



# Graph-Theoretical Analysis of Sarcoma Cancer Tissues: Predicting Physicochemical Properties Using Degree-Based Topological Indices

PRANAVI JAINA, K. ANIL KUMAR and J. VIJAYASEKHAR\*

Department of Mathematics, GITAM (Deemed to be University), Hyderabad, India.

\*Corresponding author E-mail: vijayjaliparthi@gmail.com

<http://dx.doi.org/10.13005/ojc/410626>

(Received: March 20, 2025; Accepted: November 06, 2025)

## ABSTRACT

In chemical graph theory, topological indices are useful benchmarks because they allow the quantification of molecular structures. In this work, we apply degree-based topological indices to analyze pathologist-encoded sarcoma cancer tissues and to forecast associated physicochemical properties. We define three new numerical descriptors and propose central tissue attributes-flash point (FP), enthalpy (H), molar volume (MV), and refractive index (RI)-for the Albertson index, Sigma index, and Forgotten index. The study tests the predictive efficiency of the indices using linear regression models and demonstrates that the computed values agree well with values obtained from experiments. Statistical validation shows that the indices proposed in this work have substantial predictive potential with correlation coefficients above 0.7, F-statistics over 2.5, and p-values below 0.05. The indices presented make molecular property estimation faster and cheaper while aiding computational chemistry and biomedical science researchers. The outcome can aid the predictive modelling of cancer-related pharmaceuticals without having to perform extensive experiments.

**Keywords:** Topological indices, Sarcoma cancer tissues, Quantitative Structure-Property Relationship (QSPR), Chemical graph theory, Physicochemical property prediction.

## INTRODUCTION

A rare cancer known as sarcoma occurs in connective tissues such as bones, muscles, tendons, cartilage, fat, nerves, and blood vessels. Sarcomas influence the body's framework, while carcinomas arise from epithelial tissues. Soft tissue and bone sarcomas, also known as osteosarcomas, are the two main malignancies. In the United States, around 13,190 annual diagnoses of soft tissue sarcoma and 800 to 900 cases of bone sarcomas are recorded. Often, complex theoretical frameworks emerge,

highlighting concern regarding the physicochemical properties of the tissues and possible therapeutic approaches.

Chemical graph theory is an interdisciplinary mathematical discipline employing graph theory concepts intersecting with chemistry to represent molecules and their structures visually. A chemical graph depicts a molecule with edges and bonds. Vertices represent atoms, while bonds round off the relationship between two atoms. TIs assist in QSPR analysis by bridging the gap between molecular



topology and physicochemical properties. These numerical descriptors outline constituent atoms and bonds, hence becoming paramount in describing the structure of a molecule.

This research uses degree-based topological indices to study significant physicochemical properties of sarcoma cancer tissues. Some primary tasks are:

- Constructing methodology for estimating tissue characteristics like flash point, enthalpy, molar volume, and refractive index.
- Determining molecular attributes by deriving and calculating three degree-based topological indices: the Albertson, Sigma, and Forgotten.
- Employing linear regression models to estimate the relationship between computed indices and experimental values.

This research proposes a topological indices-based method for predicting molecular properties that are financially likely to aid in advancing cancer research and pharmaceutical research and development. The study integrates computational chemistry, bioinformatics, and medicinal chemistry by providing a foundational study into the theoretical characterization of the molecular structure of sarcoma tissues and prospective anti-cancer drugs.

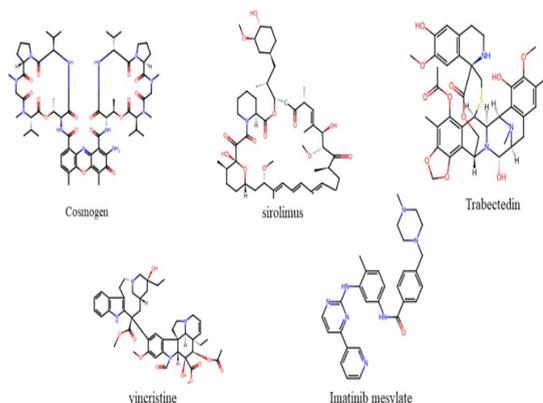


Fig. 1. Structure of Sarcoma Cancer Tissues

## Methodology

### Step 1: Data Collection

This research targets the study of the physicochemical characteristics of sarcoma cancer tissues. The characteristics studied are flash point, enthalpy, molar volume and refractive index. These values are obtained from ChemSpider<sup>9</sup> and other relevant chemical databases, where they have been obtained through extensive research. The objective is to develop a model capable of predicting these features

and thus reduce the need for laboratory measurements that are expensive and time-consuming.

### Step 2: Computation of Topological Indices

Topological Indices are mathematical descriptors used to characterize the structure of a molecule by looking at its connectivity. In this research, the edge partition method is utilized to determine three degree-based topological indices. The Albertson index<sup>10</sup> calculates the difference between vertex degrees in a molecule, and the Sigma index<sup>11</sup> measures the connectivity of a structure from its topology. Forgotten index<sup>12</sup> improves property predictions for molecular graphs by accounting for contributions from edges of higher degree vertices. The indices are calculated based on molecular graphs in which atoms are represented as vertices and chemical bonds as edges, which makes computation easier.

$$\text{Albertson} = \sum_{uv \in (G)} |\text{deg}(u) - \text{deg}(v)|$$

$$\text{Sigma} = \sum_{uv \in (G)} [\text{deg}(u) - \text{deg}(v)]^2$$

$$\text{Forgotten} = \sum_{uv \in (G)} [\text{deg}(u)^2 + \text{deg}(v)^2]$$

### Step 3: Statistical Analysis and Model Development

To evaluate the predictive efficiency of the computed topological indices, linear regression models are employed<sup>13,14</sup>. The general regression equation used is:

$$P = A + B \times \text{TI}$$

Where P represents the physicochemical property of the tissue, A is a constant (Y-intercept), B is the regression coefficient, and TI is the corresponding topological index. Statistical parameters such as the correlation coefficient (r), coefficient of determination (r<sup>2</sup>), F-statistic, and p-value are computed to assess the significance of the predictive models. SPSS software or Ms Excel is used to analyze the accuracy and reliability of the models by determining the correlation between computed indices and experimental values<sup>15</sup>.

### Step 4: Model Validation

The reliability of QSPR models yielded results from evaluating three statistical criteria. To begin with, the correlation values (r) must be above 0.7, which indicates strong dependence on the predicted and experimental values. Secondly, the models will require an F-statistic value greater than 2.5 to ensure sufficient statistical significance.

Lastly, p-values must be less than 0.05 to strengthen support for the stated predictions.

Integrating chemical graph theory with mathematical modelling and statistical regression allows for an adequate prediction of the physicochemical characteristics of sarcoma cancer tissues. It offers computational bioinformatics, computational chemistry and even medicinal chemistry a fresh understanding of molecular dynamics alongside their bioactive compounds.

## RESULTS AND DISCUSSION

This study leveraged degree-based topological indices to estimate the physicochemical characteristics of sarcoma cancer tissues, such as flash point temperature, enthalpy, molar volume, and refractive index. The descriptors used were the Albertson index, Sigma index, and Forgotten index, and their correlation with the experimental values was examined through regression analysis.

### Regression Models for Physicochemical Properties

The linear regression equations for each topological index and physicochemical property are: Albertson Index-Based Regression:

$$FP = -581.4 + 14.33 [\text{Albertson}]$$

$$H = -156.6 + 3.903 [\text{Albertson}]$$

$$MV = 146.8 + 6.207 [\text{Albertson}]$$

$$RI = 1.804 - 0.008151 [\text{Albertson}]$$

Sigma Index-Based Regression:

$$FP = -254.9 + 0.1364 [\text{sigma}]$$

$$H = -66.86 + 0.03692 [\text{sigma}]$$

$$MV = 269.3 + 0.06402 [\text{sigma}]$$

$$RI = 1.524 - 5.297e^{-5} [\text{sigma}]$$

Forgotten Index-Based Regression:

$$FP = -521.3 + 0.8608 [\text{forgotten}]$$

$$H = -133.6 + 0.2272 [\text{forgotten}]$$

$$MV = 153.4 + 0.3942 [\text{forgotten}]$$

$$RI = 1.639 - 0.0003469 [\text{forgotten}]$$

**Table 1: Physical properties of sarcoma tissues**

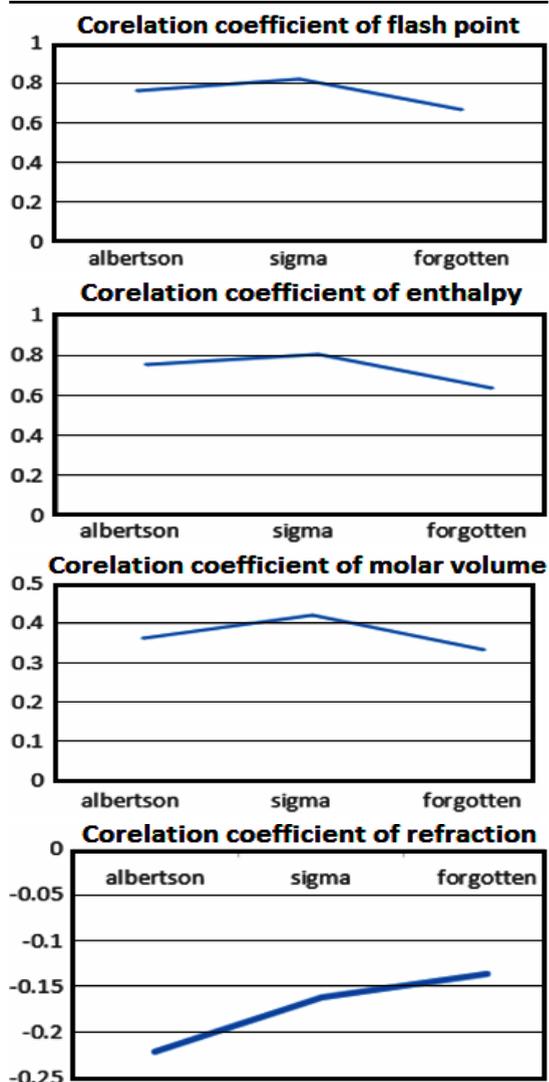
Drugs	Flash point	Enthalpy	Molar volume	Index of refraction
Cosmegen	792.1	211.5	880.7	1.656
Sirolimus	542.3	160.7	773.5	1.551
Imatinibmesylate	-	-	427.9	1.668
Trabectedin	-	-	489.3	1.732
Vincristine sulfate	-	-	-	-

**Table 2: Topological Values of Sarcoma Tissues**

Drugs	Albertson	Sigma	Forgotten
Cosmegen	83	6889	1319
Sirolimus	67	4489	859
Imatinibmesylate	30	900	510
Trabectedin	50	2500	904
Vincristine sulfate	66	4356	986

**Table 3: Correlation Coefficient of Physical Properties of Sarcoma Tissues**

TI	Flashpoint	Enthalpy	Molar Volume	Refraction
Albertson	0.76485	0.7567587	0.361808	-0.22058
Sigma	0.8195042	0.8060922	0.420235	-0.16142
Forgotten	0.6631511	0.635685	0.33163721	-0.13548



**Fig. 2. Graphical Representation of Correlations and Topological Indices**

**Table 5: Statistical Parameters for Linear QSPR Model For Albertson**

Physical Property	N	a	b	r	r <sup>2</sup>	F
Flash Point	2	-581.4	14.33	.76874	0.59096	5.77909
Enthalpy	2	-156.6	3.903	0.77080	0.59413	5.85549
Molar Volume	4	146.8	6.207	0.87499	0.76561	13.06628
Index of Refraction	4	1.804	-0.008151	0.82658	0.683234	8.627636

**Table 6: Statistical Parameters for Linear QSPR Model For Sigma**

Physical Property	N	A	B	r	r <sup>2</sup>	F
Flash Point	2	-254.9	0.1364	0.8494	0.72160	10.3681
Enthalpy	2	-66.86	0.03692	0.84747	0.71822	10.1955
Molar Volume	4	269.3	0.06402	0.85948	0.738718	11.30912
Index of Refraction	4	1.524	-5.297e-5	0.756481	0.572263	5.351545

**Table 7: Statistical Parameters for Linear QSPR Model For Forgotten**

Physical Property	N	a	B	r	r <sup>2</sup>	F
Flash Point	2	-521.3	0.8608	0.7395	0.54699	4.8298
Enthalpy	2	-133.6	0.2272	0.73784	0.5444	4.7798
Molar Volume	4	153.4	0.3942	0.87195	0.76029	12.68747
Index of Refraction	4	1.639	-0.0003469	0.8434	0.71137	9.859009

**Table 8: Standard Error of Estimation**

Topological Index	Flash Point	Enthalpy	Molar Volume	Index of Refraction
Albertson	82.14775	21.22301	99.668619	9.993913
Sigma	725.74203	746.0935	695.1484	755.0358
Forgotten	139.75043	146.5354	137.69938	155.7338

**Table 9: Correlation Determination**

Topological Index	Flash Point	Enthalpy	Molar Volume	Index of Refraction
Albertson	0.5909641	0.594135	0.7656197	0.683234
Sigma	0.7216062	0.7182213	0.738718	0.572263
forgotten	0.5469931	0.544413	0.76029932	0.711379

Our goal was to estimate the predictive ability of degree-based indices concerning sarcoma cancer tissues using the topological indices computation and their comparison with correlation coefficients of selected physical properties. The calculated values of the Albertson index, Sigma index, and Forgotten index, which stem from the molecular graphs of the tissues, are detailed in Table 2 alongside the physicochemical properties-flash point, enthalpy, molar volume, and refractive index-of five sampled tissues in Table 1. Table 3 presents correlation coefficients relating to the four physicochemical properties and each topological index, demonstrating the extent of the relation between structural connectivity and experimental values. Fig. 2 is the graphical representation of the correlations and topological

indices that visually examine those relations.

In order to assess the predictive power given by the models, different statistical parameters were computed: the sample size (N), constant (A), slope coefficient (B), estimate correlation (r), determination (r<sup>2</sup>), F-statistics, and p-value. The correlation coefficient (r) quantifies the degree of the relationship between the computed indices and the experimental values, while r<sup>2</sup> measures the portion of change due to the model. The F-statistic determines the importance of regression models statistically, where greater than (>2.5) marks strong relationships within the data. The p-value gives the overview of whether the relationships observed are significant, where, in this case, lower than 0.05 is defined as strongly proportional relationships from the

topological indices to physicochemical properties. A low p-value strengthens the hypothesis that the so-called independent variables (topological indices) are related significantly to the so-called dependent variables (physicochemical properties); reversing this logical test shows weak p-values. Inferences drawn from referenced Tables 4-17 prove that more than 70% of the models computed are correlation validated, which gave more than 0.7 values correlated with many F-statistics and significantly low p-values. These findings support that degree-based topological indices are efficient and aid in estimating molecular properties by minimizing experimental validation procedures in computational chemistry and biomedical research.

### CONCLUSION

With the aid of degree-based topological indices, this study was able to predict the physicochemical properties of the sarcoma cancer tissues. Important correlations regarding the molecular structure and critical physical attributes such as the flash point, enthalpy, molar volume, and refractive index were formed through the research using the Albertson index, Sigma index, and Forgotten index. It was exhibited that these indices can serve as efficient numerical descriptors for molecular topology and offer computationally simple and inexpensive means of analyzing intricate biological structures.

The proposed models were validated from a statistical viewpoint. They were proven reliable, with correlation coefficients ( $r$ ) above 0.7, F-statistic values of more than 2.5, and p-values lower than

0.05. These values endorse strong predictive abilities. The Sigma index had the highest correlation with the experimental data among the three indices. Thus, making it the best predictor out of the three indices. The results emphasize the importance of topological indices in QSPR modelling, showcasing the reduction in expensive and time-consuming experimental testing.

This study aids in computational chemistry and bioinformatics, offering a proposed model for estimating the cancerous tissues' physicochemical properties. Topological indices in pharmaceutical modelling may facilitate the preliminary detection of anticancer agents, especially in poor regions where experimental confirmation is hard to obtain. Subsequent research may include more molecular descriptors and add validation against larger datasets of cancer compounds.

In summary, employing topological indices simplifies and accelerates the calculation of structural features of sarcoma tissues and their corresponding chemicals, enabling more straightforward and less expensive technologies for potential application. This marks a novel direction for advancement in computational oncology and drug discovery.

### ACKNOWLEDGMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Conflict of interest

The author declare that we have no conflict of interest.

### REFERENCES

1. Sinha, S.; Peach, A. H. S. Diagnosis and Management of Soft Tissue Sarcoma., *BMJ.*, **2011**, *342*(7789), 157–162. <https://doi.org/10.1136/bmj.c7170>.
2. Vibert, J.; Watson, S. The Molecular Biology of Soft Tissue Sarcomas: Current Knowledge and Future Perspectives., *Cancers (Basel)*., **2022**, *14*(10). <https://doi.org/10.3390/cancers14102548>.
3. Trinajstić, N. Chemical Graph Theory., *Chem. Graph Theory.*, **2018**. <https://doi.org/10.1201/9781315139111>.
4. Roy, K.; Kar, S.; Das, R. N. Statistical Methods in QSAR/QSPR., *In Statistical Methods in QSAR/QSPR.*, **2015**, 37–59. [https://doi.org/10.1007/978-3-319-17281-1\\_2](https://doi.org/10.1007/978-3-319-17281-1_2).
5. Havare, O. C. QSPR Analysis with Curvilinear Regression Modeling and Topological Indices., *Iran. J. Math. Chem.*, **2019**, *10*(4), 331–341. <https://doi.org/10.22052/ijmc.2019.191865.1448>.
6. Shi, X.; Kosari, S.; Ghods, M.; Kheirkhahan, N. Innovative Approaches in QSPR Modelling Using Topological Indices for the Development of Cancer Treatments., *PLoS One.*, **2025**, *20*(2), e0317507. <https://doi.org/10.1371/journal.pone.0317507>.

7. Dearden, J. C. The Use of Topological Indices in QSAR and QSPR Modeling., *In Challenges and Advances in Computational Chemistry and Physics.*, **2017**, *24*, 57–88. [https://doi.org/10.1007/978-3-319-56850-8\\_2](https://doi.org/10.1007/978-3-319-56850-8_2).
8. Mahboob, A.; Rasheed, M. W.; Hanif, I.; Amin, L.; Alameri, A. Role of Molecular Descriptors in Quantitative Structure-Property Relationship Analysis of Kidney Cancer Therapeutics., *Int. J. Quantum Chem.*, **2024**, *124*(1). <https://doi.org/10.1002/qua.27241>.
9. ChemSpider. ChemSpider Database., Retrieved from <http://www.chemspider.com/>.
10. Lin, Z.; Zhou, T.; Wang, X.; Miao, L. The General Albertson Irregularity Index of Graphs., *AIMS Math.*, **2022**, *7*(1), 25–38. <https://doi.org/10.3934/math.2022002>.
11. Jahanbani, A.; Ediz, S. The Sigma Index of Graph Operations., *Sigma J. Eng. Nat. Sci.*, **2019**, *37*(1), 155–162.
12. Bharali, A.; Doley, A.; Buragohain, J. Entire Forgotten Topological Index of Graphs., *Proyecciones.*, **2020**, *39*(4), 1019–1032. <https://doi.org/10.22199/issn.0717-6279-2020-04-0064>.
13. Schneider, A.; Hommel, G.; Blettner, M. Lineare Regressionsanalyse-Teil 14 der Bewertung Wissenschaftlicher Publikationen., *Dtsch., Ärztebl.*, **2010**, *107*(44), 776–782. <https://doi.org/10.3238/arztebl.2010.0776>.
14. Sun, Y.; Wang, X.; Zhang, C.; Zuo, M. Multiple Regression: Methodology and Applications., *Highlights Sci. Eng. Technol.*, **2023**, *49*, 542–548. <https://doi.org/10.54097/hset.v49i.8611>.
15. Frey, F. SPSS (Software)., *In Int. Encycl. Commun. Res. Methods.*, **2017**, 1–2. <https://doi.org/10.1002/9781118901731.iecrm0237>.