

Papers

Correlation of line graph parameters with physicochemical properties of octane isomers

Minati Kuanar, Saroj K Kuanar & Bijay K Mishra*

Centre of Studies in Surface Science and Technology, Department of Chemistry, Sambalpur University,
Jyoti Vihar 768 019, India

and

Ivan Gutman

Faculty of Science, University of Kragujevac, P.O. Box 60, YU-34000 Kragujevac, Yugoslavia

Received 21 October 1998; revised 8 January 1999

Line graphs and their derivatives are constructed for hydrogen-depleted structural graphs of octane isomers, and respective topological indices, θ_i , are derived. The parameters obtained from first to fourth derivatives of the line graphs are used in obtaining quantitative structure-property relationships using several physicochemical properties of 18 octane isomers. The regression analysis indicates that θ_4 indices have the best discriminating ability for these compounds.

Octane isomers have become an important set of organic molecules to test the applicability of various topological parameters in quantitative structure-property/activity relationships (QSPR/QSAR). These compounds are structurally diverse enough to yield considerable variation in shape, branching and non-polarity¹. Since the beginning of the enumeration of molecular parameters, attempts have been made to get non-degenerate set of parameters. Though the distance connectivity index (J) proposed by Balaban² shows considerable discrimination power to differentiate between octane isomers, Bonchev³ has reported that no single topological index can discriminate among the structural graphs of octane isomers uniquely. Ray Choudhury *et al.*⁴ have defined three novel information theoretic topological indices, namely degree complexity, graph vertex complexity and graph distance complexity and have found satisfactory discriminating ability of the parameters for various isomeric graphs. However, they have not used these topological indices in QSPR/QSAR studies.

In a comprehensive study of numerous properties of octane isomers, Randic *et al.*^{1,5} have used single molecular descriptors and concluded that different physicochemical properties depend on different descriptors. Later, Randic⁶ used two variable, ^1Z and ^2Z , derived from the leading members in the path

sequence of Z matrices, in a regression equation to predict the boiling point of octane isomers with a correlation coefficient of 0.912. While correlating various physico-chemical properties of octane isomers with topological indices derived from Wiener matrix, Randic *et al.*⁷ have obtained good correlation.

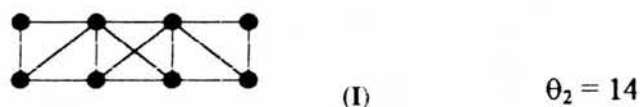
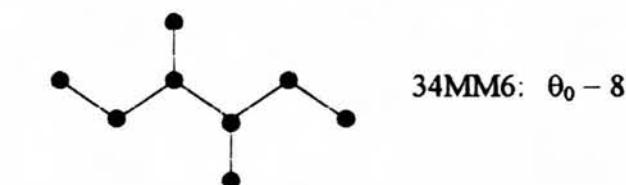
Bertz⁸ has, for the first time, proposed the possibility of the use of line (bond) graph parameters for modeling physico-chemical properties of organic molecules. Subsequently, use of first derivative of line graph in QSPR studies has been reported by Estrada and Gutman^{9,10}. Recently, Gutman *et al.*¹¹ have made an attempt to use topological parameters derived from line graph to predict surface tension of alkanes. In the present paper descriptors are obtained by using first to fourth derivatives of line graph and are used in correlating various physico-chemical properties of octane isomers.

Data base and analytical method

Eight physicochemical properties of octane isomers (Table 1) have been selected on the availability^{7,12} of a suitable body of data: boiling point (BP), critical temperature (CT), critical pressure (CP), entropy (S), density (D), mean radius (R_m)² and heat of vaporization (H_v), heat of formation (H_f). The values are compiled in Table 1.

Table 1—Line graph parameters and physicochemical properties of octane isomers

Alkane	θ_1	θ_2	θ_3	θ_4	BP	TC	PC	S	D	R^2_m	$-\Delta H_f$	$-\Delta H_v$
n-octane	6	5	4	3	125.7	296.2	24.64	111.67	0.7025	2.0449	208.6	41.49
2M	7	8	12	28	117.6	288.0	24.80	109.84	0.6980	1.8913	215.4	39.67
3M	7	9	17	54	118.9	292.0	25.60	111.26	0.7058	1.7984	212.5	39.83
4M	7	9	18	60	117.7	290.0	25.60	109.32	0.7046	1.7673	212.0	39.69
3E	7	10	23	90	118.5	292.0	25.74	109.43	0.7136	1.7673	210.7	39.64
22MM	9	17	55	324	106.8	279.0	25.60	103.42	0.6953	1.6744	224.6	37.28
23MM	8	13	35	167	115.6	293.0	26.60	108.02	0.7121	1.6464	213.8	38.78
24MM	8	12	26	92	109.4	282.0	25.80	106.98	0.7004	1.6142	219.2	37.76
25MM	8	11	20	54	109.1	279.0	25.00	105.72	0.6935	1.6449	222.5	37.85
33MM	9	19	70	464	112.0	290.8	27.20	104.74	0.7100	1.7377	220.0	37.53
34MM	8	14	40	203	117.7	298.0	27.40	106.59	0.7200	1.5230	212.8	38.97
2M3E	8	14	41	213	115.6	295.0	27.40	106.06	0.7193	1.5525	211.0	38.52
3M3E	9	21	84	606	118.3	305.0	28.90	101.48	0.7274	1.5212	214.8	37.99
223MMM	10	23	92	681	109.8	294.0	28.20	101.31	0.7161	1.4306	220.0	36.91
224MMM	10	20	65	381	99.24	271.1	25.50	104.09	0.6919	1.4010	224.0	35.14
233MMM	10	24	101	783	114.8	303.0	29.00	102.06	0.7262	1.4931	216.3	37.27
234MMM	9	17	53	297	113.5	295.0	27.60	102.39	0.7191	1.3698	217.3	37.75
2233MMMM	12	33	162	1512	106.5	270.8	24.50	93.06	0.8242	1.4612	225.6	42.90



Construction of line graph and its derivatives

Let 'G' be the hydrogen depleted molecular graph with edges $e_1, e_2, e_3, \dots, e_m$. The corresponding line graph (L_1) can be constructed with $e_1, e_2, e_3, \dots, e_m$ as the vertices with a new set of edges $e'_1, e'_2, e'_3, \dots$. Each derived edge (e') connects two vertices, when the corresponding edges, e , are incident. Similarly, other derivatives of line graph ($L_2, L_3, \dots$) can be obtained with $e'', e''' \dots$ edges. An illustration for construction of the line graphs of 34MM6 is given in (I).

An index, θ_i , for each graph is defined as the number of edges of the line graph of order i , where 'i' is the order of the derivative. Hydrogen depleted molecular graph may be treated as zeroth order line

Table 2—Cross correlation matrix of line graph parameters

	θ_1	θ_2	θ_3	θ_4
θ_1	1.0000			
θ_2	0.9795	1.0000		
θ_3	0.9478	0.9886	1.0000	
θ_4	0.9077	0.9563	0.9880	1.0000

graph. Generally, line graphs converge in case of cyclic, linear and $K_{1,3}$ (isopropane or isopropyl group) graphs. Other than these, graphs diverge on iteration and lead to complicated graphs of higher order. The θ_i of a complicated derived line graph L_i can be obtained if the degree (d_{i-1}) of the points of earlier graph L_{i-1} and θ_{i-1} are known¹³.

$$\theta_i = -\theta_{i-1} + \frac{1}{2} \sum d_{i-1}^2 \quad \dots (1)$$

The values of θ_i are given in Table 1.

Multiple regression analysis was carried out on a Pentium-155 computer. The cross correlation coefficients of the line graph parameters and the physicochemical parameters are given in Tables 2 and 3 respectively.

Results and Discussion

Cross-correlation of molecular parameters and physicochemical properties

For linear octane, the θ_i values converge with increasing number of iteration of line graph. However, with branched octanes the derived graphs become complicated with increasing iteration of the graph.

Table 3—Cross correlation matrix of physicochemical properties

	BP	TC	PC	S	D	R _m ²	−ΔH _f	−ΔH _v
BP	1.0000							
TC	0.7598	1.0000						
PC	0.1016	0.7167	1.0000					
S	0.5939	0.2722	−0.2846	1.0000				
D	−0.1241	−0.1544	0.0078	−0.7397	1.0000			
R _m ²	0.6247	0.1027	−0.5420	0.7473	−0.3615	1.0000		
−ΔH _f	−0.9116	−0.7319	−0.1469	−0.6953	0.2443	−0.5022	1.0000	
−ΔH _v	0.5172	−0.0092	−0.4914	0.0653	0.5700	0.4839	−0.3479	1.0000

Branching in the molecule also increases the complexity of the graph resulting in sharp increase in the θ_i values. Further, the extent of increase in the θ_i for geminal dimethyl substitution is more than that for the vicinal dimethyl substitution. Thus, the parameters seem to express the compactness of the molecule. With increasing number of iteration, the degeneracy in the parameter decreases. In octane isomers, θ_4 is a complete non-degenerate set of parameters. The cross correlation coefficients of these parameters are found to be high (>0.9) and decrease with increasing iteration.

Except BP and ΔH_f ($r=0.9116$), the physicochemical parameters are not well correlated with each other. The correlation coefficient values range from 0.0078 to 0.7598. Thus, these properties seem to be a good set of data for testing the application of a new set of molecular parameters like θ_i .

Multiple regression analysis

A generalized linear regression model has been proposed for the relationship of physico-chemical properties of octane isomers with the line graph parameters,

$$P = a + \sum_{i=1}^4 b_i \theta_i \quad \dots (2)$$

where P refers to physico-chemical properties, 'a' is constant, 'b_i' is the sensitivity of θ_i towards P.

The correlation coefficients of the physico-chemical properties with individual θ_i values show some significant results. The entropy values, which generally do not show good relationship with any single descriptor, are found to have good correlation coefficient (0.9260–0.9588). The critical temperature, pressure, and enthalpy of vaporization values are found to have poor correlation coefficient (0.0000 to 0.3807).

When two predictors are used almost all properties show significant predictability. Randic⁶ has used two novel indices derived from Hosoya matrix to predict the boiling points of octane isomers and achieved a correlation with $R=0.914$ having standard error of 2.658°C. In the present study, the use of descriptors θ_1 and θ_2 resulted in a R value of 0.9430 and S value of 2.168 (F=60.21). The use of θ_1 and θ_3 also yielded a good correlation coefficient of 0.9427 and standard deviation 2.174 (F=59.87). However, the addition of third descriptor does not produce any significant improvement in the regression model ($R=0.9491$, $S=2.123$, $F=42.45$).

The regression analysis of TC and ΔH_v , with the line graph descriptors reveals some interesting results. The correlation coefficients obtained by using single descriptors are found to be in the range of 0.1646 to 0.3399 for TC and 0.0000 to 0.2236 for ΔH_v . Multiple regression analysis with two descriptors leads to increase in the R value with a maximum of 0.7555 for TC (by using θ_1 and θ_2) and 0.9366 for ΔH_v (by using θ_2 and θ_3). The correlation coefficient values of single parameters for ΔH_v are 0.0000 and 0.1345 for θ_1 and θ_2 respectively. For TC, however, the R value (0.8662) obtained from θ_1 , θ_3 and θ_4 descriptors is found to be the most significant values with the maximum F (13.49) and minimum S (5.449). The R value (0.7555) obtained by using θ_1 and θ_2 is found to be higher than the values obtained from various combinations of the two-descriptors model proposed by Randic *et al.*⁷ Similarly, for ΔH_v , the regression model with θ_2 and θ_3 is the most significant ($R=0.9366$) with F value 53.57. The other properties, e.g., critical pressure (PC), density (D), heat of formation (ΔH_f) and mean radius are also found to give good correlation coefficient values ($R=0.9024$, 0.9563 and 0.9672) respectively.

The results of the QSPR studies reveal that the

simple line graph parameters can be used as candidate to represent the molecular structure for predicting physicochemical properties. Work on derivation of more descriptors from various operations on the line graph is in progress.

Acknowledgement

One of the authors (MK) thanks CSIR, New Delhi for the award of a Senior Research Fellowship.

References

- 1 Randic M & Seybold P G, *SAR QSAR Environ Res*, 1 (1983) 77.
- 2 Balaban A T, *Chem Phys Lett*, 89 (1982) 399.
- 3 Bonchev D in *Information theoretic indices for characterization of chemical structures* (Wiley, Cluchester) 1983.
- 4 Ray Choudhury C, Ray S K, Ghosh J J, Roy A B & Basak S C, *J comput Chem*, 5 (1984) 581.
- 5 Randic M, *Croat Chem Acta*, 66 (1993) 289.
- 6 Randic M, *Croat Chem Acta*, 67 (1994) 415.
- 7 Randic M, Guo X, Oxley T, Krishnapriyan T & Naylor L, *J chem Inf Comput Sci*, 34 (1994) 361.
- 8 Bertz S H, *J chem Soc, Chem Commun*, (1981) 818.
- 9 Estrada E, *J chem Inf Comput Sci*, 36 (1996) 844.
- 10 Estrada E & Gutman I, *J chem Inf Comput Sci*, 36 (1996) 850.
- 11 Gutman I, Popovic L, Mishra B K, Kuanar M, Estrada E & Guevara N, *J Serb chem Soc*, 62 (1997) 1025.
- 12 *CRC Handbook of chemistry and physics* (CRC Press, Boca Raton) 1994-95.
- 13 Harary F, in *Graph theory* (Narosa Publishing House, New Delhi) 1990, 71.