TOPS-MODE Versus DRAGON Descriptors in QSAR. 3. Soils Sorption.

Maykel Pérez González [1, 2, *], **Miguel Angel Cabrera Pérez**[2], **Reinaldo Molina Ruiz**[2] **and Ronal Ramos de Armas**[3].

[1] Unit of Services, Experimental Sugar Cane Station "Villa Clara-Cienfuegos", Ranchuelo, Cuba.

[2] Department of Drug Design, Chemical Bioactives Center, Central University of Las Villas, Santa Clara, Villa Clara. Cuba.

[3] Department of Chemistry, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Santa Clara, Villa Clara, Cuba.

Abstract

The TOPological Sub-Structural MOlecular DEsign (TOPS-MODE) approach has been applied to the study of the soil sorption coefficient of various phenylureas herbicides. A model able to describe more than 93% of the variance in the experimental soil sorption coefficient of 44 phenylureas herbicides was developed with the use of the mentioned approach. In contrast, none of eleven different approaches, including the use of Constitutional, Molecular walk counts, BCUT, Charges indices, 2D autocorrelations, Randic molecular profiles, Geometrical, RDF, 3D Morse, GETAWAY and WHIM descriptors was able to explain more than 91% of the variance in the mentioned property with the same number of descriptors. In addition the TOPS MODE allows a simple interpretation of the model in comparison with others methodologies. In addition, the TOPS-MODE approach permitted to find the contribution of different fragments to the soil sorption coefficients giving to the model a straightforward structural interpretability.

Keywords: Phenylureas, Herbicides, Soil sorption coefficient, QSPR, TOPS-MODE.

*Corresponding Author: mpgonzalez76@yahoo.es
  Tel: 53-42-281473; Fax: 53-42-281130

1 INTRODUCTION

The sorption of commercial chemicals by soil and sediment plays a very important role in their transport and mobility in the environment [1]. Furthermore, the sorption may significantly influence the chemical and biological transformation or degradation of chemicals in the aquatic environment. Thus, the measurement or accurately estimation of soil sorption coefficients for hazardous chemicals has a critical importance for evaluating their fate and potential exposure to chemicals in the environment and, consequently, for the whole process of environmental risk assessment.

Since the experimental determination of soil sorption coefficients is difficult and expensive, some theoretical methods for soil sorption coefficients determinations have been carried out, where, different regression equations between structured and parameters such as: water solubility, octanol-water partition coefficient or bioconcentration factor, have been used.

Despite of extensive experimental work carried out at numerous laboratories for more than 40 years, the measurement of soil sorption coefficients are available for less than 500 chemicals and reliable data of sorption coefficients are, in many cases, unavailable.

Thus, a considerable number of studies have been performed to develop quantitative models for estimating the soil sorption coefficients of organic chemicals [2-6]. Their analysis have shown that the majority of QSAR models [7-14], for estimating the soil sorption coefficients (log K_{oc}), use the n-octanol/water partition coefficient (log K_{ow}) as molecular descriptors. Unfortunately, these models are based on a small number of chemicals and no external validations have been performed.

In the context of *in silico* methods for modeling physicochemical and biological properties of chemicals the topological sub-structural molecular design (TOPS-MODE) approach has been introduced [15-21].

The successful application of this theoretical approach to the modeling of toxicological and ecotoxicological properties [22, 23] have inspired us to perform a more exhaustive study in order to test and/or validate the TOPS MODE applicability in assessing discoveries and chemical environmental impact. The selection of a data set of phenylureas pollutants compounds is not casual; this property was previously studied using the WHIM descriptors by Gramatica et al. [24]. Thence, it may result very interesting to test the potentialities of TOPS-MODE approach with this data set.

Therefore, the aim of this study was to investigate the role that TOPS-MODE and other molecular descriptors calculated from the molecular structure plays on the explanation of such property using a data set of 44 phenylureas herbicides.

2 MATERIALS AND METHODS

2.1 The Tops-Mode Approach

TOPS-MODE is based on the computation of the spectral moments of the bond matrix, the mathematical basis of which has been described previously [15 - 19]. The TOPS-MODE approach has been recently reviewed in the literature [25], and both the methodology and its software implementation have been described [26].

According to the authors, the application of the TOPS-MODE approach to the study of quantitative structure – property relationships (QSPR) can be summarized in the following steps:

1. To draw the hydrogen-depleted molecular graphs for each molecule of the data set,

2. To use appropriate bond weights in order to differentiate the molecular bonds, e.g., bond distance, bond dipoles, bond polarizabilities, etc.,

3. To compute the spectral moments of the bond matrix with the appropriate weights for each molecule in the data set, generating a table in which rows correspond to the compounds and columns correspond to the spectral moments of the bond matrix. Spectral moments are defined as the trace of the different powers of the bond matrix [27],

4. To find QSPR by using a suitable linear or non-linear multivariate statistical technique, such as multi-linear regression analysis (MRA), etc. to obtain an equation of the form:

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 \ldots\ldots\ldots a_k\mu_k + b \qquad\qquad (Eq.\ 1)$$

where P is the property measurement, μ_k is the kth spectral moment, and a_k's are the coefficients obtained by the MRA,

5. To test the predictive capability of the QSPR model by using cross-validation techniques.

6. To compute the contributions of different groups of interest in order to determine their quantitative contribution to the soil sorption coefficient of the molecules under study.

The computation of fragment contributions to the soil sorption coefficient being studied is probably the most important advance of the TOPS-MODE approach when compared to other traditional QSAR and QSPR methods. The procedure consists of calculating the

spectral moment for all the fragments contained in a given substructure, and by subtraction of these spectral moments to obtain the contribution of the substructure. The general algorithm for this computational approach is as follows:

First, we select the substructure whose contribution to the moments we would like to determine. Then, we generate all the fragments, which are contained in the corresponding substructure, and calculate the spectral moments for both, the substructure and all their fragments. The contribution of the substructure to the spectral moments is finally obtained as the difference between the spectral moments of the substructure and all those from their fragments. Once, the contributions of the different structural fragments are obtained, we only need to substitute these contributions into the quantitative model developed to describe the property studied.

2.2 Data Sets and Computational Strategies.

A data set of 44 compounds for which the soil sorption coefficients were reported in the literature was selected [24]. The parameter studied is log K_{oc}. This is the most common and today generally accepted quantitative measure of the sorption of organic pollutants by soil or sediment from aqueous solutions. This chemical-specific parameter provides a relative measure of mobility in aqueous/soil system. In general, compounds with higher log K_{oc} values will be less mobile than those with lower values. The names of the compounds, as well as the calculated and experimental values of log K_{oc} are shown in Table 1.

TOPS-MODE [26] and DRAGON [28] computer softwares were employed to calculate the molecular descriptors. In the case of TOPS-MODE software, the polarizability, the Gasteiger-Marsili charges and hidrophobicity were used to weigh the bond adjacency matrix. The selection of only these three types of descriptors from the whole pool of ten types included in TOPS-MODE methodology was carried out for the sake of simplicity and on the belief that hydrophobic and polarity parameters influence the soil sorption coefficient of phenylureas herbicides. The total number of descriptors used for obtaining this model was 48 spectral moments. On the other hand, we carry out geometry optimization calculations for each compound used in this study using the quantum chemical semiempirical method AM1 [29] included in MOPAC 6.0 [30]. Eleven other models were developed using the computer software Dragon [28], and calculating the Constitutional, Molecular walks counts, BCUT, Charges indices, 2D autocorrelations, Randic molecular profiles, Geometrical, RDF, 3D-MORSE, GETAWAY and WHIM descriptors [31]. The statistical processing to obtain the QSAR models was carried out by using the forward stepwise regression methods.

The statistical significance of the models was determined by examining the regression coefficient, the standard deviation, the number of variables, the cross validation leave-one-out statistics and the proportion between the cases and variables in the equation.

3 RESULTS AND DISCUSSION

3.1 Quantitative Structure Property Relations

The best QSPR model obtained with the TOPS-MODE descriptors is given below together with the statistical parameters of the regression.

$$\log K_{OC} = -0.11 - 1.02 \cdot 10^{-4} \cdot \mu_9{}^{GM} + 0.01 \cdot \mu_4{}^{H} - 2.26 \cdot 10^{-7} \mu_{13}{}^{H} + 0.03 \cdot \mu_1{}^{P} \quad \textbf{(Eq. 2)}$$
$$+ 0.08 \cdot \mu_2{}^{P}$$

$N = 44 \quad S = 0.119 \quad R^2 = 0.938 \quad F = 115.18 \quad p = 0.000 \quad q^2 = 0.917 \quad S_{cv} = 0.211$

where N is the number of compounds included in the model, R^2 is the correlation coefficient, S the standard deviation of the regression, F the Fisher ratio, q^2 the correlation coefficient of the cross − validation, p is the significance of the variables in the model and S_{cv} is the standard deviation of the cross − validation.

The variables included in the model are designated as follows: the sub-index represents the order of the spectral moment and the super-index the type of bond weight used, i.e., *GM* for Gasteiger Marsili Charges, *P* for polarizability and *H* for hydrophobicity.

From the statistical point of view this model is a robust one as can be seen from the statistical parameters of the cross-validation.

As we previously mentioned, one of the objectives of the current work was to compare the reliability of the TOPS-MODE approach to describe the property under study as compared to other different descriptors and methods. Consequently, 11 other models were developed using the same data set and the same number of variables that was included in the TOPS-MODE QSPR model. The results obtained with the use of Constitutional, Molecular walks counts, BCUT, Charges indices, 2D autocorrelations,

Randic molecular profiles, Geometrical, RDF, 3D-MORSE, GETAWAY and WHIM descriptors are given in Table 2.

Table 2 comes about here.

3.2 Comparison with other Approaches

As can be seen there are remarkable differences concerning the explanation of the experimental variance given by these models compared to the TOPS-MODE one. While the TOPS-MODE QSPR model explains more than 93% of permeability the rest of the models are unable to explain beyond 91% of such variance.

The TOPS-MODE model is higher than the other eleven models not only in the statistical parameters of the regression but also, and more importantly, in its stability upon inclusion/exclusion of compounds as measured by the correlation coefficient and standard deviation of the cross-validation. Because of the structural variability of the compounds in the data set, the statistics of the leave-one-out cross validation might be considered as a good measurement of the predictive ability of the models. As can be seen in Table 2, the value of the determination coefficient of leave-one-out cross-validation for the model obtained with the spectral moments ($q^2 = 0.917$) was the highest of all.

3.3 Interpretation of the Model

One of the most important advantages that TOPS-MODE brings to the study of QSPR and QSAR is concerned with the structural interpretability of the models. This interpretability comes from the fact that the spectral moments can be expressed as linear

combinations of structural fragments. In such a way, we can determine the fragments with a positive or negative contribution to the property under study, which can be interpreted in terms of the physicochemical or biological processes influencing its. In Table 3 and Figure 2 we show the fragments and their contributions to the soil sorption coefficient as calculated from Equation 2.

Table 3 comes about here

Figure 2 comes about here

Here we have studied only some small fragments that are present in the structures of the compounds in the data set. However, extension of this study to other fragments in such molecules or even to fragments in molecules not contained in this data set should be straightforward, as shown for other particular cases elsewhere [22, 23, 32].

According to the contributions of fragments F_{35} to F_{38} (see Figure 2), the soil sorption coefficient increase as the number of carbons in the aliphatic ring is increased. This phenomenon can be explained to when the number of carbon atoms increase in the ring, the hydrophobicity of the fragment increase too. This behavior, that not is unique of the phenylureas herbicides, has been observed for other authors [3, 5, 33]. This study evidence that an increase of the partition coefficient (octanol/water, log K_{ow}) produce high value of the soil sorption, proportionally.

Nevertheless, we have appreciated others findings that we are comment following. For example, when we compared the contribution of the fragment F_{19} and F_{37} the last fragment has the same number of carbon atom in their structure than the phenyl ring (F_{19}), but this are contribution almost two-fold higher than their aliphatic homologous. Obviously, according to the above explanation the phenyl ring not presents the same hydrophobicity that the cycle hexyls ring. Apparently, the phenyl ring can interact with

the sites of the colloid charge positively and the organic matter of the soil which facilitates this high contribution. Other factor that supports this theory is the increase of polarization atoms in the structure which contribute to increase their sorption for the soil, according to the equation 2. Also this is evidence when we compare fragments F_1 to F_3 and F_{23} to F_{25} where an increasing of the size of the halogen atom in the fragment, and therefore of their polarizability procedure high values of the sorption coefficient. This theory support that the negative fraction of the colloid forms of the soil has an important influence in this type of process for this herbicide family.

Finally, when the number of halogen is increased in a fragment of this family (F_{23}, F_{30}, F_{31}) a remarkable increase of the soil sorption coefficient is observed. This is the special interest for some time due to the pollution of the environment. We recently demonstrated how, in general sense an increasing of the halogen atoms in a chemical structure increase the herbicide property [32]. Therefore, this involve that possess a big soil sorption coefficient and for that reason present a higher trouble for their elimination and biodegradation of the soils, property not desirable for new herbicides.

For this reason combine models are necessary in the future for resolve this type of problems.

4 CONCLUDING REMARKS

We have shown that the TOPS-MODE approach is able to describe the soil sorption coefficient of phenylureas herbicides. In fact, we have developed a model for predicting the soil sorption coefficient of a data set of 44 phenylureas herbicides, which is both

statistically and chemically sound. This model explains more than 93 % of the variance in the experimental permeability coefficients and shows good predictive ability in cross-validation. These features are significantly better than those obtained for other eleven different methodologies used to predict this property. Therefore, the spectral moments show a better performance than other kind of descriptors, which suggests that they can be used in new QSPR applications.

On the other hand, the main advantage of using a TOPS-MODE approach in QSPR/QSAR has been confirmed again in this work. This approach is able to derive group contributions and simultaneously provides the means of interpreting them thus contributing to our understanding of the physicochemical or biological processes involved.

Finally, the present results were compared to others obtained in previous works and evidence was obtained on the similarity of the properties that explain the phenomenon.

5 REFERENCES

1. H. Cheng. Pesticides in the soil environment: Processes, Impact, and modeling. Editor, Soil Science Society of America, Inc., Madison, Wisconsin 1990.

2. E. Kenega and C. Goring. Relationship between water solubility, soil sorption, octanol-water partitioning and concentration chemicals in biota. In J. C. Eaton, P. R. Parrish, A. C. Hendricks, eds., Aquatic Toxicology. STP 707, (Third Symposium). American Society for Testing and Materials, Philadelphia (PA), **1980**, 78-115.

3. G. Briggs. Theoretical and Experimental Relationships Between soil Adsorption, Octanol-Water Partition Coefficients, Water Solubilities, Bioconcentration Factors and the Octanol-Water Partition Coefficients, Water Solubilities, Bioconcentration Factors and the Parachor. *J. Agric. Food Chem.* **1981**, 29, 1050-1059.

4. S. Means, G. Wood, J. Hassett and L. Banwart. Sorption of Amino-and Carboxy-Substituted polynuclear aromatic hydrocarbons by sediments and soils. *Environ. Sci. Technol*. **1982**, 16, 93-98.

5. Z. Gerstl. Estimation of organic chemicals sorption by soils. *J. Con. Hydrology*. **1990**, 6, 357-375.

6. W. Meylan, H. Howard and R. Boethling. Molecular topology/fragment contribution method for predicting soil sorption coefficients. *Environ. Sci. Technol*. **1992,** 26, 1560-1567.

7. J. Hasset, C. Means, L. Banwart and G. Wood. Sorption properties of sediments and energy-related pollutans. **1980**, EPA-600/3-80-041, 1-133.

8. S. Rao and M. Davidson. Estimation of pesticides retention and transformation parameters required in nonpoint source pollution models. In: Environmental nonpoint source pollution. *Ann Arbor Sci. Publ. Inc., Mich.*, **1980**, 23-67.

9. J. McCall, L. Swann, A. Laskowski, M. Unger, A. Vrona and J. Dishburger. Estimation of chemical mobility in soil from liquid chromatographic retention times. *Bull. Environ. Contam. Toxicol*. **1980**, 24, 190-195.

10. G. Briggs. Adsortion of pesticides by some Australian soils, *Aust. J. soil Res*. **1981**, 19, 61-68.

11. P. Lagas. Sorption of chlorophenols in the soil. *Chemosphere*. **1988**, 17, 205-216.

12. J. Kanazawa. Relationship between the soil sorption constant for pesticides and their physicochemical properties. *Environ. Toxicol. Chem.* **1989**, 8, 447-484.

13. K. Kawamoto and K. Urano. Parameters for predicting fate of organochlorine pesticides in the environment. (II) Adsorption constant to soil. *Chemosphere*. **1989**, 19, 1223-1231.

14. J. Eadie, R. Morehead and F. Landrum. Three-phase partitioning of hydrophobic organic compounds in great lakes waters. *Chemosphere*. **1990**, 20, 161-178.

15. E. Estrada. Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. *J. Chem. Inf. Comput. Sci.*, **1996**. 36, 844-849.

16. E. Estrada. Spectral moments of the edge adjacency matrix in molecular graphs. 2. Molecules containing heteroatoms and QSAR applications. *J. Chem. Inf. Comput. Sci.*, **1997**. 37, 320-328.

17. E. Estrada. Spectral moments of the edge adjacency matrix in molecular graphs. 3. Molecules containing cycles. *J. Chem. Inf. Comput. Sci.*, **1998**. 38, 23-27.

18. E. Estrada. Modeling the diamagnetic susceptibilities of organic compounds by a substructural graph theoretical approach. *J. Chem. Soc. Faraday Trans*., **1998.** 94, 1407-1411.

19. E. Estrada. Generalized Spectral Moments of the interacted line graph sequence. A novel approach to QSPR studies. *J. Chem. Inf. Comput. Sci*., **1998**. 39, 90-95.

20. E. Estrada. and Y. Gutiérrez. Modeling chromatographic parameters by a novel graph theoretical sub-structural approach. *J. Chromatogr. A*, **1999,** 858, 187-199.

21. E. Estrada, Y. Gutiérrez and H. González. Modeling diamagnetic and magnetooptic properties of organic compounds with the TOSS-MODE approach. *J. Chem. Inf. Comput. Sci*., **2000**, 40, 1386-1399.

22. E. Estrada and E. Uriarte. Quantitative structure-toxicity relationships using TOPS-MODE. 1. Nitrobenzene toxicity to *tetrahymena pyriformis. SAR and QSAR in Envirom. Research.*, **2001**, 12, 309-324.

23. E. Estrada, E. Molina and E. Uriarte. Quantitative Structure-Toxicity Relationships using TOPS-MODE. 2. Neurotoxicity of a Non-congeneric Series of Solvents*, SAR and QSAR in Environmental Research*. **2001**, 12, 445-459.

24. P. Gramatica, M. Corradi and V. Consonni. Modeling and prediction of soil sorption coefficient of non-ionic organic pesticide by molecular descriptors. *Chemosphere.* **2000**, 41, 763-777.

25. Y. Gutierrez and E. Estrada. TOSS-MODE (1997) (Topological Sub-Structural Molecular Design) for Windows Version 4.0, Universidad de Santiago de Compostela, Spain.

26. Y. Gutierrez and E. Estrada. Modes Lab ®, **2002**, version 1.0 b.

27. E. Estrada, Edge Adjacency relationship and a novel topological index related to molecular volume, *Inf. Comput. Sci*., **1995**, 35, 31-33.

28. R. Todeschini, V. Consonni and M. Pavan. 2002 Dragon. Software version 2.1.

29. J.S. Michael, E. Dewar, G. Zoebisch, F. Eamonn, and J.P. Stewart. J. Am. Chem. Soc., **1985,** 107, 3902.

30. J. J. P. Stewar. **1990** MOPAC manual, 6th ed., p 189, Frank J. Seiler Research Laboratory, U. S. Air Force academy, Colorado Springs, CO.

31. R. Todeschini and V. Consonni. Handbook of molecular descriptors. Wiley-VCH, Weinheim, Germany, **2000**.

32. M. P. González, H. D. González, R. R. Molina, M. A.Cabrera and R.A. Ramos TOPS-MODE based QSAR derived from heterogeneous series of compounds. Applications to the design of new herbicides. *J. Chem. Inf. Comput. Sci.*, **2003**, 43, 1192-1199.

33. A. Sabljic, H. Gusten, H. Verhaar and J. Hermens. QSAR modeling of soil sorption. Improvements and systematics of log K_{oc} vs. log K_{ow} correlations. *Chemosphere.* **1995**, 31, 4489-4514.

Table 1. Observed, predicted, and residual values of soil sorption coefficients for the 44 compounds used to derive the QSPR.

Number	Compounds	Observed	Predicted	Deleted Residuals
1	Phenylurea	1.500	1.335	0.165
2	2-Chlorophenylurea	1.610	1.691	-0.081
3	2-Fluorophenylurea	1.320	1.348	-0.028
4	3-Chlorophenylurea	2.010	1.916	0.094
5	3-Fluorophenylurea	1.770	1.617	0.153
6	3-Bromophenylurea	2.120	2.056	0.064
7	3-Methylphenylurea	1.560	1.573	-0.013
8	3-Trifluoromethylphenylurea	1.980	2.015	-0.035
9	4-Fluorophenylurea	1.520	1.655	-0.135
10	4-Bromophenylurea	2.060	2.082	-0.022
11	4-Phenoxyphenylurea	2.560	2.810	-0.250
12	3,4-Dichlorophenylurea	2.530	2.505	0.025
13	3-Chloro-4-methoxyphenylurea	2.000	1.898	0.102
14	3-Methyl-4-fluorophenylurea	1.750	1.748	0.002
15	3-Methyl-4-bromophenylurea	2.370	2.296	0.074
16	N-Phenyl-N'-cyclopropylurea	1.740	1.862	-0.122
17	N-Phenyl-N'-cyclopentylurea	1.930	2.055	-0.125
18	N-Phenyl-N'-cyclohexylurea	2.070	2.196	-0.126
19	N-Phenyl-N'-cycloheptylurea	2.370	2.287	0.083
20	Siduron	2.310	2.230	0.080
21	N-Phenyl-N-methylurea	1.290	1.341	-0.051
22	N-(3-Chlorophenyl)-N'-methylurea	1.930	1.957	-0.027
23	N-(3,4-Dichlorophenyl)-N'-methylurea	2.460	2.546	-0.086
24	N-(3-Chloro-4-methylphenyl)-N'-methylurea	2.100	2.160	-0.060

25	N-(3-Chloro-4-methoxyphenyl)-N'-methylurea	1.840	1.939	-0.099
26	Fenuron	1.400	1.298	0.102
27	N-(3-Chlorophenyl)-N',N'-dimethylurea	1.790	1.880	-0.090
28	N-(3-Methoxyphenyl)-N',N'-dimethylurea	1.720	1.479	0.241
29	N-(3-Fluorophenyl)-N',N'-dimethylurea	1.730	1.581	0.149
30	Fluometuron	2.000	1.979	0.021
31	N-(4-Fluorophenyl)-N',N'-dimethylurea	1.430	1.619	-0.189
32	Monuron	1.950	1.909	0.041
33	N-(4-Methylphenyl)-N',N'-dimethylurea	1.510	1.564	-0.054
34	N-(4-Methoxyphenyl)-N',N'-dimethylurea	1.400	1.521	-0.121
35	Metoxuron	1.720	1.862	-0.142
36	Chlorotoluron	2.020	2.083	-0.063
37	N-(3,5-Dimethylphenyl)-N',N'-dimethylurea	1.730	1.766	-0.036
38	N-(3,5-Dimethyl-4-bromo-phenyl)-N',N'-dimethylurea	2.530	2.489	0.041
39	Diuron	2.400	2.469	-0.069
40	Chloroxuron	3.550	3.385	0.165
41	Monolinuron	2.100	1.935	0.165
42	Metobromuron	2.100	2.071	0.029
43	Linuron	2.700	2.494	0.206
44	Chlorbromuron	2.700	2.678	0.022

Table 2. Statistical parameters of the lineal regressions models obtained for the twelve kinds of descriptors.

Descriptors	Variables[a]	S	R^2	F	q^2
Spectral moments	μ_9GM, μ_4H, μ_{13}H, μ_1P, μ_2P	0.118	0.938	115.18	0.917
Constitutional	MW, AMW, Mv, nBO, nO	0.156	0.893	63.14	0.842
Molecular walk counts	MWC03, MWC06, MWC07, MWC08, MWC09	0.301	0.603	11.565	0.521
BCUT	BEHm2, BEHm3, BEHm4, BELm4, BELe8	0.158	0.890	61.72	0.832
Charges indices	GGI3, GGI6, GGI8, GGI10, JGI8	0.294	0.621	12.47	0.524
2D autocorrelations	ATS8m, ATS1p, GATS6m, GATS7e, GATS1p	0.227	0.773	25.89	0.703
Randic molecular profiles	DP01, DP06, DP18, SP05, SP20	0.324	0.539	8.88	0.462
Geometrical	J3D, G1, G(N..Cl), G(N..Br), G(O..O)	0.167	0.877	54.04	0.821
RDF	RDF040u, RDF125u, RDF045m, RDF080m, RDF040p	0.171	0.871	51.46	0.824
3D-MORSE	Mor01m, Mor19m, Mor20m, Mor24m, Mor16p	0.155	0.894	64.32	0.831
WHIM	E2u, P1p, E1s, Tm, Kp	0.189	0.843	41.00	0.806
GETAWAY	ITH, HATS8m, H3p, HATS2p, RTu	0.139	0.915	81.64	0.884

[a] The definition of the terms appears largely explained in reference 31.

Table 3. Contributions of different groups to the soil sorption coefficients of the phenylureas herbicides in study.

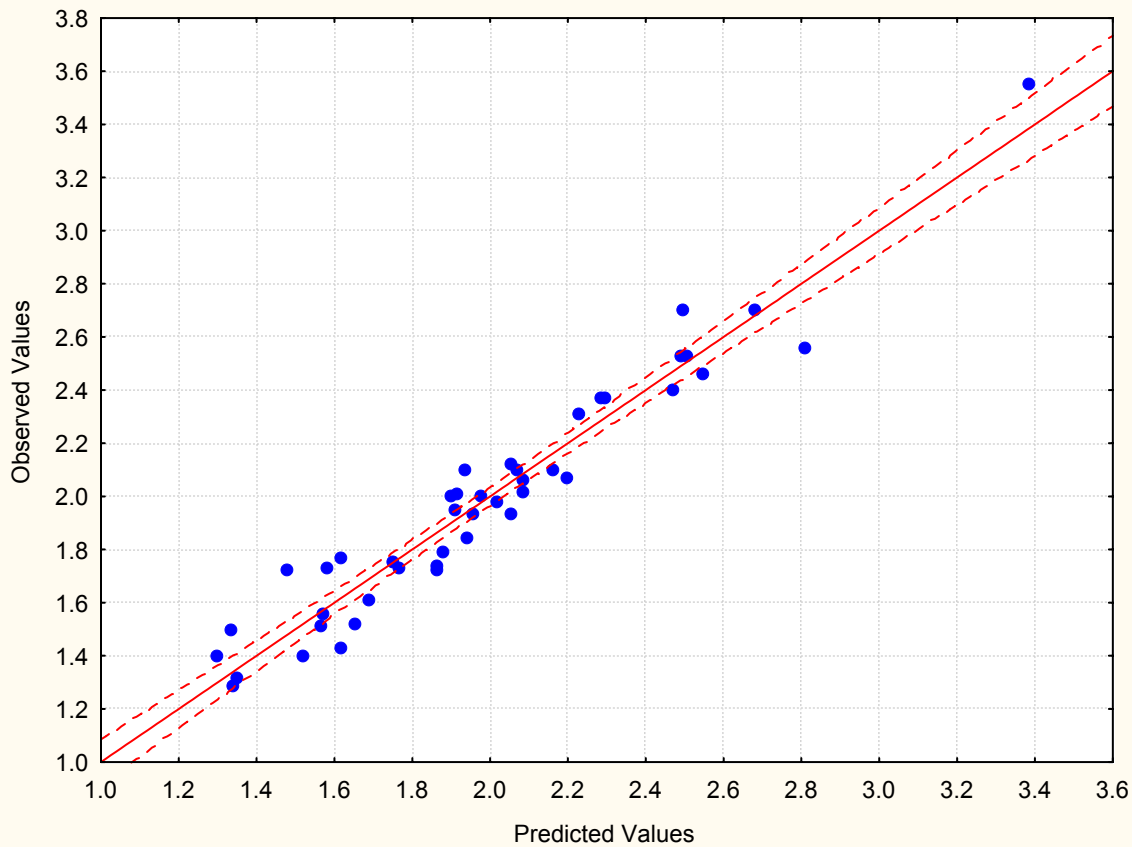| Studied Fragments	Group contribution	Studied Fragments	Group contribution
F1	-0.03	F20	-0.05
F2	0.13	F21	0.13
F3	0.20	F22	-0.04
F4	0.16	F23	0.13
F5	0.47	F24	0.25
F6	0.61	F25	0.32
F7	0.005	F26	0.03
F8	0.19	F27	0.06
F9	0.35	F28	0.21
F10	0.41	F29	0.28
F11	-0.12	F30	0.50
F12	-0.11	F31	1.13
F13	-0.247	F32	-0.19
F14	0.242	F33	0.22
F15	0.13	F34	-0.104
F16	-0.13	F35	0.25
F17	0.11	F36	0.34
F18	-0.04	F37	0.47
F19	0.83	F38	0.55

Figure 1. The linear relation between observed and predicted soil sorption for the compounds of the training set.
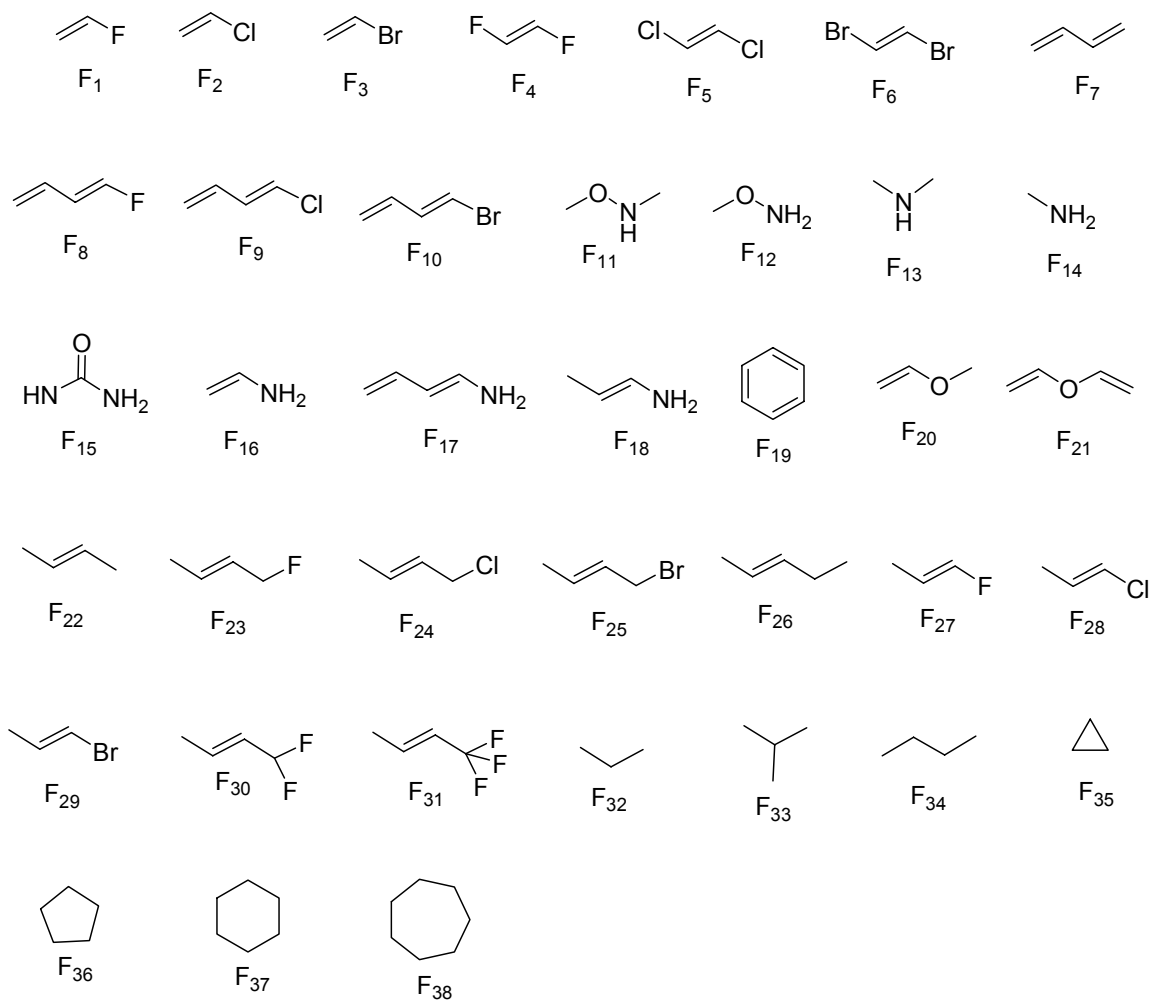
Figure 2. Structures of selected fragments for which their contributions to the soil sorption coefficient.