



OPEN ACCESS

EDITED BY

Decai Xiong,
Peking University, China

REVIEWED BY

Peichen Pan,
Zhejiang University, China
Sunliang Cui,
Zhejiang University, China

*CORRESPONDENCE

Hao Wang,
✉ hao.wang@muc.edu.cn
Zongran Pang,
✉ pangzongran@muc.edu.cn

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 20 December 2024

ACCEPTED 06 January 2025

PUBLISHED 20 January 2025

CITATION

Zhang J, Ren Y, Teng Y, Wu H, Xue J, Chen L, Song X, Li Y, Zhou Y, Pang Z and Wang H (2025) Discovery of novel PRMT1 inhibitors: a combined approach using AI classification model and traditional virtual screening. *Front. Chem.* 13:1548812. doi: 10.3389/fchem.2025.1548812

COPYRIGHT

© 2025 Zhang, Ren, Teng, Wu, Xue, Chen, Song, Li, Zhou, Pang and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Discovery of novel PRMT1 inhibitors: a combined approach using AI classification model and traditional virtual screening

Jungan Zhang^{1†}, Yixin Ren^{1,2†}, Yun Teng¹, Han Wu¹, Jingsu Xue², Lulu Chen², Xiaoyue Song¹, Yan Li¹, Ying Zhou¹, Zongran Pang^{1,3*} and Hao Wang^{1,2,3*}

¹School of Pharmacy, Minzu University of China, Beijing, China, ²Institute of National Security, Minzu University of China, Beijing, China, ³Key Laboratory of Ethnomedicine (Minzu University of China), Ministry of Education, Beijing, China

Protein arginine methyltransferases (PRMTs) play crucial roles in gene regulation, signal transduction, mRNA splicing, DNA repair, cell differentiation, and embryonic development. Due to its significant impact, PRMTs is a target for the prevention and treatment of various diseases. Among the PRMT family, PRMT1 is the most abundant and ubiquitously expressed in the human body. Although extensive research has been conducted on PRMT1, the reported inhibitors have not successfully passed clinical trials. In this study, deep learning was employed to analyze the characteristics of existing PRMTs inhibitors and to construct a classification model for PRMT1 inhibitors. Through a classification model and molecular docking, a series of potential PRMT1 inhibitors were identified. The representative compound (compound 156) demonstrates stable binding to the PRMT1 protein by molecular hybridization, molecular dynamics simulations, and binding free energy analyses. The study discovered novel scaffolds for potential PRMT1 inhibitors.

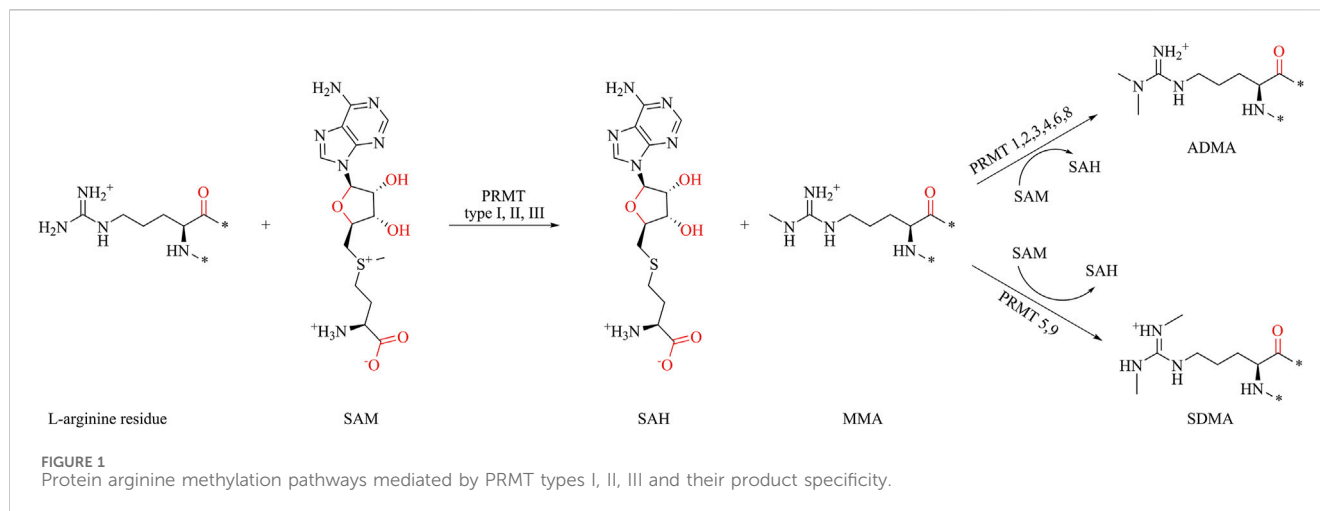
KEYWORDS

PRMT1, machine learning, molecular docking, molecular dynamics simulation, molecular hybridization

1 Introduction

Protein arginine methylation, which is catalyzed by protein arginine methyltransferases (PRMTs), is the main mechanism regulating the function of eukaryotic cells. This methylation affects epigenetic gene regulation, signal transduction, mRNA splicing, DNA repair, cell differentiation and embryonic development (Yang and Bedford, 2013; Guccione and Richard, 2019).

S-adenosyl-L-methionine (SAM) acts as a methyl group donor, transferring the methyl group to the guanidine nitrogen atom of protein arginine, resulting in the formation of methylated arginine and S-adenosyl-L-homocysteine (SAH) (Figure 1) (Tang et al., 2000b; Passos et al., 2006; Cho et al., 2012; Zhang et al., 2015). PRMTs are classified into three types based on catalytic arginine products. All PRMTs can catalyze the methyl transfer of SAM to arginine, forming monomethylated arginine (MMA). The type I PRMTs catalyze the



formation of asymmetric dimethylarginine (ADMA) from MMA, and the type II PRMT catalyzes the formation of symmetric dimethylarginine (SDMA). Type III PRMT only catalyzes the methylation of arginine, and there is no further catalysis. Until now, nine subtypes of PRMTs have been characterized, among which PRMT1, PRMT2, PRMT3, PRMT4, PRMT6, and PRMT8 belong to type I PRMTs; PRMT5 and PRMT9 belong to type II PRMTs; and PRMT7 is the only type III PRMT (Bedford and Clarke, 2009).

The arginine methylation process in *Homo sapiens* (human) cells is primarily mediated by PRMT1, followed by PRMT4 and PRMT5 (Tang et al., 2000a; Bedford and Clarke, 2009). The dysregulation of PRMT1 is associated with the occurrence and development of various diseases, including pulmonary fibrosis (Zakrzewicz et al., 2015), cardiovascular disease (Damiati, 2019), diabetes, nephropathy (Qian et al., 2018), and cancer (Cha and Jho, 2012; Stouth et al., 2017). The overexpression of PRMT1 contributes significantly to the growth, survival, metastasis, and invasion of tumor cells (Baldwin et al., 2012; Yang and Bedford, 2013; Wei et al., 2014). PRMT1 is widely distributed and expressed in human tissues and is highly expressed in cancer cells (Scorilas et al., 2000; Akter et al., 2017). Therefore, PRMT1 is considered a potential target for cancer treatment.

Currently, PRMT1 inhibitors primarily exert their effects through competitive binding with SAM or substrate competition. In addition to the competitive inhibition targeting the functional sites of PRMT1 protein itself, Spring et al. have endeavored to rationally design and synthesize PROTACs for PRMT1. However, the PROTAC developed in that study failed to induce the degradation of PRMT1 (Martin et al., 2024). GSK3368715, the most advanced PRMT1 inhibitor in clinical research, demonstrates potent inhibition of PRMT1 with an IC_{50} of 3.1 nM while showing lower activity against other PRMTs. Unfortunately, despite its potent efficacy, the clinical trial for GSK3368715 was halted due to the “overall benefit–risk profile did not support continuation of the study”, likely because 29% of the patients experienced a thromboembolic event (Fedoriw et al., 2022).

The development and clinical trial of GSK3368715 highlight the immense potential of PRMT1 as a therapeutic target (Junwei et al.,

2024). By employing a hybrid approach that integrates artificial intelligence (AI) with computer-aided drug design (CADD), virtual screening can effectively identify and eliminate potential false positives, thereby enhancing the likelihood of identifying compounds with favorable benefit–risk profiles.

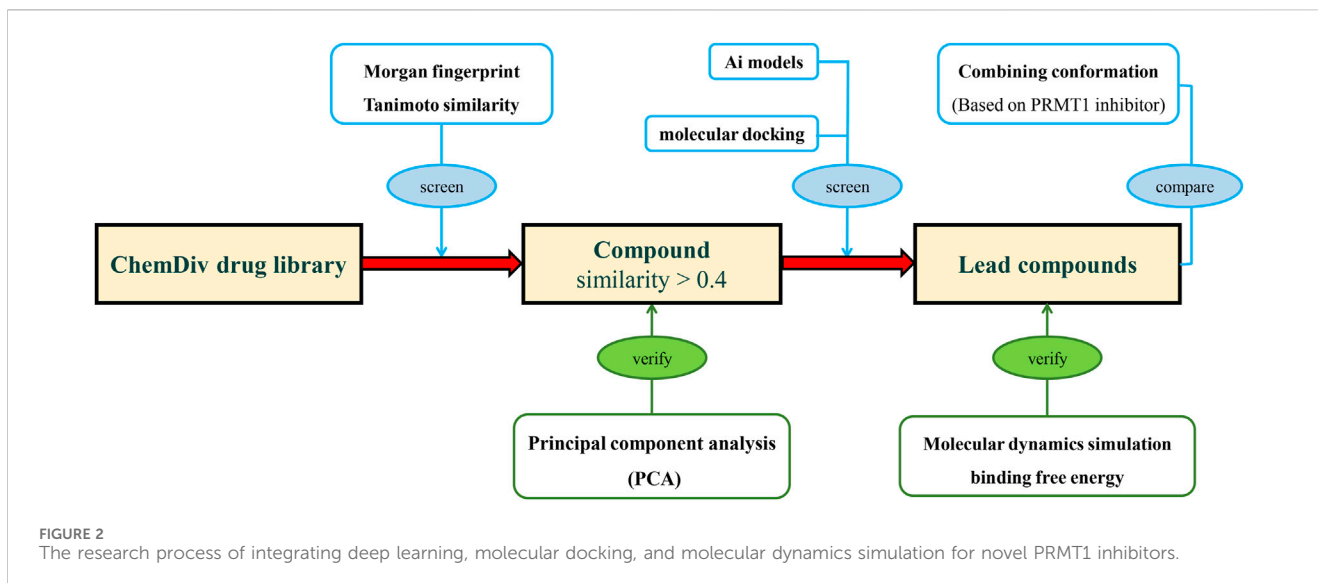
In this study, we trained an AI-based classification model based on PRMTs associated bioactivities from ChEMBL, and designed a screening process to address generalization challenges in AIDD. Initially, compounds were screened for similarity constraints to ensure their resemblance to those used in model construction, and their similarity was validated using principal component analysis (PCA). Subsequently, the classification model and molecular docking were applied to evaluate the constrained molecular database, using known PRMT1 inhibitors as a benchmark for comparison. The compounds most suitable for scaffold modification were further validated through molecular hybridization, molecular dynamics (MD) simulations, and binding free energy calculations (Figure 2).

2 Materials and methods

2.1 Preprocessing of training and screening datasets

The training dataset was collected from ChEMBL (Davies et al., 2015; Zdrazil et al., 2024). PRMTs associated bioactivities from *H. sapiens* were included in the training dataset to construct the AI screening model. Specifically, the dataset comprised bioactivity data for PRMT1, PRMT3, PRMT4, PRMT5, PRMT6, PRMT7, and PRMT8. The selected data types included IC_{50} , K_i , % inhibition, and % activity. Bioactivities without values were excluded, and compounds were annotated based on an activity threshold of 10,000 nM. For ambiguous data, such as those with greater or lesser values, a specific decision-making process (see Supplementary S1.1) was employed with manual assistance.

The SMILE of each compound in the training dataset were proofread from PubChem to obtain standard SMILE data. Salts and other extraneous components were removed to retain only the core compounds as input data. Peptides, macrocyclic molecules, and high



molecular weight compounds were removed. To address the imbalance between positive and negative samples in certain datasets, we employed stratified sampling and divided the data into training, validation, and test sets in an 8:1:1 ratio.

The screening datasets were selected from the drug-like libraries collected by our research group, primarily composed of the ChemDiv library, covering 200,000 compounds. Compounds were proofread from PubChem to obtain standard SMILES data. After removing extraneous components, the conformation of each ligand was generated. Autodock Tools 1.5.7 and OpenBabel were used to convert molecular formats (O'Boyle et al., 2011).

The PDB file of PRMT1 was obtained from the Protein Data Bank (PDB ID: 6NT2) (Jumper et al., 2021; Varadi et al., 2021). After removing extraneous components of proteins except SAH, the WHAT IF web server was used to prepare the structure of PRMT1 (Vriend, 1990).

In this study, the PCA relationships of the compounds were calculated using the following parameters: molecular weight (MW), calculated logarithm of the partition coefficient (ALOGP), polar surface area (PSA), hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), rotatable bonds (RB), and aromatic rings (AR). These parameters were calculated using RDKit (Landrum, 2024). The molecular fingerprints of the compounds were constructed using the Morgan fingerprint, which generates a 2048-bit vector representation of a molecule, capturing its structural features based on atom neighborhoods up to two bonds away. The similarity between compounds was calculated using Tanimoto similarity.

2.2 Deep learning architecture building

The single-objective and multi-objective classification models in this study were developed using the AttentiveFP from the DeepChem framework (Ramsundar et al., 2019). The model is composed of three parts: the conversion of molecules from SMILES format to graph format using MolGraphConvFeaturizer (Kearnes et al., 2016); the AttentiveFP model proposed by Xiong

et al. (2019); and a calculation scheme for AI model performance metrics in the DeepChem framework.

The MolGraphConvFeaturizer is a feature extraction tool designed for molecular graph convolution networks that transforms molecules into graphical representations for use in deep learning models. In this study, node (atom) features were constructed by concatenating various attributes to obtain a total feature length of 30. These attributes include atom type (a one-hot encoded vector for types such as "C", "N", "O", "F", "P", "S", "Cl", "Br", "I", and others), formal charge (integer electronic charge), hybridization (one-hot encoded for "sp", "sp²", "sp³"), hydrogen bonding (indicating donor or acceptor status), aromaticity (indicating aromatic ring participation), degree (a one-hot encoded vector for degrees 0–5), and number of hydrogens (a one-hot encoded vector for 0–4 hydrogens attached). Edge (bond) features are similarly constructed, with a total feature length of 11, and include bond type (one-hot encoded for "single", "double", "triple", or "aromatic"), same ring status (indicating if atoms are in the same ring), conjugation (indicating if the bond is conjugated), and stereochemistry (one-hot encoded for stereochemical configuration).

The AttentiveFP model is primarily composed of the K layer for extracting atomic features and the T layer for molecular features, with a fully connected layer for output. In the K layer, GATEConv involves atomic and edge information, and the T layer pools this information to obtain the overall molecular features. In our model, the T layer is fixed at 2. Hyperparameter optimization was performed using Optuna with 100 trial sets to target the number of K layers, graph feature size, dropout rate, learning rate, and the number of model iterations (Akiba et al., 2019).

The performance of the classification models was evaluated using the metrics AUC-ROC and AUC-PR. The activity data for PRMTs exhibit substantial missing content, as each compound is typically tested against only one or two target proteins. However, in DeepChem calculations, unannotated data are usually treated as negative. This approach conflicts with the target similarity of PRMTs. Therefore, we employed a mask to filter out invalid data, ensuring that only valid data were used in the computations.

2.3 Conformation generation by docking

Molecular docking studies were conducted using AutoDock Vina 1.2 (Trott and Olson, 2010; Eberhardt et al., 2021). The methylated arginine site was chosen as the active site and was defined as a $20 \times 20 \times 20$ Å cube box. The docking pocket centers were set as follows: with the SAH centered at $x = 8.23$, $y = 36.627$, and $z = 43.529$; without SAH centered at $x = 12.473$, $y = 30.924$, and $z = 43.529$. The exhaustiveness was set to 64. For other parameters, the default settings were used.

2.4 Molecular dynamics simulations and free binding energy calculation

Molecular dynamics (MD) simulations were performed using the AMBER99SB-ILDN force field (Lindorff-Larsen et al., 2010) implemented in the GROMACS 2019.6 program (Berendsen et al., 1995; Van Der Spoel et al., 2005; Abraham et al., 2015), and TIP3P as the water solvation model (Price and Brooks, 2004; Lu et al., 2014). The parameterization of molecules was performed with the general AMBER force field (GAFF) by Sobotop (Lu, 2024; Sousa da Silva and Vranken, 2012). A cube box (with a 0.8 nm buffer distance between the box wall and the nearest solute atoms) was created, and periodic boundary conditions were enabled. A water model was added to the container at a density of 1,000 g/L. The water was replaced by sodium and chlorine ions, aiming to electronically neutralize the system. The system was first minimized through the steepest descent minimization approach (Grubmüller et al., 1991). After that, the restricted molecular dynamics simulation was used to release any restraints. In this restricted molecular dynamics simulation, the temperature of the system was slowly increased to 298.15 K by 500 ps. Lastly, the free dynamic simulations were performed using the Verlet algorithm (Grubmüller et al., 1991). The integration step was set at 0.002 ps. The simulations were performed in an isothermal isobaric regime at 298.15 K and under 1 bar of pressure, with temperature and pressure controlled with the V-rescale and Parrinello-Rahman methods (Berendsen thermostat in annealing) (Parrinello and Rahman, 1980), respectively, and PBC (periodic boundary condition) was enabled. The root mean square deviation (RMSD) was calculated for protein–protein and protein–molecule interactions. MD trajectories were viewed using VMD software (Humphrey et al., 1996).

The binding free energy of the protein–ligand complex was performed by the gmx_MMPBSA package using the molecular mechanics/Poisson-Boltzmann (generalized-Born) surface area method (Massova and Kollman, 2000; Valdés-Tresanco et al., 2021).

2.5 ADMET predictions

The absorption, distribution, metabolism, excretion, and toxicity properties were determined using ADMETlab3, a software suite comprising models derived from 88 different datasets containing 375,187 molecules. The R^2 of each model was greater than 0.95, indicating that the prediction result was statistically significant (Dong et al., 2018).

3 Results and discussion

3.1 Deep-learning model training and evaluation

Based on the parameters of molecular weight (MW), calculated logarithm of partition coefficient (ALOGP), polar surface area (PSA), hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), rotatable bonds (RB), and aromatic rings (AR), we constructed a PCA plot to analyze the training dataset (Figure 3). In the PCA analysis, most compounds are concentrated in a central and high-density region, indicated by the yellow and green sections. The distribution of data points forms an approximately circular shape, suggesting that the variance is relatively uniform across the two principal components. Some outliers at the edges of the plot diverge from the main cluster and could unduly influence the model's training, leading to overfitting or skewed interpretations of the dataset. Outliers can introduce noise and instability in the model, detracting from its ability to generalize across the primary population of compounds. To enhance the robustness of the analysis and improve the quality of the training dataset, we removed these outliers. After this pruning process, the final training set consisted of 1,383 compounds.

We selected the Attentive FP framework as our deep learning model because of its strengths with specific compound types. However, due to the limited number of convolutional layers, this framework is not well suited for handling substructures like polypeptides and macrocyclic compounds. To enhance predictive performance, these structures were excluded from the training dataset. The same exclusion criteria were applied during the final compound screening process.

In the training dataset, activity data for additional PRMT subtypes (PRMT3, PRMT4, PRMT5, PRMT6, PRMT7, and PRMT8) were incorporated to enhance the predictive performance of the single-task model using a multi-task approach. Due to the uneven distribution of positive data and the presence of missing data in the training dataset, the improved area under the precision-recall curve (AUC-PR) metric was selected for model hyperparameter optimization to effectively evaluate the predictive capability of PRMT1. The results of the model hyperparameter optimization are presented in Table 1.

The enhancement of the multi-task model over the single-task model was examined by the area under the receiver operating characteristic curve (AUC-ROC) and AUC-PR metrics (Figure 4). It was observed that the multi-task model improved the predictive ability of PRMT1. For PRMT1 prediction, the AUC-PR increased from 0.836 in the single-task model to 0.850 in the multi-task model, an improvement of 0.014. Similarly, the AUC-ROC increased from 0.814 in the single-task model to 0.836 in the multi-task model, an improvement of 0.022. For all tasks, the multi-task model achieved an AUC-PR of 0.970 and an AUC-ROC of 0.923. Furthermore, the PCA analysis conducted on the newly introduced molecules in the multi-task model, as compared to the existing molecules in the single-task model (Figure 4), revealed that the multi-task model significantly bolstered the robustness of the single-task model. This enhancement mitigated the risk of overfitting to the data of individual tasks and reduced sensitivity to noise and outliers. The results of the AUC-PR

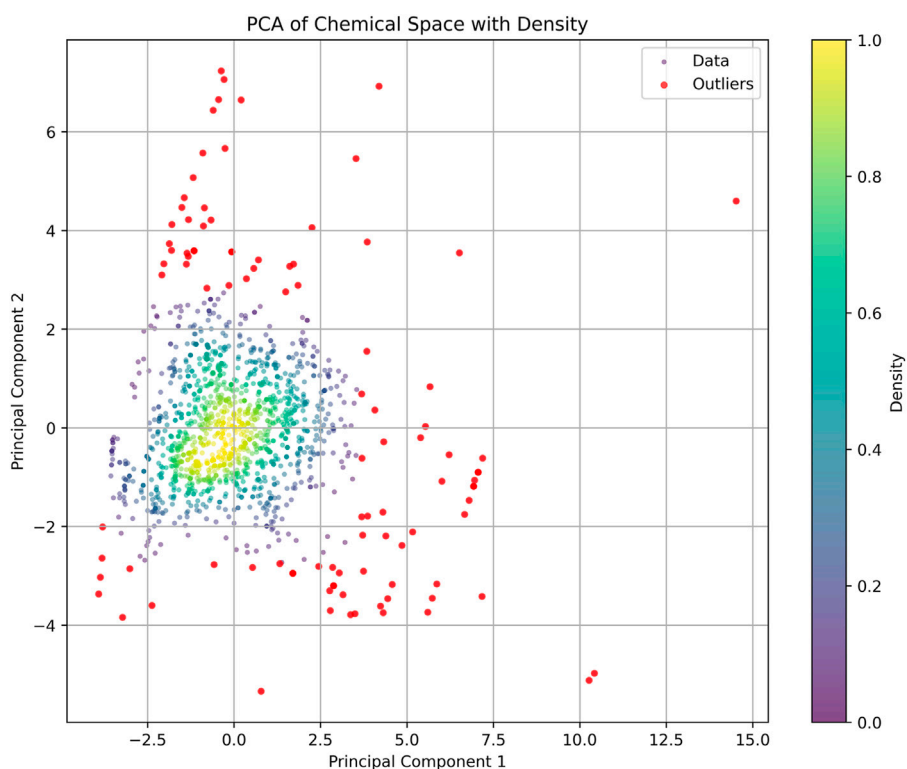


FIGURE 3

Principal component analysis (PCA) of the training dataset: The scatter plot shows the distribution of compounds based on their properties. Each point represents a compound, with color intensity indicating density. Red point: discrete compounds.

TABLE 1 The results of the model hyperparameter optimization.

Parameter	num_layers	Graph_size	Dropout	lr	Epoch
Single-Objective	2	235	0.2578	0.001768	92
Multi-Objective	3	475	0.0705	0.000685	50

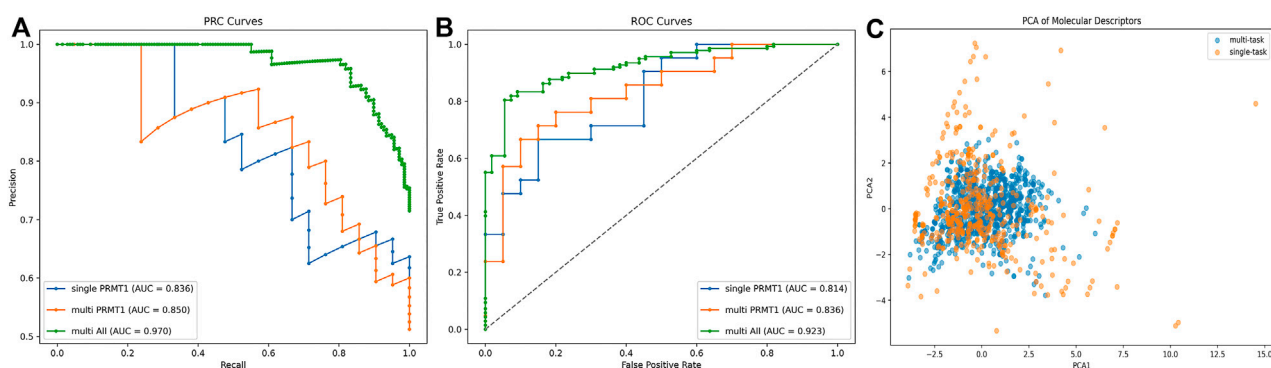
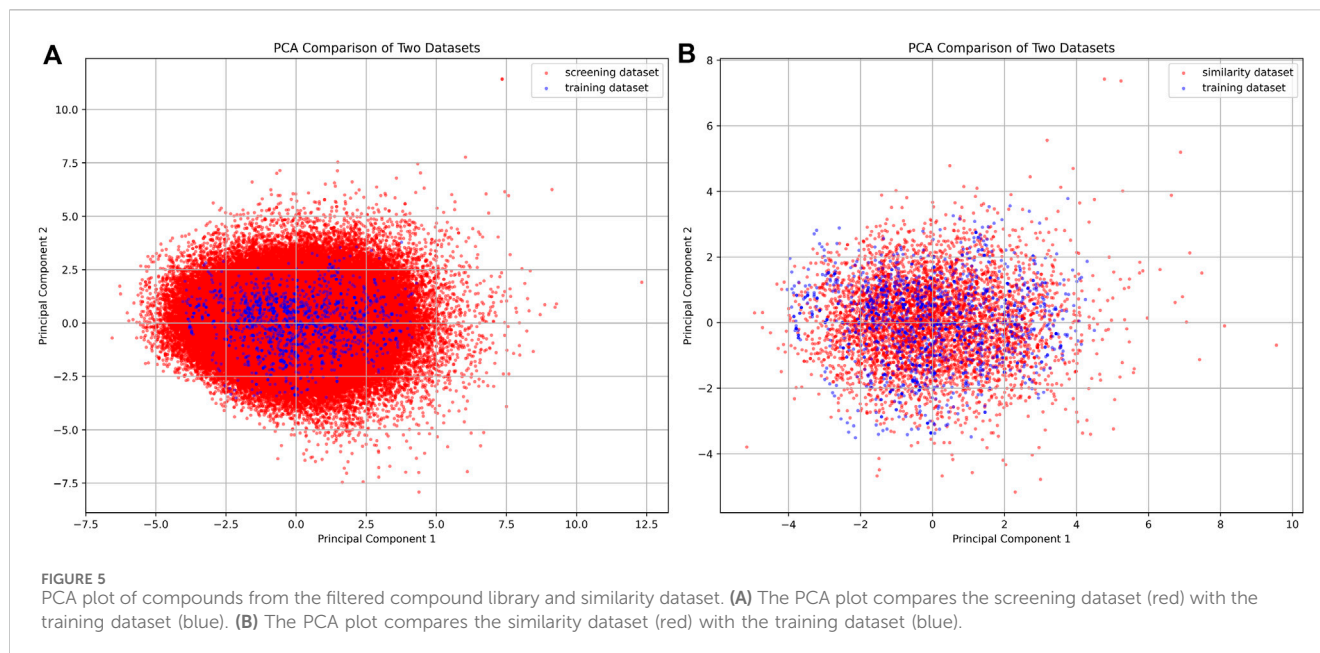


FIGURE 4

(A, B) Comparative analysis of PRC and ROC curves for single-task vs. multi-task classification models in PRMT1 prediction. (C) PCA comparison of existing molecules in single objective models and newly added compounds in multi-objective models.



demonstrate that the activity data of PRMTs can be generalized and that transfer learning effects can enhance the predictive ability of PRMT1. By accurately recognizing other tasks, the predictive capacity of the PRMT1 model was improved.

3.2 Virtual screening of PRMT1 inhibitors

AI models contain limited generalization abilities because their recognition capabilities depend on the original training dataset. To mitigate the risk of the AI model generating generalization errors, we conducted a screening of compounds using Morgan fingerprint Tanimoto similarity, selecting those with a similarity above 0.4 (Figure 5). PCA was utilized to examine the screening results, revealing that the filtered molecules are well-positioned near the training data. This suggests that the molecules identified through the similarity search have a higher likelihood of being effectively recognized by the AI model, thereby reducing the potential for generalization errors.

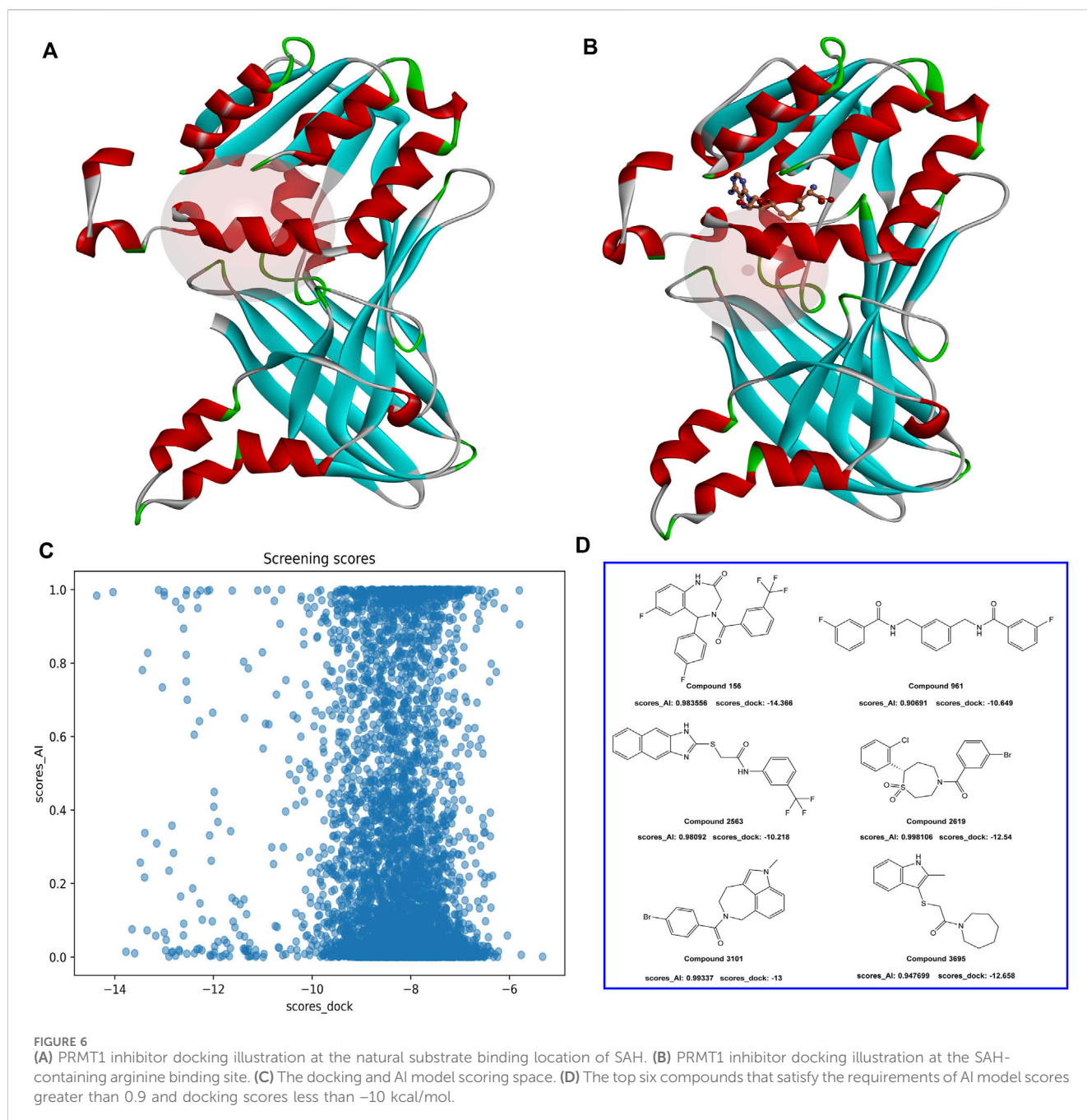
In the virtual screening process, a combined approach of docking and AI screening models was employed to evaluate the compounds from the molecular database generated by the similarity search. During the molecular docking process, we considered the multiple inhibition mechanisms of PRMT1 inhibitors, including substitution at the SAH natural substrate binding site and inhibition at the arginine binding site containing SAH. Accordingly, we performed two docking modes on the PRMT1 protein, each targeting one of the inhibitor binding mechanisms (Figure 6). In parallel, we employed our constructed multi-target classification model for PRMTs during the AI screening process. In this step, we input the compounds in SMILES format into the AI model, targeting the positive PRMT1 scoring as the primary measure for AI-based assessment. We compared the docking scores with the AI model scores. As shown in Figure 6, the evaluated molecules are well-distributed within the scoring space. Molecules with docking scores less than -10 kcal/mol and AI positive model probabilities greater than 0.9 were selected.

To investigate the binding conformation of the screened molecule within the PRMT1 protein, we aligned the docking conformation onto the reported type 1 PRMT inhibitors (PDB ID: 6NT2) (Figure 7). It is evident that the screened molecule does not act as an SAH substitute; instead, it primarily binds through the arginine pocket.

Currently, the most effective PRMT1 inhibitors reported in the literature contain a methylene-N-methylethylenediamine moiety attached to a five- or six-membered aromatic or heteroaromatic group (Hendrickson-Rebizant et al., 2024). This structural motif has demonstrated the best activity among those studied so far (Yang et al., 2017; Fedoriv et al., 2019; Wang et al., 2022). Through visual inspection, we identified compound 156 for further investigation. The selection of this compound was primarily due to its docking conformation, where it was found to be the only compound that meets the criteria for molecular hybridization involving the N-methylethylenediamine component.

Molecular hybridization was performed by connecting compound 156 with N-methylethylenediamine, which is the core structural component in PRMT1 inhibitors, to obtain compound 156-ethylenediamine (Figure 7). To assess the binding conformation of compound 156-ethylenediamine, we employed the same docking methodology as previously described for the initial screening. The binding rationalities of compound 156, compound 156-ethylenediamine, and compound GSK3368715 were further validated through MD simulations. We conducted three separate 100 ns simulations, each involving a compound in a ternary complex with SAH and PRMT1.

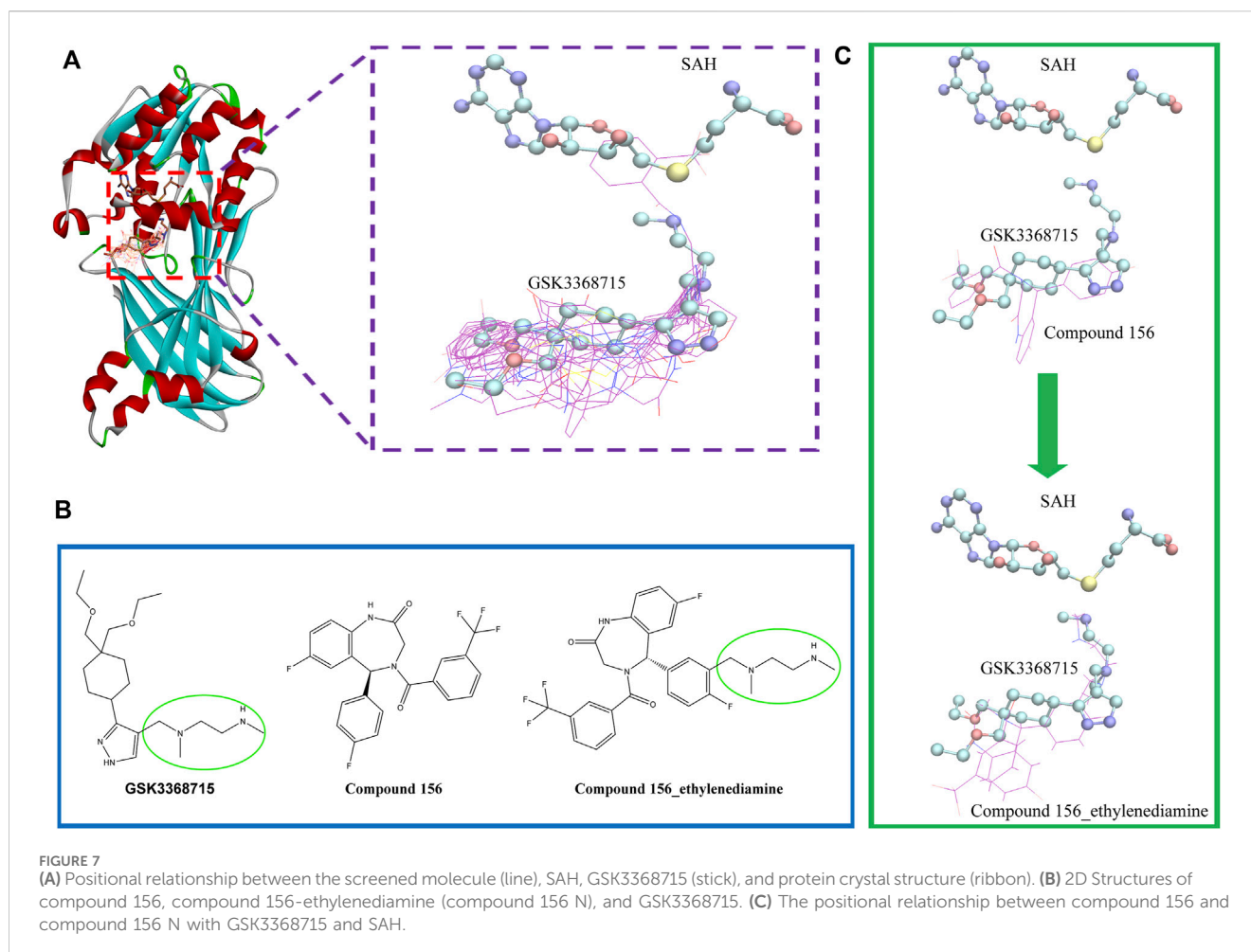
In the simulation of the positive control compound GSK3368715 (derived from the previously reported crystal structure, PDB ID: 6NT2) and the simulation of compound 156-ethylenediamine, the SAH failed to adopt a binding mode analogous to that observed in the crystal conformation. Only the simulation of compound 156 exhibited a relatively stable binding conformation for SAH, which remained near the initial binding site (Figure 8). To ascertain the simulation instability induced by GSK3368715 and



compound 156-ethylenediamine, we conducted six replicate experiments for each compound: three iterations were performed on the PRMT1 monomer, while the remaining three were executed on the PRMT1 dimer based on the crystal structure.

For the stable binding compound, as shown in **Figure 8**, compound 156 and protein underwent conformational changes at the beginning of the simulation. However, after 30 ns, as the protein conformation stabilized, the ligand maintained relative stability throughout the remaining simulation. The binding free energy was calculated using the GBSA method from 80 ns to 100 ns, which showed that the average binding free energy of compound 156 to the SAH and PRMT1 complex was -22.74 ± 0.43 kcal/mol.

For the remaining compounds with unstable binding, the binding free energy was calculated through the GBSA method over the interval of 1 ns–100 ns, revealing that the average binding free energy of compound 156 to the SAH and PRMT1 complex was lower than that of the GSK3368715 control, with values of -22.74 ± 0.43 kcal/mol for compound 156 compared to -36.05 ± 0.47 kcal/mol for GSK3368715. Compound 156-ethylenediamine exhibited comparable efficacy to GSK3368715, with values of -37.07 ± 0.44 kcal/mol, thereby indicating that compound 156-ethylenediamine can effectively bind to the inhibitor site of the PRMT1 protein.



3.3 Binding interaction analysis against PRMT1

Drawing upon the results from molecular docking studies, we meticulously investigated and validated the binding interactions of the three compounds with the active site amino acid residues of PRMT1. Through a pairwise positional comparison of the compounds (Figure 9), compound 156-ethylenediamine emerges with a novel conformation, exhibiting slight variances from its two precursor molecules.

In comparison to GSK3368715, compound 156-ethylenediamine preserves the original interacting residues Ile62 and Tyr57. The N-ethylenediamine segment remains situated within its pocket. However, the compound underwent a certain angular displacement, resulting in the forfeiture of hydrogen bonds potentially formed with Met164 while simultaneously facilitating the development of new hydrogen bonds with Tyr57.

Relative to compound 156, compound 156-ethylenediamine has experienced rotation, yet the critical residues interaction Ser56, Ile62, Tyr170, Glu171, and His311 have been retained. Moreover, the benzene ring utilized for molecular hybridization loses its π -conjugated positioning with Tyr166 due to positional dislocation.

The interactions among the residues of the three compounds are detailed in Table 2. Compound 156-ethylenediamine adeptly inherits the interactions of its two source molecules, and as the

source molecule, compound 156 reveals a promising potential for binding to the PRMT1 scaffold.

3.4 ADMET predictions

We employed the ADMETlab 3.0 computational model to forecast the absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics of the compound, which are imperative for evaluating its drug-like potential.

In terms of synthetic accessibility, GSK3368715 was assessed as “hard,” whereas compound 156 and compound 156-ethylenediamine were deemed “easy.” The molecules presented herein have been rated as relatively easy to synthesize according to AI scoring.

The compound 156 and compound 156-ethylenediamine improve the subpar Caco-2 permeability and metabolic clearance rates associated with GSK3368715 by retaining the structural framework of compound 156. However, it also inherits the latter’s elevated plasma protein binding affinity.

In predictions of acute oral toxicity in rats, compound 156 and compound 156-ethylenediamine exhibited heightened toxicity. Additionally, all three compounds tested positive for human liver toxicity and genotoxicity, although such toxicities can frequently be alleviated through optimization of the chemical structure.

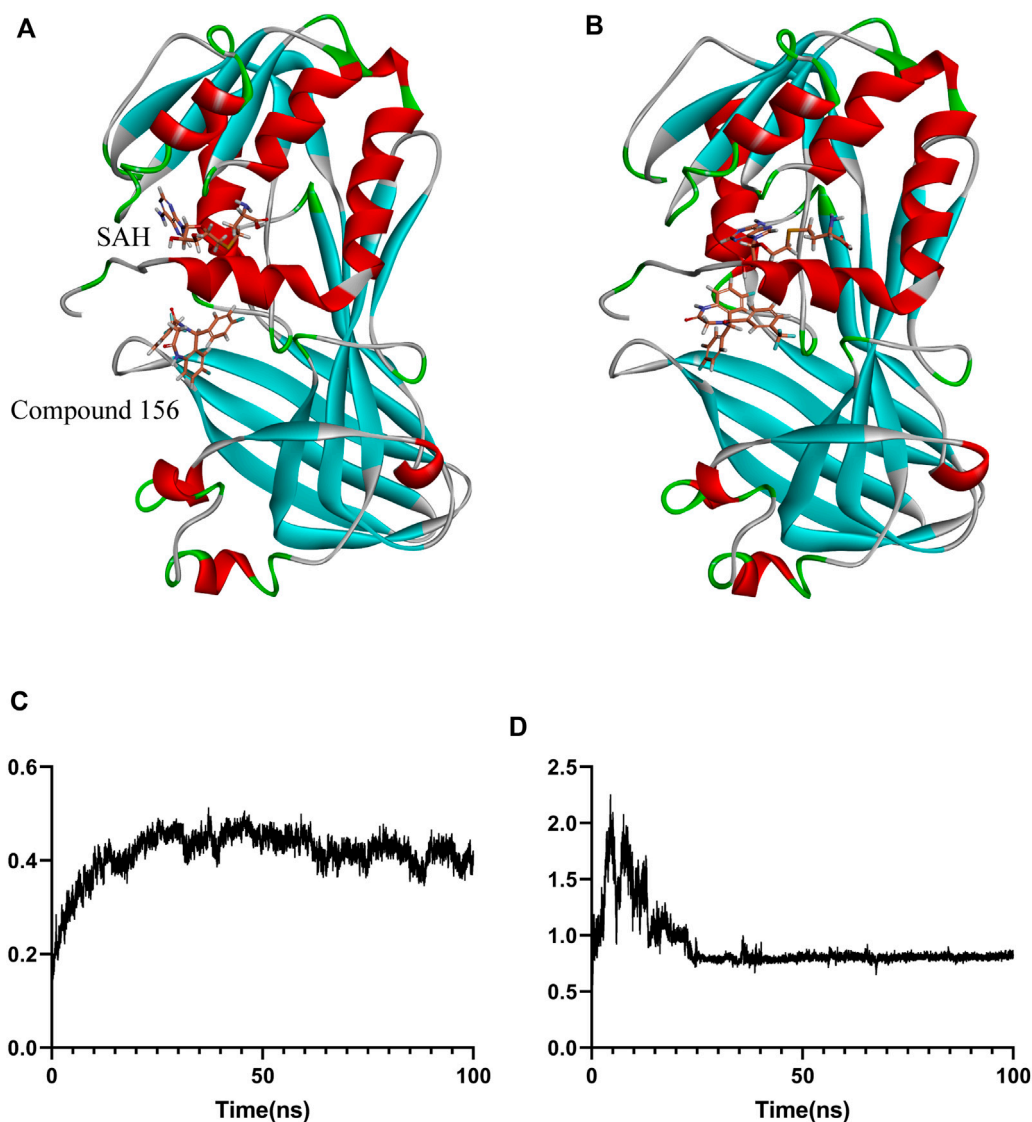


FIGURE 8
Binding positional changes and RMSD analysis of compound 156: (A) Position before simulation; (B) Position after simulation; (C) RMSD of compound 156; (D) RMSD of the protein.

The ADMET predictions affirmed that the physicochemical properties of the two compounds and GSK3368715 align with the drug-like realm delineated by the ADMETlab 3.0 platform (Figure 10).

4 Conclusion

PRMT1, a key protein responsible for arginine methylation, has received extensive research attention. Currently, PRMT-targeted drugs have not successfully completed clinical trials. This suggests that PRMT1 has a sufficient data foundation for AI design and that subtype data can enhance AI model screening capabilities. In contemporary drug development, the combination of virtual screening and AI has become an effective research strategy.

In this study, we employed a hybrid approach that combines AI screening with traditional drug screening processes. To address the

challenge of AI models recognizing out-of-distribution data that may lead to hallucination errors, we introduced fingerprint similarity to constrain the similarity between the screened and training compounds. This approach reduces the misjudgment of unrecognized molecules entering the AI scoring phase. Our final PRMT1 screening model achieved an AUC-PR of 0.850, demonstrating its ability to identify positive compounds.

Through AI model screening and molecular docking, we identified compound 156 that can cross-link with the key backbone molecule of the best PRMT1 inhibitor reported in the literature. Molecular hybridization was performed on the six-membered ring at the corresponding position of compound 156, and compound 156-ethylenediamine containing N-methylethylenediamine was obtained. Molecular docking proved that compound 156-ethylenediamine inherited the key structures of both compounds and could interact with the corresponding amino acid residues. And exhibited good stability

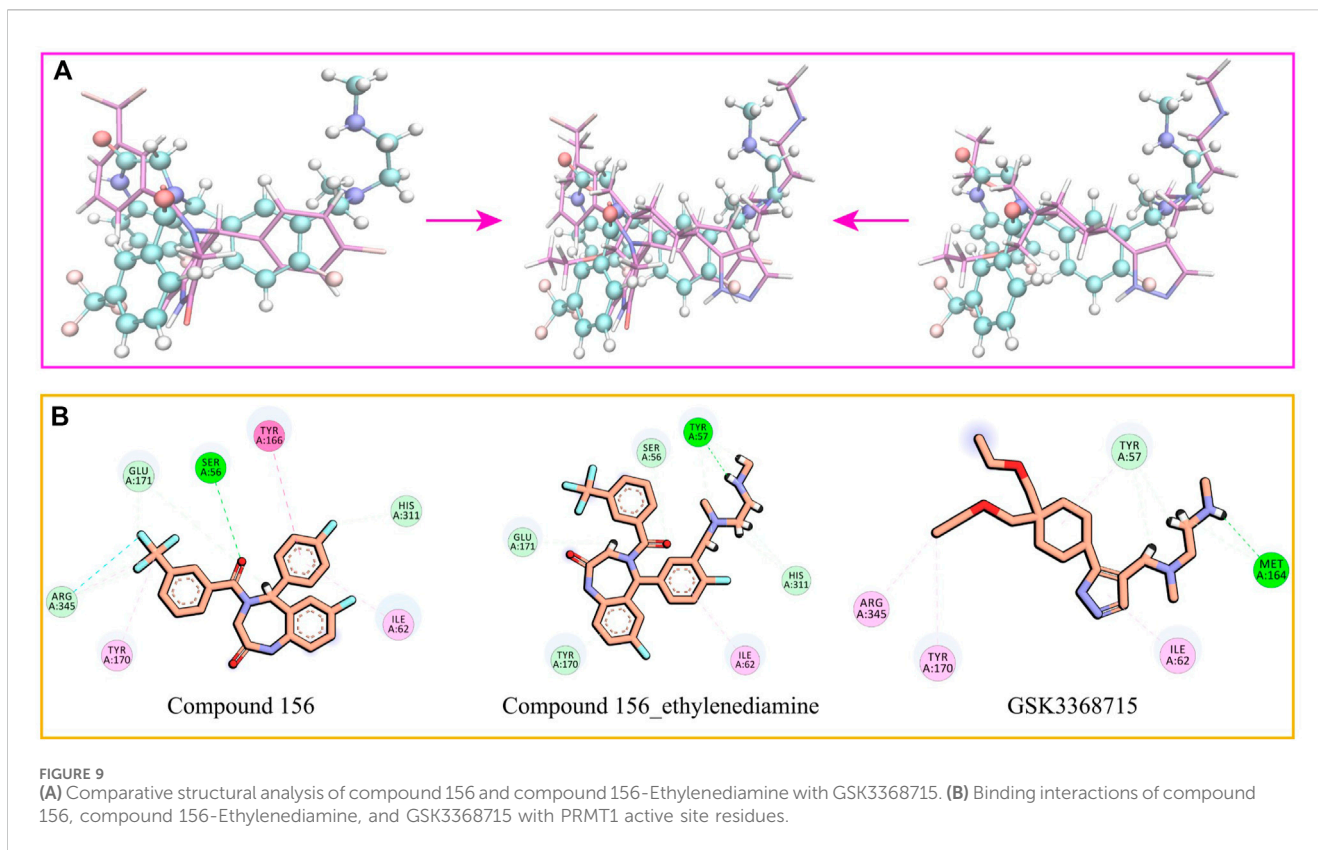
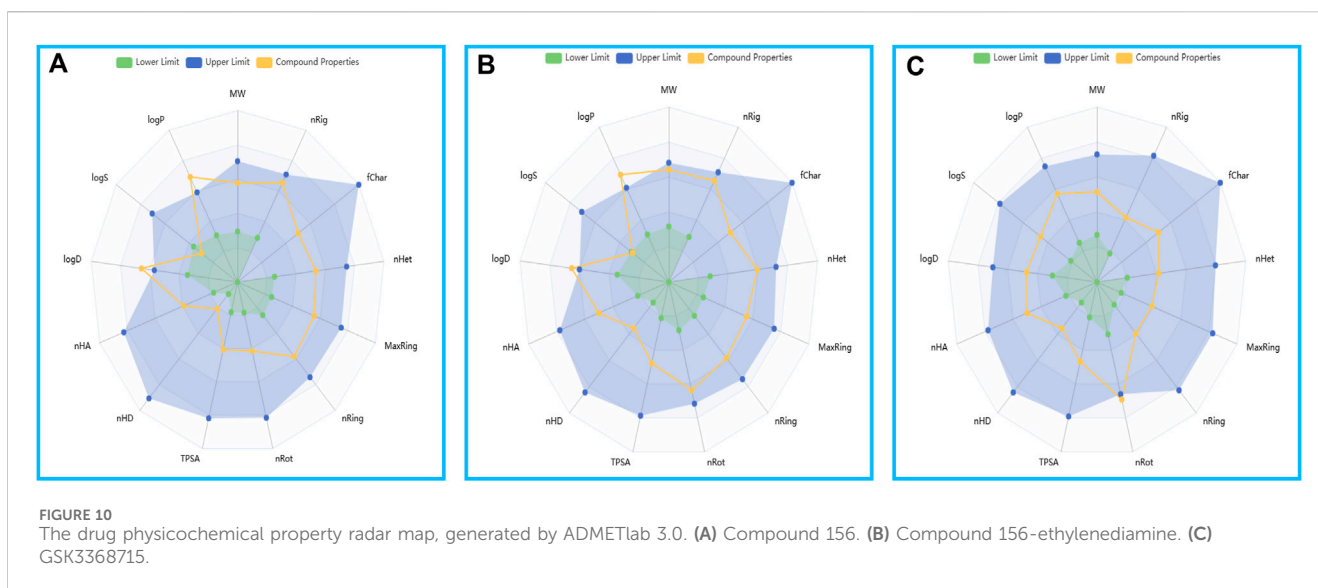


TABLE 2 Statistics of residues interacting with PRMT1 protein by three compounds.

Compounds	Hydrogen bonds and Van der Waals forces	Conjugation and hydrophobic interaction
GSK3368715	Tyr57, Met164	Ile62, Tyr170, Arg345
156-ethylenediamine	Ser56, Tyr57, Tyr170, Glu171, His311	Ile62
156	Ser56, Glu171, His311, Arg345	Ile62, Tyr166, Tyr170



during molecular dynamics simulations, proving the ability of compounds 156 to bind to the outer pocket of PRMT1 N-methylethylenediamine. The ADMET prediction indicated that the identified compounds possess favorable drug-like properties.

As a lead compound, compound 156 was shown to have the ability to bind to the PRMT1 pocket, and this ability could be transferred to the original optimal skeleton, indicating that it has the potential for further structural modification and selectivity enhancement.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding authors.

Author contributions

JZ: Writing—original draft, Writing—review and editing. YR: Writing—original draft, Writing—review and editing. YT: Writing—original draft. HWu: Writing—original draft. JJ: Writing—original draft. LC: Writing—original draft. XS: Writing—original draft. YL: Writing—original draft. YZ: Writing—original draft. ZP: Writing—review and editing. Hwa: Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1, 19–25. doi:10.1016/j.softx.2015.06.001
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: a next-generation hyperparameter optimization framework. *KDD 2019 Appl. Data Sci. track*, 2623–2631. doi:10.48550/arXiv.1907.10902
- Akter, K. A., Mansour, M. A., Hyodo, T., and Senga, T. (2017). FAM98A associates with DDX1-C14orf166-FAM98B in a novel complex involved in colorectal cancer progression. *Int. J. Biochem. Cell Biol.* 84, 1–13. doi:10.1016/j.biocel.2016.12.013
- Baldwin, R. M., Moretton, A., Paris, G., Goulet, I., and Côté, J. (2012). Alternatively spliced protein arginine methyltransferase 1 isoform PRMT1v2 promotes the survival and invasiveness of breast cancer cells. *Cell Cycle* 11 (24), 4597–4612. doi:10.4161/cc.22871
- Bedford, M. T., and Clarke, S. G. (2009). Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* 33 (1), 1–13. doi:10.1016/j.molcel.2008.12.013
- Berendsen, H. J., van der Spoel, D., and van Drunen, R. (1995). GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91 (1-3), 43–56. doi:10.1016/0010-4655(95)00042-E
- Cha, B., and Jho, E. H. (2012). Protein arginine methyltransferases (PRMTs) as therapeutic targets. *Expert Opin. Ther. Targets* 16 (7), 651–664. doi:10.1517/14728222.2012.688030
- Cho, J. H., Lee, M. K., Yoon, K. W., Lee, J., Cho, S. G., and Choi, E. J. (2012). Arginine methylation-dependent regulation of ASK1 signaling by PRMT1. *Cell Death Differ.* 19 (5), 859–870. doi:10.1038/cdd.2011.168
- Damiati, S. (2019). A pilot study to assess kidney functions and toxic dimethyl-arginines as risk biomarkers in women with low vitamin D levels. *J. Med. Biochem.* 38 (2), 145–152. doi:10.2478/jomb-2018-0025
- Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., et al. (2015). ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* 43 (1), W612–W620. doi:10.1093/nar/gkv352
- Dong, J., Wang, N. N., Yao, Z. J., Zhang, L., Cheng, Y., Ouyang, D. F., et al. (2018). ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J. Cheminform* 10, 29. doi:10.1186/s13321-018-0283-x
- Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021). AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. *J. Chem. Inf. Model.* 61 (8), 3891–3898. doi:10.1021/acs.jcim.1c00203
- Fedorow, A., Rajapurkar, S. R., O'Brien, S., Gerhart, S. V., Mitchell, L. H., Adams, N. D., et al. (2019). Anti-tumor activity of the type I PRMT inhibitor, GSK3368715, synergizes with PRMT5 inhibition through MTAP loss. *Cancer Cell* 36 (1), 100–114.e25. doi:10.1016/j.ccell.2019.05.014
- Fedorow, A., Shi, L., O'Brien, S., Smitheman, K. N., Wang, Y., Hou, J., et al. (2022). Inhibiting type I arginine methyltransferase activity promotes T cell-mediated antitumor immune responses. *Cancer Immunol. Res.* 10 (4), 420–436. doi:10.1158/2326-6066.Cir-21-0614
- Grubmüller, H., Heller, H., Windemuth, A., and Schulten, K. (1991). Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Simul.* 6 (1-3), 121–142. doi:10.1080/08927029108022142
- Guccione, E., and Richard, S. (2019). The regulation, functions and clinical relevance of arginine methylation. *Nat. Rev. Mol. Cell Biol.* 20 (10), 642–657. doi:10.1038/s41580-019-0155-x
- Hendrickson-Rebizant, T., Sudhakar, S. R. N., Rowley, M. J., Frankel, A., Davie, J. R., and Lakowski, T. M. (2024). Structure, function, and activity of small molecule and

was funded by National Natural Science Foundation of China (82173694), the Fundamental Research Funds for the Central Universities (Minzu University of China, 2023QNYL17).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Generative AI was used to grammar checker and fluency processing.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2025.1548812/full#supplementary-material>

- peptide inhibitors of protein arginine methyltransferase 1. *J. Med. Chem.* 67 (18), 15931–15946. doi:10.1021/acs.jmedchem.4c00490
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph* 14 (1), 33–38. doi:10.1016/0263-7855(96)00018-5
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Junwei, W., Deping, L., and Lifang, W. (2024). Overview of PRMT1 modulators: inhibitors and degraders. *Eur. J. Med. Chem.* 279, 116887. doi:10.1016/j.ejmech.2024.116887
- Kearnes, S., McCloskey, K., Berndt, M., Pande, V., and Riley, P. (2016). Molecular graph convolutions: moving beyond fingerprints. *J. computer-aided Mol. Des.* 30, 595–608. doi:10.1007/s10822-016-9938-8
- Landrum, G. (2024). RDKit: open-source cheminformatics.2024_09_2 (Q3 2024) release. Available at: <https://zenodo.org/records/13990314>. doi:10.5281/zenodo.13990314
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78 (8), 1950–1958. doi:10.1002/prot.22711
- Lu, J. B., Qiu, Y. Q., Baron, R., and Molinero, V. (2014). Coarse-graining of TIP4P/2005, TIP4P-ew, SPC/E, and TIP3P to monatomic anisotropic water models using relative entropy minimization. *J. Chem. Theory Comput.* 10 (9), 4104–4120. doi:10.1021/ct500487h
- Lu, T. (2024). Sobtop, version 1.0(dev3.1). Available at: <http://sobereva.com/soft/Sobtop/#citation> (Accessed September 15, 2024).
- Martin, P. L., Pérez-Areales, F. J., Rao, S. V., Walsh, S. J., Carroll, J. S., and Spring, D. R. (2024). Towards the targeted protein degradation of PRMT1. *ChemMedChem* 19 (16), e202400269. doi:10.1002/cmdc.202400269
- Massova, I., and Kollman, P. A. (2000). Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. drug Discov. Des.* 18 (1), 113–135. doi:10.1023/A:1008763014207
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: an open chemical toolbox. *J. Cheminformatics* 3 (1), 33. doi:10.1186/1758-2946-3-33
- Parrinello, M., and Rahman, A. (1980). Crystal structure and pair potentials: a molecular-dynamics study. *Phys. Rev. Lett.* 45 (14), 1196–1199. doi:10.1103/PhysRevLett.45.1196
- Passos, D. O., Bressan, G. C., Nery, F. C., and Kobarg, J. (2006). Ki-1/57 interacts with PRMT1 and is a substrate for arginine methylation. *FEBS J.* 273 (17), 3946–3961. doi:10.1111/j.1742-4658.2006.05399.x
- Price, D. J., and Brooks, C. L. (2004). A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* 121 (20), 10096–10103. doi:10.1063/1.1808117
- Qian, K., Hu, H., Xu, H., and Zheng, Y. G. (2018). Detection of PRMT1 inhibitors with stopped flow fluorescence. *Signal Transduct. Target Ther.* 3, 6. doi:10.1038/s41392-018-0009-6
- Ramsundar, B., Eastman, P., Walters, P., and Pande, V. (2019). *Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more.* O'Reilly Media, Inc.
- Scorilas, A., Black, M. H., Talieri, M., and Diamandis, E. P. (2000). Genomic organization, physical mapping, and expression analysis of the human protein arginine methyltransferase 1 gene. *Biochem. Biophys. Res. Commun.* 278 (2), 349–359. doi:10.1006/bbrc.2000.3807
- Sousa da Silva, A. W., and Vranken, W. F. (2012). ACPYPPE - AnteChamber Python parser interface. *BMC Res. Notes* 5, 367. doi:10.1186/1756-0500-5-367
- Stouth, D. W., vanLieshout, T. L., Shen, N. Y., and Ljubcic, V. (2017). Regulation of skeletal muscle plasticity by protein arginine methyltransferases and their potential roles in neuromuscular disorders. *Front. Physiol.* 8, 870. doi:10.3389/fphys.2017.00870
- Tang, J., Frankel, A., Cook, R. J., Kim, S., Paik, W. K., Williams, K. R., et al. (2000a). PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. *J. Biol. Chem.* 275 (11), 7723–7730. doi:10.1074/jbc.275.11.7723
- Tang, J., Kao, P. N., and Herschman, H. R. (2000b). Protein-arginine methyltransferase I, the predominant protein-arginine methyltransferase in cells, interacts with and is regulated by interleukin enhancer-binding factor 3. *J. Biol. Chem.* 275 (26), 19866–19876. doi:10.1074/jbc.M000023200
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31 (2), 455–461. doi:10.1002/jcc.21334
- Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., and Moreno, E. (2021). gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. *J. Chem. Theory Comput.* 17 (10), 6281–6291. doi:10.1021/acs.jctc.1c00645
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26 (16), 1701–1718. doi:10.1002/jcc.20291
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50 (D1), D439–D444. doi:10.1093/nar/gkab1061
- Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph* 8 (1), 52–56. doi:10.1016/0263-7855(90)80070-v
- Wang, C., Dong, L., Zhao, Z., Zhang, Z., Sun, Y., Li, C., et al. (2022). Design and Synthesis of Novel PRMT1 inhibitors and investigation of their effects on the migration of cancer cell. *Front. Chem.* 10, 888727. doi:10.3389/fchem.2022.888727
- Wei, H., Mundade, R., Lange, K. C., and Lu, T. (2014). Protein arginine methylation of non-histone proteins and its role in diseases. *Cell Cycle* 13 (1), 32–41. doi:10.4161/cc.27353
- Xiong, Z., Wang, D., Liu, X., Zhong, F., Wan, X., Li, X., et al. (2019). Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* 63 (16), 8749–8760. doi:10.1021/acs.jmedchem.9b00959
- Yang, H., Ouyang, Y., Ma, H., Cong, H., Zhuang, C., Lok, W.-T., et al. (2017). Design and synthesis of novel PRMT1 inhibitors and investigation of their binding preferences using molecular modelling. *J. Bioorg. Med. Chem. Lett.* 27 (20), 4635–4642. doi:10.1016/j.bmcl.2017.09.016
- Yang, Y., and Bedford, M. T. (2013). Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* 13 (1), 37–50. doi:10.1038/nrc3409
- Zakrzewicz, D., Zakrzewicz, A., Didiasova, M., Korenčak, M., Kosanovic, D., Schermuly, R. T., et al. (2015). Elevated protein arginine methyltransferase 1 expression regulates fibroblast motility in pulmonary fibrosis. *Biochim. Biophys. Acta* 1852 (12), 2678–2688. doi:10.1016/j.bbadis.2015.09.008
- Zdrzil, B., Felix, E., Hunter, F., Manners, E. J., Blackshaw, J., Corbett, S., et al. (2024). The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* 52 (D1), D1180–D1192. doi:10.1093/nar/gkad1004
- Zhang, L., Tran, N. T., Su, H., Wang, R., Lu, Y., Tang, H., et al. (2015). Cross-talk between PRMT1-mediated methylation and ubiquitylation on RBM15 controls RNA splicing. *Elife* 4, e07938. doi:10.7554/eLife.07938