

Article

Not peer-reviewed version

Chemical Space Exploration and Machine Learningbased Screening of PDE7A Inhibitors

Yuze Li , Zhe Wang , [Shengyao Ma](#) , [Xiaowen Tang](#) * , [Han-Ting Zhang](#) *

Posted Date: 20 February 2025

doi: 10.20944/preprints202502.1596.v1

Keywords: phosphodiesterase 7A inhibitor; chemical informatics; machine learning; SHAP; virtual screening



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Chemical Space Exploration and Machine Learning-based Screening of PDE7A Inhibitors

Yuze Li ¹, Zhe Wang ^{1,2}, Shengyao Ma ¹, Xiaowen Tang ^{2,3,*} and Han-Ting Zhang ^{1,2,*}

¹ Department of Pharmacology, School of Pharmacy, Qingdao University, Qingdao 266071, China

² Shandong Provincial Key Laboratory of Pathogenesis and Prevention of Brain Diseases, Qingdao University, Qingdao 266071, China

³ Department of Medical Chemistry, School of Pharmacy, Qingdao University, Qingdao 266071, China

* Correspondence: xwtang1219@qdu.edu.cn (X.T.); htzhang@qdu.edu.cn (H.T.Z.)

Abstract: Background/Objectives: Phosphodiesterase 7 (PDE7), a member of the PDE superfamily, selectively catalyzes the hydrolysis of cyclic adenosine 3',5'-monophosphate (cAMP), thereby regulating the intracellular levels of this second messenger and influencing various physiological functions and processes. There are two subtypes of PDE7, PDE7A and PDE7B, which are encoded by distinct genes. PDE7 inhibitors have been shown to exert therapeutic potentials in neurological and respiratory diseases. However, FDA-approved drugs based on PDE7A inhibitor are still absent, highlighting the need for novel compounds to advance PDE7A inhibitor development. **Methods:** To address this urgent and important issue, we conducted a comprehensive chemical informatics analysis of compounds with potential PDE7A inhibition using a curated database to elucidate the chemical characteristics of highly active PDE7A inhibitors. Specific substructures that significantly enhance the activity of PDE7A inhibitors, including benzenesulfonamido, acylamino, and phenoxy, were identified by interpretable machine learning analysis. Subsequently, a machine learning model employing the Random Forest-Morgan pattern was constructed for qualitative and quantitative prediction of PDE7A inhibitors. **Results:** As a result, 6 compounds with potential PDE7A inhibitory activity were screened out from the SPECS compound library. These identified compounds exhibited favorable molecular properties and potent binding affinities to the target protein, holding a promise as the candidates for further exploration in the development of potent PDE7A inhibitors. **Conclusions:** Results in the present study would advance the exploration of innovative PDE7A inhibitors and provide valuable insights for future endeavors in the discovery of novel PDE inhibitors.

Keywords: phosphodiesterase 7A inhibitor; chemical informatics; machine learning; SHAP; virtual screening

1. Introduction

Phosphodiesterases (PDEs) constitute a superfamily of enzymes intricately involved in the regulation of various physiological functions and processes, such as central nervous system (CNS) functions, cancer development, and inflammation[1–4]. PDEs achieve this regulatory role by specifically hydrolyzing the second messengers cyclic adenosine 3',5'-monophosphate (cAMP) and cyclic guanosine 3',5'-monophosphate (cGMP)[5,6]. The PDE superfamily, comprising 11 subfamilies classified primarily based on their structures, substrates, and distribution, exhibits diverse substrate specificities[4,6]. Notably, PDE1, 2, 3, 10, and 11 hydrolyze both cAMP and cGMP indiscriminately, whereas PDE4, 7, 8 are cAMP-specific, while PDE5, 6, 9 are cGMP-specific hydrolases[2]. PDEs have emerged as therapeutic targets for a variety of diseases, with inhibitors showcasing efficacy across various conditions[4–6]. For example, PDE1 inhibitors have been shown to mitigate adipogenesis in

mice by modulating lipolysis and adipogenic cell signaling[7]. PDE5 inhibitors are effective in treating both erectile dysfunction and heart failure[8,9]. PDE8 inhibitors have been demonstrated effective in addressing vascular dementia[10], and PDE10 inhibitors exhibit the potential to ameliorate symptoms associated with Huntington's disease[11]. Given the wide-ranging therapeutic implications of these findings, PDEs have been recognized as pivotal targets for advancements in medicinal developments.

Among the PDE subtypes, PDE4 and PDE7 are prominently expressed in the brain and play pivotal roles in diverse physiological processes associated with CNS functions. These enzymes constitute critical targets for addressing CNS disorders[12,13]. Previous investigations have highlighted the potential efficacy of PDE4 inhibitors in treating Alzheimer's disease, enhancing cognitive function, and managing alcohol addiction[13]. However, clinical utility of PDE4 inhibitors has been challenged by the adverse effects like emesis and vomiting[13]. Another cAMP-specific PDE subtype, PDE7, is emerging as a novel target in neuropharmacological drug development. The two subtypes of PDE7, PDE7A and PDE7B, which are encoded by the distinct genes, are expressed widely in the CNS but distributed variously in different brain regions. Specifically, PDE7A is broadly distributed in the cerebral cortex, hypothalamus, hippocampus, striatum, and thalamus, which makes PDE7A more prominent in the study of CNS disorders[12]. In addition, experimental findings suggest that inhibiting PDE7A produces neuroprotective effects without compromising the anesthesia duration induced by ketamine and xylazine, indicating the potential of PDE7A inhibitors in treating neurological diseases with fewer undesirable reactions such as emesis[14].

Since the identification of the first PDE7 inhibitor in 2000, diverse technological approaches have been employed in PDE7 inhibitor development[15,16]. For example, ligand-based screening methods led to the discovery of potent PDE7A inhibitors, including thiadiazine and quinazoline derivatives with notable anti-inflammatory effects in murine models[15]. Integrating pharmacophore screening and chemical modification yielded 3,4,5-trimethoxybenzyl 5-phenyl-2-furoate, a furan derivative with IC_{50} of 5.17 μ M against PDE7A, showing therapeutic efficacy in a murine model of multiple sclerosis[16]. Heretofore, numerous compounds exhibiting inhibitory activity against PDE7A have been documented, falling into diverse categories, including pyrimidinones[17] and quinazolinones[18], pyrimidine compounds[19], pyridines[20], pyridinone analogs[14], and benzenesulfonamide derivatives[21] (Figure 1). Despite the inspiring advancements, FDA-approved drugs based on PDE7 inhibitors are still absent, highlighting the needs for new screening techniques and novel compounds to advance PDE7A inhibitor development for associated diseases.

In contrast to traditional high-throughput screening, computer-aided drug design (CADD) has demonstrated rapid and efficient characteristics, becoming a vital approach in drug development[22]. Informatics analysis and artificial intelligence-based drug design (AIDD) methods have recently rejuvenated drug development[23,24]. The current study analyzed the chemical information of PDE7A inhibitors, establishing a correspondence between scaffolds, fragments/substructures, and PDE7A inhibitory activity. Subsequently, a machine learning-based predictive model for PDE7A inhibitors was constructed, leading to the screening of novel compounds with potential PDE7A inhibitory activity. This study would hold a promise for advancing the development of PDE7 inhibitors and related pharmaceuticals.

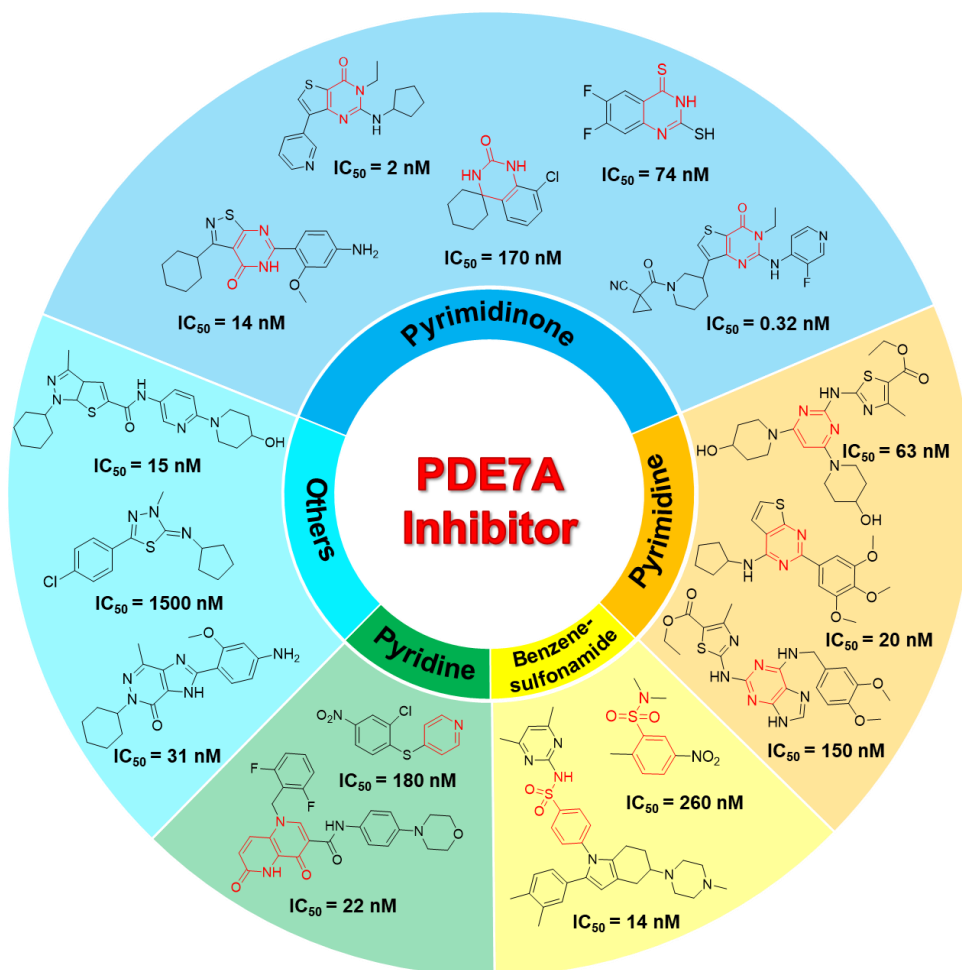


Figure 1. Exhibition and classification of PDE7A inhibitors. Characteristic skeletons were labeled in red.

2. Results and Discussion

2.1. Chemical Information Analysis

In the present study, we initiated a chemical property analysis based on the Lipinski's Rule of Five for the 596 active compounds in the database. As illustrated in Figure 2, the majority of active compounds adhered to the rules, demonstrating favorable drug-likeness for PDE7A inhibitors. Specifically, these compounds exhibited characteristics such as a molecular weight (MW) less than 500, the number of hydrogen bond donors (HBD) less than 5, the number of hydrogen bond acceptors (HBA) less than 10, a water partition coefficient (logP) less than 5, and the number of rotatable bonds (RB) less than 10.

Further exploration revealed that compounds with MW between 400 and 500 generally displayed higher inhibitory activity against PDE7A (IC₅₀, 10~100 nM). As depicted in Figure 2A, the inhibitory activity exhibited a decreasing trend as the MW decreased, with compounds having MW below 300 showing very low or even no activity (IC₅₀, 1~10 μM). In Figure 2B, the distribution of HBD for PDE7A active compounds primarily centered around 3 or below. Notably, compounds with an HBD of 2 demonstrated high inhibitory activity, while those with HBD less than 2 exhibited a bimodal distribution, indicating no clear preference for PDE7A inhibitory activity in this subset. Additionally, compounds with an HBD of 3 displayed a relatively small but significantly high activity, suggesting a potential source of potent PDE7A inhibitors. The distribution of HBA in the 596 active compounds revealed concentrations between 2~7. Higher PDE7A inhibitory activity was observed in the range of 5~7, while lower activity was observed below 4. Figure 2D illustrated a

positive correlation between logP and PDE7A inhibitory activity, emphasizing the importance of hydrophobicity for PDE7A inhibitors. Compounds with logP in the range of 4~5 were concentrated in the high activity area. As logP decreased, the distribution curve shifted to the right, and compounds with logP less than 2 were nearly inactive. Analysis of rotatable bonds (RB) in Figure 2E showed a concentration in the range of 2~7. Within this range, compounds with greater molecular flexibility (higher RB, 5~7) were more likely to exhibit higher PDE7A inhibitory activity.

In summary, compounds displaying high PDE7A inhibitory activity ($IC_{50} < 100$ nM) in the database generally possessed characteristics such as higher molecular weight (MW, 400~500) and hydrophobicity (logP, 4~5), a moderate number of hydrogen bond donor (HBD, 2~3) and acceptor (HBA, 5~7), and moderate molecular flexibility (RB, 5~7). These findings provide valuable insights into the structural and physicochemical attributes associated with potent PDE7A inhibition.

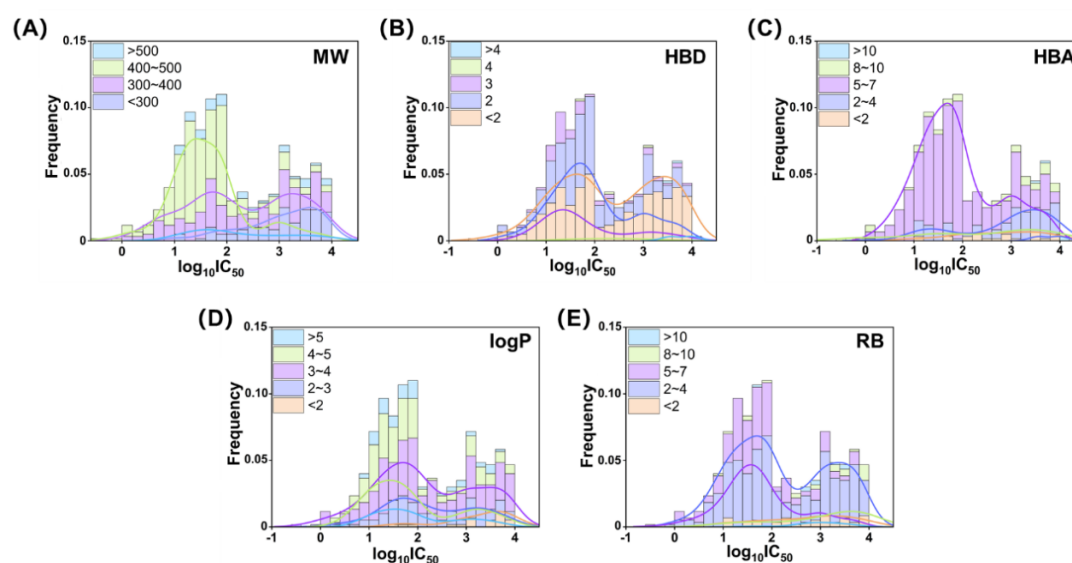


Figure 2. Distributions of PDE7A inhibitory activity against molecular weight (A), the number of hydrogen bond donor (B) and acceptor (C), logP (D), and the number of rotatable bond (E). The activity was presented in logarithmic form ($\log_{10} IC_{50}$), and the unit was used in nM.

2.2. Murcko Scaffold Analysis

Nine of the most prevalent Murcko scaffolds among the 752 compounds in the database (596 active and 156 inactive) are illustrated in Figure 3. Eight of these scaffolds (Figure 3A–H) were nitrogen-containing heterocyclic structures, such as pyrimidinone and pyridinone, aligning with previously reported PDE7A inhibitors summarized in Figure 1. Notably, diphenyl sulfide compounds, a rarity in previous studies, were discovered to exhibit slight PDE7A inhibitory activity at the micromolar scale (Figure 3I).

Among the four pyrimidinone scaffolds, isothiazopyrimidinone stood out as a prominent scaffold, with compounds containing it displaying high inhibitory activity towards PDE7A ($IC_{50} < 100$ nM), as shown in Figure 3A. Conversely, compounds containing the imidazo-pyrimidinone scaffold exhibited extremely weak or even no PDE7A inhibitory activity ($IC_{50} \sim 10$ μ M), consistent with the relatively weak activity reported for IBMX (8.10 μ M)[25] and guanine-based inhibitors (4.88 μ M)[26] in previous investigations. Additionally, two other scaffolds, thieno-pyrazole and imidazo-pyridazinone, also demonstrated remarkably high PDE7A inhibitory activity similar to isothiazopyrimidinone. Intriguingly, all three high-activity scaffolds (isothiazopyrimidinone, thieno-pyrazole, and imidazo-pyridazinone) shared a cyclohexyl-nitrogenous heterocyclic ring structure. This structural motif could potentially serve as a characteristic fragment for the further evaluation, screening, and design of PDE7A inhibitors. These findings provide valuable insights into the diversity of scaffolds contributing to PDE7A inhibitory activity, with potential implications for the

identification of novel structures and structural motifs for future drug development targeting PDE7A.

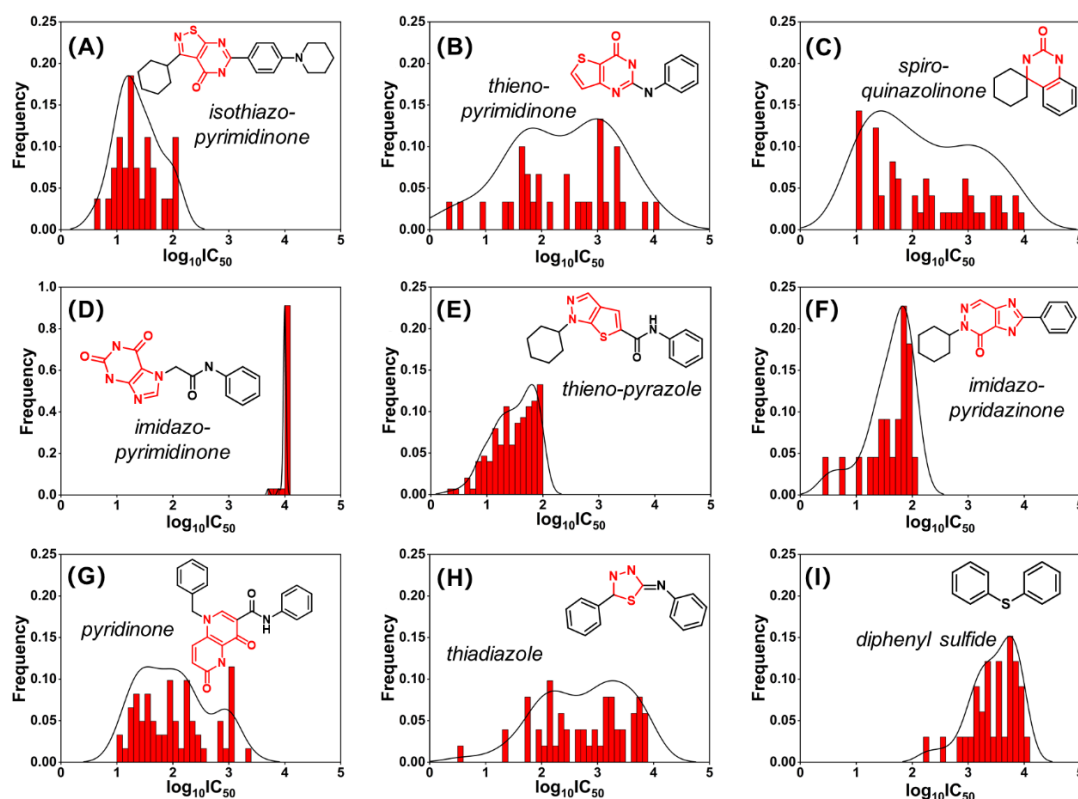


Figure 3. The most prevalent Murcko scaffolds among the active compounds in database. The scaffold feature was labeled in red. Activity was presented in logarithmic form ($\log_{10} IC_{50}$), and the unit was used in nM.

2.3. Development and Characterization of Machine Learning Models

A substantial volume of PDE7A inhibition assay data sourced from ChEMBL and PubChem facilitated the construction of a machine learning-based predictive model for PDE7A inhibitors. The present study employed a concatenated pattern, integrating qualitative prediction through a classification model and quantitative prediction through a regression model, for the construction of PDE7A inhibitors predictive model. The performance of these models was rigorously assessed through internal and external validation.

Utilizing three classical algorithms and eight molecular fingerprints, a total of 24 machine learning-based classification models were developed for the qualitative prediction of PDE7A inhibitors. Results illustrated that all 24 models exhibited remarkably high accuracy levels during both internal validation (Figure 4A and Table S3, 0.88~0.97) and external validation (Figure 4B and Table S4, 0.88~0.95). Additional evaluation criteria, including precision, recall value, and F1 score, were also consistently high as listed in Table S3~S4. The outstanding performance of these models indicated their qualification for the qualitative prediction of PDE7A inhibitors, demonstrating strong generalization capabilities.

Employing the same molecular fingerprints scale, five machine learning algorithms were introduced for constructing regression models of PDE7A inhibitors. As depicted in Figure 4C~D and Table S5~S6, R^2 of regression models constructed with XGBoost, Random Forest and Ridge algorithms (mostly more than 0.70 in both internal and external validation) outperformed those with decision tree and Lasso algorithms (mostly less than 0.50 in both validations). Model stability assessments through RMSE and MAE also underscored the superiority of XGBoost, Random Forest, and Ridge algorithms (Table S5~S6). Overall, the exceptional performance of Random Forest and XGBoost algorithms highlighted the advantage of multiple decision trees in handling complex

regression tasks. Notably, no significant disparity was observed in models using different fingerprints, indicating that the choice of machine learning algorithm may outweigh the influence of fingerprints in constructing regression models of PDE7A inhibitors.

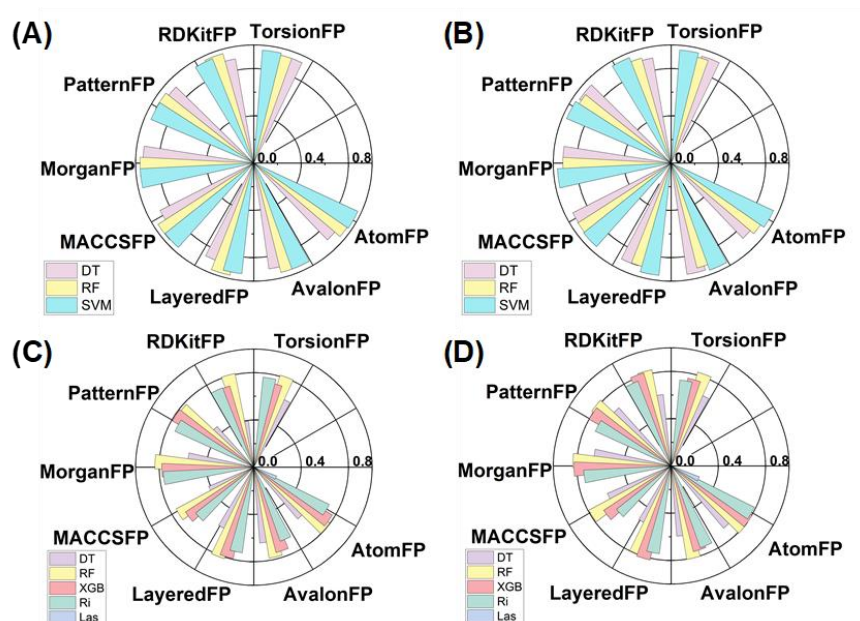


Figure 4. The performance for machine learning models trained using different algorithms based on different molecular fingerprints. Classification models in internal (A) and external (B) validation, regression models in internal (C) and external (D) validation. The performances for classification and regression models were evaluated with Accuracy and R², respectively.

Considering the comprehensive performance, the model generated by the random forest algorithm paired with Morgan fingerprint (RF-Morgan) exhibited notable accuracy (R^2 is over 0.80, the highest in all models of both internal and external validation) and high stability (shows the lowest RMSE and MAE level in the two validations). Consequently, the RF-Morgan model was selected as the quantitative model for predicting PDE7A inhibitors. Furthermore, all constructed classification models demonstrated exceptional performance in qualitatively predicting PDE7A inhibitors. Accordingly, the selected RF-Morgan model was recommended for qualitative prediction in tandem with its role in quantitative prediction. Ultimately, the RF-Morgan model was employed in subsequent interpretable machine learning analyses and further integrated into the concatenated pattern (qualitative in series with quantitative) for the prediction and screening of PDE7A inhibitors.

2.4. Interpretable Machine Learning Analysis

To enhance the interpretability of the PDE7A inhibitor prediction model, the SHapley Additive exPlanations (SHAP) method was introduced to elucidate the relationship between compound substructures and inhibitory activity on PDE7A. Figure 5 illustrates top three substructures with positive correlation to PDE7A inhibitory activity as calculated by SHAP in both the classification and regression models.

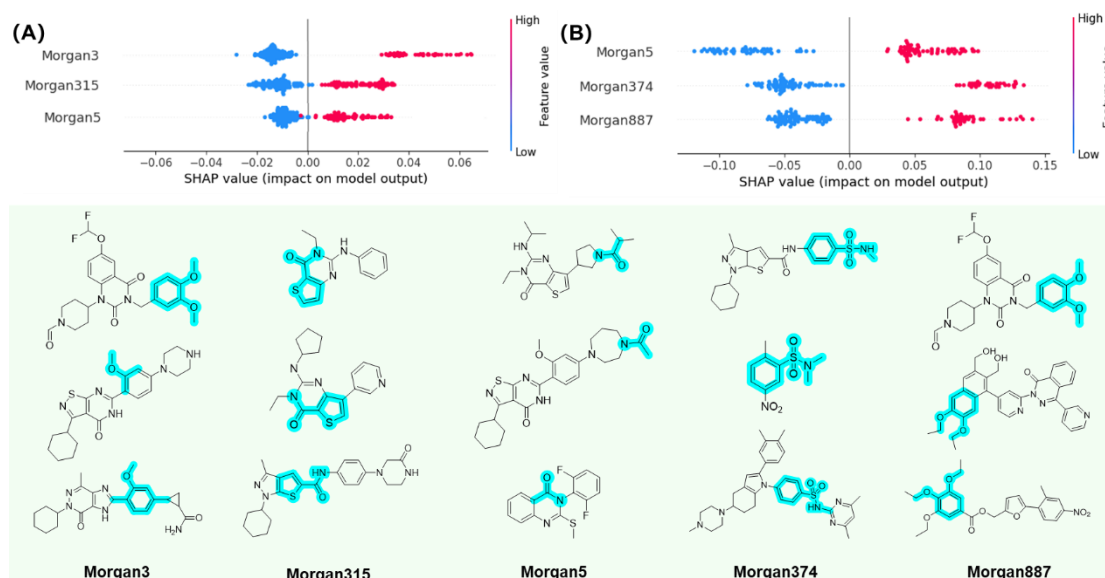


Figure 5. The performance for machine learning models trained using different algorithms based on different molecular fingerprints. Classification models in internal (A) and external (B) validation, regression models in internal (C) and external (D) validation. The performances for classification and regression models were evaluated with Accuracy and R2, respectively.

In the SHAP analysis of the classification model (Figure 5A), Morgan3, referring to phenoxyl substructure, displayed significantly high SHAP value, emphasizing the phenoxyl group as an important feature in PDE7A inhibitors. Meanwhile, Morgan315 and Morgan5, representing thiophenecarboxamide and acylamino substructure, also exhibited high SHAP values, suggesting that the acylamino group tends to be another promoter towards PDE7A inhibition. For the regression model displayed in Figure 5B, Morgan5 was identified with significant influence on PDE7A inhibitory activity once again. Actually, acylamino is the most common substructure in PDE7A inhibitors. For example, two types of conventional PDE7A inhibitors, pyrimidinone and pyridinone compounds, exactly hold acylamino as their characteristic group[14,17]. As another substructure indicated by SHAP analysis, Morgan374, representing benzenesulfonamido groups, is also commonly found in PDE7A inhibitors. For instance, BRL50481, a widely used PDE7A inhibitor with an IC₅₀ of 260 nM[21], is a typical benzenesulfonamide compound. Additionally, Morgan887 that shares structural similarities with Morgan3, also exhibited considerable impact on PDE7A inhibitory activity, which highlights the inhibitory effect of phenoxyl-based substructures on PDE7A. It is noteworthy that phenoxyl compounds have been infrequently reported in PDE7A inhibitor-related research[27], suggesting the potential to provide new insights for the design and development of future PDE7A inhibitors. The interpretability offered by the SHAP analysis enhances the understanding of key substructures influencing PDE7A inhibitory activity, facilitating more informed decisions in the design and optimization of compounds targeting PDE7A.

2.5. PDE7A Inhibitor Screening

Utilizing the explored chemical spaces and the constructed predictive model for PDE7A inhibitors, a crateriform screening pattern, incorporating the Lipinski's Rule of Five, machine learning-based qualitative and quantitative prediction, was employed for screening the SPECS commercial compound library, which consists of approximately 220,000 compounds. As displayed in Figure 6, the SPECS compound library underwent an initial filtration based on the Lipinski's Rule of Five, resulting in ~150,000 retained compounds. Subsequently, the RF-Morgan based qualitative prediction in series with quantitative prediction was utilized to predict the inhibitory activity and specific IC₅₀ values on PDE7A. This process yielded 546 compounds with predicted IC₅₀ values below 1 μ M.

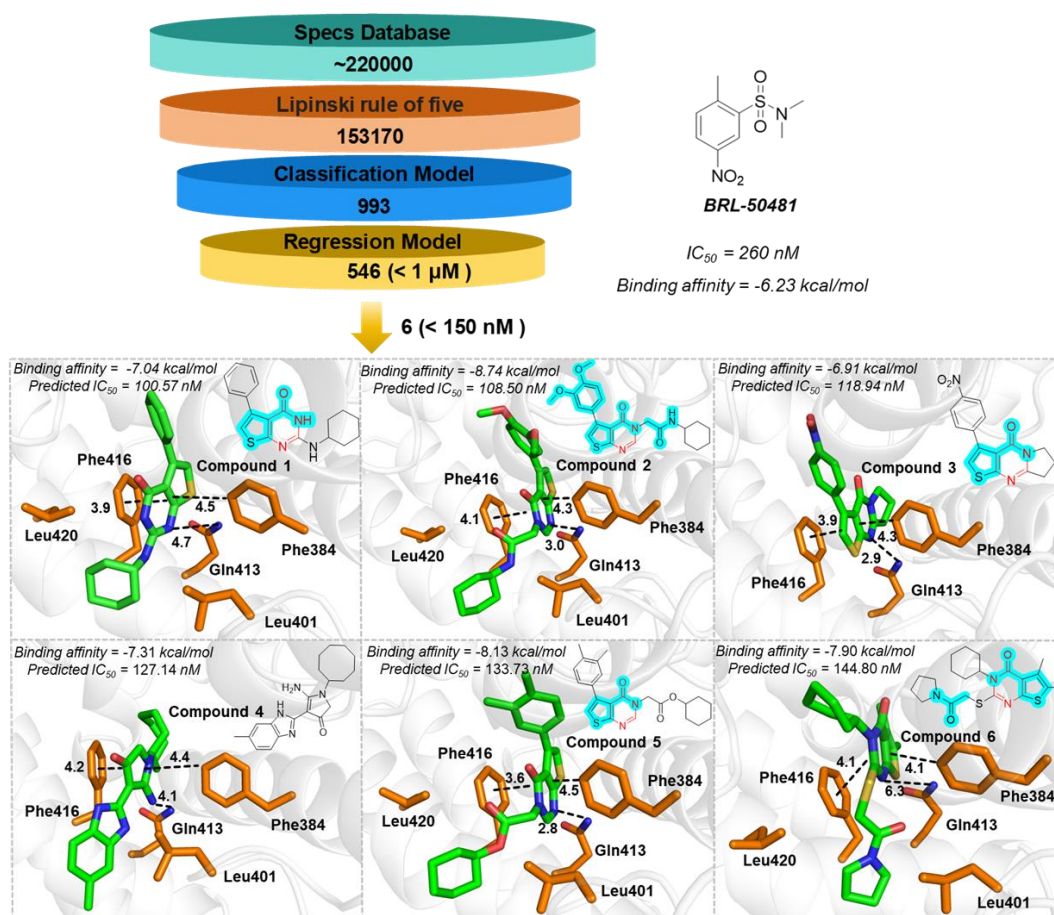


Figure 6. The crateriform screening pattern for PDE7A inhibitors and 6 identified compounds. The positive control (BRL-50481) was present in right. The pyrimidinone scaffold was labeled in red and significant substructures were highlighted in cyan.

In light of the statistical challenges presented by the abundance of compounds generated from the regression model and the pursuit of identifying more potent PDE7A inhibitors, a refined screening criterion of 150 nM was adopted, leading to the identification of 6 compounds. As depicted in Figure 6, 5 out of the 6 identified compounds were categorized as pyrimidinone derivatives, a series of compounds previously documented for their inhibitory activity against PDE7A. This observation underscores the robust generalization capability of our machine learning model in efficiently screening for PDE7A inhibitors. Additionally, compound 4 may offer novel scaffolds for the exploration of PDE7A inhibitors. Of particular note is that the 6 identified compounds encompass substructures identified as crucial contributors to PDE7A inhibitory activity, such as acylamino and phenoxy as highlighted in Figure 6. This implies once again the potential inhibitory activity against PDE7A, and molecular properties of the 6 compounds approximately adhere to the refined Lipinski's Rule of Five criteria, indicating both high PDE7A inhibitory activity and potential druggability.

Further ligand-protein interactions revealed that Phe384 and Phe416 clamp the aromatic skeleton of ligand through π - π interactions with a sandwich pattern, and Gln413 provide polar interactions with the pyrimidinone scaffold. Some aliphatic residues, like Leu401 and Leu420, form hydrophobic interactions with the aliphatic groups. The ligand-protein interactions above are consistent with the previously reported interactions of pyrimidinones with PDE7A[25,28,29]. Compared to the commonly employed PDE7A inhibitor BRL-50481, all 6 compounds exhibited higher performance on both binding affinity (docking score: about -7.00 kcal/mol vs. -6.31 kcal/mol) and predicted inhibitory activity (IC_{50} : about 100 nM vs. 260 nM) towards the target protein. This suggests the promising potential of the 6 compounds to serve as more potent PDE7A inhibitors.

3. Materials and Methods

3.1. Data Preparation

The data utilized here were sourced from PDE7A inhibition assay data available in ChEMBL and PubChem[30,31], two open source database websites that have been widely used with strong applicability[32–34]. ChEMBL played a primary role in chemical informatics analysis and in the construction of both the training and internal validation sets for machine learning, while PubChem contributed to the formation of the machine learning external validation set, supplemented by additional chemical structures obtained from previous literature[35] (detailed data were listed in Supplementary Materials_1). Before embarking on data analysis, a thorough database curation process was initiated to ensure data quality. This process involved the following steps: (1) Retention of only compound structural information (SMILES) and activity information (IC₅₀), with the presentation of activity values in logarithmic form (-log₁₀ IC₅₀)[36,37]. (2) Elimination of compounds with incomplete information, such as those lacking structural information or containing non-numeric data. In cases of compounds with repetitive information, if the activity values differed by less than ten-fold, the average value was computed; otherwise, the compound was excluded. (3) Establishment of an activity threshold of 10 μM, wherein compounds with values below 10 μM were categorized as active, while those equaled or greater than 10 μM were labeled as inactive[38]. In addition, 450 decoys obtained from DUD-E (A Database of Useful Decoys: Enhanced, <https://dude.docking.org/>)[39] were introduced, due to the imbalanced number of active and inactive data. Ultimately, a dataset with 1202 compounds (596 active and 156 inactive compounds as well as 450 decoys) was collected for the further training and testing (internal validation) of the machine learning model (Figure 7). The dataset was partitioned into training and testing sets in an 80% to 20% ratio, with the training phase further utilizing 10-fold cross-validation. Additionally, another dataset consisting of 567 compounds, comprising 525 active and 42 inactive compounds, was designated to evaluate the performance of the machine learning model (external validation, Figure 7). In addition, the active compounds extracted from training, testing and validation sets were employed for chemical informatics analysis.

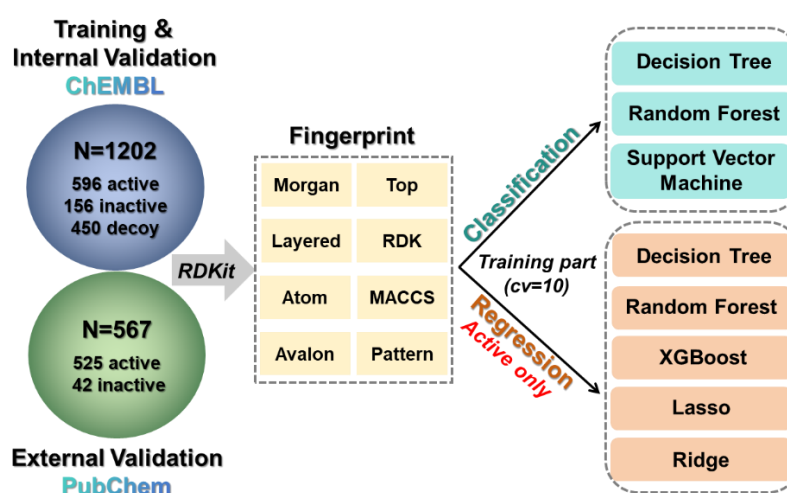


Figure 7. Composition of database and workflow of machine learning.

3.2. Molecular Feature and Fingerprint Calculation

In the present study, the Lipinski's Rule of Five and Murcko scaffold were utilized to describe the molecular features within the database[40,41]. During the extraction of Murcko scaffolds, scaffolds with a similarity greater than 0.5 were merged. Eight molecular fingerprints (RDKitFP, MorganFP, AtomFP, AvalonFP, MACCSFP, PatternFP, LayeredFP, and TorsionFP) were adopted to represent molecular structures in machine learning. During the generation of molecular fingerprints for compounds in the database, missing values were input with 0, and variance filtering was applied in cases where the variance was 0[42]. Additional information on molecular fingerprints is shown in

Table S1. In addition, RDKit was employed for the generation of all molecular features and fingerprints[32].

3.3. Machine Learning Model Construction

As depicted in Figure 7, a combined qualitative (classification model) and quantitative (regression model) scheme based on machine learning was implemented for the construction of a PDE7A inhibitor prediction model. Three algorithms (Decision Tree[43], Random Forest[44], and Support Vector Machine[45]) were used for the classification model, while five algorithms (Decision Tree, Random Forest, XGBoost[46], Lasso[47], and Ridge regression[48]) were employed for the regression model. Leveraging eight molecular fingerprints, a total of 24 classification models and 40 regression models were generated and further evaluated to establish the final PDE7A inhibitor prediction/screening scheme. The development of machine learning models utilized Scikit-learn on the Python platform (version 3.11)[49], employing a greedy strategy for optimizing hyperparameters. The best hyperparameter values for the optimal machine learning model were listed in Table S2 and that for other constructed models were summarized in Supplementary Materials_2.

3.4. Model Evaluation

The predictive ability and robustness of the constructed models were assessed through 10-fold cross-validation (internal validation) and external validation. Four evaluation criteria that refers to Precision, Recall, Accuracy, and F1 Score were introduced to evaluate the performance of the classification model. The calculation methods of them are listed as follows[50]:

$$\text{Precision} = TP / (TP + FP) \quad (1)$$

$$\text{Recall} = TP / (TP + FN) \quad (2)$$

$$\text{Accuracy} = (TN + TP) / (TN + TP + FN + FP) \quad (3)$$

$$\text{F1 Score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

where, TP (True Positive) represents correctly predicted positive data, FP (False Positive) denotes positive data with incorrect predictions, TN (True Negative) signifies correctly predicted negative data, and FN (False Negative) indicates negative data with incorrect predictions. Generally, higher Accuracy and F1 Score values tend to reflect a stronger generalization ability of the model[50].

The other three evaluation criteria that refer to coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used to assess the performance of regression model. The calculation methods of them are listed as follows[50]:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \bar{y}_i)^2} \quad (6)$$

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (7)$$

Where, m represents the total number of samples, y_i , \bar{y}_i and \hat{y}_i are the true value, average value, and predicted biological activity value of the molecule, respectively. In general, a higher R^2 (closer to 1) indicates greater precision, while a smaller RMSE or MAE (closer to 0) suggests better stability[50].

3.5. Feature Importance Analysis

The SHapley Additive exPlanations (SHAP) method was employed to enhance the interpretability of the machine learning model. SHAP values, derived from coalitional game theory, were used to analyze the contribution of each feature to the model output for various combinations of feature values. The calculation involves combining and weighting model outputs for different feature value combinations to determine the marginal contribution of each feature value to the output. This analysis aids in understanding the influence of each feature in a specific prediction and explaining the overall prediction process of the model. SHAP calculates the importance of the features by a model[51]:

$$f(x) \approx g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (8)$$

where, ϕ_0 represents the average value of all sample target variables, ϕ_i is the SHAP value of descriptor i , which represents the contribution of descriptor i to the final prediction result, $z'_i \in \{0,1\}$ indicates the presence (1) or absence (0) of descriptor i , M is the total number of molecular descriptors and $f(x)$ is the output of the original model, SHAP makes the sum of all descriptors $g(z')$ attributions approximately equal to it[51].

3.6. Molecule Docking

The initial receptor model was constructed based on the PDE7A crystal structure (PDB ID: 4Y2B)[28]. Protein and ligand parts were extracted and processed for subsequent molecular docking experiments. Compounds screened by the machine learning model underwent further optimization using the OPLS4 force field. Protonated states of ionizable groups were defined at pH 7.0 ± 0.2 , which simulated the slightly fluctuating pH conditions in the physiological environment. The protonated states of titratable residues in receptor structure were also calculated at the same pH for ligand preparation. Molecular docking analysis utilized AutoDock Vina[52], where the centroid of the ligand was defined as the center of the docking grid and the size of the grid was set to $25 \times 25 \times 25$ Å³. Finally, flexible molecular docking based on induced fit theory was executed, and results (binding mode and docking score) with the best docking score were recorded.

4. Conclusions

PDE7 inhibitors represent a novel class of drugs with promising therapeutic potential, particularly in the treatment of CNS disorders. In the present study, we conducted chemical informatics analysis to elucidate the chemical characteristics of highly active PDE7A inhibitors. We also identified scaffolds prevalent in existing PDE7A inhibitors and further recognized isothiazopyrimidinone, thieno-pyrazole and imidazo-pyridazinone derivatives as highly PDE7A inhibitory compounds. Additionally, our analysis identified specific substructures that potentially enhance the activity of PDE7A inhibitors. Subsequently, employing machine learning methods, robust RF-Morgan based classification and regression models, showcasing strong generalization capabilities, were successfully established. Utilizing these models, a refined screening of the SPECS compound library was established and used to identify 6 compounds with potential PDE7A inhibitory activity. These compounds, exhibiting favorable molecular properties and enhanced binding affinity in comparison to known PDE7 inhibitors, hold a promise as candidates for further exploration in the development of potent PDE7A inhibitors. In summary, the present work not only statistically analyzed existing PDE7A inhibitors, but also provided valuable insights and methodologies for future endeavors in the field of PDE inhibition and drug development.

Supplementary Materials: Detailed information on molecular fingerprints (Table S1), best hyperparameter of RF-Morgan in classification and regression model (Table S2), performance for classification models in internal validation (Table S3) and external validation (Table S4), performance for regression models in internal validation (Table S5) and external validation (Table S6). List of training and validation set (Supplementary Materials_1) and detailed hyperparameter of the machine learning models (Supplementary Materials_2).

Author Contributions: Y.L. contributed to data collection, initial statistical analyses, and manuscript writing and editing. Z.W. contributed to validation, data curation, investigation. S.M. contributed to formal analysis, visualization. X.T. contributed to conceptualization, methodology, resources, funding acquisition, project administration, writing-review & editing. H.T.Z. contributed to resources, supervision, funding acquisition, writing-review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by research grants of Natural Science Foundation of Shandong Province (ZR2024QH448), the Key Program of Brain Science and Technology (STI2030-Major Projects 2021ZD0202900) and Key Project by Qingdao Bureau of Sciences and Technology (22-3-3-HYGG-25-HY).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are provided within the manuscript and Supplementary Information Files.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bender, A.T.; Beavo, J.A. Cyclic nucleotide phosphodiesterases: molecular regulation to clinical use. *Pharmacological reviews* **2006**, *58*, 488-520, doi:10.1124/pr.58.3.5.
2. Zhang, H.T. Targeting phosphodiesterases (PDEs) for treatment of CNS diseases. *Current pharmaceutical design* **2015**, *21*, 271-273, doi:10.2174/138161282103141125124920.
3. Hsien Lai, S.; Zervoudakis, G.; Chou, J.; Gurney, M.E.; Quesnelle, K.M. PDE4 subtypes in cancer. *Oncogene* **2020**, *39*, 3791-3802, doi:10.1038/s41388-020-1258-8.
4. Peng, T.; Gong, J.; Jin, Y.; Zhou, Y.; Tong, R.; Wei, X.; Bai, L.; Shi, J. Inhibitors of phosphodiesterase as cancer therapeutics. *European journal of medicinal chemistry* **2018**, *150*, 742-756, doi:10.1016/j.ejmech.2018.03.046.
5. Manni, S.; Mauban, J.H.; Ward, C.W.; Bond, M. Phosphorylation of the cAMP-dependent protein kinase (PKA) regulatory subunit modulates PKA-AKAP interaction, substrate phosphorylation, and calcium signaling in cardiac cells. *The Journal of biological chemistry* **2008**, *283*, 24145-24154, doi:10.1074/jbc.M802278200.
6. Li, H.; Zuo, J.; Tang, W. Phosphodiesterase-4 Inhibitors for the Treatment of Inflammatory Diseases. *Frontiers in pharmacology* **2018**, *9*, 1048, doi:10.3389/fphar.2018.01048.
7. Kim, N.J.; Baek, J.H.; Lee, J.; Kim, H.; Song, J.K.; Chun, K.H. A PDE1 inhibitor reduces adipogenesis in mice via regulation of lipolysis and adipogenic cell signaling. *Experimental & molecular medicine* **2019**, *51*, 1-15, doi:10.1038/s12276-018-0198-7.
8. Chen, L.; Staubli, S.E.; Schneider, M.P.; Kessels, A.G.; Ivic, S.; Bachmann, L.M.; Kessler, T.M. Phosphodiesterase 5 inhibitors for the treatment of erectile dysfunction: a trade-off network meta-analysis. *European urology* **2015**, *68*, 674-680, doi:10.1016/j.eururo.2015.03.031.
9. Zhuang, X.D.; Long, M.; Li, F.; Hu, X.; Liao, X.X.; Du, Z.M. PDE5 inhibitor sildenafil in the treatment of heart failure: a meta-analysis of randomized controlled trials. *International journal of cardiology* **2014**, *172*, 581-587, doi:10.1016/j.ijcard.2014.01.102.
10. Wu, X.N.; Zhou, Q.; Huang, Y.D.; Xie, X.; Li, Z.; Wu, Y.; Luo, H.B. Structure-based discovery of orally efficient inhibitors via unique interactions with H-pocket of PDE8 for the treatment of vascular dementia. *Acta pharmaceutica Sinica. B* **2022**, *12*, 3103-3112, doi:10.1016/j.apsb.2022.02.012.
11. Beaumont, V.; Zhong, S.; Lin, H.; Xu, W.; Bradaia, A.; Steidl, E.; Gleyzes, M.; Wadel, K.; Buisson, B.; Padovan-Neto, F.E.; et al. Phosphodiesterase 10A Inhibition Improves Cortico-Basal Ganglia Function in Huntington's Disease Models. *Neuron* **2016**, *92*, 1220-1237, doi:10.1016/j.neuron.2016.10.064.
12. Chen, Y.; Wang, H.; Wang, W.Z.; Wang, D.; Skaggs, K.; Zhang, H.T. Phosphodiesterase 7 (PDE7): A unique drug target for central nervous system diseases. *Neuropharmacology* **2021**, *196*, 108694, doi:10.1016/j.neuropharm.2021.108694.
13. Peng, T.; Qi, B.; He, J.; Ke, H.; Shi, J. Advances in the Development of Phosphodiesterase-4 Inhibitors. *Journal of medicinal chemistry* **2020**, *63*, 10594-10617, doi:10.1021/acs.jmedchem.9b02170.
14. García, A.M.; Brea, J.; Morales-García, J.A.; Perez, D.I.; González, A.; Alonso-Gil, S.; Gracia-Rubio, I.; Ros-Simó, C.; Conde, S.; Cadavid, M.I.; et al. Modulation of cAMP-specific PDE without emetogenic activity: new sulfide-like PDE7 inhibitors. *Journal of medicinal chemistry* **2014**, *57*, 8590-8607, doi:10.1021/jm501090m.
15. Castaño, T.; Wang, H.; Campillo, N.E.; Ballester, S.; González-García, C.; Hernández, J.; Pérez, C.; Cuenca, J.; Pérez-Castillo, A.; Martínez, A.; et al. Synthesis, structural analysis, and biological evaluation of thioxoquinazoline derivatives as phosphodiesterase 7 inhibitors. *ChemMedChem* **2009**, *4*, 866-876, doi:10.1002/cmdc.200900043.
16. Redondo, M.; Brea, J.; Perez, D.I.; Soteras, I.; Val, C.; Perez, C.; Morales-García, J.A.; Alonso-Gil, S.; Paul-Fernandez, N.; Martin-Alvarez, R.; et al. Effect of phosphodiesterase 7 (PDE7) inhibitors in experimental autoimmune encephalomyelitis mice. Discovery of a new chemically diverse family of compounds. *Journal of medicinal chemistry* **2012**, *55*, 3274-3284, doi:10.1021/jm201720d.

17. Banerjee, A.; Yadav, P.S.; Bajpai, M.; Sangana, R.R.; Gullapalli, S.; Gudi, G.S.; Gharat, L.A. Isothiazole and isoxazole fused pyrimidones as PDE7 inhibitors: SAR and pharmacokinetic evaluation. *Bioorganic & medicinal chemistry letters* **2012**, *22*, 3223-3228, doi:10.1016/j.bmcl.2012.03.025.
18. Lorthiois, E.; Bernardelli, P.; Vergne, F.; Oliveira, C.; Mafroud, A.K.; Proust, E.; Heuze, L.; Moreau, F.; Idrissi, M.; Tertre, A.; et al. Spiroquinazolinones as novel, potent, and selective PDE7 inhibitors. Part 1. *Bioorganic & medicinal chemistry letters* **2004**, *14*, 4623-4626, doi:10.1016/j.bmcl.2004.07.011.
19. Kempson, J.; Pitts, W.J.; Barbosa, J.; Guo, J.; Omotoso, O.; Watson, A.; Stebbins, K.; Starling, G.C.; Dodd, J.H.; Barrish, J.C.; et al. Fused pyrimidine based inhibitors of phosphodiesterase 7 (PDE7): synthesis and initial structure-activity relationships. *Bioorganic & medicinal chemistry letters* **2005**, *15*, 1829-1833, doi:10.1016/j.bmcl.2005.02.025.
20. Gewald, R.; Rueger, C.; Grunwald, C.; Egerland, U.; Hoefgen, N. Synthesis and structure-activity relationship studies of dihydronaphthyridinediones as a novel structural class of potent and selective PDE7 inhibitors. *Bioorganic & medicinal chemistry letters* **2011**, *21*, 6652-6656, doi:10.1016/j.bmcl.2011.09.065.
21. Smith, S.J.; Cieslinski, L.B.; Newton, R.; Donnelly, L.E.; Fenwick, P.S.; Nicholson, A.G.; Barnes, P.J.; Barnette, M.S.; Giembycz, M.A. Discovery of BRL 50481 [3-(N,N-dimethylsulfonamido)-4-methyl-nitrobenzene], a selective inhibitor of phosphodiesterase 7: in vitro studies in human monocytes, lung macrophages, and CD8+ T-lymphocytes. *Molecular pharmacology* **2004**, *66*, 1679-1689, doi:10.1124/mol.104.002246.
22. Shao, Y.X.; Huang, M.; Cui, W.; Feng, L.J.; Wu, Y.; Cai, Y.; Li, Z.; Zhu, X.; Liu, P.; Wan, Y.; et al. Discovery of a phosphodiesterase 9A inhibitor as a potential hypoglycemic agent. *Journal of medicinal chemistry* **2014**, *57*, 10304-10313, doi:10.1021/jm500836h.
23. Wei, X.; Tang, X.; Liu, N.; Liu, Y.; Guan, G.; Liu, Y.; Wu, X.; Liu, Y.; Wang, J.; Dong, H.; et al. PyCoCa: A quantifying tool of carbon content in airway macrophage for assessment the internal dose of particles. *The Science of the total environment* **2022**, *851*, 158103, doi:10.1016/j.scitotenv.2022.158103.
24. Wei, X.; Liu, N.; Song, J.; Ren, C.; Tang, X.; Jiang, W. Effect of silica nanoparticles on cell membrane fluidity: The role of temperature and membrane composition. *The Science of the total environment* **2022**, *838*, 156552, doi:10.1016/j.scitotenv.2022.156552.
25. Wang, H.; Liu, Y.; Chen, Y.; Robinson, H.; Ke, H. Multiple elements jointly determine inhibitor selectivity of cyclic nucleotide phosphodiesterases 4 and 7. *The Journal of biological chemistry* **2005**, *280*, 30949-30955, doi:10.1074/jbc.M504398200.
26. Barnes, M.J.; Cooper, N.; Davenport, R.J.; Dyke, H.J.; Galleway, F.P.; Galvin, F.C.; Gowers, L.; Haughan, A.F.; Lowe, C.; Meissner, J.W.; et al. Synthesis and structure-activity relationships of guanine analogues as phosphodiesterase 7 (PDE7) inhibitors. *Bioorganic & medicinal chemistry letters* **2001**, *11*, 1081-1083, doi:10.1016/s0960-894x(01)00125-1.
27. Pitts, W.J.; Vaccaro, W.; Huynh, T.; Leftheris, K.; Roberge, J.Y.; Barbosa, J.; Guo, J.Q.; Brown, B.; Watson, A.; Donaldson, K.; et al. Identification of purine inhibitors of phosphodiesterase 7 (PDE7). *Bioorganic & medicinal chemistry letters* **2004**, *14*, 2955-2958, doi:10.1016/j.bmcl.2004.03.021.
28. Endo, Y.; Kawai, K.; Asano, T.; Amano, S.; Asanuma, Y.; Sawada, K.; Onodera, Y.; Ueo, N.; Takahashi, N.; Sonoda, Y.; et al. 2-(Isopropylamino)thieno[3,2-d]pyrimidin-4(3H)-one derivatives as selective phosphodiesterase 7 inhibitors with potent in vivo efficacy. *Bioorganic & medicinal chemistry letters* **2015**, *25*, 1910-1914, doi:10.1016/j.bmcl.2015.03.031.
29. Kawai, K.; Endo, Y.; Asano, T.; Amano, S.; Sawada, K.; Ueo, N.; Takahashi, N.; Sonoda, Y.; Nagai, M.; Kamei, N.; et al. Discovery of 2-(Cyclopentylamino)thieno 3,2-*d* pyrimidin-4(3*H*)-one Derivatives as a New Series of Potent Phosphodiesterase 7 Inhibitors. *Journal of medicinal chemistry* **2014**, *57*, 9844-9854, doi:10.1021/jm5008215.
30. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Motow, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic acids research* **2017**, *45*, D945-d954, doi:10.1093/nar/gkw1074.
31. Wang, Y.; Xiao, J.; Suzek, T.O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B.A.; et al. PubChem's BioAssay Database. *Nucleic acids research* **2012**, *40*, D400-412, doi:10.1093/nar/gkr1132.
32. Ding, W.; Nan, Y.; Wu, J.; Han, C.; Xin, X.; Li, S.; Liu, H.; Zhang, L. Combining multi-dimensional molecular fingerprints to predict the hERG cardiotoxicity of compounds. *Computers in biology and medicine* **2022**, *144*, 105390, doi:10.1016/j.combiomed.2022.105390.
33. Zhou, H.; Shan, M.; Qin, L.P.; Cheng, G. Reliable prediction of cannabinoid receptor 2 ligand by machine learning based on combined fingerprints. *Computers in biology and medicine* **2023**, *152*, 106379, doi:10.1016/j.combiomed.2022.106379.
34. Niu, C.; Sun, X.; Hu, F.; Tang, X.; Wang, K. Molecular determinants for the chemical activation of the warmth-sensitive TRPV3 channel by the natural monoterpenoid carvacrol. *The Journal of biological chemistry* **2022**, *298*, 101706, doi:10.1016/j.jbc.2022.101706.
35. Huang, J.X.; Zhu, B.L.; Xu, J.P.; Zhou, Z.Z. Advances in the development of phosphodiesterase 7 inhibitors. *European journal of medicinal chemistry* **2023**, *250*, 115194, doi:10.1016/j.ejmech.2023.115194.

36. Kalliokoski, T.; Kramer, C.; Vulpetti, A.; Gedeck, P. Comparability of mixed IC₅₀ data - a statistical analysis. *PLoS one* **2013**, *8*, e61007, doi:10.1371/journal.pone.0061007.
37. Feng, H.; Zhang, L.; Li, S.; Liu, L.; Yang, T.; Yang, P.; Zhao, J.; Arkin, I.T.; Liu, H. Predicting the reproductive toxicity of chemicals using ensemble learning methods and molecular fingerprints. *Toxicology letters* **2021**, *340*, 4-14, doi:10.1016/j.toxlet.2021.01.002.
38. Yan, Z.; Caldwell, G.W. Metabolism profiling, and cytochrome P450 inhibition & induction in drug discovery. *Current topics in medicinal chemistry* **2001**, *1*, 403-425, doi:10.2174/1568026013395001.
39. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of medicinal chemistry* **2012**, *55*, 6582-6594, doi:10.1021/jm300687e.
40. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews* **2001**, *46*, 3-26, doi:10.1016/s0169-409x(00)00129-0.
41. Bemis, G.W.; Murcko, M.A. The properties of known drugs. 1. Molecular frameworks. *Journal of medicinal chemistry* **1996**, *39*, 2887-2893, doi:10.1021/jm9602928.
42. Zhao, J.; Shi, X.; Wang, Z.; Xiong, S.; Lin, Y.; Wei, X.; Li, Y.; Tang, X. Hepatotoxicity assessment investigations on PFASs targeting L-FABP using binding affinity data and machine learning-based QSAR model. *Ecotoxicology and environmental safety* **2023**, *262*, 115310, doi:10.1016/j.ecoenv.2023.115310.
43. Speybroeck, N. Classification and regression trees. *International journal of public health* **2012**, *57*, 243-246, doi:10.1007/s00038-011-0315-z.
44. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5-32.
45. Nedaie, A.; Najafi, A.A. Support vector machine with Dirichlet feature mapping. *Neural networks : the official journal of the International Neural Network Society* **2018**, *98*, 87-101, doi:10.1016/j.neunet.2017.11.006.
46. Chen, T.Q.; Guestrin, C.; Assoc Comp, M. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, Aug 13-17, 2016; pp. 785-794.
47. Tibshirani, R. The lasso method for variable selection in the Cox model. *Statistics in medicine* **1997**, *16*, 385-395, doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3.
48. Tikhonov, A.N. Solution of incorrectly formulated problem and the regularization method. (trans.). **1963**.
49. Swami, A.; Jain, R.J.J.o.M.L.R. Scikit-learn: Machine Learning in Python. **2013**, *12*, 2825-2830.
50. Kirk, D.; Catal, C.; Tekinerdogan, B. Precision nutrition: A systematic literature review. *Computers in biology and medicine* **2021**, *133*, 104365, doi:10.1016/j.compbiomed.2021.104365.
51. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, Dec 04-09, 2017.
52. Trott, O.; Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010**, *31*, 455-461, doi:10.1002/jcc.21334.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.