



Short communication

Quantitative structure property relationship for Henry's law constant of some alkane isomers

Sagarika Sahoo^a, Sabita Patel^{b,*}, B.K. Mishra^{a,*}^a Centre of Studies in Surface Science and Technology, School of Chemistry, Sambalpur University, Jyoti Vihar, Burla 768019, India^b Department of Chemistry, National Institute of Technology, Rourkela 769008, India

ARTICLE INFO

Article history:

Received 24 April 2010

Received in revised form 9 October 2010

Accepted 13 October 2010

Available online 21 October 2010

Keywords:

Henry's law constant

QSPR

Topological indices

Multiple regression analysis

ABSTRACT

A novel topological parameter has been derived by a bond disconnection protocol on a vertex weighted molecular graph. The parameter was correlated with other topological indices and has been applied successfully in quantitative structure property relationship (QSPR) for Henry's law constant of alkane isomers by using single and multiparametric equations separately.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The Henry's law constant (K_H), a thermodynamic parameter, plays a fundamental role in the description of many industrial, toxicological, and environmental processes that transport chemicals between gaseous and aqueous phases. It is the limit of vapor fugacity over liquid composition for composition tends to zero (Eq. (1)) [1].

$$K_H = \lim_{x_i \rightarrow 0} \frac{f_i}{x_i} \quad (1)$$

At low pressure and at high temperature, gases approach ideal gas behavior and fugacity approaches pressure. Accordingly for dilute solution of low-volatile organic solutes, the Henry's law constant is often expressed as the ratio of the equilibrium partial pressure (P_g) of the solute and its aqueous concentration (C_w) (Eq. (2)).

$$K_H = \frac{P_g}{C_w} \quad (2)$$

Therefore, K_H is expressed in the unit of atm L mol^{-1} . The dimensionless Henry's law constant (air–water partition coefficient, K_{aw}) is related to the K_H through the ideal gas law, as follows:

$$K_{aw} = \frac{K_H}{RT} = \frac{C_g}{C_w} \quad (3)$$

where R is the ideal gas constant, T is the absolute temperature and C_g is the concentration of the solute in the gas phase.

Some quantitative structure property relationship (QSPR) models for a diverse set of experimental data of Henry's law constant of organic chemicals were developed by Modarresi et al. [2] based on four different molecular descriptor sets. Three different models based on the descriptors of CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis), Tsar and Dragon software, and a model based on a combined descriptor set from these packages, and in addition from HYBOT software, were established using the stepwise regression method. Russel et al. [3] developed models consisting of five structural descriptors, which encode information related to compound's bulk, lipophilicity, and polarity to correlate the molecular structure with the log of Henry's law constant for a diverse set of organic compounds. Duchowicz et al. [4] established a QSPR model between the Henry's Law constant in the air–water system and the molecular structure of more than hundred aliphatic hydrocarbons. The simultaneous linear regression analyses on 1086 numerical descriptors reflecting topological, geometrical, and electronic aspects led to a seven parameter equation with good calibration and cross-validated parameters.

A QSPR model for estimating the Henry's law constant in water was recently reported by English and Carroll [5]. Some QSPR methods were proposed, which couple a feature selection approach with nonlinear mapping functions. These methods were applied to the prediction of the Henry's law constant and the solubility of organic compounds in water by Yaffe et al. [6] and by Mitchell and Jurs [7], respectively. Suzuki et al. [8] used principal component analysis and found that three components are sufficient to reproduce

* Corresponding authors. Tel.: +91 663 2430114; fax: +91 663 2430158.

E-mail addresses: sabita@nitrrkl.ac.in (S. Patel), bijaym@hotmail.com (B.K. Mishra).

the Henry's law constant of 74 organic compounds having diverse structure.

With a view to optimize the QSPR for Henry's law constant of some alkanes, an attempt has been made to propose a regression model using a novel molecular descriptor derived from molecular graphs along with other molecular descriptors.

2. Data base

The Henry's law constants of alkanes were collected from literature [9] and are presented in Table 1.

3. Molecular descriptors

The following topological descriptors have been derived for C-3 to C-9 hydrocarbons by using reported methods.

- (i) Connectivity parameters ($^1\chi$, $^2\chi$, $^3\chi$) [10,11].
 (ii) Matrix derived descriptors (1W , 2W , DI, 1H , 2H , HI, R, WI, MTI) [12] and (VD, VW, VDI, VWI, VMTI) [13].

The molecular descriptors or topological indices (TIs) have been evaluated either from the molecular graph or hydrogen depleted graph. Molecular connectivity index (χ), which has found wide application in correlation analysis in chemistry, was proposed by Randic [10], Kier and Hall [11]. First order connectivity index, $^1\chi$ is expressed as follows:

$$^1\chi = \sum C_k \quad (4)$$

where C_k is the connectivity value for a bond (connecting atom i and j) is computed as follows:

$$C_k = 1 \sqrt{\delta_i \delta_j} \quad (5)$$

where a valence (degree) value, δ which is equal to the number of nonhydrogen atoms bonded to the adjacent carbon, i and j .

Second order connectivity index, $^2\chi$, is the sum of all connecting bonds $i-j-k$.

$$^2\chi = \sum C_k \quad (6)$$

where

$$C_k = 1 \sqrt{\delta_i \delta_j \delta_k} \quad (7)$$

Third order connectivity index, $^3\chi$, is the sum of all connecting bonds $i-j-k-l$

$$^3\chi = \sum C_k \quad (8)$$

where

$$C_k = 1 \sqrt{\delta_i \delta_j \delta_k \delta_l} \quad (9)$$

In order to evaluate the molecular topological indices, the following terms are calculated by using algebraic operations on distance [D], reciprocal of distance [H], adjacency [A], walk [W] and valence [V] matrices. [12]

$$A_2 = \sum \sum [A^2] \quad (10)$$

$$S_D = \sum \sum [AD]_{ij} \quad (11)$$

$$S_H = \sum \sum [AH]_{ij} \quad (12)$$

$$S_W = \sum \sum [AW]_{ij} \quad (13)$$

$$WW^{(1)} = \frac{[W]}{2} \quad (14)$$

$$DW^{(1)} = \frac{[D]}{2} \quad (15)$$

Table 1

The Henry's law constant and VWW¹ parameter data of 73 C-3 to C-9 alkane isomers.

Sl. No.	Alkane	Notation	Henry's constant	VWW ¹
1	n-Propane	3	0.685	2.828
2	n-Butane	4	0.917	5.175
3	2-Methylpropane	2M3	1.152	4.327
4	Pentane	5	0.126	7.853
5	2-Methylbutane	2M4	1.37	6.735
6	2,2-Dimethylpropane	22MM3	2.105	5.657
7	n-Hexane	6	1.291	10.759
8	2-Methylpentane	2M5	1.74	9.434
9	3-Methylpentane	3M5	1.12	9.358
10	2,2-Dimethylbutane	22MM4	1.52	8.063
11	2,3-Dimethylbutane	23MM4	1.3	8.287
12	n-Heptane	7	2.7	13.827
13	2-Methylhexane	2M6	3.42	12.33
14	3-Methylhexane	3M6	0.308	12.229
15	3-Ethylpentane	3E5	2.54	12.069
16	2,2-Dimethylpentane	22MM5	3.15	10.738
17	2,3-Dimethylpentane	23MM5	1.73	10.927
18	2,4-Dimethylpentane	24MM5	2.94	10.990
19	3,3-Dimethylpentane	33MM5	1.84	11.251
20	2,2,3-Trimethylbutane	223MMM4	2.38	9.605
21	n-Octane	8	4.93	17.019
22	2-Methylheptane	2M7	3.65	15.37
23	3-Methylheptane	3M7	3.71	15.261
24	4-Methylheptane	4M7	3.66	15.238
25	3-Ethylhexane	3E6	3.73	15.03
26	2,2-Dimethylhexane	22MM6	3.41	13.592
27	2,3-Dimethylhexane	23MM6	3.73	13.798
28	2,4-Dimethylhexane	24MM6	3.52	13.784
29	2,5-Dimethylhexane	25MM6	3.34	13.865
30	3,3-Dimethylhexane	33MM6	3.75	13.459
31	3,4-Dimethylhexane	34MM6	4.05	13.702
32	3-Ethyl-2-methylpentane	3E2M5	3.81	13.624
33	3-Ethyl-3-methylpentane	3E3M5	4.21	13.297
34	2,2,3-Trimethylpentane	223MMM5	3.81	12.226
35	2,2,4-Trimethylpentane	224MMM5	3.34	12.272
36	2,3,3-Trimethylpentane	233MMM5	4.1	12.199
37	2,3,4-Trimethylpentane	234MMM5	1.76	12.483
38	2,2,3,3-Tetramethylbutane	2233MMMM4	3.78	10.916
39	n-Nonane	9	5.93	20.305
40	2-Methyloctane	2M8	4.79	18.521
41	3-Methyloctane	3M8	5.04	18.41
42	4-Methyloctane	4M8	9.94	18.381
43	3-Ethylheptane	3E7	5.17	18.144
44	4-Ethylheptane	4E7	5.17	18.076
45	2,2-Dimethylheptane	22MM7	4.79	16.579
46	2,3-Dimethylheptane	23MM7	5.27	16.816
47	2,4-Dimethylheptane	24MM7	4.63	16.778
48	2,5-Dimethylheptane	25MM7	4.82	16.781
49	2,6-Dimethylheptane	26MM7	4.66	16.871
50	3,3-Dimethylheptane	33MM7	5.27	16.437
51	3,4-Dimethylheptane	34MM7	5.34	16.689
52	3,5-Dimethylheptane	35MM7	4.87	16.691
53	4,4-Dimethylheptane	44MM7	5.22	16.407
54	3-Ethyl-2-methylhexane	23ME6	5.19	16.565
55	4-Ethyl-2-methylhexane	24ME6	4.81	16.568
56	3-Ethyl-3-methylhexane	33ME6	5.7	16.208
57	3-Ethyl-2-methylhexane	34EM6	5.34	16.478
58	2,2,3-Trimethylhexane	223MMM6	5.24	15.065
59	2,2,4-Trimethylhexane	224MMM6	4.79	15.036
60	2,2,5-Trimethylhexane	225MMM6	5.17	15.098
61	2,3,3-Trimethylhexane	233MMM6	5.68	15.030
62	2,3,4-Trimethylhexane	234MMM6	5.6	15.260
63	2,3,5-Trimethylhexane	235MMM6	4.84	15.331
64	2,4,4-Trimethylhexane	244MMM6	5.31	15.000
65	3,3,4-Trimethylhexane	334MMM6	5.94	15.113
66	3,3-Diethylpentane	33EE5	6.55	16.008
67	2,2-Dimethyl-3-ethylpentane	223MME5	5.33	14.890
68	2,3-Dimethyl-3-ethylpentane	233MME5	6.45	14.855
69	2,4-Dimethyl-3-ethylpentane	234MEM5	5.48	15.166
70	2,2,3,3-Tetramethylpentane	2233(M)5	6.25	13.498
71	2,2,3,4-Tetramethylpentane	2234(M)5	5.69	13.767
72	2,2,4,4-Tetramethylpentane	2244(M)5	5.09	13.540
73	2,3,3,4-Tetramethylpentane	2334(M)5	6.27	13.755

Table 2
Possible disconnection for VWW¹ calculation.

Possible disconnections	^a $\zeta_{ij} = (\prod \delta_i)^{1/n}$	$C_p = \prod \zeta_{ij}$
a	$\zeta_{ij}^1 = (1 \times 2 \times 3 \times 4 \times 1 \times 1 \times 1)^{1/7} = 1.575$ $\zeta_{ij}^2 = (1)^{1/1} = 1$	$1.575 \times 1 = 1.575$
b	$\zeta_{ij}^1 = (1 \times 2 \times 3 \times 4 \times 1 \times 1 \times 1)^{1/7} = 1.575$ $\zeta_{ij}^2 = (1)^{1/1} = 1$	$1.575 \times 1 = 1.575$
c	$\zeta_{ij}^1 = (1 \times 3 \times 1)^{1/3} = 1.442$ $\zeta_{ij}^2 = (2 \times 4 \times 1 \times 1 \times 1)^{1/5} = 1.516$	$1.442 \times 1.516 = 2.186$
d	$\zeta_{ij}^1 = (1 \times 1 \times 3 \times 2)^{1/4} = 1.565$ $\zeta_{ij}^2 = (4 \times 1 \times 1 \times 1)^{1/4} = 1.414$	$1.565 \times 1.414 = 2.213$
e	$\zeta_{ij}^1 = (1 \times 1 \times 3 \times 2 \times 4 \times 1 \times 1)^{1/7} = 1.575$ $\zeta_{ij}^2 = (1)^{1/1} = 1$	$1 \times 1.575 = 1.575$
f	$\zeta_{ij}^1 = (1 \times 1 \times 3 \times 2 \times 4 \times 1 \times 1)^{1/7} = 1.575$ $\zeta_{ij}^2 = (1)^{1/1} = 1$	$1 \times 1.575 = 1.575$
g	$\zeta_{ij}^1 = (1 \times 1 \times 3 \times 2 \times 4 \times 1 \times 1)^{1/7} = 1.575$ $\zeta_{ij}^2 = (1)^{1/1} = 1$	$1 \times 1.575 = 1.575$

^a n is the no. of atoms present in the fragment.

$$HW^{(1)} = \frac{[H]}{2} \quad (16)$$

The above defined equations are used to obtain the following molecular topological indices.

$${}^1W = \frac{1}{2} \sum \sum d_{ij} \quad (17)$$

where d_{ij} refers to the elements of distance matrix.

$${}^1H = \frac{1}{2} \sum \sum h_{ij} \quad (18)$$

¹H is the reciprocal of all path distances in the molecular graph, where h_{ij} are the elements of the matrix.

$$R = \frac{1}{2} \sum \sum w_{ij} \quad (19)$$

R is the Hyper-wiener index, where w_{ij} are the elements of the walk matrix

$$DI = \sum [DW^{(1)}(A + D)] = S_D + 2x^2W \quad (20)$$

$${}^2W = \frac{\{DI - (S_D/2)\}}{2} \quad (21)$$

$$HI = \sum [HW^{(1)}(A + H)] = S_H + 2x^2H \quad (22)$$

$$WI = \sum [WW^{(1)}(A + W)] = S_W + 2x^2R \quad (23)$$

$$MTI = \sum \sum [A^2 + AD]_{ij} = A_2 + S_D \quad (24)$$

A novel set of parameters VD, VW, VDI, VWI, VMTI and VWW are obtained from various matrix-algebraic operations of [V] matrix on [D] and [W] matrices.

In a hydrogen depleted molecular graph, if the degrees of vertices are represented in a column matrix [V_c] and row matrix [V_r], then

$$[V] = [V_c] \times [V_r] \quad (25)$$

$$VD = \sum \sum (V \times D)_{ij} \quad (26)$$

$$VW = \sum \sum (V \times W)_{ij} \quad (27)$$

$$VDI = \sum \sum [W \times (V \times W)]_{ij} \quad (28)$$

$$VDI = \sum \sum [DW^1(V \times D)]_{ij} \quad (29)$$

$$VMTI = \sum \sum [V^2 + VD]_{ij} \quad (30)$$

In addition to all of the above topological indices, a new set of descriptors, vertex weighted walk parameters (VWW), which are derived from the disconnection of bonds in the hydrogen depleted vertex weighted molecular graph are taken into consideration in this paper [14].

In a vertex weighted graph a valence, δ^v value for each atom can be assigned as:

$$\delta_i^v = \sigma_i + p_i + n_i \quad (31)$$

where δ_i^v is the number of nonhydrogen valence electrons contributed by atom i [15].

Various order of VWW* can be calculated by considering generation of fragments (ζ) after disconnection of a single bond (i,j); two consecutive single bonds (i,k); three consecutive single bonds (i,l); four consecutive single bond (i,m), etc.

$$\zeta_{i,j/k/l/m} = \left(\prod \delta_i \right)^{1/n} \quad (32)$$

'n' being the number of atoms in that fragment 3

$$C_p = \prod \zeta_{i,j/k/l/m} \quad (33)$$

$$VWW^* = \sum C_p \quad (34)$$

p = number possible bond disconnection and accordingly * = 1, 2, 3 or 4 for one, two, three and four consecutive bond disconnections respectively or order of VWW. An example of calculation of VWW¹ has been presented *vide infra* (Table 2).

The possible disconnections on the valence-weighted molecular graph of 2,2,4-trimethylpentane are presented in Fig. 1.

From Table 2, VWW¹ (= $\sum C_p$) is calculated to be 12.272 for 2,2,4-trimethylpentane. The VWW¹ values have been calculated for C-3 to C-9 hydrocarbons and are listed in Table 1.

4. Monovariate regression model

The Henry's law constant (K_H) data are subjected to regression analysis with each graph theoretical descriptor. The statistical parameters like R^2 and F values are presented in Table 3.

The regression coefficient values are found to be maximum for VMTI ($r = 0.89$, $F = 280.92$) and minimum for ${}^2\chi$ ($r = 0.63$, $F = 45.790$).

Table 3
Correlation coefficient.

TIs	R^2	F	TIs	R^2	F
¹ W	0.74	204.17	VD	0.75	213.35
² W	0.69	156.77	VW	0.73	193.93
DI	0.69	158.90	VDI	0.72	185.48
¹ H	0.75	210.58	VWI	0.60	105.23
² H	0.75	210.13	VMTI	0.80	280.92
HI	0.75	208.28	1χ	0.70	163.13
R	0.64	125.93	2χ	0.39	45.790
WI	0.52	78.19	3χ	0.48	65.380
MTI	0.74	201.28	VWW	0.64	124.74

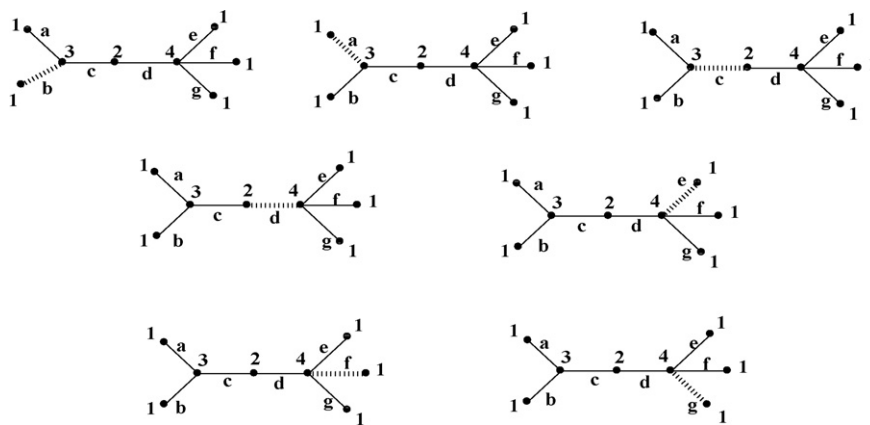


Fig. 1. First order VWW graph of 2,2,4-trimethylpentane (224MMM5).

Table 4
Optimized regression model for prediction of Henry's law constant of alkanes.^a

t_{\min}	Excluded var	R^2	F	RMS
-0.06	² H	0.9317	55.529	0.2374
-0.11	R	0.9317	60.837	0.2334
0.70	VW	0.9311	66.435	0.2314
-1.09	VWI	0.9297	72.127	0.2321
0.83	WI	0.9289	79.684	0.2309
-2.23	¹ W	0.9231	82.663	0.2458
2.47	² W	0.9155	85.310	0.2657
-2.30	VMTI	0.9084	90.639	0.2836
-0.79	DI	0.9075	106.24	0.2820
0.90	MTI	0.9063	127.68	0.2812
-1.66	VD	0.9024	154.83	0.2886
0.91	VDI	0.9012	206.69	0.2879
-5.74	¹ H	0.8532	200.61	0.4213

^a Basic regression model including eighteen descriptors ($N=72$, $R=0.9652$, $R^2=0.9317$, $F=50.9215$, $RMS=0.2416$).

5. Optimization of multivariable regression model

The monovariate regression analysis for the Henry's law constant (K_H) values of alkanes shows correlation coefficient value less than 0.90. In order to obtain a better regression model, multiparametric equations were derived and optimized by excluding variable by considering t_{\min} , F , R^2 and RMS. Increase in R^2 and F and decrease in RMS values suggest improvement of regression model (35) and hence leads to optimization. Table 4 shows the successive exclusion

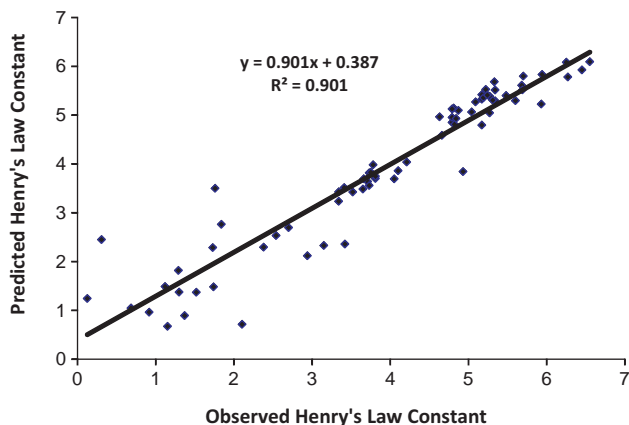


Fig. 2. Plot of predicted vs. observed Henry's law constant values of alkanes.

of variables to obtain the optimized model.

$$K_H = -(1.9922 \pm 0.3469)^1 H + (0.1420 \pm 0.0210) HI \\ + (0.4652 \pm 0.0722) VWW^1 \\ + (2.5106 \pm 0.6762) \quad (N = 72, \quad R^2 = 0.9012, \\ F = 206.69, \quad RMS = 0.2879) \quad (35)$$

The regression model (35) is found to be significant with respect to 't' and 'F' values. By using the regression model, the K_H values have been predicted and plotted against the observed values (Fig. 2). The linearity in the plot establishes the applicability of the novel topological parameter 'VWW¹' in quantitative structure property relationship.

6. Conclusion

The parameter VWW¹ is found to be a suitable parameter for quantitative structure property relationship in alkanes. The higher order VWW parameters may find applications in QSPR studies for other physico-chemical properties of alkanes.

Acknowledgements

The authors thank University Grants Commission and Department of Science and Technology, New Delhi for financial assistance through Departmental Research Support and FIST programs respectively.

References

- [1] J.M. Prausnitz, R.N. Lichtenthaler, E.G. Azevedo, Molecular Thermodynamics of Fluid-Phase Equilibria, 3rd ed., Prentice Hall, London, 1999.
- [2] H. Modarresi, H. Modarress, J.C. Dearden, QSPR model of Henry's law constant for a diverse set of organic chemicals based on genetic algorithm-radial basis function network approach, Chemosphere 66 (2007) 2067–2076.
- [3] C.J. Russell, S.L. Dixon, P.C. Jurs, Computer-assisted study of the relationship between molecular structure and Henry's law constant, Anal. Chem. 64 (1992) 1350–1355.
- [4] P.R. Duchowicz, J.C.M. Garro, E.A. Castro, QSPR study of the Henry's law constant for hydrocarbons, Chemometr. Intell. Lab. Syst. 91 (2008) 133–140.
- [5] N.J. English, D.G. Carroll, Prediction of Henry's law constants by a quantitative structure property relationship and neural networks, J. Chem. Inform. Comput. Sci. 41 (2001) 1150–1161.
- [6] D. Yaffe, Y. Cohen, G. Espinosa, A. Arenas, F. Giralt, A fuzzy ARTMAP-based quantitative structure–property relationship (QSPR) for the Henry's Law constant of organic compounds, J. Chem. Inform. Comput. Sci. 43 (2003) 85–112.
- [7] B.E. Mitchell, P.C. Jurs, Prediction of aqueous solubility of organic compounds from molecular structure, J. Chem. Inform. Comput. Sci. 38 (1998) 489–496.
- [8] T. Suzuki, K. Ohtaguchi, K. Koide, Application of principal components analysis to calculate Henry's constant from molecular structure, Comput. Chem. 16 (1992) 41–52.

- [9] J. Daugherty, *Assessment of Chemical Exposures*, Lewis Publishers, Florida, 2000.
- [10] M. Randic, *J. Am. Chem. Soc.* 97 (1975) 6609–6615.
- [11] L.B. Kier, L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Wiley, New York, 1986.
- [12] M.V. Diudea, M. Topan, A. Graovac, *Molecular topology. 17. Layer matrices of Walk degrees*, *J. Chem. Inform. Comput. Sci.* 34 (1994) 1072–1078.
- [13] M. Kuanar, B.K. Mishra, *Novel vertex-degree-weighted graph invariants*, *J. Serb. Chem. Soc.* 62 (1997) 289–294.
- [14] S. Sahoo, S.K. Dash, S. Patel, B.K. Mishra, *A molecular structure descriptor derived from bond-disconnection: application to quantitative structure property relationships*, *Indian J. Chem.* 48A (2009) 921–929.
- [15] M. Randic, A. Sabljic, S. Nikolic, N. Trinajstic, *A rational selection of graph-theoretical indices in the QSAR*, *Int. J. Quantum Chem.* 34 (1988) 267–285.