS samples. The estimated values of "ideal" K_d were used for obtaining the corresponding "calculated" values of molecular weight of HS samples from the PDX calibration curve. The obtained results, including for comparison the molecular weight values of HS estimated by all the four calibration curves, are given in Fig. 8. As it can be seen, the "calculated" molecular weight values of HS are in the range of 4000 to 15000 g mol^{-1} in comparison to the range of 5500 to 21000 g mol^{-1} which was determined from the K_d(exp) values by PDX-calibration curve.

The obtained model estimates in relation to HS are in general consistency with literature data, where the overestimation of the molecular weight values determined from the SEC-data using polydextranes as calibration standards, is considered from the point of view of the polyelectrolyte theory (De Nobili *et al.*, 1989; Chin *et al.*, 1994). However, given the limitations discussed above, the obtained results can be considered as the first approximation to such a complex system as HS. The aim of the conducted calculations was to demonstrate the potential of the model. Its further development is our immediate goal.

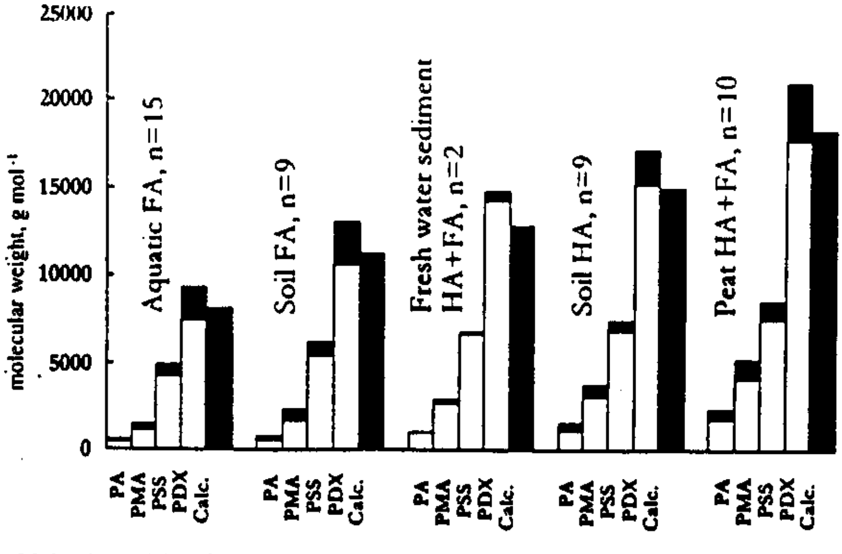

Fig. 8. Molecular weight of HS samples averaged in each group of the same origin (number of the samples (*n*) are given), determined from K_d(exp) values by the calibration curves from Fig. 2 and from calculated "ideal" K_d values (see text). The shaded areas above the bars correspond to the standard deviation of molecular weight obtained for n HS samples.

CONCLUSIONS

Thirty-five samples of HS of different origin were analysed by means of SEC. The observed data were quantified using four sets of calibration standards composed of different polymers encompassing a range of charge densities. A strong dependence of the estimated molecular weights of HS on the charge density of the molecule of the standard was established. On the basis of the SEC analysis of a broad range of low molecular weight compounds (63 representatives of aromatic and aliphatic carboxylic acids, alcohols, sugars, amino acids), which were considered as structural subunits of HS, two approaches were evaluated for calculation of the true molecular weight of the organic analyte from the experimental SEC data. The first approach, based on the consideration of electrostatic-hydrophobic interactions between an organic analyte and the gel-matrix as the only source of distortions of the ideal elution behaviour, was valid for a limited range of aliphatic acids. The second approach, based on establishing a relationship between elution characteristics of the analyte and its structural parameters, was applicable to a much broader set of the low molecular weight compounds. On the basis of the last approach a predictive model was developed, which was able to predict the SEC characteristics of an organic analyte from its known integral structural parameters (atomic ratios and distribution, either carbon or oxygen among functional groups). Such an approach is considered to be especially promising in relation to humic substances, since integral characterization is the only possible way to describe the structure of these extremely complex natural matter. Exploration of the applicability of the model for estimation of the "true" molecular weight of synthetic polyelectrolytes and humic substances allowed conclusions to be made about general consistency of the obtained estimates with the expected ones. An analysis of the observed discrepancies between calculated and experimental data made it possible to define a direction of further development of the model. It should become more applicable to polyelectrolytes in general, and to humic substances, in particular.

Acknowledgements—Dr Perminova thanks the Program of the University of Karlsruhe "International Seminar" and the German Service of Academic Exchanges (DAAD) for financial support of her stay as a visiting scientist at the Department of Water Chemistry of the Engler-Bunte-Institute during fall and spring terms of 1995/1996. She also extends acknowledgements to the J. and C. MacArthurs' Foundation for financial support of her participation at the 8th Meeting of IHSS in Poland in September of 1996 (Grant Award #96433/FSU) where the first presentation of the ideas of this article took place in the form of "the last minute poster".

REFERENCES

Belenkii B. G. and Vilenchik L. Z. (1983) Modern liquid chromatography of macromolecules. *Journal of Chromatography Library*. 25, Elsevier, Amsterdam.

Chin Yu-P., Aiken G. and O'Loughlin E. (1994) Molecular weight, polydispersity, and spectroscopic properties of aquatic humic substances. *Environ. Sci. Technol.* 28, 1853–1858.

De Haan H., Jones R. I. and Salonen K. (1987) Does ionic strength affect the configuration of aquatic humic substances, as indicated by gel filtration? *Freshwater Biol*, 17, 453–459.

De Nobili M., Gjessing E. and Sequi P. (1989) Sizes and shapes of humic substances by gel chromatography. In *Humic Substances II* (Edited by Hayes M. H. B., MacCarthy P., Malcolm R. L. and Swift R. S.), pp. 561–592. John Wiley & Sons, Chichester.

Dubin P. L. (1988) Aqueous size-exclusion chromatography. *Journal of Chromatography Library*. 40, Elsevier, Amsterdam.

Frimmel F. H., Gremm T. and Huber S. A. (1992) Liquid chromatographic characterization of refractory organic acids. *Sci. Total Environ.* 117/118, 197–206.

Hansch C. and Leo A. (1979) *Substituent constants for correlation analysis in chemistry and biology*. John Wiley, New York.

Hayes M. H. B., MacCarthy P., Malcolm R. L. and Swift R. S. (1989) Structures of humic substances: the emergence of 'forms'. In *Humic Substances II* (Edited by Hayes M. H. B., MacCarthy P., Malcolm R. L. and Swift R. S.), pp. 689–734. John Wiley & Sons, Chichester.

Leo A., Hansch C. and Elkins D. (1971) Partition coefficients and their uses. *Chem. Rev.* 71, 525–616.

Rice J. A. and MacCarthy P. (1991) Statistical evaluation of the elemental composition of humic substances. *Org. Geochem.* 17, 635–648.

Sangster J. (1989) Octanol-water partition coefficients of simple organic compounds. *J. Phys. Chem. Ref. Data.* 18, 1111–1228.

Schnitzer M. and Khan S. U. (1972) *Humic Substances in the Environment*. Marcel Decker, N.Y. pp. 37–54.

Steelink C. (1985) Implications of elemental characteristics of humic substances. In *Humic Substances in Soil, Sediment and Water* (Edited by Aiken G. R., McKnight D. M., Wershaw R. L. and MacCarthy P.), pp. 457–475. Wiley, New York.

Stevenson F. J. (1982) *Humus chemistry, genesis, composition, reactions*. Wiley Interscience, N.Y. pp. 221–237.

Sugimura Y. and Suzuki Y. (1988) A high-temperature catalytic oxidation method of determination of non-volatile dissolved organic carbon in seawater by direct injection of liquid samples. *Mar. Chem.* 24, 105–131.

Swift R. S. (1989) Molecular weight, size, shape, and charge characteristics of humic substances: some basic considerations. In *Humic Substances II* (Edited by Hayes M. H. B., MacCarthy P., Malcolm R. L. and Swift R. S.), pp. 450–465. John Wiley & Sons, Chichester.

Tenford Ch. (1965) *Physical chemistry of polymers*. Khimiya, Moscow (In Russian).

Wang E. X. and Benoit G. (1996) Mechanisms controlling the mobility of lead in the spodosols of a northern hardwood forest ecosystem. *Environ. Sci. Technol.* 30, 2211–2219.

Yau W. W., Kirkland J. J. and Bly D. D. (1979) *Modern size-exclusion liquid chromatography*. John Wiley, New York.