ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
https://www.icf.ro/rrch/

Rev. Roum. Chim.,
2026, *71*(5–6), 317–324
DOI: 10.33224/rrch.2026.71.5-6.02

PAPERS

Dedicated to Professor Alexandru T. Balaban on the occasion of his 95th anniversary

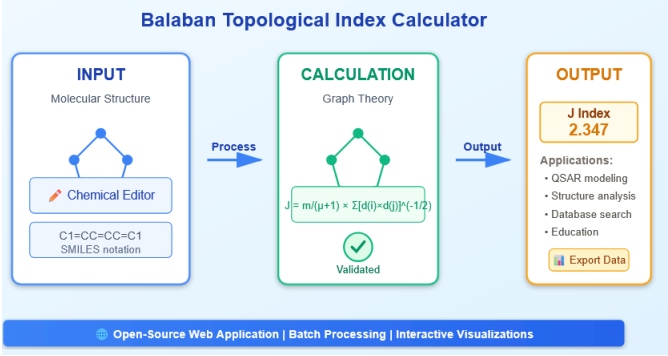
COMPUTATIONAL CHEMISTRY TOOL FOR CALCULATING THE BALABAN INDEX: DEVELOPMENT, VALIDATION, AND APPLICATIONS

Liliana HALIP, Nicoleta STOIAN and Sorin AVRAM*

Department of Computational Chemistry, "Coriolan Drăgulescu" Institute of Chemistry Timișoara, Roumanian Academy, 24th Mihai Viteazu Avenue, Timișoara, 300223, Roumania

Received November 14, 2025

The Balaban Index (***J***) is a complex molecular descriptor that quantifies structural complexity based on graph-theoretical principles. Despite its demonstrated utility in structure-property relationship studies, computational tools for routine calculation of this index are few and mostly not open-source. We present a publicly available web-based application that enables facile calculation of the Balaban Index for organic molecules. The application features an integrated chemical drawing editor, supports SMILES notation input, and provides batch processing capabilities for multiple structures. Validation using benchmark compounds from the original paper confirmed computational accuracy and demonstrated the sensitivity of the index to molecular branching, ring systems, and overall topological complexity. Interactive correlations of the Balaban Index with structural parameters are available and results can be exported in standard formats compatible with computational chemistry workflows. This freely accessible tool facilitates the use of the Balaban Index in QSAR modeling, structure elucidation, database searching, and chemical education (https://avramlab.shinyapps.io/BalabanIndexCalculator/).



INTRODUCTION

Computational chemistry has become an important component of modern chemical research, providing efficient methods for analyzing and predicting molecular properties, and thereby saving time and material resources by guiding experimental determinations. In this light, topological indices – numerical parameters derived from the molecular graph – play a central role in QSAR (Quantitative Structure–Activity Relationship) and QSPR (Quantitative Structure-Property Relationship) modeling.[1–3] Among the multitude of topological descriptors available today, the Balaban Index (***J***), introduced by Professor Alexandru T. Balaban, stands out for its robustness and discriminative power.[4] The ***J*** index captures the topology of a molecular graph by integrating both atomic connectivity and interatomic distances, thereby providing a sensitive measure of molecular complexity and structural branching.[4] Since its introduction, the Balaban Index has found broad application in the correlation of physicochemical properties, modeling of

* Corresponding author: 5orin.4vram@gmail.com

biological activities, and design of functional materials. Its ability to discriminate between structural isomers and to reflect subtle changes in molecular topology has made it one of the most informative and widely used graph-based descriptors.[3,5]

The Balaban Index: Definition, Calculation, and Chemical Significance

In chemical graph theory, a molecule is represented as a graph where vertices correspond to atoms and edges correspond to covalent bonds. The Balaban Index is defined based on the molecular distance matrix and vertex degrees, combining information on both the shortest path lengths between atoms and their local connectivity. This dual dependence enables ***J*** to reflect the degree of molecular branching as well as the overall shape of the molecular framework. Compared with other topological indices such as the Wiener index (which depends on pairwise distances)[2,3,6] or the Randić index (which emphasizes local connectivity),[2,3,7] the Balaban Index offers a more balanced and comprehensive representation of molecular topology. For example, in the series of hexane isomers (C_6H_{14}), the Balaban Index systematically increases with molecular branching, in agreement with observed trends in boiling points and other physical properties.[8]

Due to its significant correlation with many molecular properties, the Balaban Index continues to play a pivotal role in computational chemistry, cheminformatics and molecular design. Its conceptual elegance and enduring utility exemplify the deep connection between chemistry and mathematics – a connection that Professor Alexandru T. Balaban advanced throughout his scientific career.

Despite its conceptual simplicity, the practical calculation of the Balaban Index remains non-trivial, particularly for users without access to specialized cheminformatics software.[3,9] To our knowledge, a similar webtool developed by the Department of Mathematics of Gordon College[10] also includes calculation of Balaban Index together with other topological parameters, but its limitation to non-aromatic compounds restricts its general utility. To address the limitations of existing methods, the present work introduces a freely accessible online computational chemistry tool that automates and performs the rapid, accurate calculation of the Balaban Index for single molecules or larger datasets. The application features an intuitive interface, allowing users to easily compute and visualize index values directly from molecular structures in standard chemical formats. This tool is designed to facilitate both research and teaching.

MATERIALS AND METHODS

Balaban Index Calculation

The Balaban Index, ***J***, was computed following the original method proposed by Balaban.[4] Each chemical structure was represented as a hydrogen-suppressed graph, where atoms are vertices and chemical bonds are edges. The adjacency matrix was first generated to identify directly connected atoms. From this, a weighted distance matrix was constructed using the Floyd-Warshall algorithm[11] to determine the shortest paths between all vertex pairs. For unsaturated and aromatic systems, fractional edge weights were assigned to account for bond multiplicity and delocalization: 1 for single bonds, 1/2 for double bonds, 1/3 for triple bonds, and 2/3 for aromatic bonds, following Balaban's original specification. For each vertex i, the distance sum (s_i) was obtained by summing all shortest path lengths from i to the remaining vertices. The Balaban index was then evaluated using the original formula[4] shown in equation (1):

$$J = \frac{q}{\mu+1}\sum_{(i,j)} \frac{1}{\sqrt{S_i S_j}} \quad (1)$$

where q is the number of bonds, μ is the cyclomatic number ($\mu = q - n + 1$, with n being the number of atoms), and Si represents the sum of topological distances from atom i to all other atoms in the molecular graph. The summation is performed over all bonds in the structure.

Computational Procedure

Chemical structures were input as SMILES notation[12] either through an integrated molecular editor (JSME 2021–12–26[13]) or direct text entry. The computational workflow presented in Fig. 1, consists of: (1) reading chemical structures in SMILES format to automatically generate chemical graphs using the Chemistry Development Kit[14] (CDK, version 2.8), (2) generation of the adjacency matrix A where $A_{ij} = 1$ for bonded atoms, (3) calculation of the topological distance matrix D using the Floyd-Warshall algorithm, and (4) computation of the Balaban Index. Although the Kekulé structures are accepted, this input format is not recommended for aromatic compounds. When the Kekulé structure is provided, aromaticity is not correctly interpreted and an inaccurate value of ***J*** may be computed. All calculations were implemented in R (version ≥4.3.0) using the "rcdk" package [15] (version 3.8.1). The web-based application was developed using the Shiny framework[16] to provide a user-friendly

interface and ensure accessibility without requiring programming expertise.

Molecular Test Sets

The Balaban Index values calculated by ***J*** IndexCalc chemistry computational tool were validated against established reference data from two independent sources: 50 chemical structures described in the original algorithm publication[4] and 71 additional compounds from the Gordon College web-page.[10] This comprehensive approach ensured the accuracy and reliability of the implementation.

To illustrate the utility of the calculated Balaban Index, two distinct datasets were compiled from existing literature[17] and the DrugCentral database.[18] The first, a Physical Property Correlation Dataset, was constructed using SMILES codes for octane and its isomers extracted from PubChem[19] with corresponding boiling points sourced from both PubChem and literature.[17] The second, a Bioactive Compounds Dataset, containing chemical information, SMILES codes for 4,000 biologically active compounds extracted from DrugCentral[18] database, supplemented with pre-calculated property data also provided by the same source. Our web tool took only 2–3 minutes to compute the ***J*** index for 1,000 compounds and 10 minutes for the full 4,000 data set, but future work will aim to improve scalability to support much larger chemical libraries.

Data Analysis and Visualization

Correlation analyses between the Balaban Index and structural parameters (number of atoms, bonds, cyclomatic number, etc) were performed using standard linear regression. Statistical graphics were generated using the "ggplot2" package available in R statistical programming language.[20] Batch processing capabilities allowed simultaneous analysis of multiple structures with results exportable in standard formats (CSV, Excel).

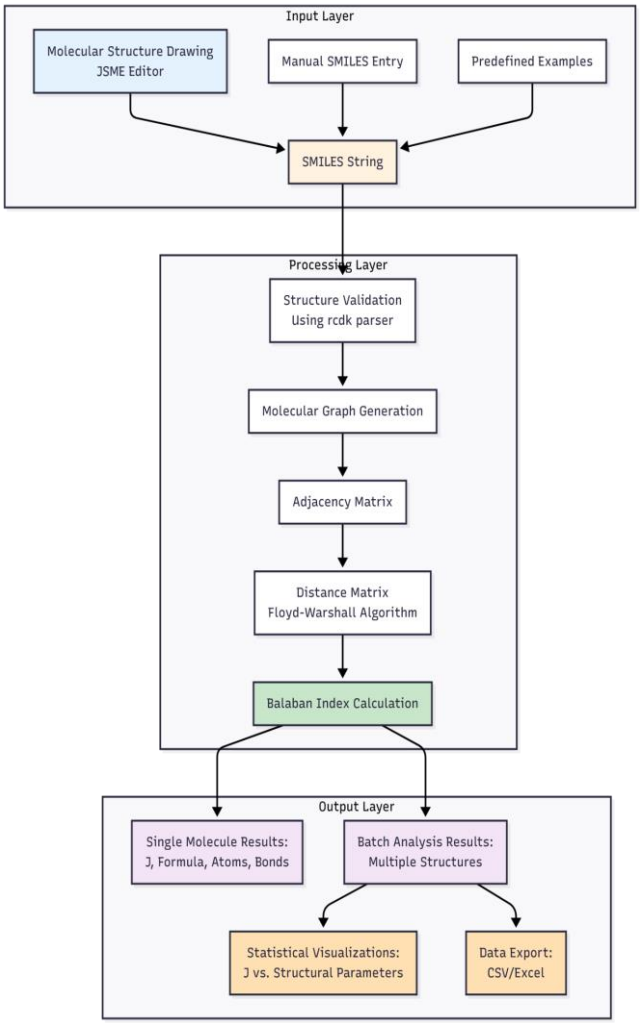


Fig. 1 – The architecture of ***J*** IndexCalc chemistry computational tool.

RESULTS AND DISCUSSION

Application Development and Functionality

The ***J***IndexCalc computational chemistry tool was developed to facilitate the calculation and analysis of the Balaban Index and related molecular descriptors for organic molecules. The application provides an intuitive interface that requires no programming knowledge, making it accessible to chemists for routine use in structural characterization and structure-property relationship studies (Fig. 2).

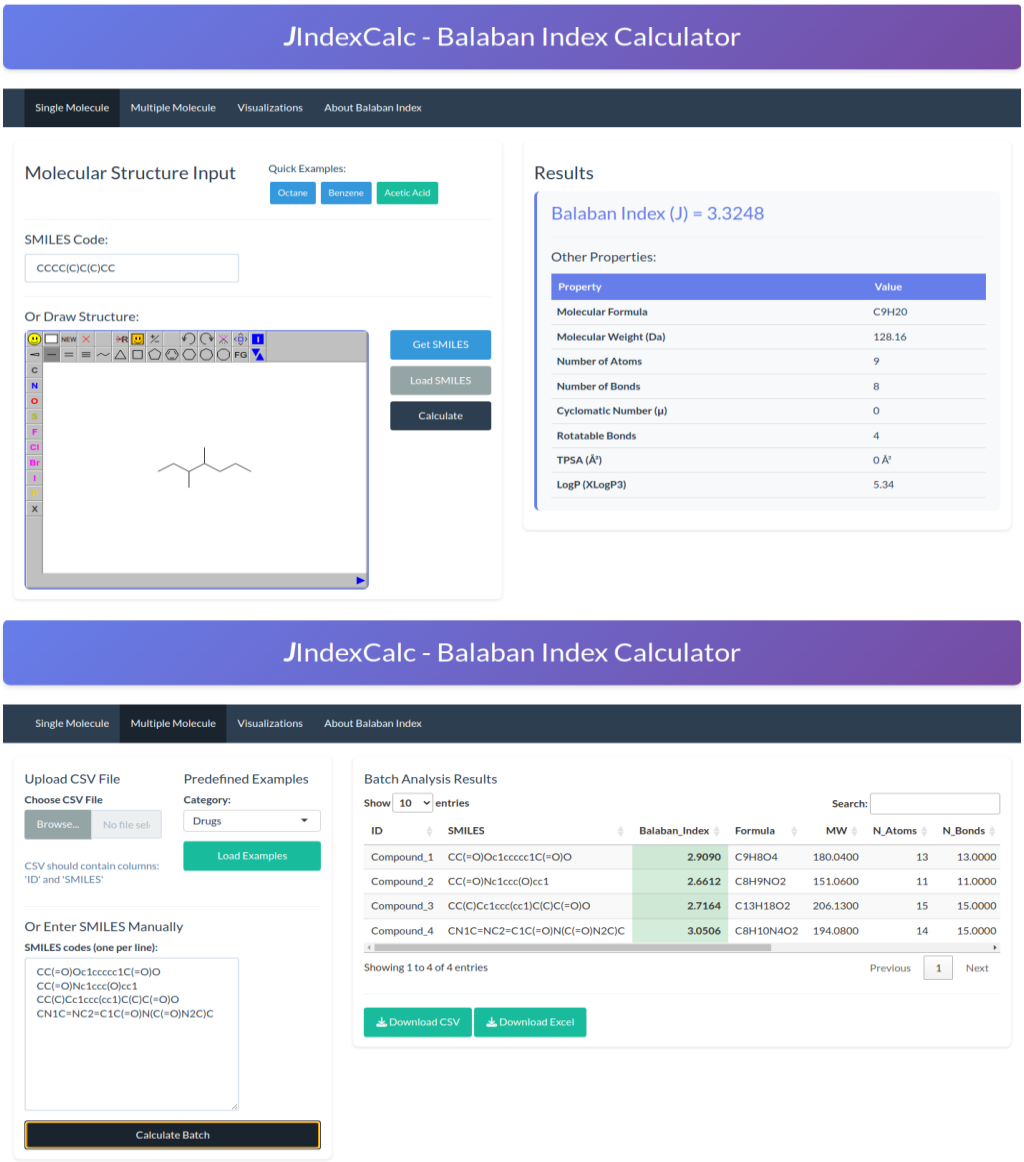


Fig. 2 – ***J***IndexCalc interface of the Balaban Index calculator. Top: Screenshot of the Single Molecule tab, used for immediate calculation and visualization of single molecule. Bottom: Screenshot of the Multiple Molecule tab, designed for batch processing of large chemical dataset.

Chemical structures are read through three input methods: (1) interactive drawing using a graphical molecular editor, (2) direct entry of SMILES notation, or (3) selection from a library of common examples. This flexibility accommodates users with varying levels of familiarity with chemical informatics while maintaining computational rigor. The integrated molecular editor allows users to construct structures atom-by-atom and bond-by-bond, with automatic conversion to machine-readable format for subsequent calculations. ***J***IndexCalc consists of following modules: i) *Single Molecule* which allows instantly calculation of Balaban

Index and molecular properties for a single molecule; ii) *Multiple Molecule* which enables simultaneous processing of multiple structures

through CSV file upload, manual input, or selection from predefined datasets (e.g., alkanes, aromatic compounds, or drugs). Results can be exported in CSV or Excel format for further analysis; iii) interactive ploting which allows comparative and exploratory analysis, including correlation plots between the Balaban Index and various molecular properties, multidimensional scatter plots with color and size encoding based on ***J*** values, and distribution histograms.

Validation with Benchmark Molecules

To ensure the accuracy and reproducibility of the application, the ***J*** index was recalculated against two distinct reference sets. The primary reference set comprised the 50 compounds originally analyzed by Balaban,[4] specifically including 22 alkanes (C_2–C_7) and 28 alkylcyclohexanes and alkylbenzenes (C_6–C_9). A secondary verification dataset consisting of 71 additional compounds, ranging in size from C_3–C_{90} atoms, was used for a broader assessment of ***J*** index calculation accuracy.[10] The ***J*** index values computed using the ***J***IndexCalc application demonstrated excellent agreement with the reported literature data. A strong linear correlation (r^2 = 0.998, $p < 0.001$) was observed between the calculated and reference ***J*** indices (Fig. 3), confirming the robustness and high reproducibility of the implemented algorithm.

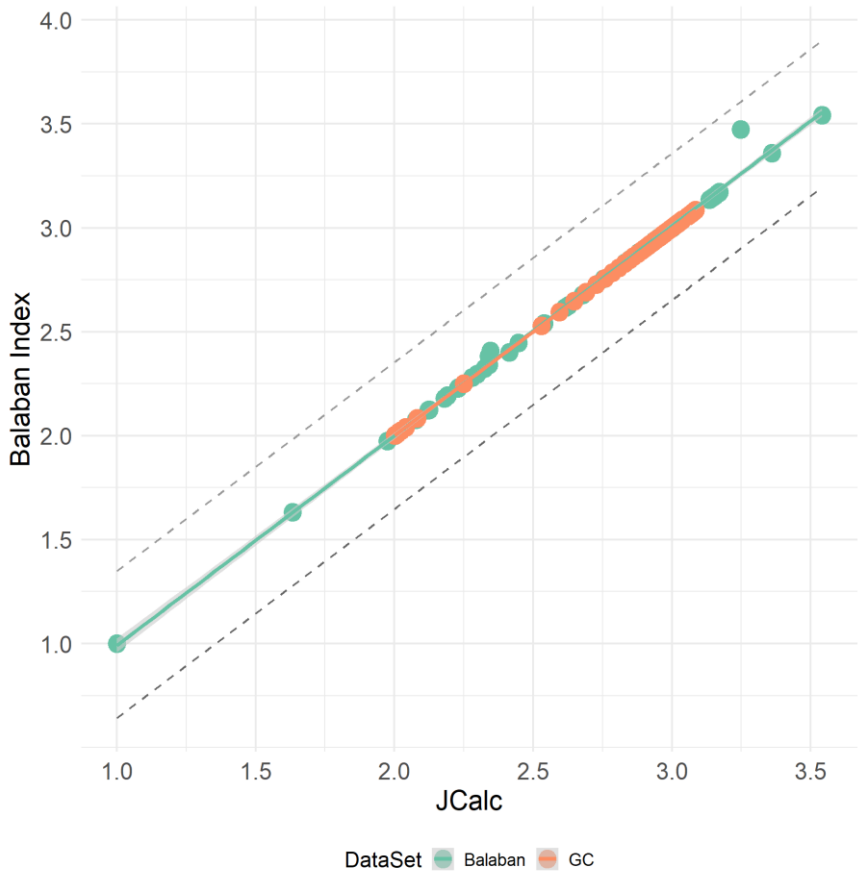


Fig. 3 – Validation of the Balaban Index Calculation. Correlation between the Balaban Index values calculated by ***J***IndexCalc and reference values obtained from two sources: the original publication (green)[4] and the Gordon College (GC) web page (orange).[10]

Potential application of JIndexCalc in chemistry research

The developed chemistry computational tool, which enables rapid calculation of the Balaban Index, offers utility across several domains of chemical research and education. In structure elucidation, quick access to values aids in distinguishing between constitutional isomers when combined with spectroscopic data. For QSAR modeling, the index serves as a valuable independent variable in regression models used to predict physicochemical properties (*e.g.*, boiling point, solubility, density, etc) or biological activities, particularly within congeneric series where topological complexity is the main key.

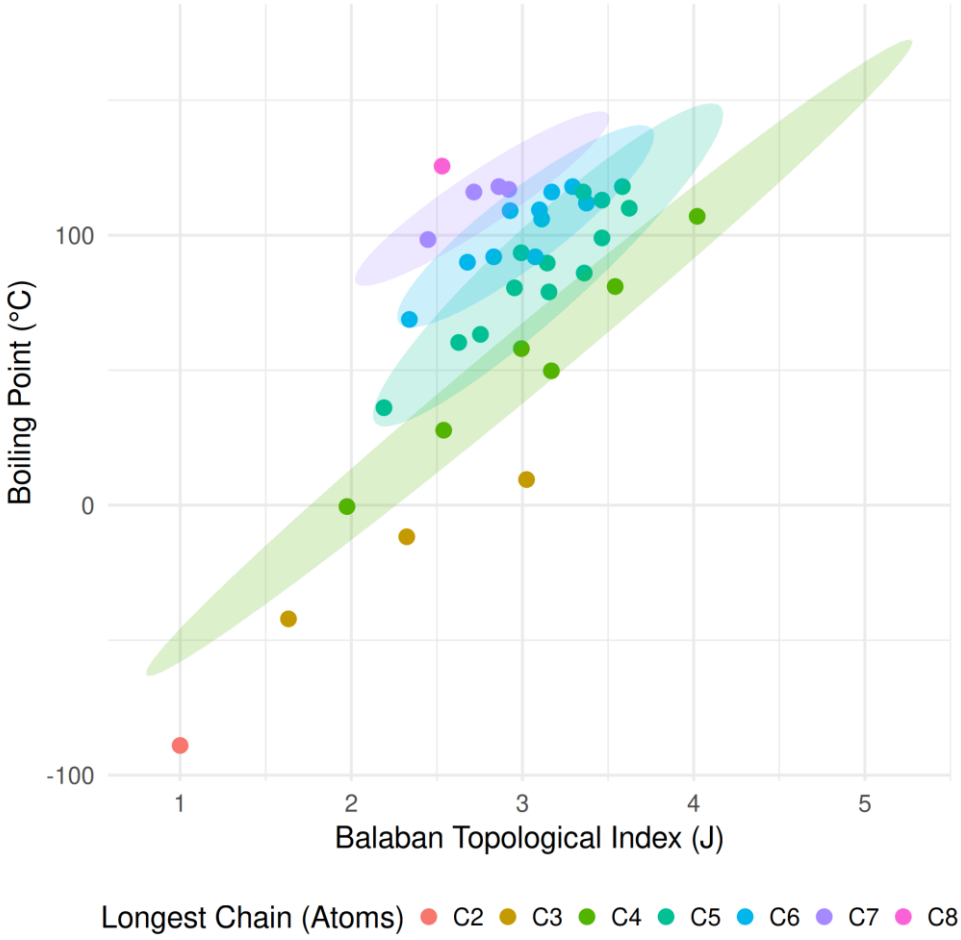


Fig. 4 – Relationship between boiling point and Balaban Index.

Figure 4 shows the relationship between the Balaban Index and the boiling points of alkanes ranging from C_2 to C_8, including their branched isomers. Each colored cluster represents chemical compounds with the same number of carbon atoms in their longest carbon chain; for example, n-heptane and 2-methylheptane are both classified as C_7. Within each cluster, an increase in the Balaban Index corresponds to greater molecular branching, which is accompanied by a decrease in boiling point due to reduced surface area and weaker intermolecular dispersion forces. Across clusters, as the total carbon count increases, the boiling points rise overall, reflecting stronger dispersion interactions in larger molecules. Together, these trends illustrate how molecular size and topology jointly influence the physical properties of alkanes.

The Balaban Index effectively differentiates molecular topologies across common hydrocarbon classes. For acyclic systems, ***J*** increases with branching as shorter average atom–atom paths arise in more compact frameworks. Ring closure, as in cycloalkanes, introduces multiple alternative paths and slightly elevates ***J*** relative to open chains of comparable size. Aromatic systems yield the highest ***J*** values, reflecting their delocalized π networks and dense connectivity. Collectively, the sequence $J_{linear} < J_{branched} < J_{cyclic} < J_{aromatic}$ parallels the chemical intuition of increasing structural compactness and connectivity.

Application to drug-like compounds

The Balaban Index, a well-established topological descriptor widely applied in molecular structure analysis, demonstrates strong potential for characterizing sets of biologically active compounds. Analysis of a dataset comprising over 4,000 approved drugs retrieved from the DrugCentral database revealed significant correlations between the Balaban Index and several key physicochemical properties, including molecular weight, structural rigidity, aqueous solubility, and lipophilicity (cLogP).

A negative Spearman correlation ($\rho = -0.66$) was observed between molecular weight and the Balaban Index, indicating that compounds with higher molecular mass generally exhibit lower topological complexity. This inverse relationship is consistent with the observation that rigid molecular frameworks tend to display higher Balaban Index values, whereas flexible or highly branched molecules – associated with greater conformational freedom – yield lower index values. Although the separation between rigid and flexible structures is less pronounced than for molecular weight, the overall trend remains consistent across the dataset.

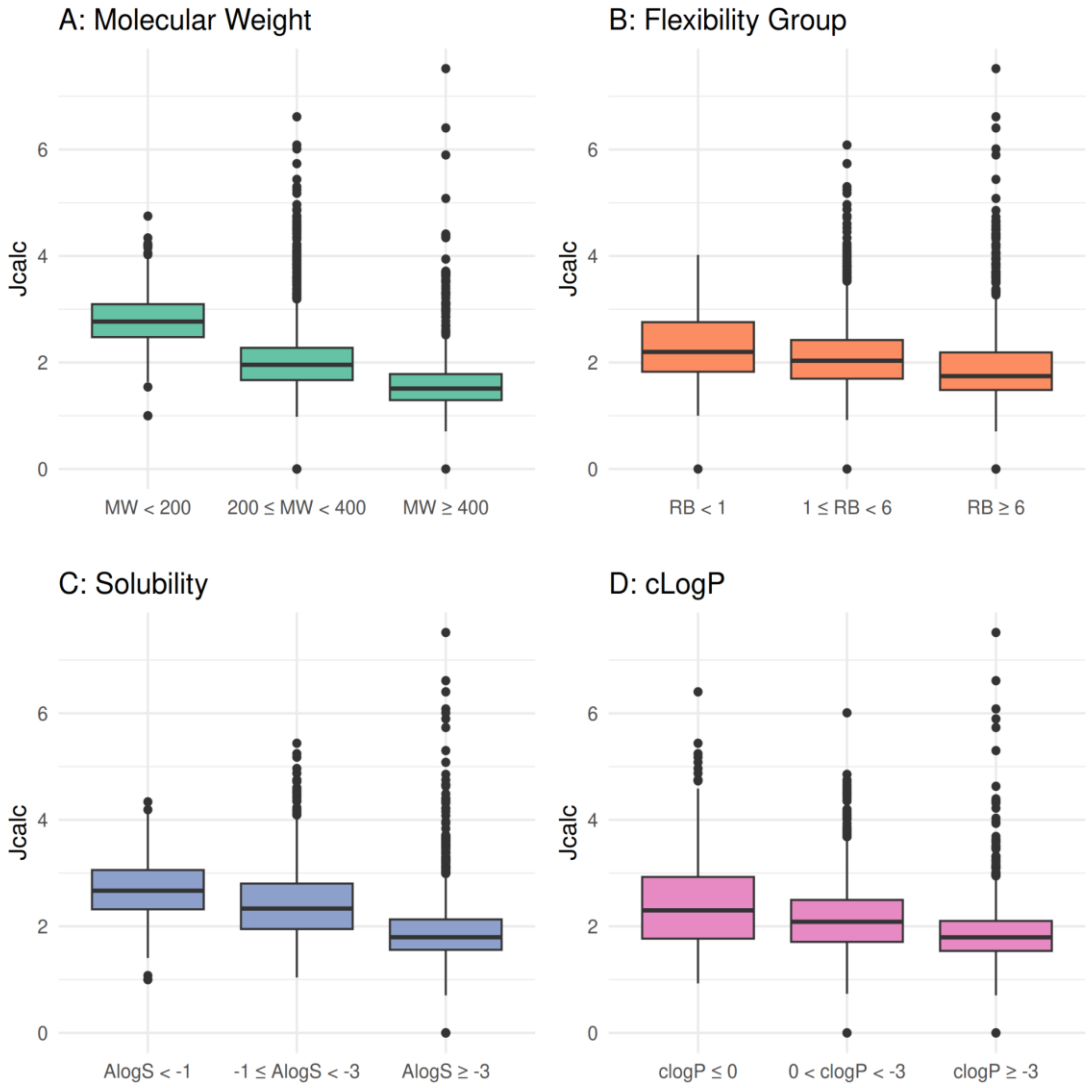


Fig. 5 – Distribution of calculated Balaban Index values across molecular property groups. (A) Molecular weight, (B) flexibility, (C) solubility, and (D) lipophilicity.

A positive correlation between aqueous solubility and the Balaban Index ($\rho = 0.55$) further suggests that more soluble compounds typically possess compact and highly connected topologies, while poorly soluble molecules display more dispersed or extended architectures, reflected in smaller index values.

In addition, a clear trend was identified for cLogP values: hydrophilic molecules exhibit higher Balaban Index values compared to lipophilic compounds. Notably, approximately 90% of lipophilic molecules show Balaban Index values lower than those of all hydrophilic compounds, reinforcing the association between topological compactness and molecular hydrophilicity. This observation aligns with the notion that denser connectivity and reduced branching facilitate stronger intermolecular interactions with polar solvents.

Collectively, these results highlight a consistent convergence between topological and physicochemical descriptors, emphasizing the relevance of the Balaban Index as a structural indicator for QSAR modeling, compound selection, and molecular profile interpretation in large-scale screening studies.

CONCLUSIONS

A freely accessible web-based application for calculating the Balaban Index has been developed and validated using structurally diverse organic compounds. The tool successfully combines computational rigor with user-friendly design, enabling chemists to calculate this important molecular descriptor without programming expertise. The ***J*** index can enhance chemical database searching by facilitating the retrieval of molecules with similar topological characteristics, and it guides drug design efforts by informing structural modifications during lead optimization. Finally, the intuitive graphical interface makes the tool an excellent instructional aid for chemical education, demonstrating concepts related to molecular graph theory.

REFERENCES

1. J. Gálvez, R. García-Domenech, J. V. de Julián-Ortiz and R. Soler, *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 272–284.
2. A. Mauri, V. Consonni and R. Todeschini in "Handbook of Computational Chemistry", Springer International Publishing, Cham, 2017, p. 2065–2093.
3. J. Devillers and A. T. Balaban (Eds.), "Topological indices and related descriptors in QSAR and QSPR", CRC Press, London, England, 2014, p. 811.
4. A. T. Balaban, *Chem. Phys. Lett.*, **1982**, *89*, 399–404.
5. E. Estrada, *Chem. Phys. Lett.*, **2000**, *319*, 713–718.
6. H. Wiener, *J. Am. Chem. Soc.*, **1947**, *69*, 17–20.
7. M. Randic, *J. Am. Chem. Soc.*, **1975**, *97*, 6609–6615.
8. A. T. Balaban, *J. Chem. Inf. Comput. Sci.*, **1985**, *25*, 334–343.
9. A. T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan in "Topics in Current Chemistry", Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, vol. 114, p. 21–55.
10. "S. Granz, "Topological Index Calculator", https://www.cs.gordon.edu/courses/organic/granz/, (accessed 5 November **2025**).
11. T. H. Cormen and C. E. Leiserson, "Introduction to Algorithms*"*, fourth edition, MIT Press, London, England, 2022, pp. 655–662.
12. D. Weininger, *J. Chem. Inf. Comput. Sci.*, **1988**, *28*, 31–36.
13. B. Bienfait and P. Ertl, *J. Cheminform.*, **2013**, *5*, 24.
14. E. L. Willighagen, J. W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliazkova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C. T. Evelo, R. Guha and C. Steinbeck, *J. Cheminform.*, **2017**, *9*, 33.
15. R. Guha, *J. Stat. Softw.*, **2007**, *18*, 1–16.
16. W. Chang, J. Cheng, J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert and B. Borges, *shiny: Web Application Framework for R. R package version*1.12.0.9000, **2025**.
17. D. R. Lide, "CRC handbook of chemistry and physics*"*, 84th edition, CRC Press, Boca Raton, FL, 84th edn., 2003, pp. 3554–3556.
18. S. Avram, T. B. Wilson, R. Curpan, L. Halip, A. Borota, A. Bora, C. G. Bologa, J. Holmes, J. Knockel, J. J. Yang and T. I. Oprea, *Nucleic Acids Res.*, **2023**, *51*, D1276–D1287.
19. S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, *Nucleic Acids Res.*, **2025**, *53*, D1516–D1525.
20. H. Wickham,W. Chang, L. Henry, T.L. Pedersen, K. Takahashi, C. Wilke, K. Woo, H. Yutani, D. Dunnington. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.* R Package Version 3.4.4, **2021**, https://CRAN.R-project.org/package=ggplot2