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, $\theta_i$, 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 $\theta_4$ 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-polarity. 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 ($I$) proposed by Balaban shows considerable discrimination power to differentiate between octane isomers, Bonchev has reported that no single topological index can discriminate among the structural graphs of octane isomers uniquely. Ray Choudhury et al. 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. have used single molecular descriptors and concluded that different physicochemical properties depend on different descriptors. Later, Randic used two variable, $I^2$ and $I^2$, 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. have obtained good correlation.

Bertz 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. Recently, Gutman et al. 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 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.
Construction of line graph and its derivatives

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

An index, \( \theta_i \), 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 graph. Generally, line graphs converge in case of cyclic, linear and \( K_{3,3} \) (isopropane or isopropyl group) graphs. Other than these, graphs diverge on iteration and lead to complicated graphs of higher order. The \( \theta_i \) of a complicated derived line graph \( L_i \) can be obtained if the degree \( d_i \) of the points of earlier graph \( L_{i-1} \) and \( \theta_{i-1} \) are known:

\[
\theta_i = \theta_{i-1} + \sum \frac{1}{2} d_i^2 \quad \ldots (1)
\]

The values of \( \theta_i \) 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 \( \theta_i \) 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.
Branching in the molecule also increases the complexity of the graph resulting in sharp increase in the $\theta_i$ values. Further, the extent of increase in the $\theta_i$ 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, $\theta_4$ 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 $\Delta H_r$ ($r=0.9116$), 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 $\theta_i$.

**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,

$$ P = a + \sum_{i=1}^{4} b_i \theta_i $$

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

The correlation coefficients of the physico-chemical properties with individual $\theta_i$ 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. Randic$^6$ 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 $\theta_1$ and $\theta_2$ resulted in a $R$ value of 0.9430 and $S$ value of 2.168 ($F=60.21$). The use of $\theta_1$ and $\theta_2$ 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 $\Delta H_r$ 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 $\Delta H_r$. Multiple regression analysis with two descriptors leads to increase in the $R$ value with a maximum of 0.7555 for TC (by using $\theta_1$ and $\theta_2$) and 0.9366 for $\Delta H_r$ (by using $\theta_2$ and $\theta_3$). The correlation coefficient values of single parameters for $\Delta H_r$ are 0.0000 and 0.1345 for $\theta_1$ and $\theta_2$ respectively. For TC, however, the $R$ value (0.8662) obtained from $\theta_1$, $\theta_2$ and $\theta_3$ 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 $\theta_1$ and $\theta_2$ 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 $\Delta H_r$, the regression model with $\theta_2$ and $\theta_3$ 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 ($\Delta H_r$) 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
3 Bonchev D in Information theoretic indices for characterization of chemical structures (Wiley, Cluchester) 1983.
13 Harrary F, in Graph theory (Narosa Publishing House, New Delhi) 1990, 71.