

## Water Activated Carbon Organics Adsorption Structure - Property Relationships

Lorentz JÄNTSCHI

*Technical University of Cluj-Napoca, Romania, <http://lori.academicdirect.org>*

### **Abstract**

Investigation (determination) of chemical compounds properties need time and many resources when is performed by classical way, or experimentations. Nowadays a number of quantitative structure-property relationships (QSPRs) were developed in order to shorting the research and analysis time of chemical properties on classes of compounds. The ability of the molecular descriptor family (MDF) was used to produce QSPRs for estimating the adsorption onto activated carbon in water. A number of sixteen organics and theirs adsorption onto activated carbon in water serves for QSPRs obtaining. The MDF methodology include the three-dimensional model of the molecules building using the HyperChem software, MDF members generating using a set of Pre Hypertext Processor (PHP) programs, storing using a MySQL database server, and finally with a set of Delphi Multiple Linear Regression programs structure-property relationships findings. A number of 105319 MDF members enter into multiple linear regressions findings. Five from our best QSPRs are presented, one mono-varied, two bi-varied and two tri-varied models. The MDF QSPR methodology has big potential in finding QSPR models and is proved for adsorption onto activated carbon in water of studied organics.

### **Keywords:**

Organics Adsorption, QSPR, MDF

## Introduction

The adsorption experimental data enable to evaluate the adsorptions for the chemicals of interest but if the numbers of chemical are large is most useful to develop quantitative structure-property relationship (QSPR) models in order to estimate the adsorptions of new organics before synthesis. Note that studies of adsorption onto activated carbon in water were previously reported [1-4].

Previous reported models use molecular descriptors indices, as proposed by Randic [5], Kier and Hall [6].

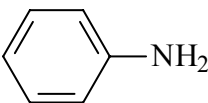
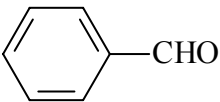
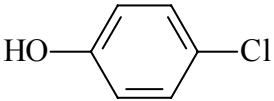
Starting from the idea that molecular topology of chemicals influence their properties, many QSPRs were developed. Pure topological indices are represented by the Wiener [7], Szeged [8, 9], and Cluj [10, 11]. The MDF differ from them by including also the topographical parameters into the calculation.

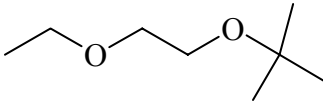
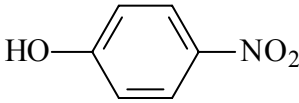
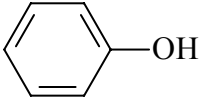
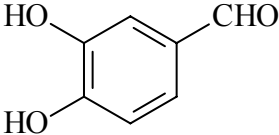
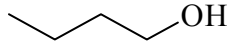
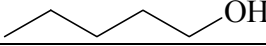
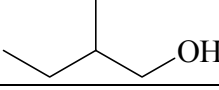
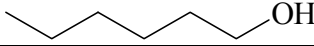
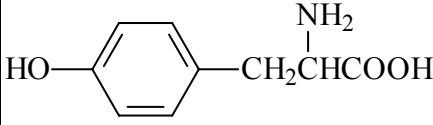
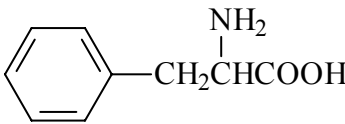
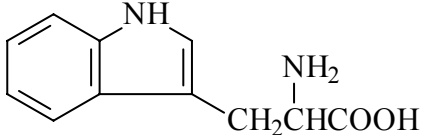
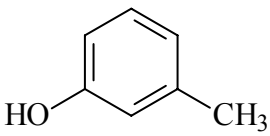
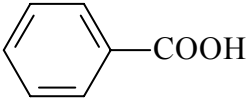
The aim of the present paper is to use the ability of molecular descriptor family on adsorption onto activated carbon in water of sixteen compounds for finding the quantitative structure-property relationships.

## Materials

The adsorption onto activated carbon in water for the sixteen organics was taken from a previous study (table 1, [12]). Table 1 contains the name of the organics, the organics planar structure and the measured adsorption.

*Table 1. The set of organics and their adsorptions into carbon water activated*

No.	Molecule	Planar structure	Adsorption
1	Aniline		1.636
2	Benzaldehyde		1.916
3	4-Chlorophenol		2.089

4	1-Etoxy-2-tertbutoxy ethane		1.663
5	4-Nitrophenol		2.242
6	Phenol		1.702
7	Vanillin		2.119
8	1-Butanol		0.910
9	1-Pentanol		1.408
10	2-Methyl-1-butanol		1.228
11	1-Hexanol		1.770
12	L-tyrosine		1.795
13	L-phenylalanin		1.787
14	L-tryptophan		2.111
15	m-Cresol		1.667
16	Benzoic Acid		1.763

The previous reported study [12] use compounds from table 1 to obtain the predicted adsorption by use of multiple linear regression (MLR) analysis and neural networks (NN) and three molecular connectivity indices [13]. The results were:

- MLR QSPR model:

$$r^2 = 0.665; s = 0.206; F = 27.3 \quad (1)$$

where  $r^2$  is the squared of correlation coefficient,  $s$  the standard error and  $F$  Fischer parameter

- NN QSPR model:

$$r^2 = 0.658; s = 0.208; F = 26.9 \quad (2)$$

Note that the proposed models does not significantly different one from other (the squared correlation coefficient between predicted values are 0.967).

## Method

First step in QSPR modeling using the molecular descriptor family was sketching up the three-dimensional structure of each compound using the HyperChem software [14], software that allows assigning standard bond lengths, bond angles, torsion angles, and stereochemistry.

Using MySQL database server and PHP programming language the molecular descriptor family database was created and gives a total number of 335657 MDF members. After filtration a total number of 105319 distinct members were included into a multiple linear regression QSPRs findings procedure. Each descriptor has a distinct name which collects the ways of his calculation. A MDF member name has seven letters with the following significances: the distance descriptor used (seventh letter), atomic property (sixth), interaction descriptor (fifth), interaction model (fourth), the fragment type (third), molecular superposing of fragmental descriptors method (second), and linearization procedure (first).

More details about MDF are in [15].

The best MDF QSPR models are recorded into the database by the client program which found the model. The MLR procedures was runs for mono-varied (simple linear regression), bi-varied and tri-varied associations of MDF members.

A client-server application provides at the end of findings a report with the statistical analysis of found QSPRs.

## Results

The calculated values of MDF members from selected best QSPR models for adsorption onto water activated carbon of sixteen organics are in table 2.

Table 2. The calculated values of MDF members for Organics set

No	Mono	Bi-varied			Three-varied			
		iSDrDQt ( $\cdot 10^3$ )	IiMMWHt ( $\cdot 10^1$ )	lPMDVQg ( $\cdot 10^{-1}$ )	IsPrVHg ( $\cdot 10^2$ )	IiMMWHt ( $\cdot 10^1$ )	lPMDVQg ( $\cdot 10^{-1}$ )	iFMdFQg ( $\cdot 10^5$ )
1	7.65	-9.29	-4.21	3.50	-9.29	-4.21	270	3.2
2	3.44	-3.62	-11.3	1.61	-3.62	-11.3	1000	11
3	-1.23	-3.51	-5.41	1.26	-3.51	-5.41	36	0.47
4	4.59	-4.50	-18.2	1.56	-4.50	-18.2	2.9	59
5	-7.15	-1.49	-8.04	0.62	-1.49	-8.04	0.023	18
6	7.89	-7.51	-9.09	3.01	-7.51	-9.09	10	330
7	2.01	-2.01	-8.45	0.77	-2.01	-8.45	6.6	19
8	16.8	-17.6	-4.38	6.80	-17.6	-4.38	19	-20
9	9.40	-11.0	-6.61	4.22	-11.0	-6.61	5500	-18
10	11.1	-13.3	-7.31	4.85	-13.3	-7.31	190	6.5
11	5.83	-7.56	-10.5	2.84	-7.56	-10.5	$1.9 \cdot 10^6$	2700
12	1.07	-1.25	-21.6	0.40	-1.25	-21.6	360	13
13	1.41	-1.56	-22.5	0.63	-1.56	-22.5	5700	19
14	0.80	-1.17	-13.1	0.42	-1.17	-13.1	$3.1 \cdot 10^5$	0.4
15	4.60	-6.81	-11.9	2.31	-6.81	-11.9	170	320
16	3.07	-2.60	-18.5	1.22	-2.60	-18.5	26	160

Table 3 contains the selected QSPRs obtained:

Table 3. Water activated carbon MDF QSPRs organics adsorption models

No	Var	QSPR
1	1	$1.99 - 57.99 \cdot \text{iSDrDQt}$
2	2	$2.58 + 2.97 \cdot 10^{-3} \cdot \text{lPMDVQg} - 22.59 \cdot \text{IsPrVHg}$
3	2	$2.58 + 8.53 \cdot 10^{-1} \cdot \text{IiMMWHt} + 2.95 \cdot 10^{-3} \cdot \text{lPMDVQg}$
4	3	$2.57 + 8.62 \cdot 10^{-1} \cdot \text{IiMMWHt} + 2.98 \cdot 10^{-3} \cdot \text{lPMDVQg} + 6.01 \cdot 10^{-5} \cdot \text{ibPMtQg}$
5	3	$2.57 + 8.57 \cdot 10^{-1} \cdot \text{IiMMWHt} + 2.95 \cdot 10^{-3} \cdot \text{lPMDVQg} + 8.02 \cdot 10^{-13} \cdot \text{iFMdFQg}$

The associated statistics with the QSPRs are in table 4 ( $r$  correlation coefficient,  $r^2$  square of correlation coefficient,  $s$  standard error,  $F$  Fisher parameter,  $p$  p-value, and  $r_{cv}^2$  leave one out cross-validation squared correlation coefficient).

Table 4. Statistics for MDF QSPRs

No	r	r <sup>2</sup>	s	F	p%	r <sup>2</sup> <sub>cv</sub>
Mono-varied						
1	0.927	0.859	0.134	86	2.4·10 <sup>-5</sup>	0.803
	p-value (t Stat) Intercept: 1.16·10 <sup>-16</sup> (45.85); iSDrDQt: 2.43·10 <sup>-7</sup> (-9.25)					
Bi-varied						
2	0.989	0.978	0.055	284	1.89·10 <sup>-9</sup>	0.97
	r <sup>2</sup> (IPMDVQg, IsPrVHg) = 0.365; r <sup>2</sup> (IPMDVQg, Ads) = 0.048; r <sup>2</sup> (IsPrVHg, Ads) = 0.195					
	p-value (t Stat) Intercept: 3.55·10 <sup>-16</sup> (49.35); IPMDVQg: 2.14·10 <sup>-7</sup> (9.85); IsPrVHg: 5.69·10 <sup>-12</sup> (-23.23)					
3	0.99	0.981	0.051	337	6.3·10 <sup>-10</sup>	0.975
	r <sup>2</sup> (IiMMWHt, IPMDVQg) = 0.362; r <sup>2</sup> (IiMMWHt, Ads) = 0.829; r <sup>2</sup> (IPMDVQg, Ads) = 0.048					
	p-value (t Stat) Intercept: 1.15·10 <sup>-16</sup> (53.85); IiMMWHt: 1.89·10 <sup>-12</sup> (25.33); IPMDVQg: 8.46·10 <sup>-8</sup> (10.67)					
Tri-varied						
4	0.997	0.995	0.027	799	4.47·10 <sup>-12</sup>	0.981
	r <sup>2</sup> (IiMMWHt, IPMDVQg) = 0.362; r <sup>2</sup> (IiMMWHt, ibPMtQg) = 0.0067; r <sup>2</sup> (ibPMtQg, IPMDVQg) = 0.0002; r <sup>2</sup> (IiMMWHt, Ads) = 0.829; r <sup>2</sup> (IPMDVQg, Ads) = 0.048; r <sup>2</sup> (ibPMtQg, Ads) = 0.175					
	p-value (t Stat) Intercept: 6.21·10 <sup>-19</sup> (100.62); IiMMWHt: 4.66·10 <sup>-15</sup> (47.75); IPMDVQg: 1.24·10 <sup>-10</sup> (20.2); ibPMtQg: 8.56·10 <sup>-5</sup> (5.79)					
5	0.997	0.994	0.029	718	8.45·10 <sup>-12</sup>	0.991
	r <sup>2</sup> (IiMMWHt, IPMDVQg) = 0.362; r <sup>2</sup> (IiMMWHt, iFMdFQg) = 0.002 r <sup>2</sup> (iFMdFQg, IPMDVQg) = 0.0006; r <sup>2</sup> (IiMMWHt, Ads) = 0.829 r <sup>2</sup> (IPMDVQg, Ads) = 0.048; r <sup>2</sup> (iFMdFQg, Ads) = 0.085					
	p-value (t Stat) Intercept: 1.22·10 <sup>-18</sup> (95.12); IiMMWHt: 9.08·10 <sup>-15</sup> (45.15); IPMDVQg: 2.59·10 <sup>-10</sup> (18.97); iFMdFQg: 1.65·10 <sup>-4</sup> (5.38)					

The plots of QSPRs are in figures 2-6.

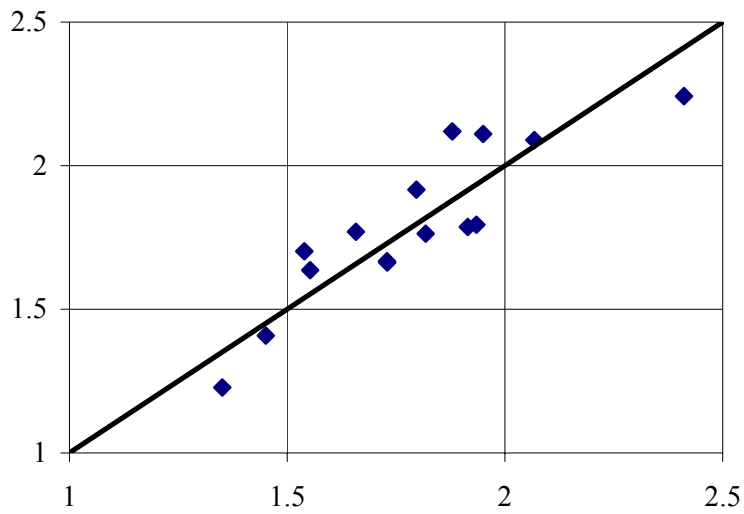


Figure 2. Adsorption =  $1.99-57.99 \cdot iSDrDQt$

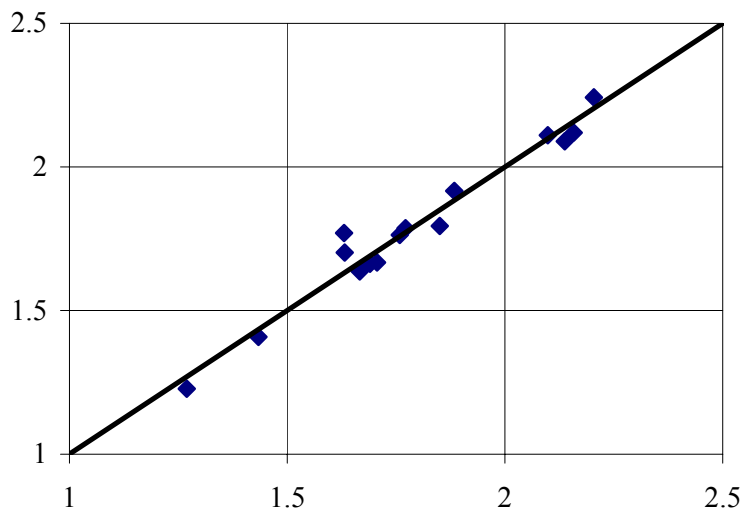


Figure 3. Adsorption =  $2.58+2.97 \cdot 10^{-3} \cdot lPMDVQg-22.59 \cdot IsPrVHg$

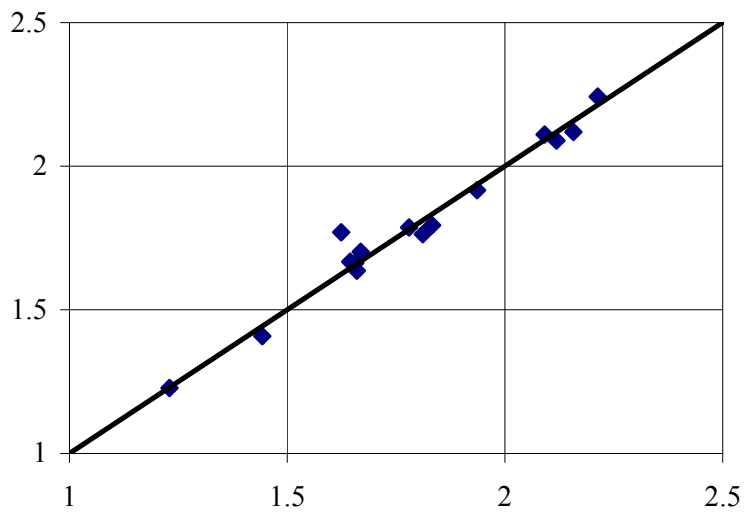


Figure 4. Adsorption =  $2.58+8.53 \cdot 10^{-1} \cdot iIMMWHt+2.95 \cdot 10^{-3} \cdot lPMDVQg$

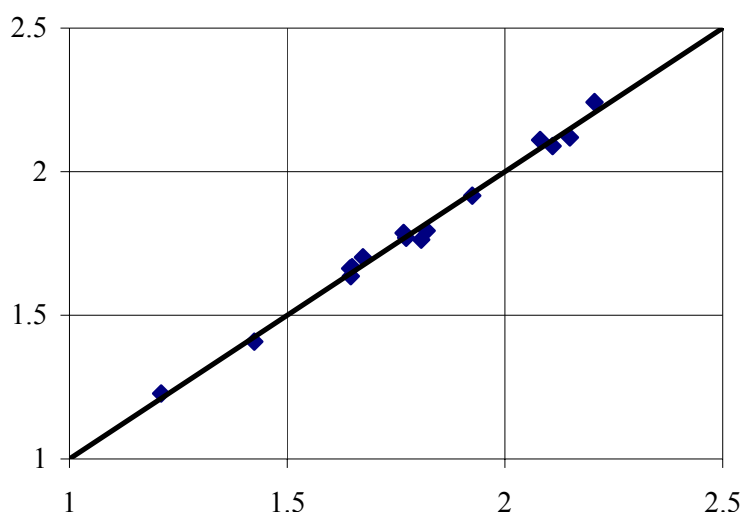


Figure 5.  $Adsorption = 2.57 + 8.62 \cdot 10^{-1} \cdot iMMWHt + 2.98 \cdot 10^{-3} \cdot lPMDVQg + 6.01 \cdot 10^{-5} \cdot ibPMtQg$

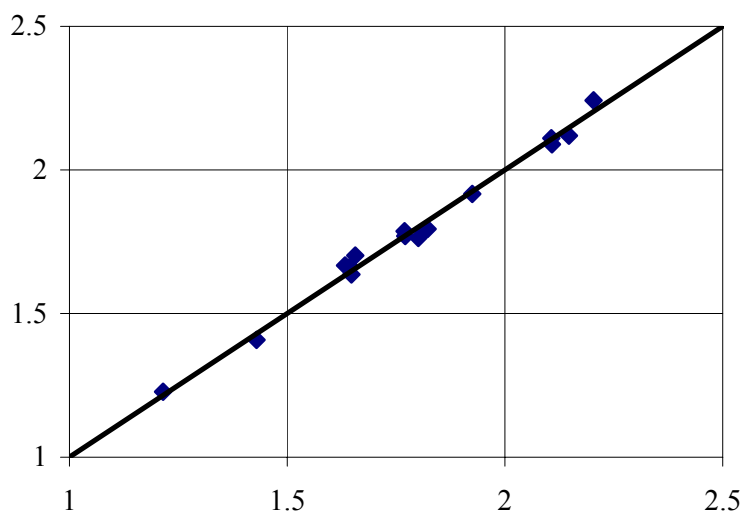


Figure 6.  $Adsorption = 2.57 + 8.57 \cdot 10^{-1} \cdot iMMWHt + 2.95 \cdot 10^{-3} \cdot lPMDVQg + 8.02 \cdot 10^{-13} \cdot iFMdFQg$

## Discussions

All five presented MDF QSPRs of organics adsorption onto activated carbon in water are statistically significant giving us probabilities of wrong model less than  $2.4 \cdot 10^{-5} \%$ .

The mono-varied MDF QSPR is based on a member that uses the topological shape (t), and the partial charge (Q) of the molecules ( $iSDrDQt$ ). Almost eighty-six percents of the variation in adsorption of organics is explainable by its linear relation with  $iSDrDQt$ . This model shows us the importance of topological shape and partial change of compounds in predicting the adsorption in mono-varied models.



First bi-varied MDF QSPR (equation 2, table 3) uses *IPMDVQg* and *IsPrVHg* members, while the second bi-varied MDF QSPR (equation 3, table 3) uses *IiMMWHt* and *IPMDVQg* members. The last letters from the members name used in both bi-varied models denote the used of topological distance on bounds (t) and the geometrical distance (g) computed using the Cartesian coordinates. The penultimate letters from the members names on both models denoted the importance of the number of directly bonded hydrogen's (H) and respectively the partial change, semi-empirical Extended Hückel model, Single Point approach (Q). Thus, we have a model in which the adsorption can be explain by the topological distance as well as the geometrical distance and a model that explain the adsorption based on pure geometrical distance descriptors. Ninety-eight percents of the variation in adsorption of organics is explainable by its linear relation with *IiMMWHt*, and *IPMDVQg* and ninety-seven percents of the variation is explainable by the linear relation with *IPMDVQg*, and *IsPrVHg*. Looking at the square of correlation coefficients between member values used in bi-varied MDF QSPRs of adsorption onto activated carbon in water of the sixteen organics, we can say that is no link between using of orthogonal descriptors (Principal and/or Dominant Component Analysis) in QSPR modeling ( $r^2 \approx 0.36$ ).

The cross validation score, with leave one out, shows that the second model (no 3, table 4) is the best bi-varied model being the best MDF QSPR in term of estimation.

First tri-varied MDF QSPR uses *IiMMWHt*, *IPMDVQg* and *ibPMtQg* members (equation 4, table 3) and second uses *IiMMWHt*, *IPMDVQg* and *iFMdFQg* members (equation 5, table 3). Both tri-varied models use one member which considers the topological distance operator (t) and two members which consider the geometrical distance (g). The penultimate letters of members names implied in the both tri-varied models shows that one member uses the number of directly bonded hydrogen's (H) and two uses the partial change from semi-empirical Extended Hückel model, Single Point approach (Q). About ninety-nine percents of the variation in adsorption of organics is explainable by its linear relation with MDF members. The square of correlation coefficients between the used members in both tri-varied MDF QSPRs (0.362, 0.006, 0.0002; 0.362, 0.002, 0.0006) prove that is no link between using of orthogonal descriptors (Principal and/or Dominant Component Analysis) in tri-varied QSPR modeling of organics adsorption onto activated carbon in water. The best tri-varied MDF QSPR is given by the equation 5, table 3, which has the best cross validation score (over 0.99); thus, this model is the best able to estimate the adsorption.

Looking at the previously reported QSPRs of organics adsorption onto activated carbon in water (expressions (1) and (2)) we can say that our results are better even if we look at the mono-, bi- or tri-varied QSPRs. The molecular descriptors family quantitative structure-property relationships are a useful tool in estimation of organics adsorption onto activated carbon in water.

## Conclusions

The molecular descriptor family methodology produces QSPRs capable to predict the adsorption of the sixteen compounds being a better method comparing with previous reported MLR and/or NN analysis. More, it enabled a discussion about the nature of the measured property (adsorption). The QSPR finding based on molecular descriptor family methodology has good resources for QSPR modeling even if it is a time consuming method.

The QSPR that has the best ability to predict and estimate the organics adsorption onto activated carbon in water is:

$$\text{Adsorption} = 2.57 + 8.57 \cdot 10^{-1} \cdot i\text{IMMWHt} + 2.95 \cdot 10^{-3} \cdot i\text{PMDVQg} + 8.02 \cdot 10^{-13} \cdot i\text{FMdFQg}$$

Two models involving *iIMMWHt*, *iPMDVQg*, and *iFMdFQg* molecular descriptors, and respectively *iIMMWHt*, *iPMDVQg*, and *ibPMtQg* found to be most relevant, contains two members being implied in both models, which denotes the stability of the models. Adsorption onto water activated carbon of the sixteen organics is strongly dependent on the partial change atomic property and number of directly bonded hydrogen's and its causality are from both molecular topology and molecular topography nature.

## References

1. Abe I., Hayashi K. and Kitagawa M. *The adsorption of aminoacids from water on activated carbon*, Bull. Chem. Soc. Jpn. 1982, 55, 687-9.
2. Abe I., Hayashi K., Kitagawa M., Hirashima T., *Adsorption of organics compounds from aqueous solution*, Bull. Chem. Soc. Jpn., 1983, 56, 1002-5.

3. Le Cloirec P., Baudu M., Martin G., Dagois G., *Membranes, toiles, fibres ou feutres: des charbons actifs d'utilisation prometteuse*, Rev. Sci. Tech. Def., 1990, 2, p. 1-3.
4. Le Cloirec P., Brasquet C., Subrenat E., *Adsorption onto fibrous activated carbon: applications to water treatment*, Energ. Fuels, 1997, 11(2), 331-6.
5. Randic M., *On characterization of molecular branching*, J. Am. Chem. Soc. 1995, 97(23), p. 6609-15.
6. Kier L. B., Hall L. H., *Molecular Connectivity in Structure-Analysis Activity*, Research Studies Press, John Wiley and Sons, Letchworth, 1986, U.K.
7. Wiener H.J., *Structural determination of paraffin boiling points*, J. Am. Chem. Soc., 1947, 69, p. 17-20.
8. Gutman I., *Graph Theory Notes*, New York 1994, 27, p. 9.
9. Khadikar P.V., Deshpande N.V., Kale P.P., Dobrynin A., Gutman I., Domotor G.J., *The Szeged Index and an Analogy with the Wiener Index*. Chem. Inf. Comput. Sci., 1995, 35, p. 547.
10. Diudea M. V., *Cluj Matrix Invariants*, J. Chem. Inf. Comput. Sci., 1997, 37, 300-305.
11. Diudea M., Gutman I., Jäntschi L., *Chapter 9 of Molecular Topology*, Nova Science, Huntington, New York, 2001.
12. Brasquet C., Le Cloirec P. *QSAR for Organics Adsorption onto Activated Carbon in Water: What about the use of Neural Networks?*, Wat. Res., 1999, 33(17), p. 3603-8.
13. Kier L.B., Hall L. H., *Molecular Connectivity in Structure-Analysis Activity*, Research Studies Press, John Wiley and Sons, Letchworth, U.K., 1986.
14. \*\*\*, *HyperChem Software*, Hypercube, Inc., 1115 NW 4th Street, Gainesville, FL 32601 USA, <http://www.hyper.com>
15. Jäntschi L., *MDF – A New QSPR/QSAR Molecular Descriptors Family*, Leonardo Journal of Sciences, 2004, 4, p. 67-85.