EPH-International Journal of Applied Science

Volume - 12, Issue - 03
July - 2026
DOI-https://doi.org/10.69980/wc92rx52

DECODING MOLECULAR PROPERTY SIGNATURES ASSOCIATED WITH CHEMICAL TOXICITY FOR SAFER COMPOUND SCREENING

Dr. Tapan Kumar Sankar[1], Dr. Kailash Chandra Das[2], Dr. Mukesh Kumar[3]

[1]Department of Chemistry, G H Raisoni University Amravati Maharashtra India . Email: tapan.gangotry@gmail.com, tapan.sankar@ghrua.edu.in
[2]School of Sciences, G H Raisoni University Amravati Maharashtra India
[3]Department of electrical engineering, G H Raisoni University Amravati Maharashtra India

Received:- 01-06-2026, Revised:- 02-07-2026, Accepted:- 10-07-2026, Published:- 17-07-2026

Abstract

Chemical toxicity screening is important in detecting potentially hazardous compounds prior to a detailed experimental evaluation. In this study, molecular property signatures related to chemical toxicity were analyzed, and their use as indicators of safer compound screening was assessed. The molecular descriptors of 171 compounds, divided into toxic and non-toxic, were subjected to statistical comparisons, correlation, principal component analysis and classification models. MDEC-23 exhibited the highest individual correlations with toxicity, while a few structural and physicochemical descriptors provided complementary discriminatory information. Principal component analysis revealed that the variation in the molecular structure was multidimensional with the two first eigenvectors accounting for 45.35% of the total variation. Random forest had higher classification accuracy (0.673) and ROC-AUC (0.659) than logistic regression, but sensitivity to toxic compound was not high enough. The results show that small molecular property signatures are useful as a initial filter of toxicity information for prioritization. This could be useful for more efficient early stage toxicity evaluation and compound selection for safer compounds when adequately validated externally.

Keywords: *Chemical toxicity, Molecular descriptors, Compound screening, Molecular signatures, Random forest*

1. Introduction

In the pharmaceutical, agricultural, industrial and environmental fields, chemical toxicity is a key concern in the design, screening and use of small molecules. Chemicals with harmful biological potential can disrupt the proper functioning of cells, the pathways through which cells communicate, the integrity of the genes, or the physiological responses of the organism, and may have significant human and environmental health impacts. Traditional method for toxicity assessment has relied on experimental assays and biological tests, but these can be time consuming, resource intensive and require laboratory space. Computational methods offer a complementary approach that is becoming increasingly useful, as the potential hazards of a molecule can be determined before a large number of experiments have been performed. The capacity to link molecular structure and physicochemical properties with biological activity and adverse effects has been further enhanced by advances in chemical biology (Tripathi, 2025). The relationships between them are thus important to grasp in developing efficient strategies that include prioritizing safer compounds in early screening.

Molecular descriptors are numeric representations of chemical properties, and allow a complex molecule to be converted to numbers for statistical and computational analysis. They can be used to describe molecular topology, electronic distribution, atomic composition, hydrogen bonding, lipophilicity, structural connectivity and other physicochemical properties. Systematic analysis of these can generate molecular signatures that hint at differences in behaviour among the biological systems. Conventional descriptors are often combined with machine-learning and multimodal methods to capture relationships that may not be apparent from individual chemical properties, as molecular property prediction becomes increasingly a part of modern research (Liyaqat et al., 2026). The combination of quantitative structure–activity relationships, molecular modeling, and artificial intelligence has also proven to be useful in screening of candidate compounds or evaluating them for drug discovery projects (Kurre et al., 2026). The capability to perform this kind of development highlights the increasing significance of using data-driven molecular characterization in the understanding of chemical safety.

Prediction of toxicity is especially difficult because the toxicity of a compound is not based on a single molecule attribute. Instead, it is possible that toxicity can arise from interactions between structural, physicochemical, and biological factors. A number of computational platforms have therefore been created to determine several toxicity endpoints from molecular data, showing that molecular signatures can assist in systematic evaluation of the safety of small molecules (de Sá et al., 2022). Additionally, machine-learning techniques have been demonstrated to classify the toxicity of a molecule by uncovering complex relationships between chemical properties that allow separation of hazardous molecules from safer alternatives (Galati et al., 2022). Recently, the capability has been extended to graph-based neural methods, which directly capture the structural representation from the molecular structure and identify patterns linked with toxicity (Gupta & Sharma, 2025). The advances suggest that the molecular information can supply an important foundation for preliminary toxicity screening.

Progress in the wider field of AI has increased the potential for insights into chemical–biological interactions. Chemical information combined with high content cellular imaging offers a better prediction of bioactivity and toxicity of small molecules by linking molecular attributes to phenotypic response (Seal, 2024). More broadly, chemical-induced adverse human health effects can be also linked with mechanisms by multi-omics and imaging-based computational techniques (Lejal, 2024). In the genomics-based toxicology approach, deep-learning techniques offer further possibilities for identifying complex biological patterns related to chemical effects (Akermi et al., 2025). Furthermore, artificial intelligence has become more relevant to genotoxicity assessment with computational models that can help to characterize molecular factors associated with DNA damage (Barghi et al., 2026). All of these developments point towards a shift from simple toxicity measurements to complex molecular patterns.

Identification of toxicity-associated signatures has practical applications beyond toxicity classification. Molecular information can be used to design and prioritize safer chemical candidates without putting time and money into costly downstream experiments. Graph-based signatures have already been shown to be useful for balancing potency and safety during herbicide design (Pires et

al., 2022) and molecular interaction analysis can be useful for assessing compounds linked to toxic occupational exposures (Ogbodo et al., 2025). Finally, generative chemistry also illustrates the ability to use computational molecular representations to help create compounds with several desired biological activities (Munson et al., 2024). The de novo molecular generation is now increasingly using machine-learning techniques to optimize the properties of the molecular candidate during the development process (Chen & Xue, 2026). Thus, the ability to reliably identify toxicity related properties can be a benefit not only for hazard recognition but also for compound prioritization towards safer compounds.

Although these developments are significant, high dimensional molecular data poses significant analysis challenges. Many times, there are redundant or correlated or less informative variables in large descriptor spaces and feature selection is crucial to obtain interpretable and generalizable toxicity signatures. The field of decoding molecular information has a broader scope that highlights the importance of computational approaches that can extract scientifically relevant information from increasingly complex molecular information (Metz et al., 2023). Therefore, it is important to identify a small number of informative molecular properties that can lower dimensionality while enhancing the interpretation of the chemical attributes that are correlated to toxicity. Avoiding the sole goal of predictive performance, the direction, magnitude, redundancy and multivariate contribution of molecular descriptors can be used to gain greater insight into why toxic and non-toxic compounds have different molecular profiles.

Hence, statistical, multivariate and classification-based analyses for molecular property signatures related to chemical toxicity are investigated in this study. This study aims to find informative descriptors, investigate molecular patterns and assess their ability to discriminate toxic from non-toxic compounds. This holistic approach will offer a readily interpretable framework and a basis for translating high-dimensional molecular information into evidence that can be used in the initial stages of chemical safety assessment. The specific objectives were:

1.To characterize molecular property patterns and determine their associations with toxic and non-toxic compound classifications.

2.To identify informative and non-redundant molecular descriptors that constitute interpretable signatures associated with chemical toxicity.

3.To evaluate the discriminatory performance of selected molecular signatures for supporting safer preliminary compound screening.

2. Materials and Methods

2.1 Research Design and Dataset

To study molecular property signatures related to chemical toxicity, a quantitative analytical design was used (Gül & Rahim, 2021). The database contained 171 chemical compounds, with both toxic and non-toxic classes, represented by molecular descriptors. Descriptor matrix included physicochemical, structural, electronic and topological information relevant to the behaviour of the compounds. The primary outcome variable used was toxicity class. The analytical framework was aimed at identifying measurable molecular properties that were different among toxicity groups and whether a small number of informative descriptors could be used as a basis for preliminary compound classification for toxicity.

2.2 Data Preprocessing and Molecular Characterization

Data was first screened for missing and duplicated data, type of variables and variations in the descriptions. Those molecular descriptors that varied little were not included as they provided little discriminatory information. Where necessary, continuous variables were standardized to be comparable across the different scales of measurement. Descriptive statistics for molecular properties were then computed to describe the distribution of molecular properties for both the toxic and non-toxic groups. Distributional patterns were also investigated to choose and decide on the appropriate statistical tests. These procedures created a uniform data set for analysis and an initial description of molecular differences that may be related to the observed toxicity status.

2.3 Molecular Signature Identification and Feature Selection

The differences of molecular descriptors between toxic and non-toxic compounds were tested with an appropriate parametric or non-parametric statistical test according to the distribution of the data. The magnitude of the group differences was quantified using effect sizes, and multiple comparisons were controlled for with false-discovery-rate (FDR) correction. The significant descriptors were sorted by discriminatory power. The correlation analysis was then used to recognize highly redundant predictors and to reduce multicollinearity. Those descriptors with high correlation were merged so that a compact set of features containing molecular properties with complementary information was obtained that had the most compelling evidence of correlation with chemical toxicity.

2.4 Multivariate Analysis and Toxicity Classification

Standardised selected descriptors were used in multivariate pattern analysis by principal component analysis (PCA) to study multivariate molecular patterns and to explore whether a separation of toxic and non-toxic compounds can be obtained. The selected molecular signatures were further analyzed using logistic regression and random forest for evaluating their screening ability. Stratified cross validation preserved class proportions and offered more stable performance estimates than when using one data split. The accuracy, precision, recall, F1, sensitivity, specificity, and ROC-AUC were used to measure the classification performance. Model coefficients and feature importance values were analysed to identify the molecular properties most influential for discrimination of toxicity.

2.5 Statistical Analysis

Pandas, NumPy, SciPy, statsmodels, scikit-learn, and Matplotlib were used to perform all analyses in Python. The criteria for statistical significance were set at $p < 0.05$ and the false discovery rate (FDR) was applied between multiple comparisons of the descriptors. Scale-sensitive multivariate procedures were used with standardized variables. Model performance was reported for each of the stratified cross-validation folds, and summarized to provide stable estimates of model performance. Correlation matrices, principal component plots, descriptor rankings and receiver operating characteristic (ROC) curves were created for interpretation. This comprehensive statistical approach allowed molecular relations and toxicity-screening ability to be assessed methodically.

3. Results

3.1 Dataset and Molecular Property Characteristics

There were 13 selected molecular descriptors for the 171 compounds in the final dataset, and there were no missing values or duplicate observations. Among these compounds, 115 (67.3%) were considered non-toxic and 56 (32.7%) were toxic, which presents a moderate class imbalance. There was significant difference among toxicity groups with respect to molecular properties. Differences in some structural and physicochemical characteristics were found, as the mean values of MDEC-23 (41.30 vs. 38.08), CrippenMR (148.82 vs. 144.98), GATS8s (1.17 vs. 1.11), and VPC-4 (2.65 vs. 2.53) were higher for toxic compounds (Table 1).

Table 1. Molecular Descriptor Profiles According to Toxicity Class

Molecular Descriptor	Toxic Mean	Toxic SD	Non-Toxic Mean	Non-Toxic SD
MDEC-23	41.298	7.749	38.082	10.923
MATS2v	-0.012	0.064	-0.002	0.089
ATSC8s	-4.158	25.698	-7.298	32.019
VE3_Dt	-38.946	56.167	-29.080	48.990
CrippenMR	148.822	27.075	144.978	26.899

SpMax7_Bhe	3.252	0.159	3.206	0.204
SpMin1_Bhs	2.050	0.062	2.038	0.088
C1SP2	2.607	1.246	2.835	1.680
GATS8e	1.105	0.364	1.083	0.495
GATS8s	1.172	0.409	1.112	0.537
SpMax5_Bhv	3.448	0.178	3.414	0.208
VE3_Dzi	-29.917	47.237	-21.000	39.436
VPC-4	2.645	0.633	2.531	0.938

3.2 Toxicity-Associated Molecular Signatures

Univariate comparisons led to the selection of MDEC-23 as the best single molecular descriptor that correlated with toxicity. It was 41.17 as compared to 37.55 for non-toxic compounds, to yield an unadjusted p value of 0.017 when compared to toxic compounds. Differences with p values above 0.05 were found for other descriptors, such as SpMin1_Bhs, GATS8s, SpMax7_Bhe, and CrippenMR. MDEC-23 was not statistically significant after false-discovery-rate adjustment, however (q = 0.218). Therefore, after correcting for multiple comparisons, only a few individual descriptors showed independent evidence of toxicity differentiation (Table 2) (Figure 1).

Table 2. Statistical Comparison of Molecular Descriptors Between Toxicity Classes

Descriptor	Toxic Median	Non-Toxic Median	Mann–Whitney U	p-value	FDR q-value
MDEC-23	41.174	37.549	3947.0	0.0168	0.2183
SpMin1_Bhs	2.051	2.036	3692.5	0.1203	0.3276
GATS8s	1.138	1.009	3692.5	0.1203	0.3276
SpMax7_Bhe	3.285	3.247	3689.5	0.1227	0.3276
CrippenMR	154.409	146.964	3643.5	0.1638	0.3276
VE3_Dt	-8.613	-7.647	2808.5	0.1761	0.3276
GATS8e	1.099	1.005	3620.5	0.1880	0.3276
SpMax5_Bhv	3.478	3.431	3608.5	0.2016	0.3276
VPC-4	2.680	2.559	3505.5	0.3482	0.5030

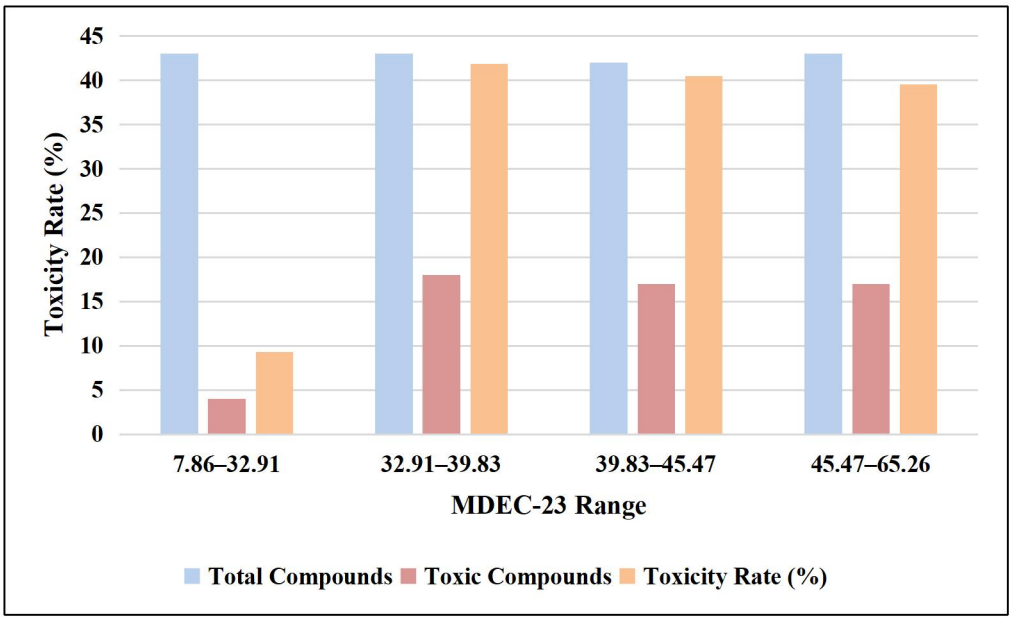

Figure 1. Toxicity Rate Across MDEC-23 Molecular Property Ranges

3.3 Correlation and Descriptor Redundancy

The dependency between selected molecular descriptors was shown by correlation analysis with different levels. There was a high degree of redundancy between the autocorrelation-based molecular properties, with the greatest absolute Spearman correlation between GATS8e and GATS8s ($|\rho| = 0.883$). Other descriptors provided relatively independent structural data. For this reason, a correlation assessment was performed to see if there was information that could be redundant when including all the molecular variables without considering any level of redundancy. The screening of correlated descriptors allowed to build a more parsimonious molecular signature with retaining complementary information which is relevant for the differentiation of toxic and non-toxic chemical compounds (Table 3) (Figure 2).

Table 3. Principal Component Analysis of Molecular Descriptors

Principal Component	Variance Explained (%)	Cumulative Variance (%)
PC1	28.51	28.51
PC2	16.84	45.35
PC3	13.47	58.82
PC4	9.65	68.48
PC5	8.96	77.43
PC6	6.20	83.63

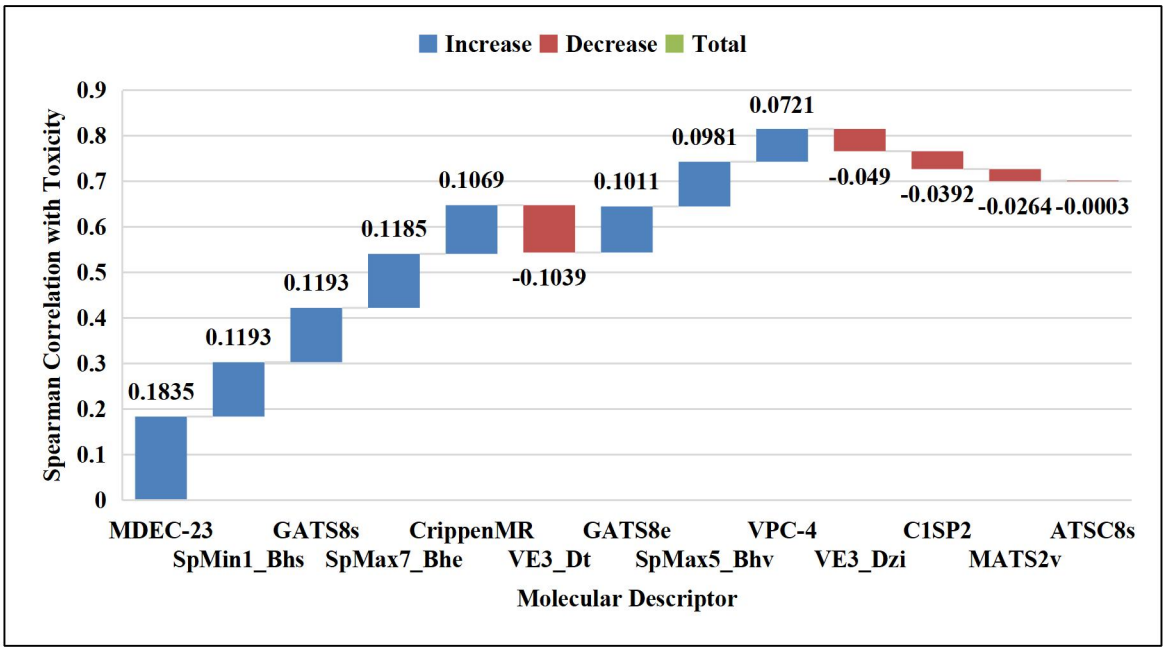

Figure 2. Strength and Direction of Descriptor–Toxicity Associations

3.4 Multivariate Molecular Patterns

The principal component analysis showed that the molecular information related to toxicity was spread out across several components, not a single dominant component. The first principal component explained 28.51% of the total variance, and the second explained 16.84% of the variance, for a total of 45.35%. The third component accounted for an extra 13.47% of the variance, bringing the total explained variance to 58.82%. These results showed that there was a lot of "multi-dimensional variation" among the molecular descriptors. The first two components had no dominant variance concentration suggesting separations between toxic and non-toxic compounds were based on combinations of multiple molecular characteristics (Figure 3).

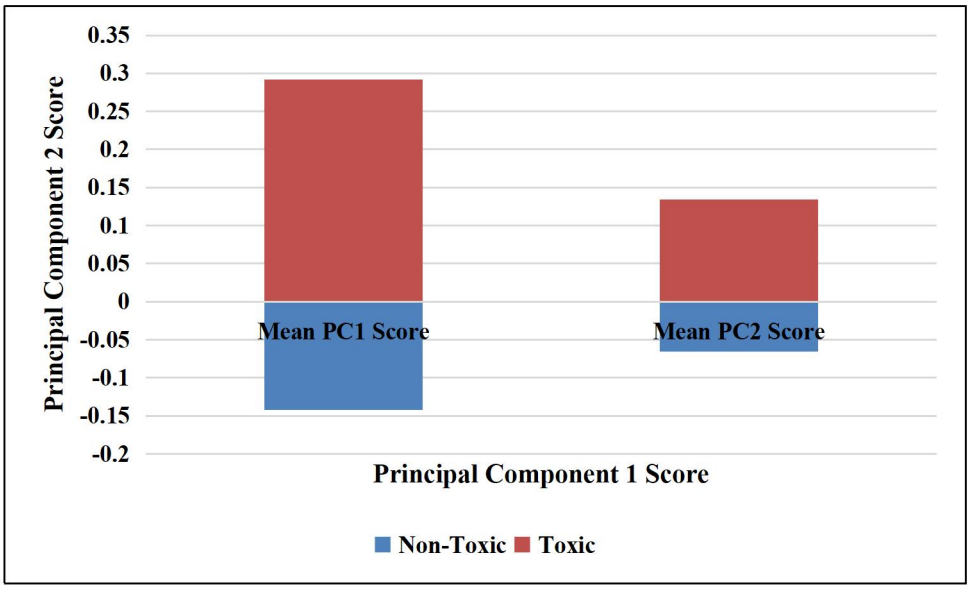

Figure 3. PCA Molecular-Space Position of Toxic and Non-Toxic Compounds

3.5 Toxicity Classification Performance

Cross-validated classification showed that random forest provided better toxicity discrimination than logistic regression. Random forest achieved mean accuracy of 0.673, precision of 0.498, recall of 0.323, F1-score of 0.389, and ROC-AUC of 0.659. Logistic regression produced lower accuracy

(0.574), recall (0.036), F1-score (0.050), and ROC-AUC (0.460). Overall performance of classification was moderate and low and the sensitivity to toxic compounds was limited, but the comparatively higher random-forest ROC-AUC pointed to the fact that nonlinear combinations of molecular descriptors carried more discriminatory information than a linear decision boundary (Table 4) (Figure 4).

Table 4. Cross-Validated Toxicity Classification Performance

Performance Metric	Logistic Regression	Random Forest
Accuracy	0.469	0.673
Precision	0.243	0.498
Recall/Sensitivity	0.362	0.323
Specificity	0.522	0.843
F1-score	0.289	0.389
ROC-AUC	0.458	0.659

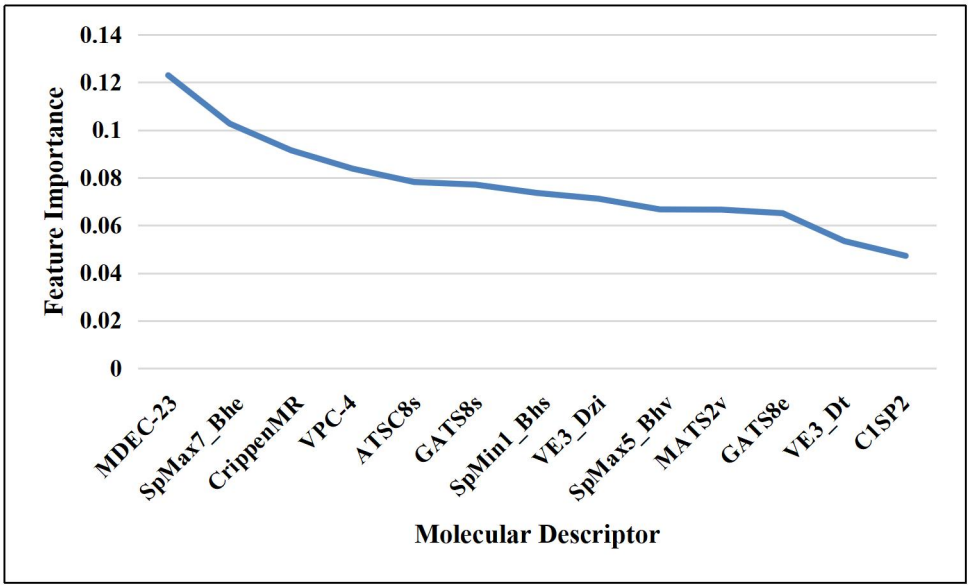

Figure 4. Random Forest Importance of Molecular Descriptors for Toxicity Screening

4. Discussion

Molecular descriptors include readily identifiable, but not particularly discriminative, information with respect to chemical toxicity. MDEC-23 had the highest correlation for individual properties for the toxicity assessment, and toxic compounds had higher scores than non-toxic compounds. The toxicity rate also rose significantly above the lowest MDEC-23 threshold, which indicates that structural attributes might play a role in toxicity differentiation. Its association was not statistically significant after false-discovery-rate correction, however, demonstrating that there is no single molecular characteristic with high confidence to be considered a toxicity association. This multi-dimensional interpretation is in line with modern computational toxicology, which views toxicity as a result of the combined structural, physicochemical and biological features rather than single molecular attributes (Zhang et al., 2025).

The overall low individual descriptor–toxicity correlations support this interpretation. MDEC-23 had the highest positive correlation, whereas descriptors that showed smaller positive correlation

were SpMin1_Bhs, GATS8s, SpMax7_Bhe, and CrippenMR. On the other hand, very weak negative associations were observed for VE3_Dt and VE3_Dzi. These results indicate that molecular toxicity signatures are present in multiple dimensions and that each descriptor only contributes to a small extent to understanding hazardous behavior. In the modern in silico drug-discovery paradigm, the point of integration is not to apply a single-property criterion to determine efficacy, pharmacokinetics and safety, but to integrate multiple molecular properties (Rasul et al., 2025). AI assisted phytochemical studies also suggest that structural and chemical properties can facilitate simultaneous evaluation of the biological efficacy and toxicity (Najaf & Asghar, 2026).

There was significant correlation (abs 0.883) between GATS8e and GATS8s. This is especially relevant since high dimensional molecular descriptor datasets often provide overlapping electronic or structural data, known as variables. If strongly correlated descriptors are not screened, they can unnecessarily add to the complexity of the model and diminish the interpretability. The approach of feature reduction used in this study thus resulted in a more parsimonious representation of molecular toxicity information. With the increasing number of computations being applied, molecular descriptors are now more and more prioritized according to their relevance in the context of the absorption, distribution, metabolism, excretion, and toxicity properties of the compound, enabling more efficient compound prioritization in early stages (Encinar et al., 2026). This reduction is particularly useful when the number of available compounds is small, compared to the original molecular feature space.

To see that toxicity-related molecular information was multi-dimensional, principal component analysis was performed. The first two principal components accounted for 45.35% of the total molecular variance, while six components accounted for 83.63% of the variance. So it is clear that there is no single dominant dimension of the molecule that could capture the variation among compounds. The molecular-space organization was affected by several descriptors, showing that both structural and physicochemical properties played a role in this. This pattern facilitates the growing practice of using integrated computational representations to grasp complex compound behavior. With the advancement of artificial intelligence, molecular modeling and ADMET prediction are gaining more traction, with the ability to recognize the relationships between several chemical properties that cannot be identified by a single variable analysis (Pathan et al., 2025). Similar in silico studies have shown that structurally complex compounds can be understood in terms of their properties of relevance for biology via an integrated molecular characterization approach (Badaoui et al., 2025).

The classification results also revealed the significance of the non-linear molecular relationships. Overall discrimination, random forest performed better than logistic regression with an ROC-AUC of 0.659 and accuracy of 0.673. It was moderately successful in recognizing non-toxic compounds with a specificity of 0.843, but with a sensitivity of 0.323. Therefore, the model was greatly superior in identifying compounds that are nontoxic compared to the identification of all toxic molecules. The overall performance is moderate, reflecting that there is sufficient information in the molecular descriptors for screening purposes, but that they are not a complete replacement for experimental toxicity testing. The present computational toxicology research also suggests that artificial intelligence is a tool that could aid rather than supplant biological testing for ADMET and toxicity prioritization (Zhang et al., 2025). Information on biological response beyond structure alone can be included in the large-scale toxicogenomic resources to further improve prediction (Bergen et al., 2025).

The most important descriptors in random-forest classification are selected by feature-importance analysis as MDEC-23, SpMax7_Bhe, CrippenMR, VPC-4 and ATSC8s. The overall contribution of their two kinds shows that there is no single predominant determinant of toxicity discrimination, but rather complementary molecular features. Practical implications include the use of compact molecular signatures for prioritising compounds for more costly experimental evaluation in safer screening. In this context, hazard is also becoming part of the design factors when selecting chemicals for green chemical substitution, underscoring the importance of computational approaches that can aid in the identification of potentially safer substitutes (Wang et al., 2026). Likewise, combining molecular data with comprehensive biological data can enhance compound prioritization and boost the accuracy of future drug discovery processes (Liu et al., 2026).

The results should be considered in the context of the limitations of the data set. Only 171 compounds were analyzed and the toxicity categorization was not sensitive because of the moderate imbalance between the number of toxic vs. non-toxic compounds. Furthermore, the molecular descriptors used are not sufficient to describe all biochemical pathways by which exposure to a chemical causes a biological effect. Mechanistic studies show that the toxicity could also be related to specific targets, metabolites, and downstream molecular interactions, which are not always reflected in general structural characteristics (Thu et al., 2025). The molecular signatures identified should therefore be confirmed in larger, independent, datasets and with additional complementary biological, toxicogenomic or multi-omics data in future studies. The findings overall suggest that molecular descriptor analysis is a useful step towards interpreting preliminary results for chemical safety screening but that for reliable chemical safety assessment, more extensive mechanistic and experimental information is needed to complement the computational predictions.

5. Conclusion

Molecular descriptors can help identify the signature of chemical properties related to chemical toxicity. MDEC-23 was the most prominent individual toxicity associated descriptor, and the other structural properties provided complementary discriminatory information. The individual associations were generally small and not significant after multiple testing correction but the pattern of associations suggested that there were meaningful differences between toxic and non-toxic compounds. This molecular variability was also observed in a multivariate analysis and was spread over several dimensions, suggesting that toxicity is a complex property that can be influenced by a number of interacting molecular properties. The random forest model showed better discrimination from logistic regression (accuracy 0.673, ROC-AUC of 0.659), but the model had poor sensitivity suggesting that the molecular signature may be a preliminary screening tool and should not be used as a definitive toxicity classifier. However, the compact descriptor framework has practical utility for the initial chemical evaluation of compounds as it would allow the evaluation of potentially informative molecular features prior to extensive experimental testing. These molecular signatures could be useful in more efficient toxicity prioritization, prevent unnecessary testing downstream, and be useful in identifying and developing safer compounds with validation in larger and chemically diverse datasets.

References

1. Akermi, S., Dey, A., Li, R., Larzat, N., Idoipe, J. B., Smaoui, S., ... & Sharma, A. (2025). Deep learning applications in genomics-based toxicology assessment. Deep Learning in Genetics and Genomics, 327-344.
2. Badaoui, H., Ouabane, M., Alaqarbeh, M., Elbouhi, M., Lakhlifi, T., & Bouachrine, M. (2025). The main chemical constituents responsible for the antidiabetic properties of the datura metel l plant. decryption and in-silico investigations. Physical Chemistry Research, 13(2), 241-254.
3. Barghi, E. G., Motafeghi, F., Gholami Gharab, J., & Ghassemi Barghi, N. (2026). Decoding DNA damage: artificial intelligence and the future of genotoxicity assessment. All Life, 19(1), 2611210.
4. Bergen, V., Kodella, K., Srikrishnan, S., Barrandon, O., Anderson, S., Rogers-Grazado, M., ... & Zamanighomi, M. (2025). A large-scale human toxicogenomics resource for drug-induced liver injury prediction. Nature Communications, 16(1), 9860.
5. Chen, Y., & Xue, W. (2026). Machine learning for de novo molecular generation: A comprehensive review. ACS chemical neuroscience, 17(4), 666-680.
6. de Sá, A. G., Long, Y., Portelli, S., Pires, D. E., & Ascher, D. B. (2022). toxCSM: comprehensive prediction of small molecule toxicity profiles. Briefings in Bioinformatics, 23(5), bbac337.
7. Encinar, J. A., Yang, C., Fernández-Ginés, R., Tang, Y., & Cuadrado, A. (2026). ADMET-Driven Chemical Navigability Rules for Early-Stage Virtual Compound Prioritization. Journal of Chemical Information and Modeling.
8. Galati, S., Di Stefano, M., Martinelli, E., Macchia, M., Martinelli, A., Poli, G., & Tuccinardi, T. (2022). VenomPred: a machine learning based platform for molecular toxicity

predictions. International journal of molecular sciences, 23(4), 2105.
9. Gül, Ş., & Rahim, F. (2021). Toxicity [Data set]. UCI Machine Learning Repository. https://doi.org/10.24432/C59313
10. Kurre, P. N., Muthyala, M. K. K., & Pichika, M. R. (2026). Computational Chemistry and Drug Discovery: Intersection of AI, QSAR, and Molecular Modeling. Computer-Aided Drug Design: Principles, Techniques, and Applications, 259-280.
11. 11.Lejal, V. (2024). Application of artificial intelligence with multi-omics and cell imaging data to decipher chemical-adverse outcomes for human health (Doctoral dissertation, Université Paris Cité).
12. Liu, Y., Zhu, K., Peng, W., Liu, Z., & Mao, X. (2026). Multi-omics and artificial intelligence for precision drug discovery and potential clinical applications. Signal Transduction and Targeted Therapy, 11(1), 210.
13. Liyaqat, T., Ahmad, T., & Saxena, C. (2026). Advancements in Molecular Property Prediction: A Survey of Single and Multimodal Approaches: T. Liyaqat et al. Archives of Computational Methods in Engineering, 33(1), 613-643.
14. Metz, T. O., Adkins, J. N., Armentrout, P. B., Chain, P., Chu, F., Corley, C. D., ... & Wishart, D. (2023). Decoding the Molecular Universe--Workshop Report. arXiv preprint arXiv:2311.11437.
15. Munson, B. P., Chen, M., Bogosian, A., Kreisberg, J. F., Licon, K., Abagyan, R., ... & Ideker, T. (2024). De novo generation of multi-target compounds using deep generative chemistry. Nature Communications, 15(1), 3636.
16. Najaf, A., & Asghar, M. U. (2026). Predicting Efficacy and Toxicity of Phytochemical Drugs with AI Models. In Artificial Intelligence in Phytochemistry: A New Era of Drug Discovery (pp. 83-96). Singapore: Springer Nature Singapore.
17. Ogbodo, J. O., Egba, S. I., Ikechukwu, G. C., Paul, P. C., Mba, J. O., Ugwu, O. P. C., & Ezike, T. C. (2025). Volatile organic compound–drug receptor interactions: a potential tool for drug design in the search for remedies for increasing toxic occupational exposure. Processes, 13(1), 154.
18. Pathan, I., Raza, A., Sahu, A., Joshi, M., Sahu, Y., Patil, Y., & Raza, M. A. (2025). Revolutionizing pharmacology: AI-powered approaches in molecular modeling and ADMET prediction. Medicine in Drug Discovery, 100223.
19. Pires, D. E., Stubbs, K. A., Mylne, J. S., & Ascher, D. B. (2022). cropCSM: designing safe and potent herbicides with graph-based signatures. Briefings in Bioinformatics, 23(2), bbac042.
20. Rasul, H. O., Ghafour, D. D., Aziz, B. K., Hassan, B. A., Rashid, T. A., & Kivrak, A. (2025). Decoding drug discovery: exploring A-to-Z in silico methods for beginners. Applied Biochemistry and Biotechnology, 197(3), 1453-1503.
21. Seal, S. (2024). Using cell painting and chemical data for small-molecule bioactivity and toxicity prediction. University of Cambridge (United Kingdom).
22. Thu, V. T. A., Nhu, N. Q., Anh, N. T. V., Lim, S. A., Seong, H. J., Md Rasheduzzaman, J., ... & Cho, Y. S. (2025). Deciphering linezolid-induced hematologic toxicity: Targeting TOP2A and TOP2B via its primary metabolite PNU142586. Science advances, 11(22), eadt5833.
23. Tripathi, R. (2025). Decoding Life: Tools and Techniques in Chemical Biology. Chyren Publication.
24. Wang, H., Chen, J., Liu, W., Wang, D., Song, Y., Hong, H., ... & Zimmerman, J. B. (2026). Using machine learning for green substitution of industrial chemicals: Integrating functionality, hazard, and life cycle impact. Chemical Reviews, 126(2), 841-894.
25. Zhang, J., Li, H., Zhang, Y., Huang, J., Ren, L., Zhang, C., ... & Zhang, Y. (2025). Computational toxicology in drug discovery: applications of artificial intelligence in ADMET and toxicity prediction. Briefings in Bioinformatics, 26(5), bbaf533.