6th International Conference on
Combinatorics, Cryptography, Computer Science and Computing
November. 17-18, 2021

On the Boiling Point of Potential Drug Candidates against SARS-CoV-2 by Curvilinear Regression Modeling

Özge Çolakoğlu Havare[1]
Mersin University, Mersin, Turkey

Abstract

The new disease caused by the family of the coronavirus, SARS-CoV-2 disease, has spread rapidly all over the world and causes deaths. Therefore, it has been declared a pandemic. No drug has yet been discovered to end this pandemic. Some drugs are used to slow down the pandemic. However, since these drugs are not sufficient, combinations and analogs of these drugs are being studied. With topological indices, physicochemical and bioactivity properties of chemical structures can be predicted. The topological index is a numerical descriptor obtained from the molecular graph of chemical structures. In this study, degree-based topological indices of analogs of lopinavir and favipiravir are computed and these results are used to predict Boiling point property with curvilinear regression.

Keywords: Chemical graph theory, Topological index, SARS-CoV-2, Drug, Boiling point
Mathematics Subject Classification [2010]: 05C07, 92E10

1 Introduction

Coronavirus family is enveloped, positive-sense, single-stranded RNA viruses [1]. SARS-CoV-2 (COVID-19) is a positive single-stranded RNA virus containing proteins[2]. This virus, which emerged in China in 2019, spread to the whole world, and was declared a pandemic by the WHO on March 20, 2020, and approximately three million people have now lost their lives due to this disease[3]. Drugs such as Chloroquine, Hydroxychloroquine, Azithromycin, Remdesivir, Lopinavir, Ritonavir have been used to treat this disease[4]. Vaccines have been produced for COVID-19, but many of these vaccines may be insufficient for new variants of COVID-19. Therefore, medicine to treat COVID-19 disease should be produced. This is a time-consuming and costly process. To overcome this, chemical graph theory focuses on topological indices.

The topological index is a numerical descriptor of the molecular graph of chemical structure. The molecular graph is represented by non-hydrocarbon skeletons of chemical structure. The edges of the graph correspond to covalent bonds between the atoms and the vertices of the graph correspond to atoms[5].Topological indices are generally used in the quantitative structure-property/structure-activity relationship studies to predict physicochemical and bioactivity properties ([6],[7]). Recently, many articles have been published on drugs used in the treatment of COVID-19 disease using the topological index (See [8]-[14]).

Rafi et al. has worked on drugs that may be more suitable than drugs used for the treatment of COVID-19 disease. It has been shown that these drugs contain less toxic and highly curable properties[15].This drugs are CID10009410, CID44271905, CID3010243, and CID271958 and CID89869520, CID71749833.

In this paper, degree-based topological indices of these chemical structures in the paper of Rafi are computed and obtained the best predictor model using linear regression for the boiling point property of these structures.

[1]speaker

2 Material and Method

The structure of all selected compounds was downloaded from the PubChem database [16].

Lopinavir (see Figure 1) and ritonavir (see Figure 1) is an inhibitor of human immunodeficiency virus-1 (HIV-1) aspartate protease. Since the previous SARS-CoV major protease has 96.1 similarities to the SARS-CoV-2 major protease, these two drugs can be used as a homologous target [17].

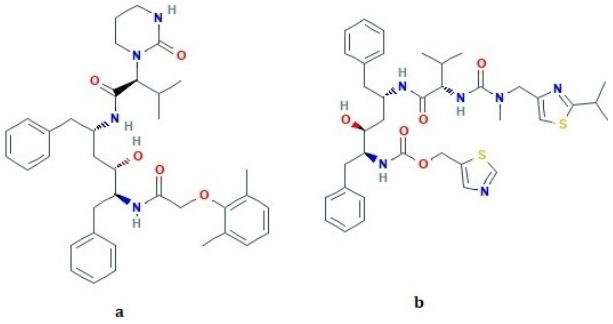

Figure 1: The chemical structures of a)Lopinavir and b) Ritonavir

Favipiravir is a pyrazine carboxamide derivative with activity against RNA viruses (see figure 2). It is an antiviral drug developed against influenza (flu virus). It was approved for the treatment of pandemic influenza emerging in Japan in 2014. It is used to treat moderate to mild COVID-19 patients. It is being studied for the treatment of COVID-19 disease ([4],[18]).

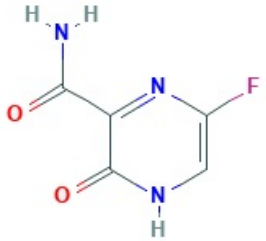

Figure 2: The chemical structures of Favipiravir

The structure of CID10009410 is a lopinavir analogue and was generated by adding –F groups at the end of their two dimensional (2D) structure (see Figure 3) [15].

The structure of CID44271905 which is lopinavir analogue was generated by removing trimethyl-benzene fragment into the 2D structure of lopinavir (see Figure 3) [15].

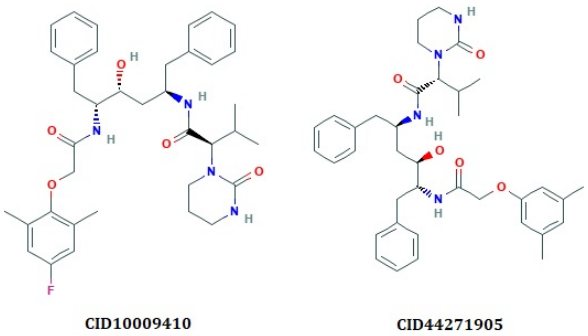

Figure 3: The chemical structures of CID10009410 and CID44271905

The structure of CID44271958 was generated by adding 1,3,5- trimethyl-benzene and benzene fragments into the 2D structure of lopinavir (see Figure 4) [15].

The structure of CID3010243 was generated by removing tetrahydro-pyrimidionepropylene urea fragment and adding 2-imidazolidone fragments into lopinavir (see Figure 4) [15].

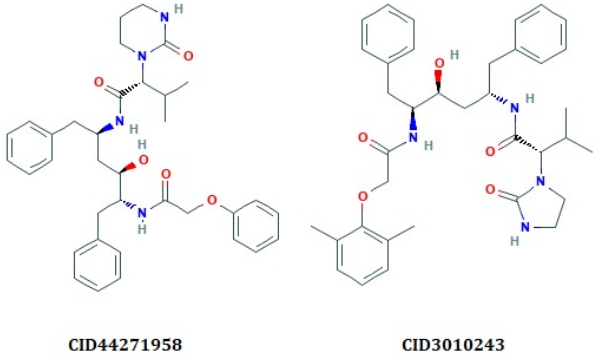

Figure 4: The chemical structures of CID44271958 and CID3010243

CID89869520 structure which is the favipiravir analogue was generated by adding –CH3 groups at the end of its 2D structure (See Figure 5) [15].

Lopinavir-d8 is a labelled selective HIV-1 protease inhibiting drug which is an analogue of ritonavir, this drug may act against COVID-19 (See Figure 5) [15].

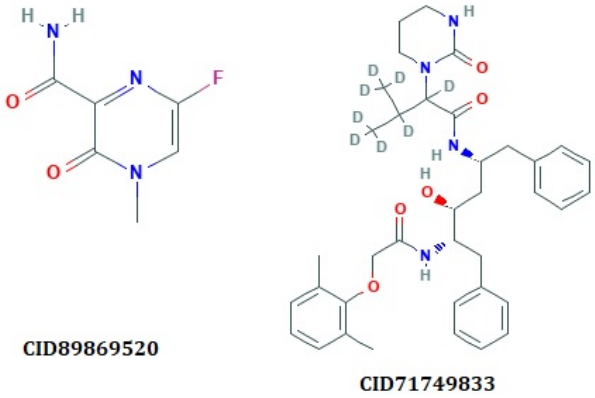

Figure 5: The chemical structures of CID89869520 and Lopinavir-d8

The degree-based topological indices used in this study are given in the table below.

Table 1: Topological Indices and Theirs Mathematical Expressions

Vertex-Degree-based topological indices	Mathematical expressions
First Zagreb index[19]	$M_1(G) = \sum_{uv \in E(G)} (d(u) + d(v))$
Second Zagreb index[19]	$M_2(G) = \sum_{uv \in E(G)} (d(u)d(v))$
Hyper Zagreb index[20]	$HM(G) = \sum_{uv \in E(G)} (d(u) + d(v))^2$
Max-min Rodeg index[21]	$Mm_{sde}(G) = \sum_{uv \in E(G)} \sqrt{\frac{\max\{d(u),d(v)\}}{\min\{d(u),d(v)\}}}$
Min-max Rodeg index[21]	$mM_{sde}(G) = \sum_{uv \in E(G)} \sqrt{\frac{\min\{d(u),d(v)\}}{\max\{d(u),d(v)\}}}$
Albertson index[22]	$irr(G) = \sum_{uv \in E(G)} \lvert d(u) - d(v) \rvert$
Sigma index[23]	$\sigma(G) = \sum_{uv \in E(G)} (d(u) - d(v))^2$
Inverse symmetric deg index[24]	$ISDD(G) = \sum_{uv \in E(G)} \frac{d(u)d(v)}{d(u)^2 + d(v)^2}$
Atom bond connectivity index[25]	$ABC(G) = \sum_{uv \in E(G)} \sqrt{\frac{d(u)+d(v)-2}{d(u)d(v)}}$
Inverse sum indeg index[21]	$ISI(G) = \sum_{uv \in E(G)} \frac{d(u)d(v)}{d(u)+d(v)}$

The Boiling point (BP) properties of these potential drugs against to COVID-19 are taken from Chem-Spider [26]. The values of Boiling point properties of these structures are given in the table below.

Table 2: The Boiling point properties of potential drugs to be used in the treatment of COVID-19 disease

PubChem ID	CID71749833	CID10009410	CID44271905	CID3010243	CID44271958	CID89869520
BP	924,2	925,8	924,2	918,8	911,6	383

In this study, the following equations are tested;

$$Y = a + b_1 X; n, R^2, F$$
$$Y = a + b_1 X + b_2 X^2; n, R^2, F$$

where Y is the response or dependent variable, a is the regression model constant,$b_i (i = 1, 2)$ are the coefficients for the individual descriptor, X, X are independent variables. n is the number of samples used for building the regression equation, R^2 is the square of the correlation coefficient, R is the correlation coefficient.

When the experimental and theoretical results are close to each other, the correlation coefficient is close to 1. Moreover, the observed values and model predictions must be compared to measure the predictive quality of the model (see detail [27],[28]). Therefore, we consider the RMSE (Root Mean Square Error) metric for the predictive power of the model. It is clear that the best predictive model is the minimum error, i.e. the minimum RMSE is defined as

$$RMSE = \sqrt{(\sum_{i=1}^{n} (y_i - z_i)^2)/n}$$

where y_i is the observed value of the independent variable in the test set, z_i is the predicted value of the independent variables in the test set, n is the number of samples in the test [28]. R^2, R, F, and RMSE is considered for the goodness of fit of the model i.e. $max([R]^2)$, $max(R), max(F)$, and $min(RMSE)$. The curvilinear regression analyses are obtained by using the SPSS statistical software. The independent variables in the curvilinear regression models are the values of the topological indices, which are described above, of various drugs used in the treatment of COVID-19.

3 Main results

Let $E_{i,j} = \{i = d_u, j = d_v \,||uv \in E(G)|\}$. Topological indices of chemical structures using edge partition technique are computed.

The molecular graph of CID71749833 has 54 vertices and 57 edges. Its edges can be partitions as $|E_{1,3}| = 6, |E_{1,4}| = 8, |E_{2,2}| = 14, |E_{2,3}| = 22, |E_{3,3}| = 2, |E_{3,4}| = 2, |E_{4,4}| = 3$.

The molecular graph of CID10009410 has 47 vertices and 50 edges. Its edges can be partitions as $|E_{1,3}| = 9, |E_{2,2}| = 12, |E_{2,3}| = 22, |E_{3,3}| = 7$.

The molecular graph of CID44271905 has 46 vertices and 49 edges. Its edges can be partitions as $|E_{1,3}| = 8, |E_{2,2}| = 12, |E_{2,3}| = 24, |E_{3,3}| = 5$.

The molecular graph of CID3010243 has 45 vertices and 48 edges. Its edges can be partitions as $|E_{1,3}| = 8, |E_{2,2}| = 13, |E_{2,3}| = 20, |E_{3,3}| = 7$.

The molecular graph of CID44271958 has 44 vertices and 47 edges. Its edges can be partitions as $|E_{1,3}| = 6, |E_{2,2}| = 16, |E_{2,3}| = 20, |E_{3,3}| = 5$.

The molecular graph of CID89869520 has 12 vertices and 12 edges. Its edges can be partitions as $|E_{1,3}| = 5, |E_{2,3}| = 4, |E_{3,3}| = 3$.

The topological index values of these structures are given in Table 3.

Table 3: The topological index values of potential drugs to be used in the treatment of COVID-19 disease

PubChem ID	M_1	M_2	HM	Mm_{sde}	mM_{sde}	irr	σ	$ISDD$	ABC	ISI
71749833	282	328	1432	74,646	46,159	60	120	24,296	24,916	63,72
10009410	236	270	1138	61,532	42,159	40	58	22,353	36,056	55,65
44271905	230	261	1100	60,250	41,214	40	56	21,976	35,321	54,3
3010243	226	259	1088	58,351	40,948	36	52	15,130	34,533	58,75
44271958	218	247	1032	55,887	40,794	32	44	19,030	33,688	52
89869520	58	66	288	16,559	9,152	14	24	4,846	8,910	13,05

The linear and quadratic models are obtained by using the data in Table 2 and Table 3 with the SPSS program.

linear regression models for Boiling point of potential drugs against COVID-19 disease are

$$Bp = 255.570 + 10.556[Mm]_{sde}, R = 0.991, R^2 = 0.898, F = 35.263, SE = 78.390, RMSE = 64.004$$

$$Bp = 261.143 + 2.737M_1, R = 0.960, R^2 = 0.921, F = 46.551, SE = 69.085, RMSE = 56.407$$

$$Bp = 272.235 + 2.344M_2, R = 0.951, R^2 = 0.905, F = 38.260, SE = 75.559, RMSE = 61.693$$

$$Bp = 287.675 + 0.537HM, R = 0.934, R^2 = 0.872, F = 27.177, SE = 87.970, RMSE = 71.827$$

$$Bp = 246.190 + 15.926mM_{sde}, R = 0.991, R^2 = 0.981, F = 208.384, SE = 33.705, RMSE = 27.519$$

$$Bp = 410.877 + 11.362irr, R = 0.768, R^2 = 0.590, F = 5.747, SE = 157.328, RMSE = 128.457$$

$$Bp = 615.040 + 3.665\sigma, R = 0.540, R^2 = 0.291, F = 1.643, SE = 206.775, RMSE = 168.830$$

$$Bp = 335.519 + 27.636ISDD, R = 0.901, R^2 = 0.813, F = 17.344, SE = 106.319, RMSE = 86.809$$

$$Bp = 279.379 + 19.094ABC, R = 0.922R^2 = 0.850, F = 22.680, SE = 95.094, RMSE = 77.644$$

$$Bp = 251.168 + 11.700ISI, R = 0.977R^2 = 0.955, F = 85.549, SE = 51.906, RMSE = 42.381$$

The mM_{sde} index is the best estimator index for boiling point in linear regression models from $max(R)$.

Quadratic regression models for Boiling point of potential drugs against COVID-19 disease are

$$Bp = 15.889 + 7.134M_1 - 0.014[M_1]^2, R^2 = 1, F = 102634.62, SE = 1.084, RMSE = 0.766$$

$$Bp = 20.363 + 6.186M_2 - 0.010[M_2]^2, R^2 = 1, F = 24671.497, SE = 2.211, RMSE = 1.563$$

$$Bp = 7.572 + 1.471HM - 0.001[HM]^2, R^2 = 1, F = 9296.834, SE = 3.602, RMSE = 2.546$$

$$Bp = 70.494 + 38.008mM_{sde} - 0.422[mM_{sde}]^2, R^2 = 1, F = 4650.447, SE = 5.092, RMSE = 3.600$$

$$Bp = -37.869 + 29.001[Mm]_{sde} - 0.216[[Mm]_{sde}]^2, R^2 = 1, F = 16334.478, SE = 2.717, RMSE = 1.921$$

$$Bp = 32.875 + 30.152ISI - 0.254[ISI]^2, R = 1, R^2 = 1, F = 3986.803, SE = 5.500, RMSE = 3.888$$

$$Bp = -194.912 + 49.225irr - 0.512[irr]^2, R^2 = 0.979, R = 0.989F = 69.437, SE = 41.238, RMSE = 29.159$$

$$Bp = -230.178 + 31.418\sigma - 0.182\sigma^2, R^2 = 0.926, R = 0.962F = 18.656, SE = 77.364, RMSE = 54.704$$

$$Bp = -24.250{+}96.670ISDD{-}2.410[ISDD]^2, R^2 = 0.989, R = 0.994F = 132.257, SE = 30.031, RMSE = 21.235$$

$$Bp = -188.070{+}75.358ABC{-}1.247[ABC]^2, R^2 = 0.996, R = 0.998F = 334.961, SE = 18.935, RMSE = 13.389$$

From the above equations, M_1 is the best predictor index for the boiling point in quadratic regression models from $min(RMSE)$.

Figure 6 shows the plots of the linear regression equation of the Boiling point (Bp) with the Min-max Rodeg index (mM_{sde}), and the quadratic regression equations of the Boiling point (Bp) with the first Zagreb index (M_1), respectively.

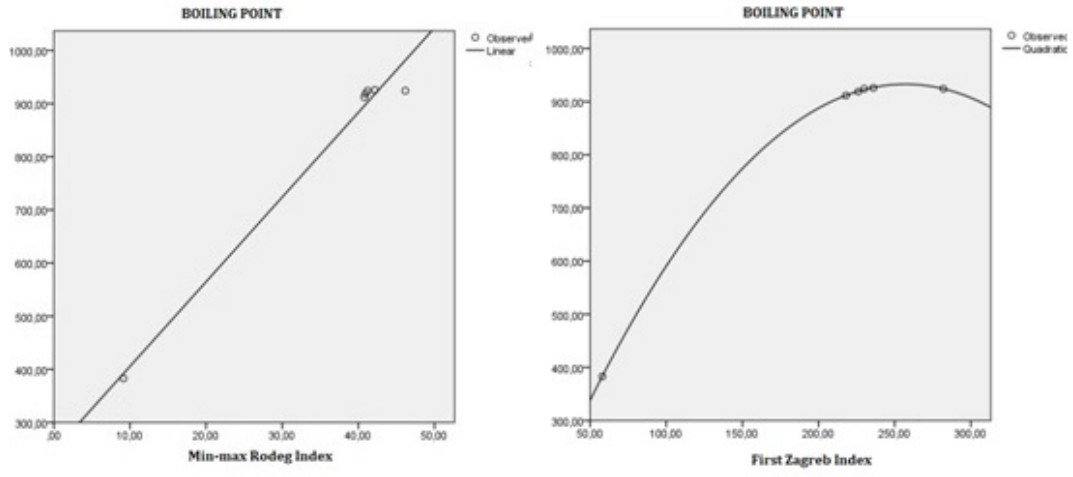

Figure 6: The plots of the linear regression model of the Boiling point (Bp) with the Min-max Rodeg index and the quadratic regression model of the Boiling point with the first Zagreb index

Models were studied with 6 descriptors and 10 topological indices. QSPR modeling shows that the best predictive topological index is the min-max rodeg index, the first Zagreb index for boiling point in linear regression, quadratic regression models, respectively.

References

[1] G.W. Ejuh, C. Fonkem, Y. Tadjouteu Assatse, R.A. Yossa Kamsi, Tchangnwa Nya, L.P. Ndukum , J.M.B. Ndjaka, *Study of the structural, chemical descriptors and optoelectronic properties of the drugs Hydroxychloroquine and Azithromycin*, Heliyon, 6(2020), e04647.

[2] B. Nutho, P. Mahalapbutr, K. Hengphasatporn, N. C. Pattaranggoon, N. Simanon, Y. Shigeta, S. Hannongbua, and T. Rungrotmongkol, *Why Are Lopinavir and Ritonavir Effective against the Newly Emerged Coronavirus 2019? Atomistic Insights into the Inhibitory Mechanisms*, Biochemistry. 59(18) (2020), 1769–1779.

[3] World Health Organization, Coronavirus disease (COVID-19) pandemic, last modified March 11, 2020, accessed July 20, 2021, https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19—11-march-2020.

[4] Janik, E., Niemcewicz, M., Podogrocki, M., Saluk-Bijak, J., Bijak, M. *Existing Drugs Considered as Promising in COVID-19 Therapy* Int. J. Mol. Sci. 22 (2021), 5434. https://doi.org/10.3390/ijms22115434

[5] I. Gutman, *A property of the Simple Topological Index*, MATCH Commun. Math. Comput. Chem., 25(1990), 131-140.

[6] H. Wiener, *Structural determination of paraffin boiling points*, J. Am. Chem. Soc., 69(1947), 17-20.

[7] S. Hussain, F. Afzal, D. Afzal, M. Cancan, S. Ediz, MR. Farahani, *Analyzing the boron triangular nanotube through topological indices via M-polynomial*, J Disc Math Sci Cryp., 24:2 (2021), 415-426, DOI: 10.1080/09720529.2021.1882158

[8] RMK. Omar, AM. Najar, E. Bobtaina, AF. Elsheikh, *Pryazolylpyridine and Triazolylpyridine derivative of hydroxychloroquine as Potential Therapeutic against COVID-19: Theoretical Evaluation*, J. drug deliv. ther, 10(2020), 181-186. http://dx.doi.org/10.22270/jddt.v10i4-s.4298

[9] S.A. K. Kirmani, P. Ali, F. Azam, *Topological indices and QSPR/QSAR analysis of some antiviral drugs being investigated for the treatment of COVID-19 patients*, Int. J. Quantum Chem, 2020, 121(9), e26594.

[10] Ö. Çolakoğlu Havare, *Quantitative Structure Analysis of Some Molecules in Drugs Used in the Treatment of COVID-19 with Topological Indices*, Polycycl. Aromat. Compd. (2021) DOI: 10.1080/10406638.2021.1934045

[11] J. F. Zhong, A. Rauf, M. Naeem, J. Rahman, A. Aslam, *Quantitative structure-property relationships (QSPR) of valency based topological indices with COVID-19 drugs and application*, Arab. J. Chem., 14 (2021), 103240.

[12] B. Chaluvaraju, A. B. Shaikh, *Different Versions of Atom-Bond Connectivity Indices of Some Molecular Structures: Applied for the Treatment and Prevention of COVID-19*, Polycycl. Aromat. Compd. (2021). DOI: 10.1080/10406638.2021.1872655.

[13] J.B. Liu, M. Arockiaraj, M. Arulperumjothi, S. Prabhu, *Distance based and bond additive topological indices of certain repurposed antiviral drug compounds tested for treating COVID-19*, Int. J. Quantum Chem., 121 (2021) ,e26617. https://doi.org/10.1002/qua.26617.

[14] J. Wei , M. Cancan , A. Ur Rehman, M. K. Siddiqui , M. Nasir, M. Tayyab Younas and M. F. Hanif, *On Topological Indices of Remdesivir Compound Used in Treatment of Corona Virus (COVID-19)*, Polycycl. Aromat. Compd. (2021) DOI: 10.1080/10406638.2021.1887299.

[15] Md. O. Rafi, G. Bhattacharje, K. Al-Khafaji, T. Taskın Tok, Md. A. Alfasane, A. K. Das, Md. A. K. Parvez, Md. S. Rahman, *Combination of QSAR, molecular docking, molecular dynamic simulation and MM-PBSA: analogues of lopinavir and favipiravir as potential drug candidates against COVID-19*, J. Biomol. Struct. Dyn. (2020) DOI: 10.1080/07391102.2020.1850355.

[16] PubChem, an open chemistry database at the National Institutes of Health (NIH), accessed 20 July 2021, https://pubchem.ncbi.nlm.nih.gov/.

[17] G. Bolcato, M. Bissaro, M. Pavan, M. Sturlese, S. Moro, *Targeting the coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors lopinavir, ritonavir and nelfinavir*, Sci. Rep. , 10 (2020), Article number: 20927.

[18] K. Shiraki, T. Daikoku, Favipiravir, *an anti-influenza drug against life-threatening RNA virus infections.* Pharmacol. Ther., 209 (2020), 107512.

[19] I. Gutman, B. Ruscic, N. Trinajstic, Jr. C.F. Wilson, *Graph theory and molecular orbitals. XII. Acyclic polyenes*, J. Chem. Phys., 62 (1975), 3399-3405.

[20] G. H. Shirdel, H. Rezapour, A. M. Sayadi, *The hyper–Zagreb index of graph operations*, Iranian J. Math. Chem., 4 (2013), 213–220.

[21] D. Vukiccevic, M. Gasparov, *Bond additive modeling 1. Adriatic indices*, Croat. Chem. Acta., 83 (2010), 243–260.

[22] M. O. Albertson, *The irregularity of a graph*, Ars Comb.,46(1997), 219-225.

[23] I. Gutman, M. Togan, A. Yurttas, A. S. Cevik, I. N. Cangul, *Inverse Problem for Sigma Index*, MATCH Commun. Math. Comput. Chem., 79 (2018), 491-508.

[24] M. Ghorbani, S. Zangi, N. Amraei, *New results on symmetric division deg index*, J. Appl Math Comput., 65 (2021), 161–176. Doi:10.1007/s12190-020-01386-9.

[25] E. Estrada, L. Torres, L. Rodríguez, I. Gutman, *An atom-bond connectivity index: modelling the enthalpy of formation of alkanes*, Indian J. Chem., 37A (1998), 849-855.

[26] Chemspider, Search and share chemistry, http://www.chemspider.com/AboutUs.aspx. (Accessed on 20 July 2021)

[27] L. Wang, P. Xing, C. Wang, X. Zhou, Z. Dai, L. Bai, *Maximal Information Coefficient and Support Vector Regression Based Nonlinear Feature Selection and QSAR Modeling on Toxicity of Alcohol Compounds to Tadpoles of Rana temporaria*, J Brazilian Chem Soc, 30(2), (2019) 279-285. Doi:http://dx.doi.org/10.21577/0103-5053.20180176

[28] V.Consonni, D. Ballabio, R. Todeschini, *Comments on the Definition of the Q2 Parameter for QSAR Validation*, J Cheml Inf Modeling, 49(7) (2009), 1669-1678.

Email: ozgecolakoglu@mersin.edu.tr
ozgeeclkgl@gmail.com