Journal of Mathematical Chemistry 11(1992)271–280

SOIL SORPTION AND CHEMICAL TOPOLOGY

Sanja SEKUŠAK and Aleksandar SABLJIĆ
Rugjer Bošković Institute, P.O. Box 1016, 41001 Zagreb, The Republic of Croatia

Abstract

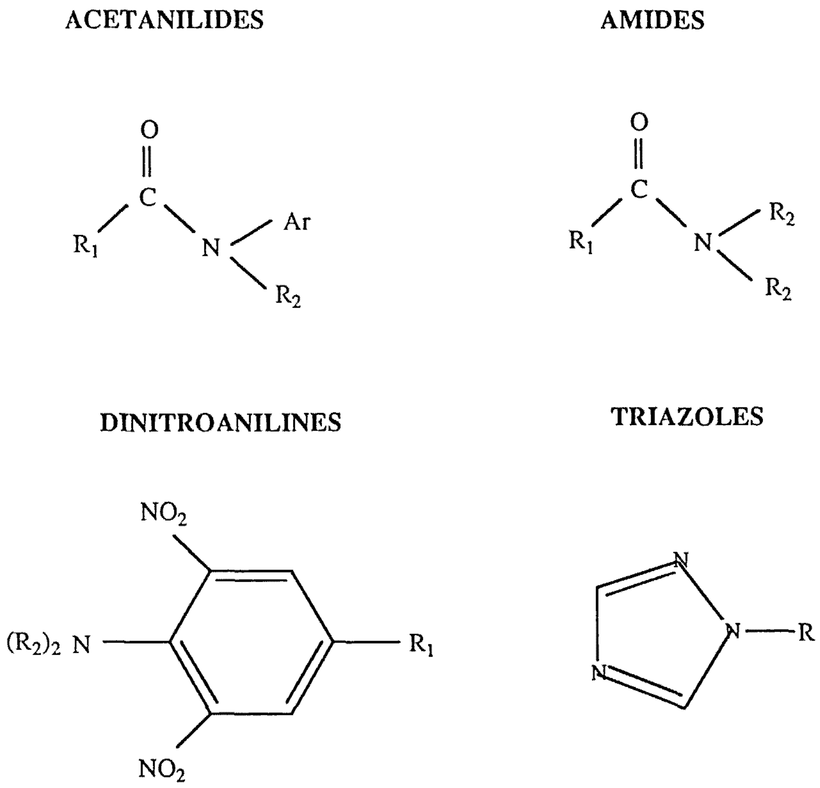
This study was undertaken to test the ability of the molecular connectivity model to predict the soil sorption coefficients of polar organic compounds. Quantitative models, based on the first- and second-order molecular connectivity indices, were developed for estimating soil sorption coefficients of 65 polar compounds arranged into four classes by their functional groups: acetanilides, amides, dinitroanilines, and triazoles.

1. Introduction

The contamination of groundwater by agricultural chemicals, by hazardous chemicals from waste disposal sites, by gasoline and chemicals from underground storage tanks presents a major environmental problem. The adsorption of those chemicals on soil and sediment plays an important role in their transport and mobility in surface and subsurface systems. In addition, the adsorption of agricultural chemicals by soil from aqueous solution strongly influences their performance and residue problems.

Sorption of chemicals in soil is a complex phenomenon and it is influenced by sorption-relevant properties of soils, e.g. organic carbon content [1–3], clay content [4], pH value [5], temperature [5,6], etc., as well as by structural properties of sorbed chemicals. Normalization of the distribution coefficient between soil and water (K_d), measured under standard conditions, on the organic carbon content of soil results in a soil sorption coefficient K_{oc}, which is assumed to be independent of soil characteristics. It is still discussed [2,3] under which conditions it is justified to use K_{oc} as a soil-independent parameter for nonpolar as well as for polar compounds.

The molecular connectivity approach has been shown to be a very useful structural parameter for calculating soil sorption coefficients of nonpolar organic chemicals [7–11]. Our primary objective in this study is to examine the possibility to quantitatively model soil sorption coefficients of polar agricultural chemicals. Molecular structures representing pesticides included in the data set are given in fig. 1.


ACETANILIDES

AMIDES

DINITROANILINES

TRIAZOLES

Fig. 1. Sample of molecular structures of pesticides included in the data set.

2. Method of calculation and experimental data

Molecular connectivity indices are nonempirical parameters which can be easily computed from the structural formula of the molecule. Information used in calculations of molecular connectivity indices are the number and type of atoms and bonds, as well as the number of all electrons and valence electrons. The concept of molecular connectivity indices was introduced by Randić [12] and further developed and extensively exploited by Kier and Hall [13,14], and many others [15,16]. Simple molecular connectivity indices are calculated by the atomic delta values (δ), which are defined for each non-hydrogen atom as the number of adjacent non-hydrogen atoms. The valence molecular connectivity indices are calculated by the valence atomic delta values (δ^V). The valence atomic delta values are calculated from the electron configuration of an atom by

$$\delta^V = (Z^V - h)/(Z - Z^V - 1), \tag{1}$$

where Z^V is the number of valence electrons in an atom, Z is its atomic number, and h is the number of hydrogen atoms bound to the same atom.

The first-order molecular connectivity indices are calculated by

$$^1\chi = \sum\left(\delta_i * \delta_j\right)^{-0.5}, \tag{2}$$

where i and j correspond to the pairs of adjacent non-hydrogen atoms and summations are over all bonds between non-hydrogen atoms, and the second-order molecular connectivity indices by

$$^2\chi = \sum\left(\delta_i * \delta_j * \delta_k\right)^{-0.5}, \tag{3}$$

where i, j, and k correspond to three consecutive non-hydrogen atoms and summation is over all pairs of adjacent bonds between non-hydrogen atoms. Their valence analogs are calculated in the same fashion using valence atomic delta values (δ^V).

Molecular connectivity indices used in this study were calculated by the GRAPH III computer program for microcomputers on an Apple Macintosh SE/30 personal computer [16]. The GRAPH III computer program can calculate molecular connectivity indices up to tenth order for molecules with 36 non-hydrogen atoms or less.

Statistical analysis was carried out in three stages: descriptor elimination, model generation, and model evaluation. Descriptors with more than 80% identical values are removed. All remaining descriptors were examined to see if intercorrelations between them exist. This was done by pairwise correlation and one of each pair exceeding 0.7 was rejected. The problem with the set of molecular connectivity indices is that they are highly intercorrelated, so most of them, especially the higher-order molecular connectivity indices, have to be removed from the analysis. In addition to molecular connectivity indices, we have included in the parameter set some indicator and structural parameters which seem to be adequate descriptors of specific structural features important for the extent of soil sorption of examined polar compounds. The principal component analysis was performed to reduce independent variable space, and to examine multicollinearities and detect patterns among a set of variables. Examination of component loadings gave information about the variance accounted for by each descriptor. Cluster analysis was employed to detect possible groupings within each class of polar compounds, and within independent variables used in the modelling process. Using these procedures, 35 descriptors were excluded from further analysis. The remaining descriptors, about 10 for each class of polar compounds, were analyzed by stepwise and multiple linear regression analysis. The analysis of residuals was performed to detect and identify outliers.

To test the quality of the generated regression equations, the following criteria were taken into consideration: multiple regression coefficient (R), standard error of estimate (s), the test of the null hypothesis (F-test), and the amount of explained variance (EV).

Statistical analysis was carried out using a statistical analysis system (SYSTAT, version 5.0) on an Apple Macintosh SE/30 personal computer.

The soil sorption coefficients of acetanilides, amides, dinitroanilines, and triazoles are taken from the data base described in the study by Gerstl [17]. The distribution coefficients (K_d) used in this study were normalized to the organic carbon content of soil. Logarithmically transformed soil sorption coefficients ($\log K_{oc}$) were used in the modelling process.

3. Results and discussion

In this study, we have tested the ability of the molecular connectivity model to predict the soil sorption coefficients of four groups of polar organic compounds: acetanilides, amides, dinitroanilines, and triazoles. First, the single variable linear models were calculated for the simple and valence first- and second-order molecular connectivity indices. It was found that those indices alone were not able to predict soil sorption coefficients of polar agricultural compounds within the desired accuracy. Thus, we have tested other structural descriptors, specific for each group of chemicals, which describe polar characteristics for the compounds of interest. Variables used in the final models were: ${}^1\chi$, ${}^1\chi^V$, ${}^2\chi^V$, the number of nonpolar parts in a molecule (No^{NP}), the number of polar parts in a molecule (No^{PP}), the number of rings in a molecule (No^{RING}), the number of discrete functional groups in a molecule normalized to the number of atoms in a molecule (FG), the difference between simple and valence first-order molecular connectivity indices (Δ_1), and the polarity index (${}^1_F\chi^V$). The polarity index is the ${}^1\chi^V$ index normalized to the number of discrete functional groups. A functional group is defined as an ensemble of atoms with non-sigma electrons that engages, as a unit, in an intermolecular interaction [14]. Thus, by this definition, the two oxygen atoms and the nitrogen atom of a nitro-group constitute a functional group. Aromatic ring and halogen atoms are not treated as functional groups because they do not make a significant contribution to a molecule in terms of enhancing its polarity. The number of nonpolar parts and the number of polar parts in a molecule were derived under the assumption that only bonds between carbon atoms and between carbon atoms and halogen atoms are nonpolar.

Two or three variable models were developed, depending on the size of each group of polar compounds. Quantitative models describing the soil sorption of acetanilides, amides, dinitroanilines and triazoles are given by the following equations and their statistics:

acetanilides

$$\log K_{oc} = 0.28\ {}^1\chi^V - 0.49\ \Delta_1 + 3.33\ \mathrm{FG} + 0.70, \tag{4}$$

$$n = 21,\quad R = 0.92,\quad F^{3,17} = 33,\quad s = 0.15,\quad \mathrm{EV} = 82.8\%;$$

amides

$$\log K_{oc} = 0.44\ {}^{2}\chi^{V} - 0.24\ \mathrm{No}^{NP} - 0.45, \qquad (5)$$

$$n = 11,\quad R = 0.97,\quad F^{2,8} = 68,\quad s = 0.09,\quad \mathrm{EV} = 93.1\%;$$

dinitroanilines

$$\log K_{oc} = 0.25\ {}^{1}\chi - 0.99\ \mathrm{No}^{PP} - 4.23, \qquad (6)$$

$$n = 16,\quad R = 0.97,\quad F^{2,13} = 88,\quad s = 0.15,\quad \mathrm{EV} = 92.1\%;$$

triazoles

$$\log K_{oc} = 1.27\ {}^{1}_{F}\chi^{V} - 14.18\ \mathrm{No}^{RING} - 0.25\ \mathrm{No}^{PP} - 2.32, \qquad (7)$$

$$n = 15,\quad R = 0.95,\quad F^{3,11} = 33,\quad s = 0.28,\quad \mathrm{EV} = 88.0\%.$$

The above models are also graphically represented in fig. 2.

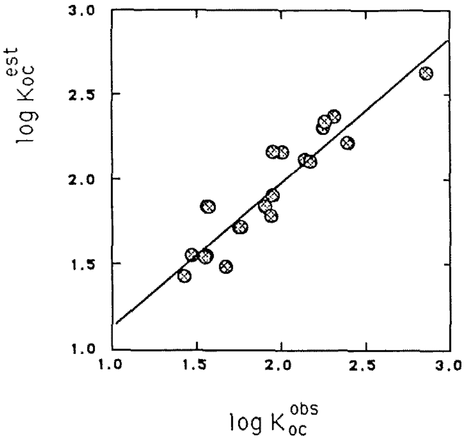
The statistical parameters show that eqs. (4)–(7) are statistically significant above the 99% level. They account for 82–93% of the variations in the log K_{oc} data. Standard errors of estimate are within experimental errors, which are between 0.1 and 0.4 logarithmic units.

The molecular connectivity model was shown to be very successful in modelling soil sorption coefficients of non-polar compounds, but it is not sufficient to accurately predict soil sorption coefficients of polar compounds [7,8]. Since molecular connectivity indices are topological indices, they express in numerical form the topology of chemical species and reflect in varying degrees their shape and size. This information is insufficient for a description of the soil sorption process of polar compounds. Thus, some local structural descriptors and indicator variables are needed. No^{NP}, No^{PP}, and FG were used to distinguish between the polar and nonpolar part of a molecule. Separation between the polar and non-polar part of a molecule was needed because the hydrophobic part of a molecule influences soil sorption, while the polar part tends to be dissolved in water. The Δ_1 value is a count of pi and lone pair electrons [14], so it seems reasonable to assume that it can give us some measure of polarity of compounds of interest. The polarity index (${}^{1}_{F}\chi^{V}$) is introduced to account for the effect of multiple interaction sites [14]. This is a rather rough scheme for the description of polar characteristics of compounds and some calculations of bond polarity, atomic charges, polarizability or some other electronic parameter of examined compounds should be made. The advantage of our models is their simplicity and the fact that parameters can be estimated very easily and quickly.

One difficulty in modelling soil sorption coefficients represents the quality of measured soil sorption data. Since the distribution coefficient between soil and water is influenced by soil properties [1–7] and experimental conditions under

ACETANILIDES

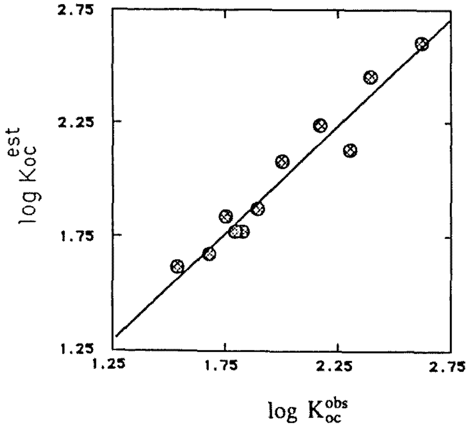
$\log K_{oc} = a\,{}^{1}\chi^{V} + b\,\mathit{DELTA1} + c\,\mathit{FG} + \text{const.}$



a = 0.28
b = - 0.49
c = 3.33
const. = 0.70
R = 0.92

AMIDES

$\log K_{oc} = a\,{}^{2}\chi^{V} + b\,\mathrm{No}^{NP} + \text{const}$



a = 0.44
b = -0.24
const. = -0.45
R = 0.97

Fig. 2. Plots of observed versus calculated soil sorption coefficients (acetanilides, amides).

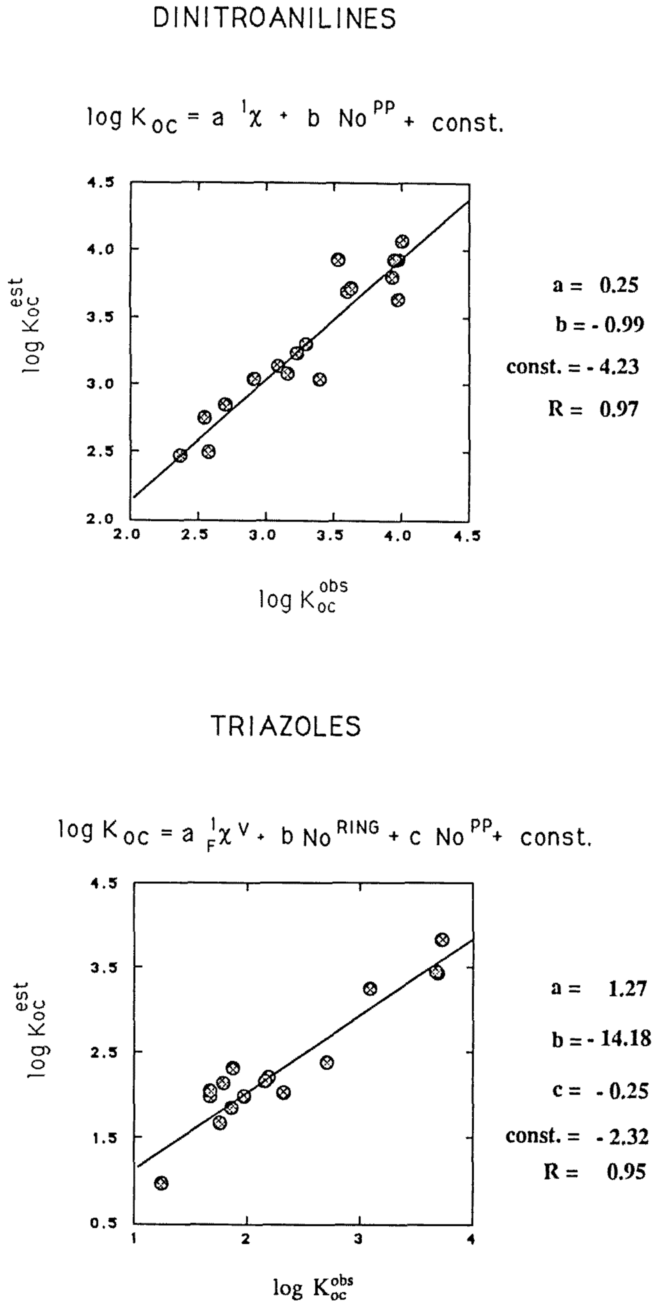


Fig. 2. Plots of observed versus calculated soil sorption coefficients (dinitroanilines, triazoles).

Table 1

Comparison of observed and calculated soil sorption coefficients of acetanilides, their first-order valence molecular connectivity indices (${}^1\chi^V$), the difference between simple and valence first-order molecular connectivity indices (Δ_1), and the number of discrete functional groups normalized to the number of atoms in a molecule (FG).

Compound		${}^1\chi^V$	Δ_1	FG	log K_{oc}	
					observed	predicted
1.	acetochlor	6.711	1.940	0.222	2.320	2.373
2.	alachlor	6.684	2.005	0.222	2.260	2.333
3.	butachlor	8.207	2.071	0.190	2.860	2.623
4.	butyranilide	4.675	1.651	0.154	1.770	1.717
5.	metholachlor	6.986	2.165	0.211	2.250	2.303
6.	metazachlor	6.565	2.565	0.250	2.140	2.118
7.	metalaxyl	6.387	3.047	0.250	1.570	1.832
8.	propachlor	5.129	1.535	0.214	2.170	2.101
9.	acetanilide	3.115	1.673	0.200	1.430	1.422
10.	2-chloroacetanilide	3.598	1.600	0.273	1.580	1.836
11.	3-chloroacetanilide	3.592	1.590	0.273	1.910	1.839
12.	3-bromoacetanilide	4.006	1.176	0.273	2.010	2.158
13.	4-bromoacetanilide	4.006	1.176	0.273	1.950	2.158
14.	3-fluoroacetanilide	3.214	1.968	0.273	1.570	1.548
15.	4-fluoroacetanilide	3.214	1.968	0.273	1.480	1.548
16.	3-methylacetanilide	3.525	1.657	0.182	1.680	1.485
17.	3-nitroacetanilide	3.614	2.478	0.385	1.940	1.783
18.	4-methoxyacetanilide	3.638	2.082	0.250	1.550	1.534
19.	3-trifluoromethylacetanilide	3.842	2.551	0.357	1.750	1.718
20.	3-chloro-4-methoxyacetanilide	4.121	2.009	0.308	1.950	1.899
21.	3,4-dichloroacetanilide	4.074	1.518	0.333	2.400	2.210

Table 2

Comparison of observed and calculated soil sorption coefficients of amides, their second-order valence molecular connectivity indices (${}^2\chi^V$), and the number of nonpolar parts in a molecule (No^{NP}).

Compound		${}^2\chi^V$	No^{NP}	log K_{oc}	
				observed	predicted
1.	isoxaben	5.499	2.000	2.400	2.455
2.	pronamid	4.753	2.000	2.310	2.126
3.	fluoropronamid	3.730	2.000	1.680	1.674
4.	chloropronamid	4.166	2.000	1.900	1.866
5.	bromopronamid	4.644	2.000	2.010	2.077
6.	methylpronamid	4.089	2.000	1.760	1.832
7.	methoxypronamid	3.952	2.000	1.830	1.772
8.	ipropylpronamid	5.000	2.000	2.170	2.235
9.	H-pronamid	3.589	2.000	1.540	1.611
10.	napropamid	4.748	4.000	2.620	2.604
11.	diphenamid	4.487	1.000	1.800	1.768

Table 3

Comparison of observed and calculated soil sorption coefficients of dinitroanilines, their first-order molecular connectivity indices (${}^{1}\chi$), and the number of polar parts in a molecule (No^{PP}).

Compound	${}^{1}\chi$	No^{PP}	log K_{oc} observed	log K_{oc} predicted
1. nitralin	10.628	4.000	2.920	2.905
2. SD11830	8.552	4.000	2.370	2.385
3. SD12030	11.189	4.000	3.100	3.046
4. SD12346	11.689	4.000	3.300	3.171
5. SD12400	9.496	4.000	2.580	2.621
6. SD12639	9.628	4.000	2.560	2.654
7. SD13207	10.189	4.000	2.710	2.795
8. GS-38496	11.482	4.000	3.170	3.119
9. GS-39985	11.982	4.000	3.230	3.244
10. DNnPTFpT*	9.307	3.000	3.610	3.572
11. chlornidine	9.417	3.000	3.940	3.599
12. benfluralin	10.628	3.000	3.985	3.903
13. fluchloralin	10.628	3.000	3.545	3.903
14. trifluralin	10.628	3.000	3.950	3.903
15. profluralin	11.146	3.000	4.010	4.032
16. dinitramin	10.056	3.000	3.630	3.759

*DNnPTFpT: 2,6-dinitro-N-*n*-propyl-trifluoro-*p*-toluidine.

Table 4

Comparison of observed and calculated soil sorption coefficients of triazoles, their polarity indices (${}^{1}_{F}\chi^{V}$), the number of rings in a molecule (No^{RING}), and the number of polar parts in a molecule (No^{PP}).

Compound	${}^{1}_{F}\chi^{V}$	No^{RING}	No^{PP}	log K_{oc} observed	log K_{oc} predicted
1. benzyltriazole	1.271	1.000	2.000	1.690	1.761
substituted benzyltriazoles					
2. (4-fluorobenzyl)	1.304	1.000	2.000	1.870	1.779
3. 1-(4-methoxybenzyl)	1.445	2.000	2.000	1.800	2.232
4. 1-(4-t-butylbenzyl)	1.928	1.000	2.000	2.160	2.122
5. 1-(4-methylbenzyl)	1.505	1.000	2.000	1.770	1.889
6. 1-(4-chlorobenzyl)	1.430	1.000	2.000	1.980	1.848
7. 1-(4,4-dichlorobenzyl)	1.591	1.000	2.000	2.330	1.936
8. buthidazole	2.061	1.000	2.000	2.190	2.195
9. NIA23486	1.834	1.000	2.000	1.690	2.070
10. imazalil	3.381	2.000	2.000	3.730	3.296
11. methazole	5.007	1.000	2.000	3.660	3.813
12. propiconazol	2.660	2.000	3.000	3.700	3.975
13. thiabendazol	1.735	2.000	3.000	3.670	3.467
14. triadimefon	2.187	2.000	2.000	2.710	2.640
15. tricyclazole	1.618	1.000	3.000	3.100	3.027

which it is measured [5,6], it will be advantageous that the soil sorption data are from a single source. Unfortunately, this is usually not the case. Even in this study, data were collected from many sources. Consequently, the limitations on the amount and quality of available data reduce the precision with which we can fit models and limit the precision at which a particular model can be applied. A priority task in future investigations should be to develop a unique scheme for analyzing, sorting and standardizing reported data.

4. Conclusions

Simple models, based on topological properties of organic molecules, can be used to predict, within the experimental error, the soil sorption coefficients of polar organic chemicals. The first- and second-order molecular connectivity indices, which describe the molecular surface and the size of the molecule, are used as main variables, while some local variables are added to describe specific structural features of polar compounds. These structural models will enable reliable predictions to be made for acetanilides, amides, dinitroanilines, and triazoles whose soil sorption coefficients have not yet been measured.

Acknowledgements

This work was supported by the National Institutes of Health, DHHS, and RZAMTES of Croatia through funds made available to the US–Yugoslav Joint Board on Scientific and Technological Cooperation (Grant No. JF 918), and by Grant No. 1-07-159 awarded by the Ministry of Science, Technology, and Informatics of the Republic of Croatia.

References

[1] P. Grathwohl, Environ. Sci. Technol. 20(1990)1687.
[2] H. Kishi, N. Kogure and Y. Hashimoto, Chemosphere 21(1990)867.
[3] M. Schrap and A. Opperhuizen, Chemosphere 18(1990)1883.
[4] R.S. Adams and P. Li, Soil Sci. Soc. Amer. Proc. 35(1971)78.
[5] D.M. McGlamery and F.W. Slife, Weeds 14(1966)237.
[6] T.H. Dao and T.L. Lavy, Weed Science 26(1978)303.
[7] A. Sabljić, Environ. Health Perspec. 83(1989)179.
[8] A. Sabljić, Environ. Sci. Technol. 21(1987)358.
[9] A. Sabljić and M. Protić, Bull. Environ. Contam. Toxicol. 28(1982)162.
[10] A. Sabljić, J. Agric. Food Chem. 32(1984)243.
[11] A. Sabljić, *QSAR in Environmental Toxicology*, Vol. 2 (Reidel, Dordrecht, 1987), p. 309.
[12] M. Randić, J. Amer. Chem.Soc. 97(1975)6609.
[13] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemical and Drug Research* (Academic Press, New York, 1976).
[14] L.B. Kier and L.H. Hall, *Molecular Connectivity in Structure–Activity Analysis* (Research Studies Press, Chichester, 1986).
[15] N. Trinajstić, *Chemical Graph Theory*, Vols. 1, 2 (CRC Press, Boca Raton, 1983).
[16] A. Sabljić and N. Trinajstić, Acta Pharm. Yugosl. 31(1981)189.
[17] Z. Gerstl, J. Contam. Hydrol., in press.

Program GRAPH III is now fully operational and is distributed by Apple Center Ljubljana, Parmova 41, 61000 Ljubljana, The Republic of Slovenia.