Araştırma Makalesi / Research Article

Çoklu Doğrusal Regresyon Kullanılarak Monokarboksilik Asitlerin Bazı Termodinamik Özelliklerin Belirlenmesi

Özge ÇOLAKOĞLU HAVARE*

Mersin Üniversitesi, Matematik Bölümü, Mersin
(ORCID: 0000-0003-4094-3380)

Öz
Topolojik indeksler, moleküler G grafı aracılığıyla elde edilmiş moleküler yapının nümeriksel tanımlayıcısıdır. Topolojik indeksler, yapı-özellik ilişkisi (QSPR), yapı-aktivite ilişkisi (QSAR) ve kimya, nanoteknoloji ve farmakolojideki yapısal tasarım gibi moleküllerin özelliklerini incelemek için kullanılır. Başlıca görevi QSAR QSPR modellerinde sayısal bir moleküler tanımlayıcı olarak çalışmaktır. Üstelik kaynama noktası, buharlaşma ve kararlılık entalpisi gibi fizikokimyasal özellikler, QSAR/QSPR modelleri ile öngörülebilir. Bu çalışmada, tek karbonlu asitlerin termodinamik özelliklerini tahmin etmek için Forgotten topolojik indeks, Forgotten coindex, max-min Rodeg indeksi ve ters toplam indeg indeksi kullanılarak QSPR (Nicel Yapı - Özellik İlişkisi) modeli tasarlanmıştır. Termodinamik özellikler ile topolojik indeksler arasındaki ilişki analizi çoklu doğrusal regresyon yöntemi kullanılarak yapıldı.

Anahtar kelimeler: Topolojik indeks, Forgotten topolojik indeks, Maks-min Rodeg topolojik indeks, Ters Toplam indeg indeks, Tek karbonlu asitler, QSPR.

Determination of Some Thermodynamic Properties of Monocarboxylic Acids using Multiple Linear Regression

Abstract
Topological indices are the numerical descriptor of a molecular structure obtained via molecular graph G. Topological indices are used for studying the properties of molecules such as structure-property relationship (QSPR), structure-activity relationship (QSAR) and structural design in chemistry, nanotechnology, and pharmacology. Its main role is to work as a numerical molecular descriptor in QSAR/QSPR models. Moreover, physicochemical properties such as boiling point, enthalpy of vaporization, and stability could be envisaged by QSAR/QSPR models. In this study, the QSPR (Quantitative Structure Property Relationship) model is designed using the Forgotten topological index, the Forgotten coindex, the max-min Rodeg index, and the inverse sum indeg index to predict the thermodynamic properties of monocarboxylic acids. The relationship analysis between the thermodynamic properties and the topological indices was done by using the multiple linear regression method.

Keywords: Topological index, Forgotten topological index, Max-min Rodeg index, Inverse sum indeg index, Monocarboxylic Acids, QSPR.

1. Introduction

Computing the connectivity indices of molecular graphs is a meaningful branch in chemical graph theory. Chemical graph theory focuses on finding topological indices which are numerical quantity related to a graph that does not depend on a labeling of its vertices. It is interested in finding topological indices which are well correlated with the chemical properties of the chemical molecules. Topological indices are also used for studying the properties of molecules such as the structure property relationship (QSPR), the structure-activity relationship (QSAR), and the structural design in chemistry,

*Sorumlu yazar: ozgeeclkgl@gmail.com
Geliş Tarihi: 21.12.2018, Kabul Tarihi: 07.05.2019

nanotechnology, and pharmacology. The first topological index is the Wiener index, which was introduced by Harold Wiener. It was used for the correlation of the measured properties of molecules with their structural features by H. Wiener in 1947 [1]. In the last seventy years, many topological indices have been defined and many applications in nanotechnology and pharmacology have been made [2].

Let G be a simple connected graph with a vertex set V(G) and edge set E(G) in chemical graph theory. An edge of G is e, which connects the vertices u and v. And also it writes $e = uv$, u and v are adjacent. A simple graph is an unweighted, undirected graph without loops or multiple edges. A connected graph is a graph which is a path between all the pairs of vertices. We refer to this for an undefined term and a notation [3].

A molecular graph is a simple graph. Its vertices and edges represent the atoms and the bonds, respectively. Note that hydrogen atoms are omitted. The topological index, which is known as a graph based molecular descriptor or graph invariant, is a graph-theoretic invariant characterizing numerically the topological structure of a molecule [4].

In 1972, within a study of the structure dependence of total *p*-electron energy (E), Gutman and Trinajstic used the Forgotten index for the first time, but this topological index was not named [5]. Furtula and Gutman called this an index for the first time in 2015 and examined its properties. In 1972, they used the first Zagreb index, but 10 years later, they introduced this index and explored its properties. Recently, the forgotten index F(G) or F-index was introduced by Furtula and Gutman as the revised version of the first and second Zagreb index in 2015. The Forgotten index is defined as

$$F(G) = \sum_{u \in V(G)} d(u)^3 = \sum_{uv \in E(G)} (d(u)^2 + d(v)^2) \quad (1)$$

where d(u) is denoted as the degree of vertex u [6]. Furtula and Gutman proved that the F(G) index is quite similar to the first Zagreb index and is also better to use for testing the chemical and pharmacological properties of drug molecular structures. Thus, this topological index attracts attention in terms of academic and industrial areas. F(G) of a special molecular graph has recently been studied in chemical, nanomaterials and pharmaceutical engineering [7].

Like the Zagreb coincides corresponding to F(G), Nilanjan De et al., (2016) introduced a new invariant, the Co-F(G) which is defined as follows

$$Co - F(G) = \bar{F}(G) = \sum_{uv \in E(\bar{G})} (d(u)^2 + d(v)^2) = \sum_{uv \notin E(G)} (d(u)^2 + d(v)^2) \quad (2)$$

Like Zagreb coincides, F-coindex of G [Co-F(G)] is not the F-index of G [F(G)]. Here the sum runs over E(G), but the degree is with respect to G [8]. De et al. showed that the logarithm of the octanol water partition coefficient values with high accuracy using Co-F(G) values of octane isomers.

In 2010, D. Vukicevic and M. Gasperov introduced Adriatic indices, which were obtained by the analysis of well-known indices such as the Randic and Wiener index. In addition, QSAR and QSPR studies of them have been performed. There have been three classes of Adriatic descriptors defined. One of these descriptors is the Discrete Adriatic, which consists of 148 descriptors. They have very good predictive properties. Thus, many scientists have studied these indices. Two of the Discrete Adriatic descriptors are the max-min Rodeg index and the inverse sum indeg index [9].

The Max-min Rodeg index is the best predictor for the enthalpy of the vaporization and standard enthalpy of the vaporization in the set of octane isomers and also for the log water activity coefficient in the set of polychlorobiphenyles. This index is defined as:

$$Mm_{sde}(G) = \sum_{uv \in E(G)} \sqrt{\frac{max\{d_u, d_v\}}{min\{d_u, d_v\}}} \quad (3)$$

where d(u) is denoted as the degree of vertex u (see [9]-[11]).

The inverse sum indeg index, ISI(G), which is the best predictor for the total surface area octane isomers, is defined as:

$$ISI(G) = \sum_{uv \in E(G)} \frac{d_u d_v}{d_u + d_v} \quad ([9], [12]). \quad (4)$$

Various regression methods are implemented for the QSPR and QSAR models (see [13, 14]). Shafiei studied the thermodynamic properties of the monocarboxylic acids in QSPR with linear models using Randic, Balaban, Szeged indices, and Harary numbers [15]. Hosseini and Shafiei represented the QSPR model with multiple linear regression analysis for the prediction of entropy and of the gas heat capacity of benzene derivatives using topological indices [16, 17]. Mohammadinasab determined the critical properties of alkenes' derivatives using multiple linear regressions [18]. Khadikar and Karmarkar obtained the best results using curvilinear correlations [19].

In this study, some topological indices of monocarboxylic acids via the graph theory were calculated. Then the most appropriate topological indices with the model of multiple linear regressions for predicting the thermodynamic properties of monocarboxylic acids were found.

2. Materials and Methods

The benchmark sets and structural formula of the thermodynamic properties of the monocarboxylic acids were taken from [15], [20], and [21]. Table 1 shows the thermodynamic properties of the monocarboxylic acids.

Table 1. The values of Enthalpies of formation of liquid (ΔH^o_f kJ/mol), Enthalpies of combustion of liquid (ΔH^o_C kJ/mol), Enthalpies of vaporization (ΔH^o_{vap} kJ/mol) and Enthalpies of sublimation (ΔH^o_{sub} kJ/mol) of monocarboxylic acids ($C_2H_4O_2$-$C_{20}H_{40}O_2$) at conditions (normally 298.15 K, 1 atm).

Name of compounds	Formula	Enthalpies of formation of liquid	Enthalpies of combustion of liquid	Enthalpies of vaporization	Enthalpies of sublimation
Acetic acid	$C_2H_4O_2$	483,50	875,16	46,3	49,7
Propanoic acid	$C_3H_6O_2$	510,8	1527,3	50	56,1
Butanoic acid	$C_4H_8O_2$	533,9	2183,5	54,9	62,9
Pentanoic acid	$C_5H_{10}O_2$	558,9	2837,8	58,2	69
Hexanoic acid	$C_6H_{12}O_2$	581,8	3494,3	63	75
Heptanoic acid	$C_7H_{14}O_2$	608,5	4146,9	64,8	81,7
Octanoic acid	$C_8H_{16}O_2$	634,8	4799,9	69,4	86,9
Nonanoic acid	$C_9H_{18}O_2$	658	5456,1	72,3	93,6
Decanoic acid	$C_{10}H_{20}O_2$	713,7	6079,3	76,3	100,8
Undecanoic acid	$C_{11}H_{22}O_2$	736,2	6736,5	78,9	106,7
Dodecanoic acid	$C_{12}H_{24}O_2$	775,1	7377	82,2	115,9
Tridecanoic acid	$C_{13}H_{26}O_2$	807,2	8024,2	84,9	121,2
Tetradecanoic acid	$C_{14}H_{28}O_2$	834,1	8676,7	87,7	130,2
Pentadecanoic acid	$C_{15}H_{30}O_2$	862,4	9327,7	91,4	136,5
Hexadecanoic acid	$C_{16}H_{32}O_2$	892,2	9977,2	94,5	144,3
Heptadecanoic acid	$C_{17}H_{34}O_2$	924,4	10624,4	100,7	159,6
Octadecanoic acid	$C_{18}H_{36}O_2$	947,2	11280,1	102,8	164,7
Nonadecanoic acid	$C_{19}H_{38}O_2$	984,1	11923,4	105	172,9
Eicosanoic acid	$C_{20}H_{40}O_2$	1012,6	12574,2	109,9	179,2

Multiple linear regression models can be defined as a linear equation of the form:

$$Y = a + b_1X_1 + \cdots + b_pX_p; \qquad n, R^2, s, F$$

where Y is the response or dependent variable, a is the regression model constant, $b_i \quad (i = 1, \dots, p)$ are the coefficients for the individual descriptor, $X_i \quad (i = 1, \dots, p)$, $X_i \quad (i = 1, \dots, p)$ are independent variables. n is the number of samples used for building the regression equation, R^2 is correlation coefficient, s is the standard error deviation, and F is the calculated value of the F-ration test. For

detailed information see [22]. R. Todeschini in [23] represented to select the best model in a population of models that can be used with any of the parameter max(R^2), max(F).

In [23] and [24] it was reported that the best regression model was the multiple linear regression model. During this study, the multiple linear regression analysis showed better results than the linear regression analysis. Hence, the multiple linear regression analysis is studied in this paper.

The multiple linear regression analyses were implemented using SPSS statistical software. The independent variables in the multiple regression models are the Forgotten index, the Forgotten coindex, the max-min Rodeg index, and the inverse sum indeg index of 19 monocarboxylic acids.

3. Results and Discussion

By (1) -(4) formulas, the topological indices of 19 monocarboxylic acids firstly were computed to design the QSPR. Table 2 shows the value of the topological indices of 19 monocarboxylic acids.

Table 2. The value of topological indices of 19 monocarboxylic acids

Formula	$Mm_{sde}(G)$	$F(G)$	**Co-** $F(G)$	$ISI(G)$
$C_2H_4O_2$	5	30	6	2
$C_3H_6O_2$	6,1	38	26	3,3666
$C_4H_8O_2$	7,1	46	54	4,6666
$C_5H_{10}O_2$	8,1	54	90	5,6666
$C_6H_{12}O_2$	9,1	62	105	6,6666
$C_7H_{14}O_2$	10,1	70	186	7,6666
$C_8H_{16}O_2$	11,1	78	246	8,6666
$C_9H_{18}O_2$	12,1	86	214	9,6666
$C_{10}H_{20}O_2$	13,1	94	390	10,6666
$C_{11}H_{22}O_2$	14,1	102	474	11,6666
$C_{12}H_{24}O_2$	15,1	110	566	12,6666
$C_{13}H_{26}O_2$	16,1	118	666	13,6666
$C_{14}H_{28}O_2$	17,1	126	774	14,6666
$C_{15}H_{30}O_2$	18,1	134	890	15,6666
$C_{16}H_{32}O_2$	19,1	142	1014	16,6666
$C_{17}H_{34}O_2$	20,1	150	1146	17,6666
$C_{18}H_{36}O_2$	21,1	158	1286	18,6666
$C_{19}H_{38}O_2$	22,1	166	1439	19,6666
$C_{20}H_{40}O_2$	23,1	174	1590	20,6666

The following structure-property models were obtained for the Forgotten Topological index, the Co-forgotten index, the Max-min Rodeg index, and the inverse sum indeg index.

ΔH^o_f =343,097+25,577$Mm_{sde}(G)$+0,159$Mm_{sde}(G)^2$ n=19, R^2=0.998, F=3581.557 (5)

ΔH^o_f = 375,701+ 3,293$F(G)$ + 0,002$F(G)^2$ n=19, R^2=0.998, F=3718.134 (6)

ΔH^o_f = 511,652+ 0,537(Co- $F(G)$)+ 0,000(Co- $F(G)$)^2 n=19, R^2=0.992, F=995.346 (7)

ΔH^o_f = 343,097+ 25,577 $ISI(G)$ + 0,159 $ISI(G)^2$ n=19, R^2=0.997, F=2600.798 (8)

From equations (5) -(8) and, according to the correlation coefficients and F, equation (6) describes the best model for ΔH^o_f.

ΔH^{o}_{C}=-2386.650+ 645.388$Mm_{sde}(G)$+0.1$Mm_{sde}(G)^2$ n=19, R^2=1, F=694396.497 (9)

ΔH^{o}_{C}=-1564.468+81.624$F(G)$-0.002$F(G)^2$ n=19, R^2=1, F=1731510.573 (10)

ΔH^{o}_{C}=1817.865+ 11.635Co- $F(G)$-0.003(Co- $F(G)$)^2 n=19, R^2=0.983, F=457.894 (11)

ΔH^{o}_{C}=-494.359+593.910 $ISI(G)$+1.980 $ISI(G)^2$ n=19, R^2=1, F=26478.053 (12)

From equations (9)-(12) and, according to the correlation coefficients and F, the ideal model is equation (10).

ΔH^{o}_{vap}=28.877+3.672 $Mm_{sde}(G)$-0.009$Mm_{sde}(G)^2$, n=19, R^2=0.998, F=3675.803 (13)

ΔH^{o}_{vap}=33.562+0.461 $F(G)$+0.000 $F(G)^2$, n=19, R^2=0.998, F=3531.400 (14)

ΔH^{o}_{vap}=52.752+0.061Co- $F(G)$-1.97E-5(Co- $F(G)$)^2, n=19, R^2=0.973, F=293.849 (15)

ΔH^{o}_{vap}=39.438+3.361 $ISI(G)$+0.001 $ISI(G)^2$, n=19, R^2=0.998, F=4572.565 (16)

From equations (13)-(16), the best ideal model is equation (16).

ΔH^{o}_{sup}=26.503+4.311$Mm_{sde}(G)$+0.103$Mm_{sde}(G)^2$, n=19, R^2=0.999, F=5348.880 (17)

ΔH^{o}_{sup}=32.127+0.581$F(G)$+0.002 $F(G)^2$, n=19, R^2=0.998, F=5210.228 (18)

ΔH^{o}_{sup}=58.396+0.114 Co- $F(G)$+0.114 Co- $F(G)^2$, n=19, R^2=0.988, F=666.779 (19)

ΔH^{o}_{sup}=40.156+4.406 $ISI(G)$+0.118 $ISI(G)^2$, n=19, R^2=0.999, F=5744.892 (20)

From equations (17)-(20), the ideal model is equation (20).

We know that the best QSPR models must have correlation coefficient R^2=1. Consequently, according to the above results, the Forgotten Topological index can be used for predicting the enthalpies of the formation of the liquid and the enthalpies of the combustion of a liquid of the monocarboxylic acids. The inverse sum indeg index can be used to predict the enthalpies of vaporization and the enthalpies of the sublimation of the monocarboxylic acids.

4. Conclusions

The Forgotten index, the max-min Rodeg index, the inverse sum indeg index, and the Forgotten Coindex have been described and studied in recent years. They are frequently used in chemistry and pharmacology. Through these indices, multiple linear regression models were created to predict some thermodynamic properties of monoacids.

The most accurate results for the prediction of the enthalpies of the formation of the liquid (ΔH^{o}_{f} kJ/mol), the enthalpies of the combustion of liquid (ΔH^{o}_{C} kJ/mol), the enthalpies of the vaporization (ΔH^{o}_{vap} kJ/mol), and the enthalpies of the sublimation (ΔH^{o}_{sub} kJ/mol) of monocarboxylic acids can be used with the Forgotten index, the Forgotten index, the inverse sum indeg index, and the inverse sum indeg topological index, respectively.

References

[1] Wiener H. 1947. Structural Determination of Paraffin Boiling Points. Journal of the American Chemical Society, 69 (1): 17-20.

[2] Gao W., Siddiqui M.K., Imran M., Jamil M.K., Farahani M.R. 2016. Forgotten Topological Index of Chemical Structure in Drugs. Saudi Pharmaceutical Journal, 24 (3): 258-264.

[3] Chartrand G., Lesniak L. 2005. Graphs and Digraphs. the CRS Press, 386p. USA.

[4] Gutman I. 1990. A Property of the Simple Topological Index. MATCH Communications in Mathematical and in Computer Chemistry, 25: 131-140.

[5] Gutman I., Trinajstic N. 1972. Graph Theory and Molecular Orbitals. Total ϕ-electron Energy of Alternant Hydrocarbons. Chemical Physics Letters, 17 (4): 535-538.

[6] Furtula B., Gutman I. 2015. A Forgotten Topological Index. Journal of Mathematical Chemistry, 53 (4): 1184-1190.

[7] Khaksari A., Ghorbani M. 2017. On the Forgotten Topological Index. Iranian Journal of Mathematical Chemistry, 8 (3): 327-338.

[8] De N., Nayeem S.M.A. 2016. Computing the F-index of Nanostar Dendrimers. Pacific Science Review A: Natural Science and Engineering, 18 (1): 14-21.
[9] Vukicevic D., Gasperov M. 2010. Bond Additive Modeling 1. Adriatic Indices, Croatica Chemica Acta, 83 (3): 243-260.
[10] Vukicevic D. 2010. Bond Additive Modeling 2. Mathematical Properties of Max-min Rodeg Index. Croatica Chemica Acta, 83 (3): 261-273.
[11] Çolakoglu Havare Ö. 2018. Max-min Rodeg Index of Bridge Graphs and Fullerenes. Malaysian Journal of Fundamental and Applied Scıences, 14 (1): 48-51.
[12] Falahati-Nezhad F., Azari M. 2016. The Inverse Sum Indeg Index of Some Nanotubes. Studia UBB Chemia, 61 (1): 63-70.
[13] Hosamani S., Perigidad D., Jamagoud S., Maled Y., Gavade S. 2017. QSPR Analysis of Certain Degree Based Topological Indices. Journal of Statistics Applications and Probability, 6 (2): 361-371.
[14] Sivaraman N., Srinivasan T.G., Vasudeva Rao P.R. 2001. QSPR Modeling for Solubility of Fullerene (C_{60}) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41 (4): 1067-1074.
[15] Shafiei F. 2015. Relationship between Topological Indices and Thermodynamic properties and of the Monocarboxylic Acids Applications in QSPR. Iranian Journal of Mathematical Chemistry, 6 (1): 15-28.
[16] Hosseini H., Shafiei F. 2016. Quantitative Structure Property Relationship Models for the Prediction of Gas Heat Capacity of Benzene Derivatives Using Topological Indices, MATCH Communications in Mathematical and in Computer Chemistry, 75(3): 583-592.
[17] Hosseini H., Shafiei F. 2017. Entropy Prediction of Benzene Derivatives Using topological Indices. Studia UBB Chemia, LXII (2): 297-310.
[18] Mohammadinasab E. 2017. Determination of Critical Properties of Alkanes Derivatives Using Multiple Linear Regression. Iranian Journal of Mathematical Chemistry, 8 (2): 199-220.
[19] Khadikar P.V., Karmarkar S. 2001. A Novel PI Index and Its Applications to QSPR/QSAR Studies. Journal of Chemical Information and Computer Sciences, 41 (4): 934-949.
[20] Baev A.K. 2012. Specific Intermolecular Interaction of Organic Compounds. Springer, 430p. Heidelberg.
[21] National Institute of Standards and Technology U.S. Department of Commerce, NIST Chemisty webBook. Search for Species Data by Chemical Name. http://webbook.nist.gov/chemistry/name-ser/ (Access date: 23.06.2018).
[22] Dearden J.C. 2017. Advances in QSAR Modeling. Springer international publishing, 555p. Switzerland.
[23] Todeschini R. 2010. Useful and Unuseful Summaries of Regression Models. http://www.moleculardescriptors.eu/tutorials/T5_moleculardescriptors_models.pdf. (Access date: 23.06.2018).
[24] Cornwell E. 2006. New Idea for the Topological Index Evaluation and Treatise Multiple Regression with Three Independent Variables: Saturated Hydrocarbons Used Like a Model. Journal of the Chilean Chemical Society, 51 (1): 765-768.