Journal of Mathematical Problems, Equations and Statistics

E-ISSN: 2709-9407
P-ISSN: 2709-9393
Impact Factor (RJIF): 5.94
JMPES 2025; 6(2): 227-232

www.mathematicaljournal.com
Received: 20-06-2025
Accepted: 19-07-2025

S Ranjitham
1) Research Scholar, Department of Mathematics, Sree Saraswathi Thyagaraja College, Bharathiar University, Coimbatore, Pollachi, Coimbatore, Tamil Nadu, India
2) Lecturer, Department of Mathematics, Shree Venkateshwara Hi-Tech Polytechnic College, Othakuthirai, Gobichettipalayam, Erode, Tamil Nadu, India

OV Shanmuga Sundaram
Associate Professor, Department of Mathematics, Sree Saraswathi Thyagaraja College (Autonomous), Pollachi, Coimbatore, Tamil Nadu, India

G Priyadharsini
Assistant Professor, Department of Mathematics, Kongu Arts & Science College, Erode, Tamil Nadu, India

Corresponding Author:
S Ranjitham
1) Research Scholar, Department of Mathematics, Sree Saraswathi Thyagaraja College, Bharathiar University, Coimbatore, Pollachi, Coimbatore, Tamil Nadu, India
2) Lecturer, Department of Mathematics, Shree Venkateshwara Hi-Tech Polytechnic College, Othakuthirai, Gobichettipalayam, Erode, Tamil Nadu, India

QSPR modelling based on topological indices and regression analysis for predicting the physicochemical properties of herbal glycosides

S Ranjitham, OV Shanmuga Sundaram and G Priyadharsini

DOI: https://www.doi.org/10.22271/math.2025.v6.i2b.236

Abstract
Quantitative Structure Property Relationship (QSPR) modeling is shown to be an interesting method to predict the physicochemical properties of herbal glycosides that are compounds known in terms of their medicinal importance, though in many cases, they may not have a lot of experimental data published about them. The paper aims at investigating the use of degree-based topological indices as molecular descriptors in the construction of regressions to predict several important physicochemical properties, including boiling point, flash point, polarizability, surface tension, and molar volume, in a selection of herbal glycosides. Topological indices like First and Second Zagreb, Forgotten, Yemen and sum-connectivity were used in doing a linear regression analysis. These statistical tests indicated that boiling point, polarizability, and molar volume had the highest correlations and prediction accuracies with some of the topological indices ($R^2 > 0.90$, $p<0.001$), specifically the Forgotten and the Yemen one. Conversely, the flash point and surface tension predictions exhibited poorer model performance suggesting a certain degree of lack of degree-based descriptors describing these properties. These findings indicate that topological indices can be used as effective predictors of some important properties of herbal glycosides, and in silico search and property estimation can be performed very efficiently with them. It has also been established that QSPR methodologies have the potential to contribute to the development of phytopharmaceutical development, though more extensive sets of descriptors and the application of more sophisticated model development methods are required to handle more complex property predictions. Future works must be concentrated on the enlargement of set of data and the further development of additional computational methods that will allow to increase the robustness and range of QSPR models on natural product research.

Keywords: QSPR, topological indices, regression analysis, physicochemical properties, herbal glycosides, molecular descriptors

1. Introduction
Quantitative structure-property relationship (QSPR) modelling has become an important research tool in the field of computational chemistry and drug design helping to predict the physicochemical class of molecules on the basis of structural descriptors. The method is especially noteworthy in reference to the herbal glycosides-naturally derived compounds that are known to have a great variety of pharmacofungal effects, yet with experimental data concerning properties being either sparse or expensive to acquire. The structure of herbal glycosides, present in medicinal plants, has direct effects on the solubility, thermal stability, bioactivity of the drug as well as its pharmacokinetics. Prediction of these physicochemical properties is essential for the estimation of their therapeutic utility and safety characterization [1, 2]. Topological indices are the numerical descriptors based on the molecular graph representation of compounds, and have been widely used in QSPR and QSAR (Quantitative Structure Activity Relationship) studies. These indices reflect key molecular features of branching, size and connectivity, and have been used in correlations with physicochemical and biological properties with a great deal of success [3-5]. The introduction of the pioneering indices like the Wiener and Zagreb indices established the foundations in which they can be used in many chemical datasets including drug and bioactive natural compounds [3, 4].
Modelling boiling point, polarizability, surface tension, and molar volume Recent studies have noted the application of degree-based topped making indices and eccentricity-based topologic indices as models in some cases with the predictivity classification of certain models very high using statistical models (linear regression and more complex examples) [4-6].

Although QSPR models have been established and verified by a different variety of different kinds of molecules, varying between synthetic drugs to natural product such as flavonoid and anthraquinones, little has been done in relation to herbal glycoside with regards to developing and validating these models [4].

Current literature evidence shows that using topological indices in QSPR models possesses the potential of greatly simplifying the process of virtual screening and prediction of properties at an early discovery stage of drugs, so that they may help in selecting candidate chemicals that have a good chance of being synthesized and tested in biological systems [3, 6, 7]. Even taking into account all the achievements in the QSPR method and the proved actuality of employing topological indices, there is still a significant research gap in the application to herbal glycosides. The majority of previous works are dealing with synthesis of synthetic or small natural molecules and have little extension into the wider area of glycosidic natural products. Consequently, there is a solid need to conduct comprehensive research in this area using regression analysis and degree-based topological indices so as to be able to predict significant physicochemical properties of the herbal glycosides which includes the boiling point, flash point, polarizability and the molar volume. In this way, the study mains to build and confirm QSPR models using the topological indices employed as structural indicators to foresee the vital physicochemical properties of herbal glycosides.

2. Terminology

In theoretical chemistry and cheminformatics, drug compounds can be conveniently models as graphical representations denoted by G=(V, E), where V represents atoms in the compounds and E represents the chemical bond between any pair of atoms. The molecular graphs involved in this study are simple graphs, i.e. they do not contain any loops (edges connecting a vertex to itself) or repeated edges between any pair of vertices. The graphs will have the form of acyclic molecules, typical in many drug molecules [8]. In this work, they have been applied a number of degree-based topological indices that have been shown to be capable of characterizing molecular structure. Their indices are given as;

Definition 1

The first and second Zagreb indices are proposed by Gutman & Trinajstic (1972), [9], as

$$M_1(G) = \sum_{e\in\varphi(G)} (r_a + r_b)$$

$$M_2(G) = \sum_{e\in\varphi(G)} (r_a r_b)$$

Definition 2: The Forgotten index is proposed by Furtula & Gutman (2015) [10], as

$$F(G) = \sum_{e\in\varphi(G)} (r_a^2 + r_b^2)$$

Definition 3: The Sum-Connectivity index is proposed by Zhou & Trinajstic (2010) [11], and

$$S(G) = \sum_{e\in\varphi(G)} (r_a^4 + r_b^4)$$

Definition 4: Yemen index have been introduced more than thirty years ago by A.A Naggar [12].

$$Y(G) = \sum_{e\in\varphi(G)} (r_a^3 + r_b^3)$$

Definition 5: Degree index D(G) is proposed by Nagaraja & Durga (2024) [13]

$$D(G) = \sum_{e\in\varphi(G)} (r_a^6 + r_b^6)$$

3. Topological indices in QSPR degree based topological

This paper examined six-degree topological indices consisting of the First Zagreb index M_1(G), Second Zagreb M_2(G), Forgotten index F(G), Sum-connectivity index S(G) and Yemen index Y(G) and Correspondence degree index D(G). They used these indices to predict five important physicochemical features of the chosen herbal glycosides which are Boiling Point (BP), Flash Point (FP), Polarizability (P), Surface Tension (ST), and Molar Volume (MV). These properties were obtained using the ChemSpider database; the experimentally derived values were used. The calculated values of the considered topological indices together with the experimental values of the six examined herbal glycosides (Figure 1) are tabulated in Table 1.

The results in table 1 report the experimental physicochemical investigation of physicochemical characteristics of six medicinal herbal glycosides such as Aloe-emodin, Chrysophanol, Emodin, Rhein, Hypericin and Sennoside-A. The corresponding properties boiling points, flash points, polarizability, surface tension, and molar volume are the same ones mentioned above. These are the properties required to know the thermal stability, molecular interaction properties and general stability of compounds which also influence pharmacological activity and environmental fate. Hypericin and sennoside A, it can be seen that they have higher values on most of the properties as they have bigger and more complicated structures.

Table 1: Experimental physicochemical properties of selected herbal glycoside

Physico-chemical properties	Drugs					
	Aloe-emodin	**Chrysophanol**	**Emodin**	**Rhein**	**Hypericin**	**Sennoside-A**
Boiling Point (°C at 760 mmHg)	568.80	489.50	586.90	597.80	930.10	1144.80
Flash Point (°C)	311.90	263.90	322.80	329.40	530.10	348.60
Polarizability (cm^3)	27.30	26.70	27.40	27.50	56.50	80.90
Surface Tension (dyne/cm)	88.50	73.10	85.40	94.90	150.30	107.20
Molar Volume (cm^3)	169.70	172.20	170.60	168.50	263.40	494.90

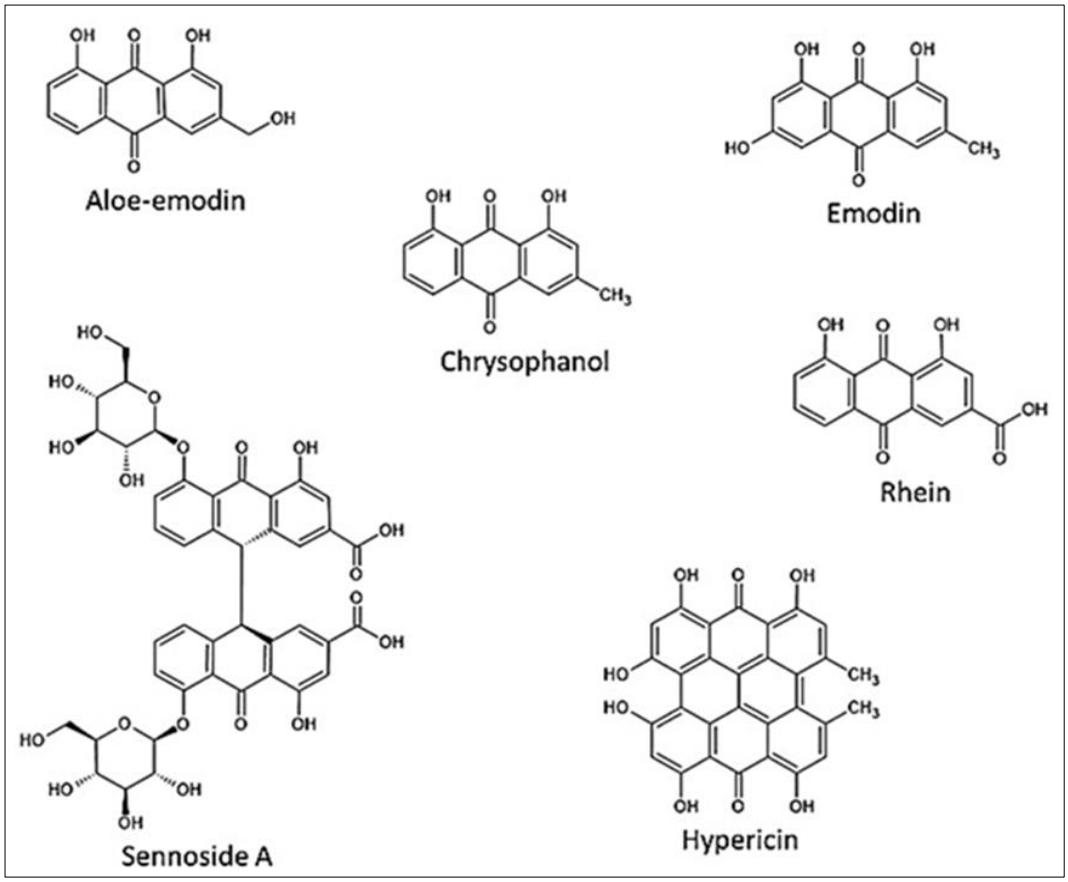

Fig 1: Molecular structures of herbal glycosides

Table 2: Computed Degree-based topological indices of selected Herbal Glucoside

Topological Indices values	Drugs					
	Aloe-emodin	**Chrysophanol**	**Emodin**	**Rhein**	**Hypericin**	**Sennoside-A**
$M_1(G)$	107	108	109	116	242	341
$M_2(G)$	181	138	132	143	321	415
F(G)	291	296	293	316	690	933
Y(G)	811	840	823	896	2018	2611
D(G)	18461	21056	20393	23716	53010	66753
S(G)	1926	1847	2369	2596	5970	7689

Source: Calculated value

The results of calculation of topological indicators on a 6-degrees level on the foundation of the same series of herbs glycosides are presented in Table 2. Some of these numbers are: they hold M1(G) and M2(G) which describe the first and second Zagreb index, F(G) which is the forgotten index, Y(G) which is the Yemen index, D(G), and S(G) are other degree based descriptors and reflect various properties of connectivity and branching of the molecule. They are topologic indices, which are numerical scores that define the molecular structures, in terms of the numbers which represent the complexity of a molecule, its branching and connectivity, and therefore molecular descriptors and can therefore be utilized to QSPR model.

Table 3: Correlation coefficients

Indices / Properties	Boiling Point	Flash Point	Polarizability	Surface Tension	Molar Volume
M1(G)	.990**	.510	.999**	.627	.962**
M2(G)	.981**	.558	.986**	.674	.932**
F(G)	.990**	.541	.998**	.656	.950**
Y(G)	.989**	.573	.994**	.683	.937**
D(G)	.987**	.595	.988**	.703	.924**
S(G)	.995**	.604	.987**	.710	.924**

Source: Calculated value, ** Correlation is significant at the 0.01 level (2-tailed).

Table 3 gives the correlation coefficients between degree independent topological indices and the physicochemical properties of herbal glycosides such as boiling point, polarizability, surface tension and molar volume. Correlations were found to be strong positive (coefficients>0.9) as in the case of boiling point, polarizability, molar volume with $M_1(G)$, $M_2(G)$, F(G), Y(G), D(G), and S(G). This indicates that these descriptors of topology are significant in structural representation to these properties. Particularly, $M_1(G)$ and polarizability were highly correlated with a value of 0.99. Comparatively, the correlation of flash point and surface tension was moderate (0.5 to 0.7). This implies that other

intermolecular forces may influence them.

4. Regression analysis

Results in correlation matrix show that the data are close to normal distribution. The observation is the reason to employ linear regression analysis in the exploration of an observed relationship between degree-based topologic indices and the physicochemical properties of liver cancer drugs. Linear regression model applied in the research is generalized as below:

$$PP=\alpha+\beta(t_i) \quad (1)$$

In this equation, PP represents the physicochemical property, 'α' is the regression intercept (constant), 'β' is the regression coefficient, and 't_i' refers to the value of the degree-based topological index. With the help of the SPSS software, the regression models were developed in accordance with the values of five physicochemical properties and six degree-based topological parameters of the used liver cancer drugs. Using equation (1), the study achieved a series of linear equations of regression that addressed each of the parameters defined topological index. Tables and models to the results are given below.

Table 4: Statistical specifications for the linear QSPR model for $M_1(G)$

Physical properties	BP	FP	P	ST	MV
α	279.028	270.061	1.814	70.556	23.590
β	2.584*	0.475	0.230*	0.172	1.269*
SE	41.320	88.620	0.922	23.607	39.803
R	0.990	0.510	0.999	0.627	0.962
R^2	0.980	0.260	0.999	0.394	0.925
Adj R^2	0.974	0.075	0.998	0.242	0.907
F	191.138*	1.406	3045.541*	2.598	49.645*
p	0.000	0.301	0.000	0.182	0.002
N	6	6	6	6	6
Results	S	NS	S	NS	S

Source: Calculated value, * Significant at 1% level

The table 4 shows that the regression model is boiling point (BP)=279.028+ 2.584($M_1(G)$) The relationship model between $M_1(G)$ and boiling point explained 98 (R^2=0.980) percent of the variation and an extremely significant F-statistic (F=191.14, p<0.001), and there was a strong association between $M_1(G)$ and boiling point. Similarly, polarizability (P=1.814+0.230($M_1(G)$) and the molar volume (MV=23.590+1.269($M_1(G)$) resulted in equally very high R^2 values, at (R^2=0.999 and R^2=0.925) and that the p-value was also highly significant, that $M_1(G)$ serves an excellent predictor of these properties as well. Poor model fit is obtained though (R^2 < 0.4, p>0.05) in the case of flash point (FP), surface tension (ST) and no predictions can be made.

Table 5: Statistical specifications for the linear QSPR model for $M_2(G)$

Physical properties	BP	FP	P	ST	MV
α	245.577	254.924	-0.954	65.708	12.348
β	2.139*	0.434	0.189*	0.154	1.026*
SE	55.854	85.518	4.287	22.407	52.676
R	0.981	0.558	0.986	0.674	0.932
R^2	0.963	0.311	0.972	0.454	0.869
Adj R^2	0.953	0.139	0.965	0.317	0.837
F	102.817*	1.806	137.026*	3.323	26.629*
p	0.001	0.250	0.000	0.142	0.007
N	6	6	6	6	6
Results	S	NS	S	NS	S

Source: Calculated value, * Significant at 1% level

The table 5 shows that the regression model is boiling point *(BP)=245.577+ 2.139($M_2(G)$).* This relationship model between model $M_2(G)$ and boiling point explained 96.3 (R^2=0.963) percent of the variation and a highly significant F-statistic (F=102.817, p<0.001), and there was a strong association between $M_2(G)$ and boiling point. Similarly, polarizability (P=-0.954+0.189($M_2(G)$) and the molar volume (MV=12.348+1.026($M_2(G)$) resulted in exceptionally high R^2 values (R^2=0.972 and R^2=0.869), and that p-values was also highly significant, that $M_2(G)$ serves an excellent predictor of these properties as well. However, the model fit is poor (R^2 < 0.311, p>0.05) in case of flash point (FP), surface tension (ST) and no prediction can be given.

Table 6 shows that the regression equation is boiling point (BP)=284.025+ 0.927(F(G)). This model explained 98.1 (R^2=0.981) percent of the variability and an extremely significant F-Statistic (F=202.806, p<0.001); there was a very strong relationship between F(G) and boiling point. More or less the same, the polarizability (P=2.351+ 0.082(F(G)) and the molar volume (MV=28.726+ 0.449(F(G)) also had incredibly high R^2 values (R^2=0.995 and $R^{2=}$0.903, respectively). The p-values are also extremely significant, a trend that has held through to these properties as well-being great predictors of F(G). However, the model fit is poor (R^2 < 0.293, p>0.05) in the case of flash point (FP), surface tension (ST), and no prediction can be offered.

Table 6: Statistical specifications for the linear QSPR model for F(G)

Physical properties	BP	FP	P	ST	MV
α	284.025	266.127	2.351	69.608	28.726
β	0.927	0.181	0.082	0.064	0.449
SE	40.142	86.632	1.780	22.895	45.297
R	0.990	0.541	0.998	0.656	0.950
R^2	0.981	0.293	0.995	0.430	0.903
Adj R^2	0.976	0.116	0.994	0.287	0.879
F	202.806	1.657	814.224	3.014	37.421
p	0.000	0.267	0.000	0.158	0.004
N	6	6	6	6	6
Results	S	NS	S	NS	S

Source: Calculated value, * Significant at 1% level

Table 7: Statistical specifications for the linear QSPR model for Y(G)

Physical properties	BP	FP	P	ST	MV
α	284.982	261.292	2.544	68.365	32.009
β	0.326	0.067	0.029	0.024	0.156
SE	41.773	84.449	2.794	22.134	50.918
R	0.989	0.573	0.994	0.683	0.937
R^2	0.979	0.328	0.988	0.467	0.878
Adj R^2	0.974	0.160	0.985	0.334	0.847
F	186.969	1.954	328.033	3.505	28.781
p	0.000	0.235	0.000	0.134	0.006
N	6	6	6	6	6
Results	S	NS	S	NS	S

Source: Calculated value, * Significant at 1% level

The table 7 indicates that the regression model is boiling point (BP)=284.982+ 0.326(Y(G)). This model accounted for 97.9 (R^2=0.979) percent of the variation and a highly significant F-statistic (F=186.969, p<0.001), a strong relationship existed between Y(G) and boiling point. Much the same, the polarizability (P=2.544+ 0.029(Y(G)) and the molar volume (MV=32.009+ 0.024(Y(G)) also had exceptionally high R^2 values (R^2=0.988 and R^2=0.878, respectively), and p-values that are also highly significant, continuing the result that Y(G) is an exceptional predictor of these properties. The model fit is however poor (R^2 < 0.328, p>0.05) in case of flash point (FP), surface tension (ST) and no prediction can be given.

Table 8: Statistical specifications for the linear QSPR model for D(G)

Physical properties	BP	FP	P	ST	MV
α	301.749	261.097	4.143	68.607	42.226
β	0.012	0.003	0.001	0.001	0.006
SE	47.187	82.780	3.937	21.554	55.755
R	0.987	0.595	0.988	0.703	0.924
R^2	0.973	0.354	0.976	0.495	0.854
Adj R^2	0.967	0.193	0.970	0.368	0.817
F	145.659	2.196	163.166	3.914	23.340
p	0.000	0.212	0.000	0.119	0.008
N	6	6	6	6	6
Results	S	NS	S	NS	S

Source: Calculated value, * Significant at 1% level

Table 8 shows that the predictive model is boiling point (BP)=301.749+ 0.012(D(G)). With exceptionally high F-statistic (F=145.659, p<0.001) and a high% of the variation explained by the model (R^2=0.973) or (97.93 percent), there existed a very strong relationship between D(G) and boiling point. Similarly, polarizability (P=4.143+ 0.001(D(G)) and the molar volume (MV=42.226+ 0.006(D(G)) also produced an exceptionally high R^2 value (R^2=0.976 and R^2=0.854, respectively) and p-value too, which is highly significant, and a trend goes on that indicates that D(G) is an excellent predictor of these properties. Nevertheless, flash point (FP), surface tension (ST) does not fit very well and it is unable to make any prediction (R^2 < 0.354, p>0.05).

Table 9: Statistical specifications for the linear QSPR model for S(G)

Physical properties	BP	FP	P	ST	MV
α	332.233	267.136	7.153	70.867	58.103
β	0.104	0.022	0.009	0.008	0.049
SE	28.753	82.092	4.057	21.354	55.570
R	0.995	0.604	0.987	0.710	0.924
R^2	0.990	0.365	0.975	0.504	0.855
Adj R^2	0.988	0.206	0.968	0.380	0.818
F	399.088	2.301	153.428	4.063	23.522
p	0.000	0.204	0.000	0.114	0.008
N	6	6	6	6	6
Results	S	NS	S	NS	S

Source: Calculated value, * Significant at 1% level

The table 9 indicates that the regression model is boiling point (BP)=332.233+ 0.104(S(G)). This model accounted for 99.0 (R^2=0.990) percent of the variation and a highly significant F-statistic (F=399.088, p<0.001), a strong relationship existed between S(G) and boiling point. Much the same, the polarizability (P=7.153+ 0.009 (S(G)) and the molar volume (MV=58.103+ 0.049(S(G)) also had exceptionally high R^2 values (R^2=0.975 and R^2=0.855, respectively), and p-values that are also highly significant, continuing the result that S(G) is an exceptional predictor of these properties. The model fit is however poor (R^2 < 0.354, p>0.05) in case of flash point (FP), surface tension (ST) and no prediction can be given.

5. Conclusion

The study shows how the degree-constructed topological indices have a great potential to be useful and efficient molecular descriptions of important physicochemical properties of the glycosides in herbs using QSPR modelling and linear relationship analysis. Strong correlations and well-regressed models were produced concerning boiling point, polarizability, and molar volumes especially when the use of the indices was applied, such as the Second Zagreb, Forgotten index, and the Yemen index. These results indicate that the thermal and spatial properties of glycosidic herbal compounds are heavily determined by the molecular branching and connectivity which is reflected in the metric scale of selected topological descriptors. Nonetheless, it is also established during the study that some of the properties, especially the flash point and surface tension, are much harder to predict with the utilized set of indices, which implies the necessity of an extended range of molecular descriptors to take a look at or more sophisticated modelling methods which might not be linear. Nonetheless, the findings indicate an equally persuasive argument to consider QSPR strategies within phytopharmaceutical studies and development as an efficient method of studying property values and molecular screening using resources that are time and resource efficient. Further improvements in the generalizability and predictive capacity of these models with QSPR should be pursued in the future by applying larger datasets and by adding more types of descriptors, and by cross-validation to experimentally-derived data.

References

1. Liu Y, Lawless M, Li M, Fairman K, Embry MR, Mitchell CA. Prediction of physicochemical and pharmacokinetic properties of botanical constituents by computational models. J Appl Toxicol. 2024;44(8):1236-45. DOI: 10.1002/jat.4617.
2. Wang Z, Ma P, Xu L, He C, Peng Y, Xiao P. Evaluation of the content variation of anthraquinone glycosides in rhubarb by UPLC-PDA. Chem Cent J. 2013;7:170. DOI: 10.1186/1752-153X-7-170.
3. Kirmani SAK, Ali P, Azam F. Topological indices and QSPR/QSAR analysis of some antiviral drugs being investigated for the treatment of COVID-19 patients. Int J Quantum Chem. 2021;121(9):e26594. DOI: 10.1002/qua.26594.
4. Priyadharsini G, Amutha R, Sangeetha N, Amudhamalar V. Degree based topological indices in QSPR analysis of flavonoids. Libr Prog Int. 2024;44(3):16279-16289. Available from: www.bpasjournals.com
5. Rasheed MW, Mahboob A, Hanif I. Uses of degree-based topological indices in QSPR analysis of alkaloids with poisonous and healthful nature. Front Phys. 2024;12:1381887. DOI: 10.3389/fphy.2024.1381887.
6. Kara Y, Özkan YS, Ullah A, Hamed YS, Belay MB. QSPR modeling of some COVID-19 drugs using neighborhood eccentricity-based topological indices: a comparative analysis. PLOS One. 2016;11(5):e0155707. DOI: 10.1371/journal.pone.0155707.
7. Can PFJ, Sánchez SJ, Rivera LI, Tlahuext H, Pastor N, Razo-Hernández RS. Identification of a family of glycoside derivatives biologically active against *Acinetobacter baumannii* and other MDR bacteria using a QSPR model. Pharmaceuticals. 2023;16(2):250. DOI: 10.3390/ph16020250.
8. Priyadharsini G, Amutha R, Sangeetha N, Amudhamalar V. Degree based topological indices in QSPR analysis of flavonoids. Libr Prog Int. 2024;44(3):16279-16289. Available from: https://bpasjournals.com/library-science/index.php/journal/article/view/1665/2084
9. Gutman I, Trinajstić N. Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons. Chem Phys Lett. 1972;17(4):535-8. DOI: 10.1016/0009-2614(72)85099-1.
10. Furtula B, Gutman I. A forgotten topological index. J Math Chem. 2015;53(4):1184-90. DOI: 10.1007/s10910-015-0480-z.
11. Zhou B, Trinajstić N. On general sum-connectivity index. J Math Chem. 2010;47(1):210-8. DOI: 10.1007/s10910-009-9542-4.
12. Nagarajan S, Durga M. Computing Y-index of different corona products of graphs. Asian Res J Math. 2023;19(10):67-74. DOI: 10.9734/arjom/2023/v19i10729.
13. Nagarajan S, Durga M. A computational approach on fenofibrate drug using degree-based topological indices and M-polynomials. Asian J Chem Sci. 2024;14(2):43-57. DOI: 10.9734/ajocs/2024/v14i2293.