ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Gap-Δenergy, a New Metric of the Bond Energy State, Assisting to Predict Molecular Toxicity

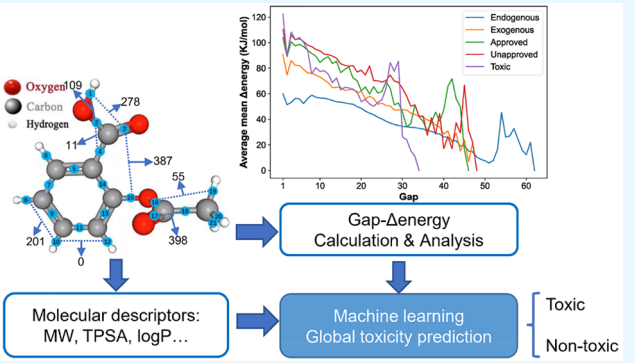
Senpeng Zhang, Dongyu Zhao,* and Qinghua Cui*

Cite This: *ACS Omega* 2024, 9, 17839−17847

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Molecular toxicity is a critical feature of drug development. It is thus very important to develop computational models to evaluate the toxicity of small molecules. The accuracy of toxicity prediction largely depends on the quality of molecular representation; however, current methods for this purpose do not address this issue well. Here, we introduce a new metric, gap-Δenergy, which is designed to quantify the intermolecular bond energy difference with atom distance. We next find significant variations in the gap-Δenergy distribution among different types of molecules. Moreover, we show that this metric is able to distinguish the toxic small molecules. We collected data sets of toxic and exogenous small molecules and presented a novel index, namely, global toxicity, to evaluate the overall toxicity of molecules. Based on molecular descriptors and the proposed gap-Δenergy metric, we further constructed machine learning models that were trained with 7816 small molecules. The XGBoost-based model achieved the best performance with an AUC score of 0.965 and an F1 score of 0.849 on the test set (1954 small molecules), which outperformed the model that did not use gap-Δenergy features, with a sensitivity score increase of 3.2%.

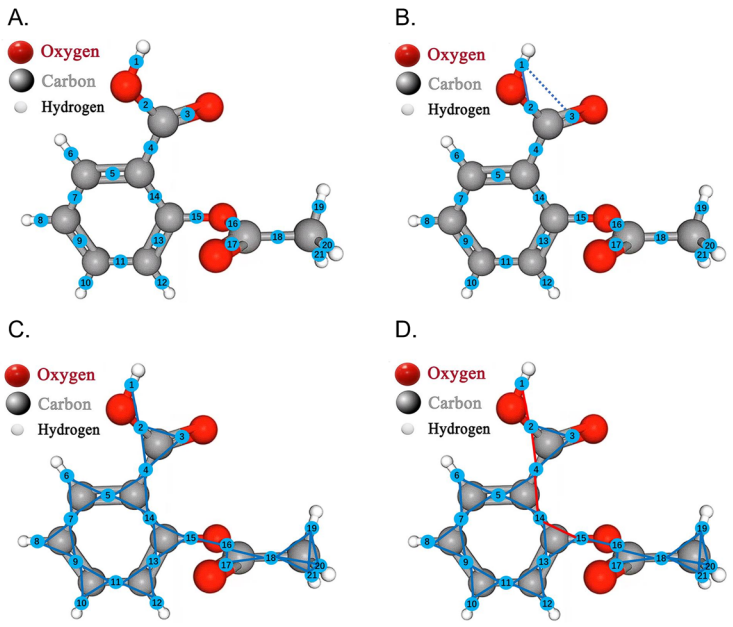


INTRODUCTION

Drug discovery is a long and challenging process with considerable resource demand.[1,2] The average capitalized development cost of a new pharmaceutical molecule is estimated to range from $314 million to $2.8 billion,[3] and the complete workflow can take over 12 years.[4] About 90% of experimental drugs fail to advance to clinical trials, and only 1% are eventually approved.[1] The high attrition rate of small-molecule drug candidates is primarily attributed to safety and toxicology in the preclinical or clinical phase.[5] Numerous *in vitro* and *in vivo* experiments are required to assess drug toxicity, which are time-consuming and expensive, and *in vivo* animal tests may arouse ethical concerns.[6] Furthermore, the differences in physiology and genetics between humans and animal models can cause the inapplicability of toxicity prediction to humans.[7,8] Studies have shown that the absence of toxicity observations from preclinical animal tests does not imply harmlessness for humans.[9,10] Therefore, it is crucial to utilize computational toxicology to assist in high-throughput drug toxicity prediction. Computational toxicology constructs quantitative models of chemical structures and biological toxicity, which can be used to eliminate potentially toxic molecules in early drug development stages and reduce the expenses of cell and animal experiments.

Machine learning is a prevalent and powerful modeling approach in drug development that involves learning the relationship between the abundant characteristics of molecules and the desired bioactivity. For drug toxicity prediction, the most commonly used algorithms include support vector machine (SVM), random forest (RF), K-nearest neighbors (KNN), and deep learning.[11] There are various molecular descriptors and fingerprints to quantitatively characterize a compound, which encode the physicochemical information and can be processed by these algorithms.[7,12] Molecules can also be represented as images, natural language, or graphs so that molecular features can be captured by deep learning architectures such as convolutional neural networks, transformers, and graph neural networks.[13−16] Toxicity prediction has various end points to be investigated, which can be divided into acute toxicity and chronic toxicity or can depend on the specific target proteins and organs. The Tox21 Data Challenge provided a high-throughput toxicity data set of more than 12000 environmental compounds targeted at 12 different toxic effects.[17] DeepTox, an algorithm based on deep learning, won the competition by using 2500 toxicophore features,[18] and many studies have since improved the prediction performance

Received: October 6, 2023
Revised: April 2, 2024
Accepted: April 5, 2024
Published: April 12, 2024



E. Gap and Δenergy pairs

Node1	Node2	Gap	Δenergy (KJ/mol)
1	2	1	109
1	3	2	278
1	4	2	120
1	5	3	147
…	…	…	…
20	21	1	0

F. Gap-mean Δenergy and gap-std Δenergy

Gap	Mean-Δenergy (KJ/mol)	Std-Δenergy (KJ/mol)
1	159.47	133.64
2	137.52	127.15
3	150.80	121.03
…	…	…
7	27.0	27.0

Figure 1. Using aspirin molecules as an example, a molecular graph is constructed where points represent chemical bonds and are labeled with numbers. Then gap-Δenergy can be calculated on this graph. (A) Turn all chemical bonds into nodes, so all 21 bonds in aspirin become 21 nodes, which are named 1−21 in the figure. (B) If there is only one atom between two nodes, then the two nodes should be connected. For example, there is only the oxygen atom between nodes 1 and 2, so we should connect nodes 1 and 2. There are oxygen and carbon atoms between nodes 1 and 3, so nodes 1 and 3 should not be connected. (C) Referring to the rules of Figure 1B, we connected all the nodes that should be connected. (D) In aspirin, the shortest distance between nodes 1 and 15 is marked with a red line, which is 4 steps and passes through 4 atoms, so the gap is 4. Δenergy is the absolute value of the difference between the O−H bond energy corresponding to point 1 and the C−H bond energy corresponding to point 15. (E) Calculate the gap and Δenergy between all pairs of points on the graph. (F) Compute the mean and variance of different Δenergy values corresponding to the same gap.

with diverse models.[19,20] Organ toxicity is also a common area of concern, and machine learning has been applied to predict hepatotoxicity and hemolytic toxicity of chemicals using molecular descriptors or fingerprints.[21,22] Several studies have focused more on the effects of compounds on cell lines and animal models, and algorithms have been developed to predict the cytotoxicity and the median lethal dose of animals.[23,24] Current prediction models mainly focus on a specific aspect of toxicity with small data sets, which may result in inaccurate predictions and limited application. The prediction of molecular toxicity also relies on the extraction of molecular characteristics. We previously found that normalized bond energy is able to describe molecular properties,[25] but it remains unclear whether it can be utilized for predicting molecular toxicity. Furthermore, multiple toxicities need to be considered in actual drug development, and a unified metric is required to assess the probability of a molecule being a toxic molecule.

Here we propose gap-Δenergy, a novel descriptor to characterize the diversity of bond-energy distribution inside molecules. Gap-Δenergy can reflect the average relationship between the distances of chemical bonds and the difference in bond energy in a molecule. We next reveal that there are significant differences in gap-Δenergy values among small molecules in different groups. Toxic small molecules and exogenous small molecules are collected, and we propose a metric named global toxicity to quantify the overall toxicity of a molecule. Finally, gap-Δenergy is found to be useful for predicting global toxicity and improving the prediction performance of molecular descriptors. Our gap-Δenergy feature and machine learning model can serve as tools to assist in predicting the potential harm of new small molecules to humans in the drug development process.

MATERIALS AND METHODS

Calculation of Gap-Δenergy. Here, molecules are conceptualized as undirected graphs composed of atoms and chemical bonds. However, in contrast to the traditional understanding, we treat chemical bonds as nodes in the graph (Figure 1A), and if there is only one atom between two nodes, we then connect the two nodes (Figure 1B). So that we can represent a molecule as a graph, and we need to combine the nodes of the graph in pairs in a permutation arrangement. For each pair of nodes, we calculate the shortest path length between them (Figure 1C,D), which is defined as "gap", and then we calculate the absolute value of the difference between the bond energies corresponding to the corresponding node pair (Figure 1D), which is defined as "Δenergy". We iterate through all of the chemical bonds within a molecule, calculate the gap and Δenergy for each pair of nodes (Figure 1E), and then aggregate all the corresponding Δenergy values for the same gap (Figure 1F). The mean value and the standard deviation of Δenergy (std Δenergy) are calculated at each gap, and a small molecule can be featured as a set of gap-mean Δenergy and gap-std Δenergy pairs (for the sake of simplicity, we named them that as gap-Δenergy).

We used the RDKit package to extract the chemical bonds and atoms in small molecules and then used the NetworkX (Hagberg et al., 2008) (Version 2.6.3) package to build molecular graphs. The energy (kJ/mol) of each bond in a molecule was obtained from the bond energy table (see Table S1), which depends on the bond type and the two

https://doi.org/10.1021/acsomega.3c07682
ACS Omega 2024, 9, 17839−17847

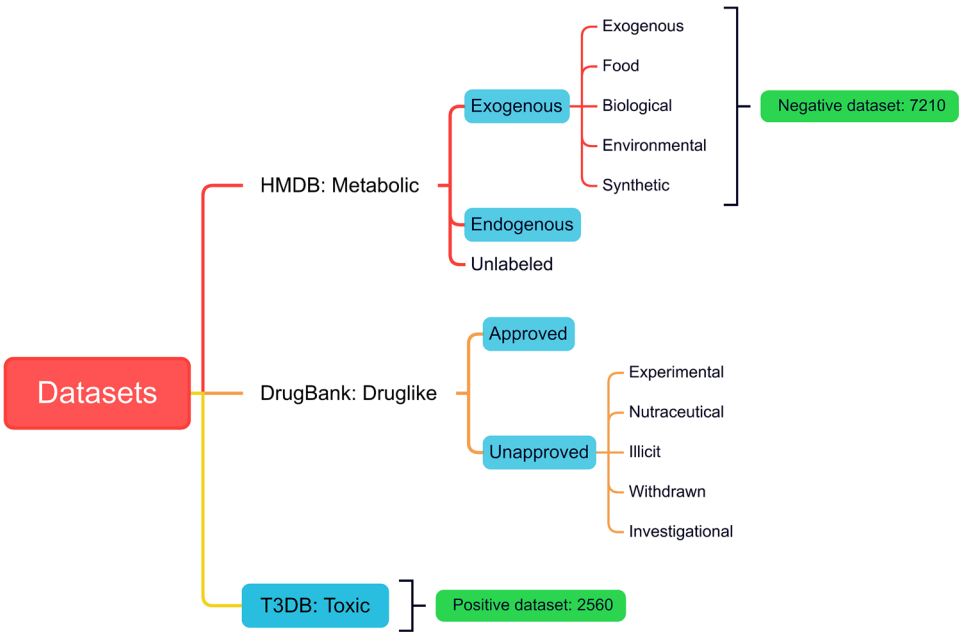


Figure 2. Data sets used in the study. We selected small molecules from three data sets, with the categories of interest marked in blue and the green ones representing positive and negative sample data.

corresponding atoms. Finally, we calculated gap-Δenergy for 120568 metabolic small molecules, 10206 druglike small molecules, and 2560 toxic small molecules.

Data Set Preparation. We downloaded the structures of small molecules in SDF format from the following three data sets: the Human Metabolome Database (HMDB)[26] (Version 5.0), the DrugBank database[27] (Version 5.1.8) and the Toxin and Toxin Target Database (T3DB).[28] In this study, we focused on and selected the small molecules (molecular weight ≤1000) in these databases. Metabolite molecules were obtained from HDMB, including seven categories: endogenous, food, biological, environmental, synthetic, exogenous, and unlabeled. And we classified food, biological, environmental, and synthetic molecules as exogenous small molecules as well. So, the small molecules in the HDMB database were mainly divided into two categories, including exogenous and endogenous molecules. Druglike molecules were obtained from DrugBank, including six categories: approved, experimental, nutraceutical, illicit, withdrawn, and investigational molecules. All classes of druglike molecules except approved were classified as unapproved molecules. Therefore, the small molecules in the DrugBank database were divided into two categories including approved and unapproved. In addition, we included the toxic small molecules from T3DB as well. We used the RDKit (Version 2022.03.2) package in Python[28] (Version 3.9.13) to sanitize the input molecules in order to standardize nonstandard valence states and to remove fragments and metal ions. Molecules with fewer than 3 chemical bonds were excluded because molecules of such small size contain too few gaps. And the duplicate molecules were respectively dropped to one in metabolite, druglike, and toxic small molecules.

Data-Splitting. We selected five classes of molecules in Data Set Preparation, including exogenous and endogenous molecules from HDMB, approved and unapproved molecules from DrugBank, and toxic molecules from T3DB. Toxic small molecules were used as the positive data set, and exogenous small molecules were used as the negative data set. The remaining three categories of small molecules were used for demonstration and comparison of the gap-mean Δenergy only. The overlapping parts between the exogenous molecules and toxic molecules were removed from the exogenous molecules. So there were a total of 9770 small molecules in the data set including 2560 toxic small molecules and 7210 exogenous small molecules. The data set was randomly partitioned based on label categories, with 80% (7816 instances, comprising 2048 toxic small molecules and 5768 exogenous small molecules) assigned to the training set and 20% (1954 instances, including 512 toxic small molecules and 1442 exogenous small molecules) designated as the test set. Subsequently, a 5-fold cross-validation was conducted within the training set to identify optimal parameters. The model's performance was then assessed on the test set using the best combination of parameters. The partitioning of the data sets is shown in Figure 2.

Feature Engineering. Max-gap is defined as the number of atoms passing through the shortest path between two bonds. Max-gap can be a measure of molecular size. Given that the maximum-gap for each molecule was inconsistent, using the full-length gap-Δenergy as input features could result in sparse features. Therefore, we used the first 10 gap-mean Δenergy of each molecule and used feature engineering to expand the number of features. First, the mean Δenergies corresponding to 10 gaps from the largest to the smallest were used as features, since larger molecules may have more than 10 gaps. If the max-gap of a small molecule is less than 10, then the Δenergy value corresponding to the nonexistent gap is filled with None. The mean value and standard deviation of the Δenergy corresponding to each gap were both included as features. The interaction of the Δenergy under different gaps was also taken into account, which was the pairwise product between two different Δenergies. In addition, we added the size and some overall energy characteristics of the molecule, including max-gap, normalized bond energy (NBE, eq 1), standard deviation of bond energy (SBE, eq 2), newNBE (eq 3), and newSBE (eq 4). Gap-Δenergy feature combinations are

https://doi.org/10.1021/acsomega.3c07682
ACS Omega 2024, 9, 17839−17847

listed in Table S2. Before the model training, all the features of the data set were standardized using the mean and variance of the training set, and then the None values existing in certain features were replaced with 0.

$$\text{NBE} = \frac{\sum_{i=1}^{n} \text{bond energy}(i)}{\text{MW}} \quad (1)$$

$$\text{SBE} = \sqrt{\frac{\sum_{i=1}^{n} (\text{bond energy}(i) - \overline{\text{bond energy}})^2}{n-1}} \quad (2)$$

$$\text{new NBE} = \frac{\sum_{i=1}^{n} \frac{\text{bond energy}(i)}{\text{bond weight}(i)}}{\text{bond num}} \quad (3)$$

$$\text{new SBE} = \sqrt{\frac{\sum_{i=1}^{n} \left[\frac{\text{bond energy}(i)}{\text{bond weight}(i)} - \left(\frac{\overline{\text{bond energy}}}{\text{bond weight}}\right)\right]^2}{n-1}} \quad (4)$$

Molecular Descriptors. Molecular descriptors can be used to describe the properties of the molecule in various aspects, and RDkit can calculate 208 molecular descriptors (such as MW, number of H acceptors, TPSA, and log *P*) based on the Simplified Molecular Input Entry System (SMILES) format of the input molecules. In this study, the descriptors with zero variance on the molecules of the data set were removed, and 204 descriptors in total were used as features.

Machine Learning Methods. K-Nearest Neighbors (KNN), Support Vector Classification (SVM), Random Forest (RF), and Logistic Regression (LR) algorithms from the Scikit-learn[29] (Version 1.1.3) package were used to construct binary classifier models. We also applied the Multilayer Perceptron (MLP) model in Pytorch[30] (Version 1.13.0), the XGBoost (XGB)[31] (Version 1.7.2) model, and the LightGBM (LGBM)[32] (Version 3.3.4) model. Because the positive and negative samples of the data set are not balanced, greater weight should be given to the category with a small sample size in training, so as to increase the accuracy of the results. Therefore, we set the parameter "class_weight" equal to "balance" in SVM, RF, and LR algorithms, set the parameter "scale_pos_weight" equal to the ratio of positive and negative samples in XGB and LGBM algorithms, and set focal loss[33] in the MLP algorithm. Recursive feature elimination with cross-validation (RFECV) in Scikit-learn was performed on the training set to select features. The features of gap-Δenergy and molecular descriptors with low importance were sequentially eliminated using the RFECV algorithm in the wrapper method. We applied the stacking method for ensemble learning. The output of the model using the XGB algorithm and gap-Δenergy features and the output of the model using the XGB algorithm and molecular descriptor features were combined as new features and then put in the LR classifier. Through cross-validation on the training set, the output toxic probabilities of the molecules in the last fold were predicted by the two models that were trained on the data of the first four folds, and then all the probabilities of the training set were obtained in turn. For the test set, we trained the two models with the training set to predict the output probabilities of the molecules in the test set. Finally, the LR classifier was trained with the probabilities of the training set and verified on the test set. All of the procedures above are shown in Figure 3.

Hyperparameter Selection. There are many hyperparameters in every machine learning algorithm. In order to optimize the performance of each classifier for the toxicity prediction task and to better compare the performance of different classifiers, we used Ray Tune[34] (Version 2.2.0) to search for the best value of hyperparameters. After the range of the hyperparameters was set, the random search strategy was applied to select the best hyperparameters according to the balanced accuracy scores of the training set from the 64 hyperparameter combinations. The hyperparameter search scope for each model is provided in Table S3.

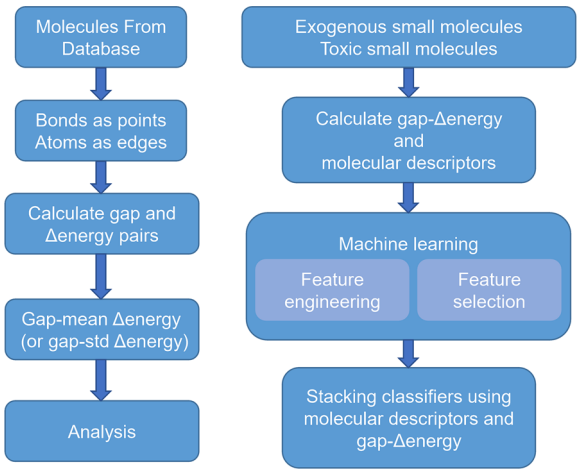


Figure 3. Workflow for calculating gap-Δenergy and toxicity prediction.

Evaluation of the Model. We employed statistical metrics to evaluate the model's ability to predict toxicity molecules, including accuracy (ACC, eq 5), sensitivity (SE, eq 6), specificity (SP, eq 7), balanced accuracy (BA, eq 8), F1 (eq 9), the area under the receiver operating characteristic (ROC) curve (AUC), and Matthew's correlation coefficient (MCC, eq 10), which are defined as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

$$\text{SE} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (7)$$

$$\text{BA} = \frac{1}{2} \times (\text{SE} + \text{SP}) \quad (8)$$

$$\text{F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FN} + \text{FP}} \quad (9)$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (10)$$

RESULTS AND DISCUSSION

Mean Δenergy Shows an Overall Decreasing Trend with the Gap Value Increasing. We first calculated the gap-mean Δenergy on all druglike small molecules from DrugBank and all metabolic small molecules from HMDB. The distributions of mean Δenergy on each gap are displayed in Figure 4. We observed that the median mean Δenergy on each

https://doi.org/10.1021/acsomega.3c07682
ACS Omega 2024, 9, 17839−17847

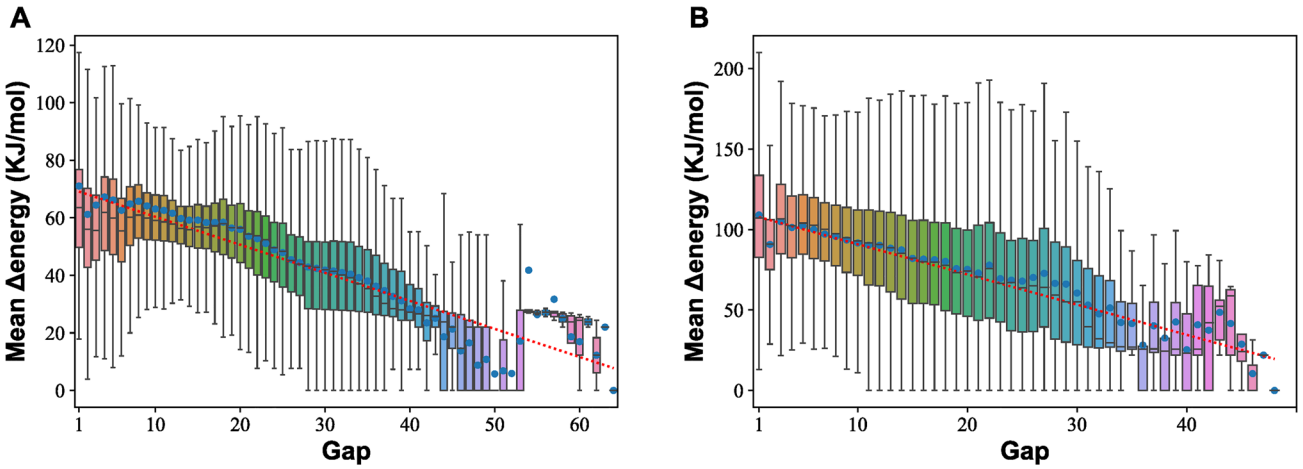


Figure 4. Distribution of mean Δenergy under different gaps. The entirety of the mean Δenergy decreases as the gap increases in metabolic small molecules (A) and druglike small molecules (B). The blue dots are the average mean Δenergy, and the red dashed lines show the correlation of average mean Δenergy with gaps in metabolic small molecules and druglike small molecules.

gap shows a decreasing trend as the gap increases. Then we averaged the mean Δenergy values of all molecules at each gap to construct a gap-average mean Δenergy plot, which shows that the average mean Δenergy has a significant negative correlation with the gap in both the druglike small molecules (Spearman $\rho = -0.936$, $P = 7.46\text{e-}30$) and the metabolic small molecules (Spearman $\rho = -0.977$, $P = 2.02\text{e-}32$).

The distribution of the molecular sizes was investigated as well. As a result, we found that 87.0% of the metabolic small molecules had max-gap less than 45, and 92.4% of the druglike small molecules had max-gap less than 20 (Figure S1). We also counted the number of molecules at each gap, which showed that 90.3% of the metabolic small molecules and 74.6% of the druglike small molecules had the property of Δenergy at gap 10. Therefore, the mean Δenergy of larger gaps can be biased by a few molecules, and we should focus on the Δenergy of smaller gaps in molecules.

Significant Differences in Gap-Mean Δenergy among Various Classes of Small Molecules. We compared the differences in mean Δenergy of these five classes of molecules at each gap and found significant differences ($P \leq 3.1\text{e-}163$, Kruskal−Wallis test) for all classes at the first 30 gaps, implying that at least one class was significantly different from the others. We then pairwise calculated the classes of interest (Mann−Whitney U test), and the results showed significant differences in mean Δenergy for toxic and exogenous small molecules at the first 7 gaps ($P \leq 6.0\text{e-}97$), for approved and unapproved druglike small molecules at the first 16 gaps ($P \leq 5.2\text{e-}03$), and for endogenous and exogenous small molecules at all 30 gaps (see Table S4). In addition, the average mean Δenergy at each gap was calculated for different classes of molecules and is shown in Figure 5. Intuitively, the average mean Δenergy of each class of small molecules tends to decrease as the gap increases. Endogenous molecules had the lowest average mean Δenergy in the first 20 gaps, followed by exogenous molecules. Approved and unapproved drugs had similar gap-averaged Δenergy curves. The average mean Δenergy of toxic molecules was similar to that of approved and unapproved drugs at the first 7 gaps, but the average mean Δenergy at 8−20 gaps was more like that of exogenous molecules. Considering that the estimated average value of gap-mean Δenergy of each class of molecules may fluctuate with the change of the number of molecules, we also sampled

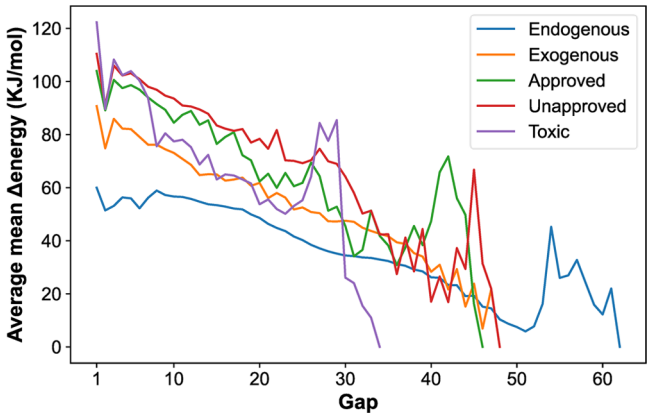


Figure 5. Relationship between the average mean Δenergy and gap in endogenous, exogenous, approved, unapproved, and toxic small molecules. Different classes of molecules show diverse gap-averaged mean Δenergy curves, which may be applied to molecular property prediction.

200, 400, 800, and 1600 molecules in each class to observe the fluctuation of mean Δenergy. We found that the average mean Δenergy curves remained stable at the first 10 gaps, even if only a small fraction of the molecules were sampled. With the increase in the number of sampling molecules, the average gap-mean Δenergy converged to the empirical average values and the average gap-mean Δenergy characteristics of different classes were consistent with the overall pattern (see Figure S2).

Model Performance for Global Toxicity Prediction Using Gap-Δenergy. Global toxicity differs from target-specific toxicity and signifies the overall toxicity of human intake or exposure from the outside world. Exogenous small molecules are considered relatively harmless, while toxic small molecules are absolutely harmful to humans; therefore, the prediction task was to make the distinction between exogenous and toxic small molecules. Since we found that the average mean Δenergy of exogenous and toxic small molecules differed significantly at the first 7 gaps, we conjectured that the gap-Δenergy of molecules could be used as a feature to predict the global toxicity of molecules. Considering that the mean Δenergy of larger gaps is less significantly different between exogenous and toxic small molecules, we used the mean

https://doi.org/10.1021/acsomega.3c07682
ACS Omega 2024, 9, 17839−17847

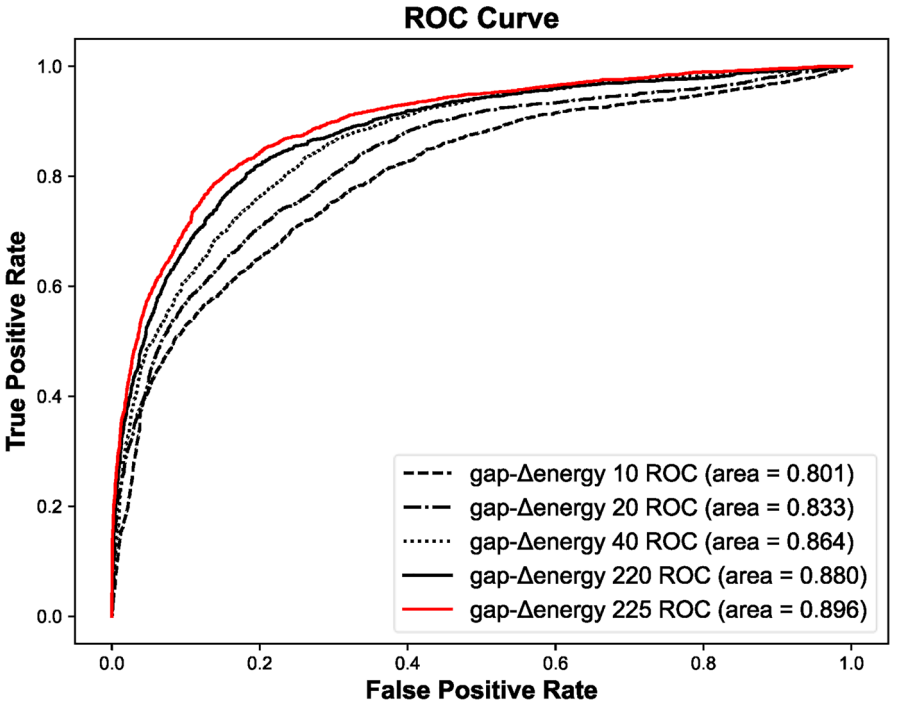


Figure 6. Performance of the SVM models with different feature combinations. Gap-Δenergy 10 contains only the mean Δenergy of the first 10 gaps. Gap-Δenergy 20 contains the mean Δenergy of the first and last 10 gaps. Gap-Δenergy 40 contains the mean and std Δenergy of the first and last 10 gaps. On the basis of gap-Δenergy 40, gap-Δenergy 220 includes the interaction of Δenergy under different gaps. Gap-Δenergy 225 includes gap-Δenergy 220, max-gap, and overall energy characteristics. The SVM model with gap-Δenergy 225 performed the best and reached 0.896 AUC on the validation set.

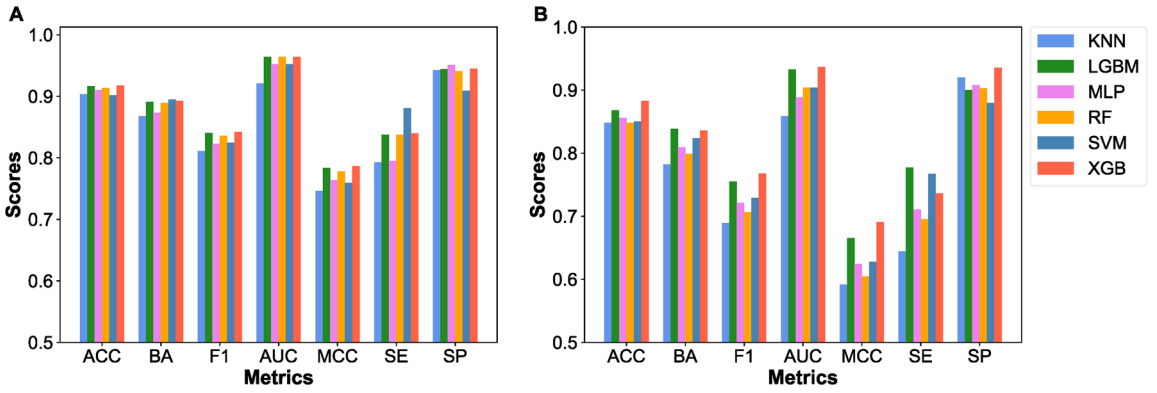


Figure 7. Performance of different models on molecular descriptors (A) and gap-Δenergy (B). The XGB model reached four top-ranked metrics using the molecular descriptors with BA of 0.893, F1 of 0.842, and AUC of 0.964 on the test set. For gap-Δenergy, the XGB model had five top-ranked metrics, with BA of 0.836, F1 of 0.878, and AUC of 0.937 on the test set.

Δenergy corresponding to the first 10 gaps as features (gap-Δenergy 10) and applied feature engineering to expand the number of features (gap-Δenergy 20, 40, 220, and 225). The new gap-Δenergy features took into account not only the mean **ΔX**energy, but also the std Δenergy at each gap and the overall energy of the molecules. We first trained the SVM models with different feature combinations and evaluated their prediction performance with AUC scores by 5-fold cross-validation. As a result, gap-Δenergy 225 combined all of the features and had the best performance (Figure 6). SVM, RF, XGB, LGBM, KNN, and MLP models were trained with hyperparameter search respectively on gap-Δenergy 225 and molecular descriptors, and the models with their best hyperparameters were used to predict the global toxicity. The results showed that for gap-Δenergy and molecular descriptors, the XGB model achieved the best performance with the most top-ranked metric accounts (Figure 7). The gap-Δenergy was less effective in predicting global toxicity than the molecular descriptors. Even if gap-Δenergy and molecular descriptors were combined, the performance of these six classifiers rarely exceeded that of using only molecular descriptors (see Table S5).

In addition, we also took subsets of exogenous and toxic small molecules with a maximum gap of no more than 30 and trained the classification task on the XGB classifier with the same hyperparameter search method and features, because the size of toxic small molecules is smaller on the whole, and the size of molecules may leak information about molecular toxicity. It was found that the performance of the classifier was slightly reduced compared with the XGB model using the full

https://doi.org/10.1021/acsomega.3c07682
ACS Omega 2024, 9, 17839−17847

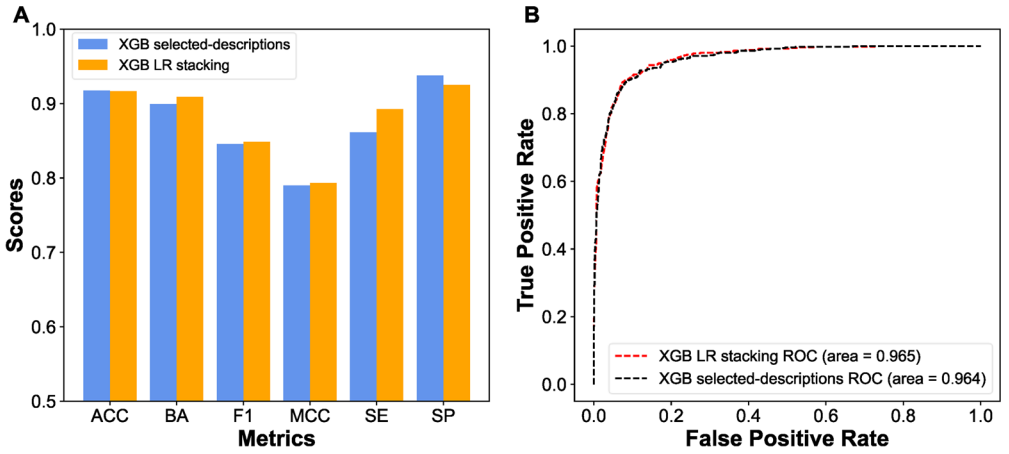


Figure 8. Comparison of models' performance before and after stacking (A). After stacking, the overall performance was improved. The SE score was the most significantly increased metric, along with the improvement of BA, F1, and MCC scores, while the SN score decreased. The ROC curve of the XGB model using selected molecular descriptors and the LR model after stacking (B).

set of exogenous molecules and toxic molecules (see Table S6), possibly because of the decrease in the number of molecules available for training, which also indicated that the molecular size had little effect on the classifier.

Gap-Δenergy Improved the Sensitivity of Toxicity Prediction Models. We used the best-performing algorithm, XGBoost, for the follow-up study. Considering that features may have redundancy, feature selection was performed on the molecular descriptors and the gap-Δenergy. For molecular descriptors, using 109 out of 204 features reached the highest balanced accuracy on the validation set, while 50 gap-Δenergy features should be selected from gap-Δenergy 225 (see Figure S3). After feature selection, the performance of the classifiers in the test set was similar to that before. The newly selected gap-Δenergy features were put into the XGB classifier for training, and it was found that compared with the original features of gap-Δenergy 225, BA, F1, and SN scores on the test set were improved. However, the selected molecular descriptors had higher BA, F1, MCC, and SN scores on the test set (see Table S7).

Since directly combining the two types of features will not bring performance improvement, the stacking method was used to combine the selected gap-Δenergy and molecular descriptors by integrating the output probabilities of XGB classifiers, which were put into the LR classifier for training and prediction. After stacking, the BA score of the XGB classifier increased by 1% compared to using only molecular descriptors, and the F1 and MCC scores increased slightly by 0.3%. Among these indicators, the SE score increased the most by 3.2%. In contrast, the SP score was reduced by 1.3% (see Figure 8 and Table S8). This indicates that the gap-Δenergy features can improve the sensitivity and balanced accuracy of the classifier by using molecular descriptors for predicting toxic molecules by stacking.

To sum up, the workflow for predicting the toxicity of a small molecule should first calculate the gap-Δenergy and descriptors of the molecule and calculate the gap-Δenergy 225 through feature engineering. Then the 50 gap-Δenergy and 109 descriptor features should be selected according to the important features list and put into the trained XGB-selected-gap-Δenergy and XGB-selected-descriptor models, respectively, to acquire two prediction probabilities. At last, the probabilities are input into the LR model to get the final probability of whether the molecule is toxic or not. We found several cases to implement our model that are not included in our training and testing sets. Benfluorex is a small molecule as a hypolipidemic and hypoglycemic drug, which was withdrawn from Europe in 2010 due to cardiovascular side effects.[35] When its SMILES was input into the final model, the global toxicity probability given was 0.953, indicating that this molecule is likely harmful to the human body. Bexagliflozin is an approved small molecule, a sodium-glucose cotransporter 2 (SGLT2) inhibitor used for treating adults with type 2 diabetes.[36] Inputting its SMILES into the model gave a global toxicity probability of 0.206, indicating that this molecule may have a weaker toxicity.

CONCLUSIONS

In this study, we defined a new metric, gap-Δenergy, to reflect the topological relationship of the bond energy in a molecule. We found that the entirety of mean Δenergy decreases as the gap increases in metabolite and druglike small molecules. Our hypothesis is that in organic compounds bonds with a larger distance are more likely to be C−H bonds that are typically located on the periphery of the molecule, and then, the Δenergy tends to be zero. Different groups of molecules possess different gap-Δenergy characteristics, which indicates that gap-Δenergy can reflect some properties of small molecules. Toxic small molecules show significantly higher average mean Δenergy than exogenous small molecules at the first 7 gaps, and the average mean Δenergy of unapproved drugs in the first 15 gaps is also significantly higher than that of approved drugs. Considering that endogenous small molecules have the lowest average mean Δenergy at the first 20 gaps, we speculate that a higher gap-mean Δenergy may be associated with toxicity or adverse reactions. Since that we have already applied gap-Δenergy to predict toxicity, this method provides a potential direction for assessing the drug-likeness of small molecules. However, there are limitations in gap-Δenergy, as some chemical bonds in molecules are not included in the bond energy table, making it impossible to calculate gap-Δenergy for those molecules.

T3DB is a toxic exposure database, encompassing various categories such as pollutants, toxic drugs, and carcinogens, as well as naturally occurring or chronically toxic compounds.[28] The T3DB database has been utilized for supplementing drug

https://doi.org/10.1021/acsomega.3c07682
ACS Omega 2024, 9, 17839−17847

development with hepatotoxicity libraries,[37] predicting acute oral toxicity,[38] and even screening anticancer drugs based on the target protein.[39] In this study, toxic small molecules from T3DB have global toxicity and are regarded as positive samples. Several classifiers based on machine learning have been constructed to distinguish explicitly toxic small molecules from exogenous small molecules and to predict the global toxicity of molecules. We found that in simple toxicity classification tasks with hundreds of features neural networks are not superior to traditional machine learning methods, while XGB and LGBM, based on gradient boosting trees, are simpler and more effective. Both gap-Δenergy and molecular descriptors can distinguish global toxicity, but the performance of gap-Δenergy is not as good as that of molecular descriptors. Stacking the two classifiers is required to increase the sensitivity of global toxicity prediction, and the final model can be used to eliminate molecules that may be harmful to humans. In the future, we hope to have more labeled data and develop new computational methods to improve the performance of global toxicity prediction and assist in drug design to generate molecules with reduced toxicity.

■ ASSOCIATED CONTENT

Data Availability Statement

To facilitate the use and verification of our proposed method, we have made our source code publicly available on GitHub at the following link: https://github.com/zsp999/Gap-delta-energy.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c07682.

Bond energy table, the combination of gap-Δenergy features, hyperparameter search scope and the best parameters for different models, *P*-values of mean Δenergy in different classes of molecules, the performance of different models with values, the distribution of molecular size, the process of stepwise feature elimination, and the average gap-mean Δenergy graphs using randomly selected molecules (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Qinghua Cui − *Department of Biomedical Informatics, State Key Laboratory of Vascular Homeostasis and Remodeling, School of Basic Medical Sciences, Peking University, Beijing 100191, People's Republic of China;* Present Address: School of Sports Medicine, Wuhan Institute of Physical Education, No. 461 Luoyu Rd. Wuchang District, Wuhan 430079, Hubei Province, People's Republic of China; orcid.org/0000-0003-3018-5221; Email: cuiqinghua@bjmu.edu.cn

Dongyu Zhao − *Department of Biomedical Informatics, State Key Laboratory of Vascular Homeostasis and Remodeling, School of Basic Medical Sciences, Peking University, Beijing 100191, People's Republic of China*; Email: zhaodongyu@bjmu.edu.cn

Author

Senpeng Zhang − *Department of Biomedical Informatics, State Key Laboratory of Vascular Homeostasis and Remodeling, School of Basic Medical Sciences, Peking University, Beijing 100191, People's Republic of China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c07682

Funding

This study was supported by grants from the National Natural Science Foundation of China (62025102, 81921001) and the Scientific and Technological Research Project of Xinjiang Production and Construction Corps (2023AB057).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank members of Prof. Cui's lab for helpful discussions.

■ REFERENCES

(1) Chi, L. H.; Burrows, A. D.; Anderson, R. L. Can preclinical drug development help to predict adverse events in clinical trials? *Drug Discovery Today* **2022**, *27* (1), 257−268.

(2) Vo, A. H.; Van Vleet, T. R.; Gupta, R. R.; Liguori, M. J.; Rao, M. S. An Overview of Machine Learning and Big Data for Drug Toxicity Evaluation. *Chem. Res. Toxicol.* **2020**, *33* (1), 20−37.

(3) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009−2018. *JAMA* **2020**, *323* (9), 844−853.

(4) Chan, H. C. S.; Shan, H.; Dahoun, T.; Vogel, H.; Yuan, S. Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* **2019**, *40* (8), 592−604.

(5) Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov* **2015**, *14* (7), 475−86.

(6) Limbu, S.; Zakka, C.; Dakshanamurthy, S. Predicting Dose-Range Chemical Toxicity using Novel Hybrid Deep Machine-Learning Method. *Toxics* **2022**, *10* (11), 706.

(7) Wu, Y.; Wang, G. Machine Learning Based Toxicity Prediction: From Chemical Structural Description to Transcriptome Analysis. *Int. J. Mol. Sci.* **2018**, *19* (8), 2358.

(8) Akhtar, A. The flaws and human harms of animal experimentation. *Camb Q Healthc Ethics* **2015**, *24* (4), 407−19.

(9) Clark, M.; Steger-Hartmann, T. A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans. *Regul. Toxicol. Pharmacol.* **2018**, *96*, 94−105.

(10) Clark, M. Prediction of clinical risks by analysis of preclinical and clinical adverse events. *J. Biomed Inform* **2015**, *54*, 167−73.

(11) Cavasotto, C. N.; Scardino, V. Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point. *ACS Omega* **2022**, *7* (51), 47536−47546.

(12) Carracedo-Reboredo, P.; Linares-Blanco, J.; Rodriguez-Fernandez, N.; Cedron, F.; Novoa, F. J.; Carballal, A.; Maojo, V.; Pazos, A.; Fernandez-Lozano, C. A review on machine learning approaches and trends in drug discovery. *Comput. Struct Biotechnol J.* **2021**, *19*, 4538−4558.

(13) Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R. K. Artificial intelligence in drug discovery and development. *Drug Discov Today* **2021**, *26* (1), 80−93.

(14) Yoshimori, A. Prediction of Molecular Properties Using Molecular Topographic Map. *Molecules* **2021**, *26* (15), 4475.

(15) Chen, D.; Gao, K.; Nguyen, D. D.; Chen, X.; Jiang, Y.; Wei, G. W.; Pan, F. Algebraic graph-assisted bidirectional transformers for molecular property prediction. *Nat. Commun.* **2021**, *12* (1), 3521.

(16) Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; Zheng, M. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *J. Med. Chem.* **2020**, *63* (16), 8749−8760.

(17) Tice, R. R.; Austin, C. P.; Kavlock, R. J.; Bucher, J. R. Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect* **2013**, *121* (7), 756−65.
(18) Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. DeepTox: Toxicity prediction using deep learning. *Toxicol. Lett.* **2017**, *280*, S69−S69.
(19) Yuan, Q.; Wei, Z.; Guan, X.; Jiang, M.; Wang, S.; Zhang, S.; Li, Z. Toxicity Prediction Method Based on Multi-Channel Convolutional Neural Network. *Molecules* **2019**, *24* (18), 3383.
(20) Jiang, J.; Wang, R.; Wei, G. W. GGL-Tox: Geometric Graph Learning for Toxicity Prediction. *J. Chem. Inf Model* **2021**, *61* (4), 1691−1700.
(21) Jaganathan, K.; Tayara, H.; Chong, K. T. Prediction of Drug-Induced Liver Toxicity Using SVM and Optimal Descriptor Sets. *Int. J. Mol. Sci.* **2021**, *22* (15), 8073.
(22) Zheng, S.; Wang, Y.; Liu, W.; Chang, W.; Liang, G.; Xu, Y.; Lin, F. In Silico Prediction of Hemolytic Toxicity on the Human Erythrocytes for Small Molecules by Machine-Learning and Genetic Algorithm. *J. Med. Chem.* **2020**, *63* (12), 6499−6512.
(23) Varsou, D. D.; Melagraki, G.; Sarimveis, H.; Afantitis, A. MouseTox: An online toxicity assessment tool for small molecules through Enalos Cloud platform. *Food Chem. Toxicol.* **2017**, *110*, 83−93.
(24) Wu, K.; Wei, G. W. Quantitative Toxicity Prediction Using Topology Based Multitask Deep Neural Networks. *J. Chem. Inf Model* **2018**, *58* (2), 520−531.
(25) Huang, C.; Zhou, Y.; Yang, J.; Cui, Q.; Li, Y. A New Metric Quantifying Chemical and Biological Property of Small Molecule Metabolites and Drugs. *Front Mol. Biosci* **2020**, *7*, 594800.
(26) Wishart, D. S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schioth, H. B.; Greiner, R.; Gautam, V. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50* (D1), D622−D631.
(27) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D1074−D1082.
(28) Wishart, D.; Arndt, D.; Pon, A.; Sajed, T.; Guo, A. C.; Djoumbou, Y.; Knox, C.; Wilson, M.; Liang, Y.; Grant, J.; Liu, Y.; Goldansaz, S. A.; Rappaport, S. M. T3DB: the toxic exposome database. *Nucleic Acids Res.* **2015**, *43* (D1), D928.
(29) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (null), 2825−2830.
(30) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Köpf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: an imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*; Curran Associates Inc.: 2019; p Article 721.
(31) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
(32) Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*; Curran Associates Inc.: 2017; pp 3149−3157.
(33) Lin, T. Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. In *Focal Loss for Dense Object Detection*; IEEE International Conference on Computer Vision (ICCV), 22−29 Oct 2017; 2017; pp 2999−3007.
(34) Liaw, R.; Liang, E.; Nishihara, R.; Moritz, P.; Gonzalez, J. E.; Stoica, I. Tune: A Research Platform for Distributed Model Selection and Training. *ArXiv* 2018, abs/1807.05118.
(35) Tribouilloy, C.; Rusinaru, D.; Marechaux, S.; Jeu, A.; Ederhy, S.; Donal, E.; Reant, P.; Arnalsteen, E.; Boulanger, J.; Ennezat, P. V.; Garban, T.; Jobic, Y. Increased risk of left heart valve regurgitation associated with benfluorex use in patients with diabetes mellitus: a multicenter study. *Circulation* **2012**, *126* (24), 2852−8.
(36) Azzam, O.; Carnagarin, R.; Lugo-Gavidia, L. M.; Nolde, J.; Matthews, V. B.; Schlaich, M. P. Bexagliflozin for type 2 diabetes: an overview of the data. *Expert Opin Pharmacother* **2021**, *22* (16), 2095−2103.
(37) Luo, G.; Shen, Y.; Yang, L.; Lu, A.; Xiang, Z. A review of drug-induced liver injury databases. *Arch. Toxicol.* **2017**, *91* (9), 3039−3049.
(38) García-Jacas, C. R.; Marrero-Ponce, Y.; Cortés-Guzmán, F.; Suárez-Lezcano, J.; Martinez-Rios, F. O.; García-González, L. A.; Pupo-Meriño, M.; Martinez-Mayorga, K. Enhancing Acute Oral Toxicity Predictions by using Consensus Modeling and Algebraic Form-Based 0D-to-2D Molecular Encodes. *Chem. Res. Toxicol.* **2019**, *32* (6), 1178−1192.
(39) Ibrahim, M. A. A.; Abdeljawaad, K. A. A.; Abdelrahman, A. H. M.; Sidhom, P. A.; Tawfeek, A. M.; Mekhemer, G. A. H.; Abd El-Rahman, M. K.; Dabbish, E.; Shoeib, T. In-Silico Mining of the Toxins Database (T3DB) towards Hunting Prospective Candidates as ABCB1 Inhibitors: Integrated Molecular Docking and Lipid Bilayer-Enhanced Molecular Dynamics Study. *Pharmaceuticals-Base* **2023**, *16* (7), 1019.

https://doi.org/10.1021/acsomega.3c07682
ACS Omega 2024, 9, 17839−17847