

Discovery of Novel Anti-Acetylcholinesterase Peptides Using a Machine Learning and Molecular Docking Approach

Wei Xiao, Liu-Zhen Chen, Jun Chang, Yi-Wen Xiao

School of Life Science, Jiangxi Science & Technology Normal University, Nanchang, Jiangxi, People's Republic of China

Correspondence: Jun Chang, School of Life Science, Jiangxi Science & Technology Normal University, Nanchang, Jiangxi, People's Republic of China, Email changjun@jxstnu.edu.cn

Objective: Alzheimer's disease poses a significant threat to human health. Current therapeutic medicines, while alleviate symptoms, fail to reverse the disease progression or reduce its harmful effects, and exhibit toxicity and side effects such as gastrointestinal discomfort and cardiovascular disorders. The major challenge in developing machine learning models for anti-acetylcholinesterase peptides discovery is the limited availability of active peptide data in public databases. This study primarily aims to address this challenge and secondarily to discover novel, safer, and less toxic anti-acetylcholinesterase peptides for better Alzheimer's disease treatment.

Methods: A Random Forest Classifier model was constructed from a hybrid dataset of non-peptide small molecules and peptides. It was applied to screen a custom peptide library. The binding affinities of the predicted peptides to acetylcholinesterase were assessed via molecular docking, and top ranked peptides were selected for experimental assay.

Results: The top six peptides (IFLSMC, WCWIYN, WIGCWD, LHTMELL, WHLCVLF, and VWIIGFEHM) were selected for experimental validation. Their inhibitory effects on acetylcholinesterase were determined to be 0.007, 3.4, 1.9, 10.6, 1.5, and 3.9 $\mu\text{mol/L}$, respectively.

Discussion: Predicting anti-acetylcholinesterase peptides is challenging due to the absence of a comprehensive, publicly accessible peptide database. Traditional approaches using only non-peptide small molecules for model construction often have poor performance on predicting active peptides. Here, we developed a machine-learning model from a hybrid dataset of non-peptide small molecules and peptides, which find six potent peptides. This model was as/superior accuracy compared to small-molecule-only models reported before, but has a significant higher capability of discriminating active peptides. Our work shows that hybrid datasets can boost machine-learning model prediction in peptide drug discovery.

Keywords: Alzheimer's disease, acetylcholinesterase, machine learning, random forest classifier, Peptides

Introduction

Alzheimer's disease (AD), the most prevalent form of dementia,¹ is characterized by memory impairment, declined cognitive functions, and decreased intellect. Acetylcholinesterase (AChE), a key player in acetylcholine hydrolysis,² is closely associated with the pathophysiology of AD.³ Consequently, AChE has become a primary target for designing and screening of novel acetylcholinesterase inhibitors (AChEIs).⁴ These inhibitors alleviates the symptoms by slowing the breakdown of acetylcholine, thereby increasing its concentration and duration of action in the central nervous system, neuromuscular junctions, and autonomic ganglia.⁵ However, the growing use of AChEIs has led to a significant rise in adverse effects such as cardiovascular and gastrointestinal disorders.⁶ Thus, the development of novel AChEIs with high efficacy and minimal side effects remains an urgent need.

Machine learning (ML) has been a powerful and high-throughput tool in drug discovery,⁷ with applications in new drug development,⁸ enzyme activity prediction,⁹ and disease diagnosis.¹⁰ In AD research, ML models have been used to identify patient neuroimages,¹¹ predict endophenotypes,¹² diagnose cognitive status,¹³ and discover novel AChEIs.^{14,15} Peptides, as therapeutic agents, offer several advantages over non-peptide small molecules (NPSMs). Their higher molecule weights offer

a larger surface area for interacting protein targets and potentially reducing toxicity and side effects.¹⁶ With more flexible backbones,¹⁷ peptides exhibit superior solubility and stability. Consequently, they have lower immunogenicity, and better safety.¹⁸ Additionally, their structural similarity to natural peptides in the body enhances tolerability and reduce the risk of adverse reactions.¹⁹ Peptides are also more readily modifiable than small-molecule compounds. Their sequences can be changed or specific amino acid replaced to achieve desired modifications and precise targeting.²⁰ These attributes make peptides a valuable resource for the developing novel drugs targeting various biological functions.¹⁰

Despite the extensive use of ML for identifying non-peptide AChEIs, the discovery of anti-AChE peptides (AAPs) has primarily relied on molecular docking (MD),²¹ molecular dynamics simulations,²² and traditional web-lab experiments.²³ No ML models have been specifically developed for AAPs. The scarcity of ML-based reports on AAPs may stem from the limited availability of peptide databases and the paucity of literature on AAPs. The significant differences in the physicochemical properties between NPSMs and peptides further complicate matters, as NPSM-based ML models generally exhibit poor predictive capabilities for peptides. This limitation has impeded AAP-related research and development. Therefore, overcoming this hurdle to discover novel AAPs is crucial.

In this study, we successfully developed a ML model based on a hybrid dataset of NPSMs and a limited number of AAPs (see Figure 1 for the model architecture). This model was employed to predict potent AAPs from a self-constructed peptide library. The predicted peptides were then validated through in vitro experiments, demonstrating their inhibitory activities against AChE. Our work highlights the potential of integrating peptides into ML-based drug discovery pipelines, addressing the limitations of existing methods and paving the way for the development of novel, effective AAPs with improved therapeutic profiles.

Materials and Methods

Dataset

The datasets of AChE from multiple species, including *Homo sapiens* (ChEMBL220), *Mus musculus* (ChEMBL3198), *Rattus norvegicus* (ChEMBL3199), *Electrophorus electricus* (ChEMBL4078), *Bos taurus* (ChEMBL4768), *Torpedo californica* (ChEMBL4780), *Musca domestica* (ChEMBL5752), *Anopheles gambiae* (ChEMBL2046266), *Plutella xylostella* (ChEMBL2242729), *Drosophila melanogaster* (ChEMBL2242744), *Spodoptera litura* (ChEMBL2366422), *Streptococcus*

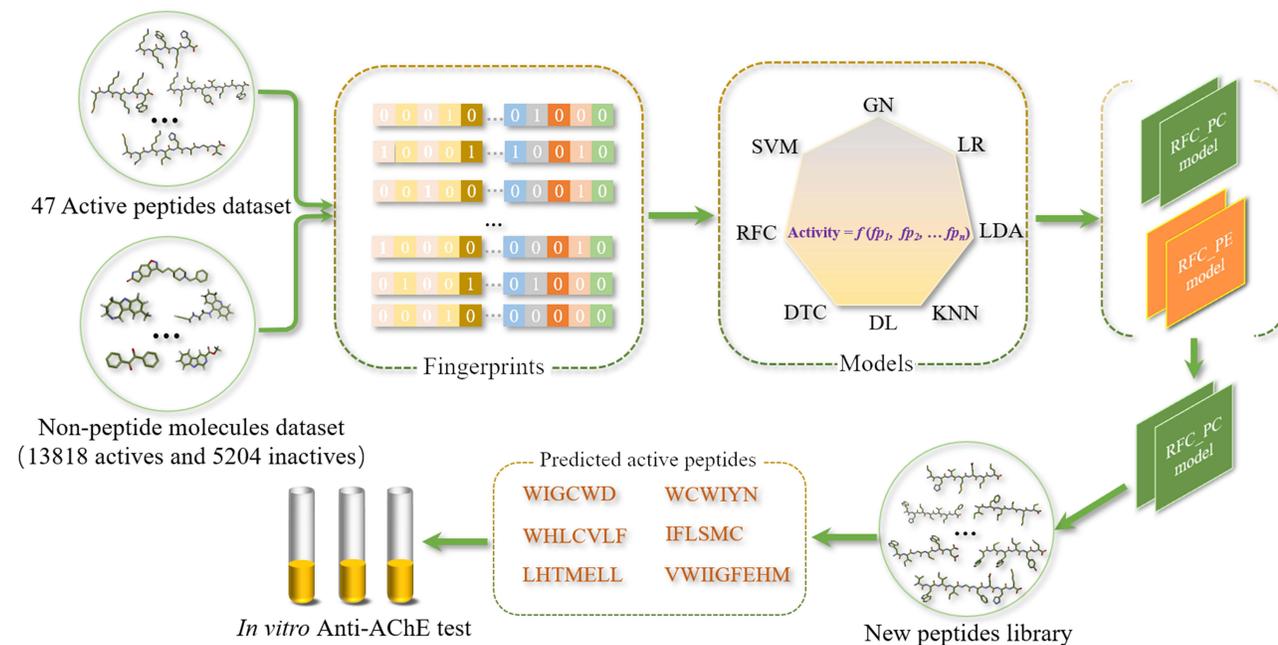


Figure 1 The architecture of screening anti-AChE peptides. GN, LR, LDA, DTC, KNN, RFC, SVM, and DL were the algorithms of GaussianNB, Logistic Regression, Linear Discriminant Analysis, Decision Trees Classifier, KNeighbors, Random Forest Classifier, Support Vector Machine, and Deep Learning, respectively. The RFC_PC and RFC_PE models were variants of the Random Forest Classifier, trained on datasets that included and excluded peptides, respectively.

Table 1 The Definition for AChEIs and Non-AChEIs

Activity Type	AChEIs	Non-AChEIs
IC ₅₀ , ED ₅₀ , EC ₅₀ , or Ki	<1000 nM	>10000 nM
Inhibition or Activity	>70%	≤70%
Num of non-peptide compounds	8614	5204
Num of AAPs	47	0

Notes: AChEIs and non-AChEIs referenced active and inactive AChE inhibitors, respectively. AAPs was anti-AChE peptides.

mutans serotypec (CHEMBL2366466), *Leptinotarsa decemlineata* (CHEMBL2366490), *Schizaphis graminum* (CHEMBL2366511), *Nephotettix cincticeps* (CHEMBL2366514), *Gallus gallus* (CHEMBL3227914), *Danio rerio* (CHEMBL3308995), and *Aedes aegypti* (CHEMBL4295607) were collected from the ChEMBL database.²⁴ Entries with reported IC₅₀, ED₅₀, EC₅₀, Ki, logIC₅₀ (converted into IC₅₀), Inhibition, or Activity values were retained. A new “Activity” feature as defined in Table 1 was created, resulting in a final dataset of 8614 AChE inhibitors (AChEIs) and 5204 non-AChEIs. For external validation of non-peptide AChEIs, a dataset from the BindingDB database (<https://www.bindingdb.org/>) was used, with a 60μM threshold to differentiate active and inactive AChEIs.

Anti-AChE peptides (AAPs) were manually curated from the literature. Peptides reported as active AChEIs or ED₅₀, EC₅₀, Ki, IC₅₀, or logIC₅₀ (converted to IC₅₀) values below 60μM were classed as AChEIs. Of 47 AAPs collected, 80% (38 peptides) were used to construct the peptide-containing model, and the remaining 20% (9 peptides) served as an external validation dataset. For validation of the peptide-excluded model, all 47 peptides were used.

Models

Non-peptide molecules were represented in SMILES, and peptide by amino acid sequences. Morgan fingerprints (2048 bits, radius 2) generated via RDKit²⁵ served as input features for training. Models were built using algorithms like Logistic Regression (LR), Linear Discriminant Analysis (LDA), KNeighbors (KNN), Decision Tree Classifier (DTC), GaussianNB (GN), Random Forest Classifier (RFC), and Support Vector Machine (SVM), and Deep Learning, with ten-fold cross validation. To evaluate the impact of peptides on models’ performance, two datasets were employed: one with peptides (8614 active AChEIs, 5204 non-AChEIs, and 47 peptides) and another without peptides (8614 active AChEIs and 5204 non-AChEIs). Additionally, the peptide dataset was duplicated 1, 2, 3, and 4 times to boost peptide weight in training data. The optimal algorithm and the peptide weight were chosen based on the model performance across these variations.

Molecular Docking

Molecular docking (MD) was employed to assess the binding affinities of the identified hits to AChE. The 3D structure of AChE²⁶ (PDB ID 1E66, resolution 2.1Å, from *Tetronarce californica*) was retrieved from the PDB database (<https://www.rcsb.org/>). The docking pocket was defined based on the active residuals interacting with huprine X, the ligand in the AChE crystal structure. MD was performed using Autodock vina 1.2.3.²⁷ The structural files of lead compounds (SDF format) and AChE (PDB format) were converted into PDBQT format using OpenBabel 2.4.1.²⁸ Docking parameters were as follows: center_x: 4.3973, center_y: 68.6326, center_z: 65.5042, size_x: 25.106, size_y: 25.106, size_z: 25. All other parameters were default.

In vitro Anti-AChE Assay

The anti-AChE activity was assessed as described by Ingkaninan²⁹ with the following modifications: a 180-μL reaction mixture was prepared by mixing 80μL DTNB (5, 5'-dithiobis (2-nitrobenzoic acid), MedChemExpress, China), 20μL of the tested AChEIs (NJPeptide, China) at various concentrations, 25μL of 0.01M PBS buffer (pH 7.4), and 20μL of AChE (0.2 U/mL). The reaction mixture was pre-incubated at 37°C for 10 minutes. Subsequently, 35μL of 7.5mM ATCI (acetylthiocholine iodide) was added to initiate the reaction for an additional 10 minutes at the same temperature. The reaction was terminated by adding 20μL of SDS, and the ODs were measured at 405 nm. For the control experiments, instead of the AChEIs solutions, 20 μL of PBS buffer was added to the reaction mixture under the same experimental

conditions. The percentage inhibitions of AChE were calculated using Equation (1). The AChE was purchased from Sigma (China), and all other reagents used in the assays were obtained from MedChemExpress (China).

$$\text{Inhibition}(\%) = \frac{OD_{\text{Control}} - OD_{\text{test}}}{OD_{\text{Control}}} \times 100 \quad (1)$$

Results

Datasets

To identify peptides or peptide analogs within the dataset, an RFC model was constructed. The training dataset included 1475 NPSMs from the ZINC database (<http://files.docking.org/2D/BA/BAAA.smi>) and 1921 randomly generated peptides (2–9 natural amino acids) using Python. The resultant model distinguished peptides from NPSMs with 0.996 accuracy. Ultimately, 85 peptide analogs ([Supplement 1](#)) were identified, but no peptides were found in the dataset. Detailed information of the model is provided in [Supplement 2](#).

In drug discovery, understanding the physicochemical properties of molecules is crucial for predicting their interactions with biological targets. Molecular properties analysis revealed significant differences between peptides and NPSMs. Peptides exhibited lower MolLogP values ([Figure 2a](#)), indicating less hydrophobicity. Thus, compared to NPSMs, peptides are more likely to bind the membrane-bound proteins or receptors embedded in lipid bilayers and less able to enter non-polar environments. Moreover, peptides have significantly higher topological polar surface area (TPSA) values ([Figure 2b](#)) than NPSMs, conferring greater polarity or solubility in the biological fluids due to more outward-facing polar atoms and functional groups. Collectively, these findings suggested that the peptides in our dataset generally have higher solubility but lower membrane permeability than NPSMs, and these peptides properties are particularly significant for medical applications. Higher solubility means peptides can interact more effectively with biological systems and remain stable in aqueous environments. However, the lower membrane permeability may restrict their oral bioavailability. These insights provide a foundation for grasping the unique peptide behavior relative to NPSMs and highlight the therapeutic importance of peptides.

[Figure 2c](#) showed the molecular weight distribution. Most molecules have molecular weights ranging from 250 to 600Da, peaking around 425Da. This suggests that the compounds have a good overall bioavailability and cell permeability. Specifically, lower-molecular-weight compounds (below 300Da) are expected to demonstrate better cell permeability and oral bioavailability. In contrast, higher-molecular-weight peptides (above 500Da) may offer high specificity and binding affinity to protein targets.

Models

The Performance of Algorithms on the Dataset

Our previous study³⁰ and reported researches^{31,32} demonstrated that the Morgan fingerprints outperformed molecular descriptors and the MACC keys in performance across various algorithms such as GN, LR, LDA, KNN, DTC, RFC, and SVM. This superior of Morgan fingerprints may stem from their ability to capture both local and global structural information of a molecule, which is crucial for accurately predicting its activity. Consequently, Morgan fingerprints were chosen as the input features of our models. Additionally, we explored the impact of the weight of data on model performance, particularly focusing on the duplication of peptide entries. The results ([Table 2](#)) revealed that the RFC model achieved the highest accuracy of 0.88, with a RSD of 0.012. Notably, no remarkable differences in model accuracies were observed across different duplication times of peptide data. Therefore, we select the RFC model for further development and set the duplication time to 1. The resulting peptide-containing RFC model was then utilized to predict peptide activities.

For a classification task, the AUC (area under the curve of the receiver operating characteristic, ROC) server a crucial metric and indicator of model performance. It comprehensively measure a model's ability to distinguish between different classes. The AUC values ([Figure 3](#)) suggested that the peptide-containing Random Forest Classifier (RFC_PC) model performed best on the test dataset, achieving the highest AUC value of 0.94 ($P < 0.01$). This was followed by KNN (AUC=0.93), LR (AUC=0.90), LDA (AUC=0.89), and DTC (AUC=0.81) models, which correlated

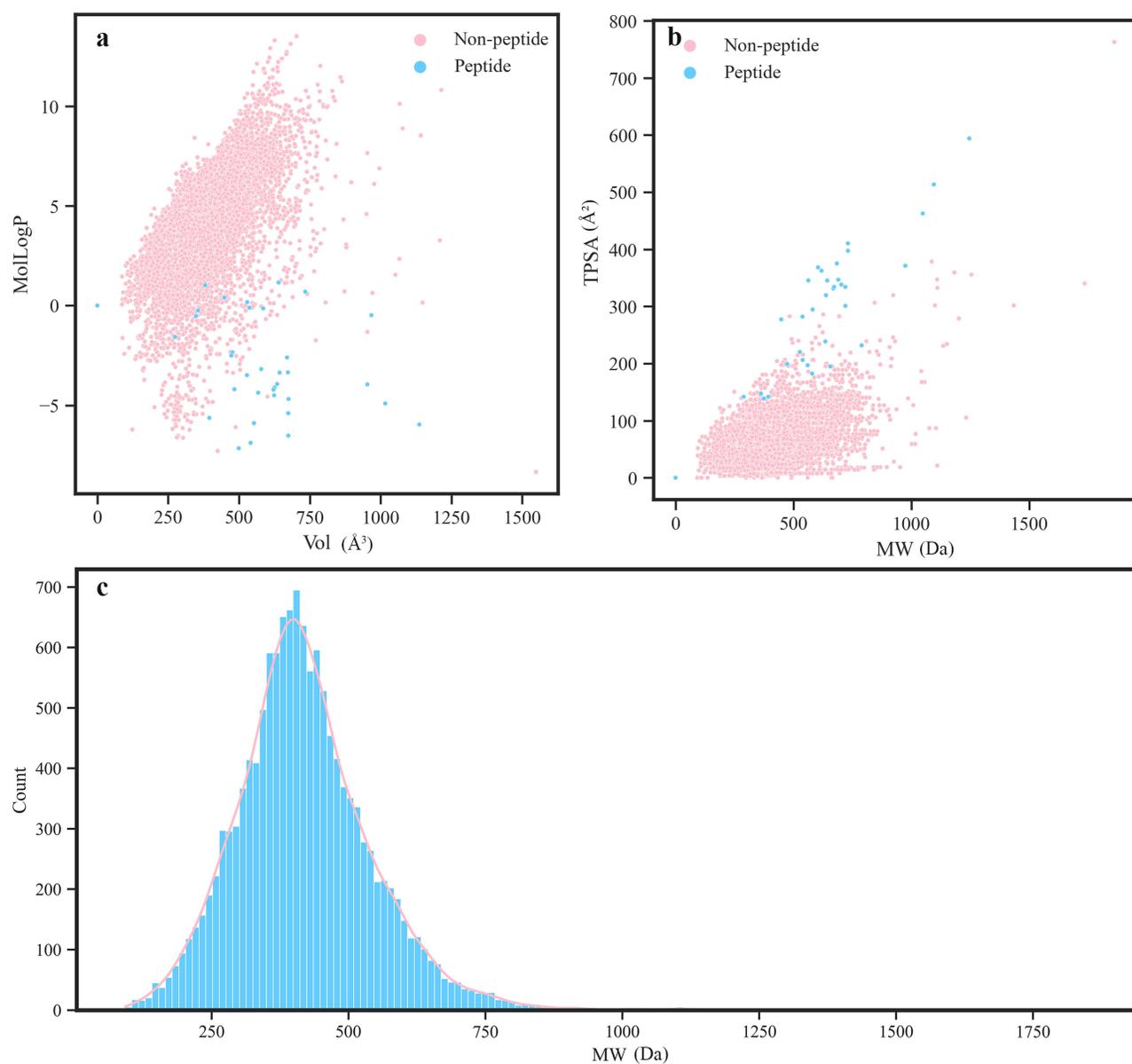


Figure 2 The relationship of MolLogP vs Vol (a), correlation of TPSA vs MW (b), and the distribution of the molecular weights (c) for peptide and non-peptide compounds in the dataset. MolLogP, TPSA, MW, Vol were \log_{10} (Octanol water partition), topological polar surface area (\AA^2), molecular weight of molecules (Da), and the volume of molecule (\AA^3), respectively. The MolLogP, TPSA, and MW were calculated using RDKit.

with their respective accuracies. In contrast, GN demonstrated the lowest AUC value of 0.75 ($P < 0.001$). The precision-recall curves of models (Figure 4) were consistent with the results of AUC values.

To further evaluate the impact of peptide inclusion, a peptide-excluded RFC (RFC_PE) model was constructed. Both the RFC_PC and RFC_PE models were subsequently tested on the training and testing datasets. The confusion matrixes of the RFC_PC model revealed F1-scores of 0.996 (precision 0.996, recall 0.997) and 0.898 (precision 0.909, recall 0.888) for the training (Figure 5a) and testing (Figure 5b) datasets, respectively. In comparison, the confusion matrixes of the RFC_PE model showed F1-scores of 0.997 (precision 0.996, recall 0.998) and 0.905 (precision 0.899, recall 0.912) for the training (Figure 5c) and testing (Figure 5d) datasets, respectively. When applied to 1836 non-peptide molecules, the RFC_PC and RFC_PE models correctly predicted 1632 (88.9%) and 1630 (88.8%) molecules, respectively. The distribution of prediction probabilities for AChEIs was compared between the two models (Figure 6a), indicating no significant difference in their performance for predicting NPSMs as AChEIs. However, a notable distinction emerged when evaluating their abilities to

Table 2 The Influence of Weights of Peptide Data on Model Performance

Algorithms	Accuracies			
	The Duplication Times of Peptide Data (n)			
	1	2	3	4
GN	0.76 (0.012*)	0.76 (0.011)	0.76 (0.012)	0.76 (0.022)
LR	0.83 (0.011)	0.83 (0.007)	0.83 (0.009)	0.83 (0.007)
LDA	0.82 (0.009)	0.82 (0.007)	0.83 (0.008)	0.83 (0.011)
KNN	0.85 (0.013)	0.86 (0.008)	0.86 (0.008)	0.86 (0.013)
DTC	0.82 (0.013)	0.82 (0.010)	0.82 (0.012)	0.82 (0.012)
RFC	0.88 (0.012)	0.87 (0.008)	0.87 (0.012)	0.87 (0.012)
SVM	0.76 (0.011)	0.76 (0.008)	0.76 (0.013)	0.76 (0.013)
Deep learning	0.86 (0.013)	0.87 (0.008)	0.87 (0.005)	0.86 (0.015)

Notes: *RSDs of the models constructed on the related algorithm. "n" is the the number of peptide duplications. GN, LR, LDA, DTC, KNN, RFC, and SVM were the algorithms of GaussianNB, Logistic Regression, Linear Discriminant Analysis, Decision Trees Classifier, KNeighbors, Random Forest Classifier, and Support Vector Machine, respectively.

predict active peptides. The RFC_PC model successfully identified all active peptides (100%) in the external validation dataset, whereas the RFC_PE model only correctly predicted 10 out of 47 active peptides (21%) (Figure 6b). Moreover, the prediction probabilities generated by the RFC_PC model were significantly higher than those of the RFC_PE model. This finding highlights the critical role of peptide data in enhancing the RFC model's performance. In summary, while both models performed similarly for NPSMs, the inclusion of peptide data in the RFC_PC model markedly improved its ability to discriminate active peptides, underscoring its superiority in this context.

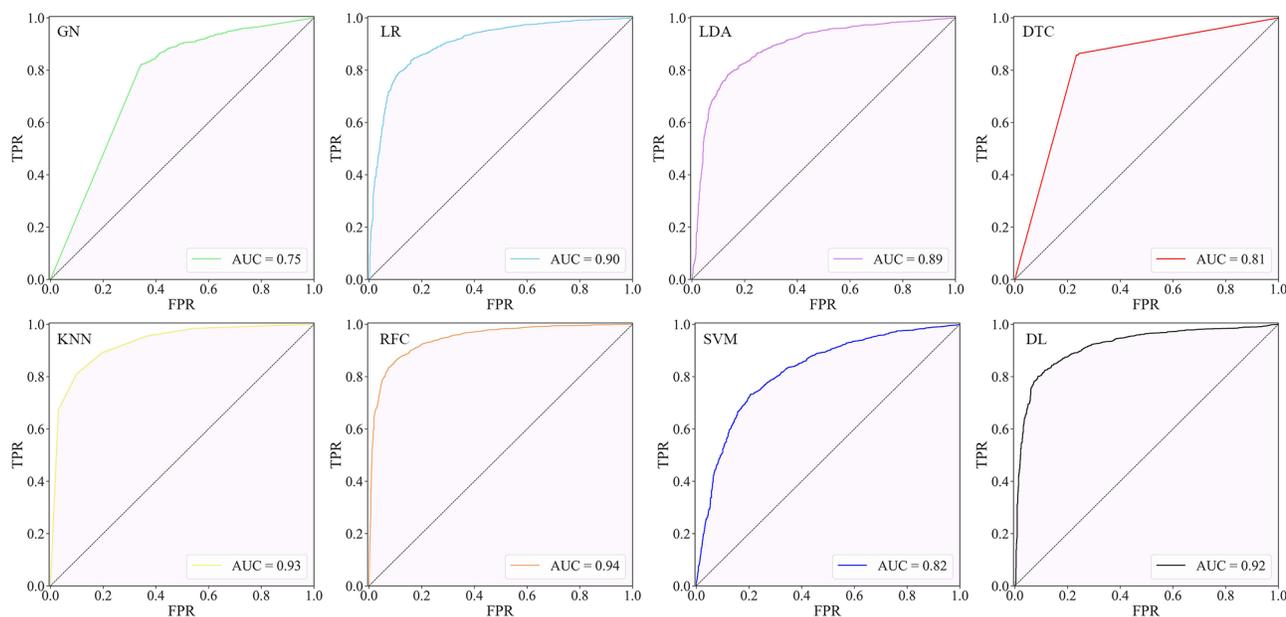


Figure 3 The ROCs of models on the test dataset ($n = 1$). GN, LR, LDA, DTC, KNN, RFC, SVM, and DL were algorithms of GaussianNB, Logistic Regression, Linear Discriminant Analysis, Decision Trees Classifier, KNeighbors, Random Forest Classifier, Support Vector Machine, and Deep Learning, respectively. The true positive rate (TPR) and false positive rate (FPR) were calculated based on the predicted and actual activities, respectively. The variable n represented the weight of peptide data. The areas under the receiver operating characteristic (ROC) curves were denoted as AUCs. The black lines in the figures served as reference lines corresponding to the equation $y = x$. The total number of samples used were 2744. The F1, precision, and recall for GN were 0.82, 0.80, and 0.83, respectively. The F1, precision, and recall for LR were 0.87, 0.87, and 0.88, respectively. The F1, precision, and recall for LDA were 0.86, 0.85, and 0.87, respectively. The F1, precision, and recall for KNN were 0.89, 0.89, and 0.89, respectively. The F1, precision, and recall for DTC were 0.86, 0.86, and 0.86, respectively. The F1, precision, and recall for RFC were 0.92, 0.91, and 0.90, respectively. The F1, precision, and recall for SVM were 0.82, 0.79, and 0.85, respectively. The F1, precision, and recall for DL were 0.90, 0.00, and 0.89, respectively.

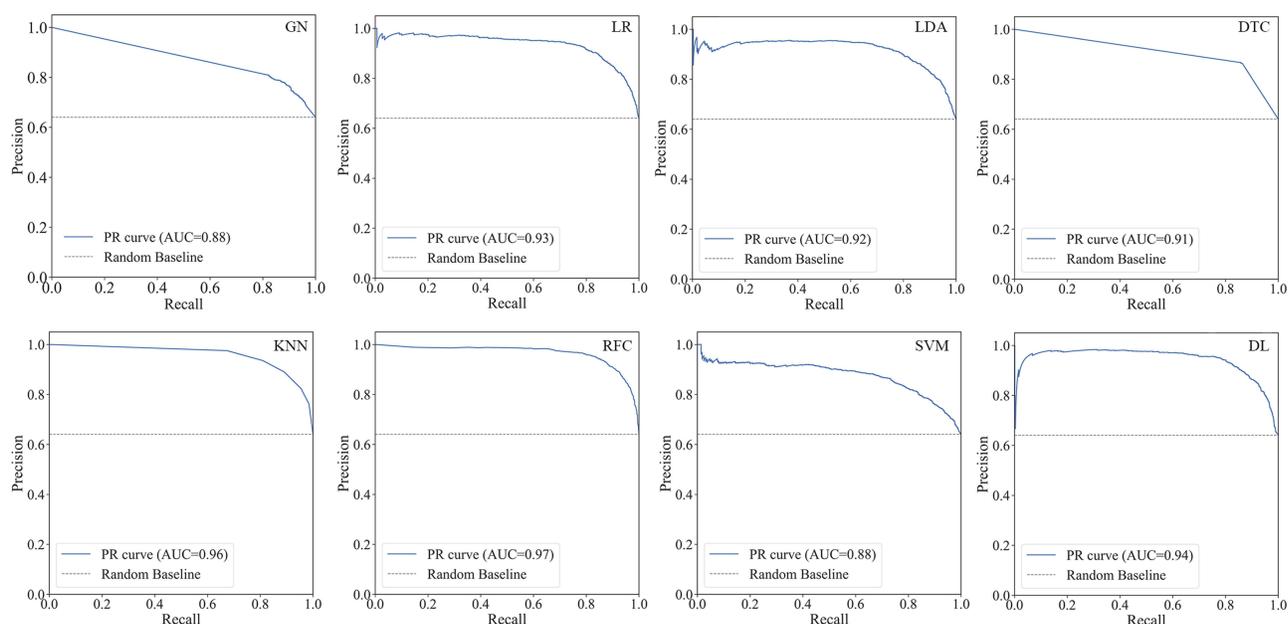


Figure 4 The precision-recall curves of models on the test dataset ($n = 1$). GN, LR, LDA, DTC, KNN, RFC, SVM, and DL were algorithms of GaussianNB, Logistic Regression, Linear Discriminant Analysis, Decision Trees Classifier, KNeighbors, Random Forest Classifier, Support Vector Machine, and Deep Learning respectively. The true positive rate (TPR) and false positive rate (FPR) were calculated based on the predicted and actual activities, respectively. The variable n represented the weight of peptide data.

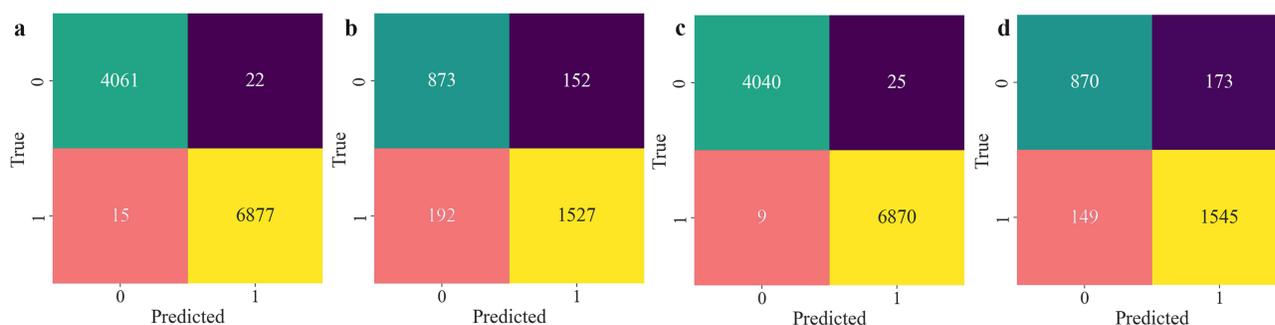


Figure 5 The confusion matrixes of the RFC_PC and RFC_PE models on the training and testing datasets. (a and b) Are the confusion matrixes of the RFC_PC model on the training and testing datasets, respectively. (c and d) Display the confusion matrixes for the RFC_PE model on the training and testing datasets, respectively.

Predictions of AAPs

The RFC_PC model was subsequently utilized to predict the potential activities of novel peptides against AChE, with a focus on peptides comprising amino acid sequences of 3–10 residues in length. A total of 1396 peptides were identified as putative AAPs and were subjected to further evaluation through molecular docking to assess their binding affinities to AChE. Among these, the top six peptides exhibiting the highest binding affinities were selected as lead candidates for experimental validation (Table 3).

Molecular Docking Validation

The active pocket of acetylcholinesterase (AChE) consists of 17 residues, six of which, TRP432, ILE439, MET436, PHE330, TRP84, and TYR121, interact strongly with huprine X (Figure 7g and Supplement 3), the active ligand exists in the crystal structure of AChE. MD revealed that the peptides WCWIYN (Figure 7a), WIGCWD (Figure 7b), LHTMELL (Figure 7c), WHLCVLF (Figure 7d), IFLSMC (Figure 7e), and VWIIGFEHM (Figure 7f) all form strong interactions with these key active-site residues. Supplement 3 details the interactions between AChE and AAPs. The binding affinities of these lead peptides are all below -8.0 kcal/mol, with WIGCWD demonstrating the lowest affinity of -11.7 kcal/mol

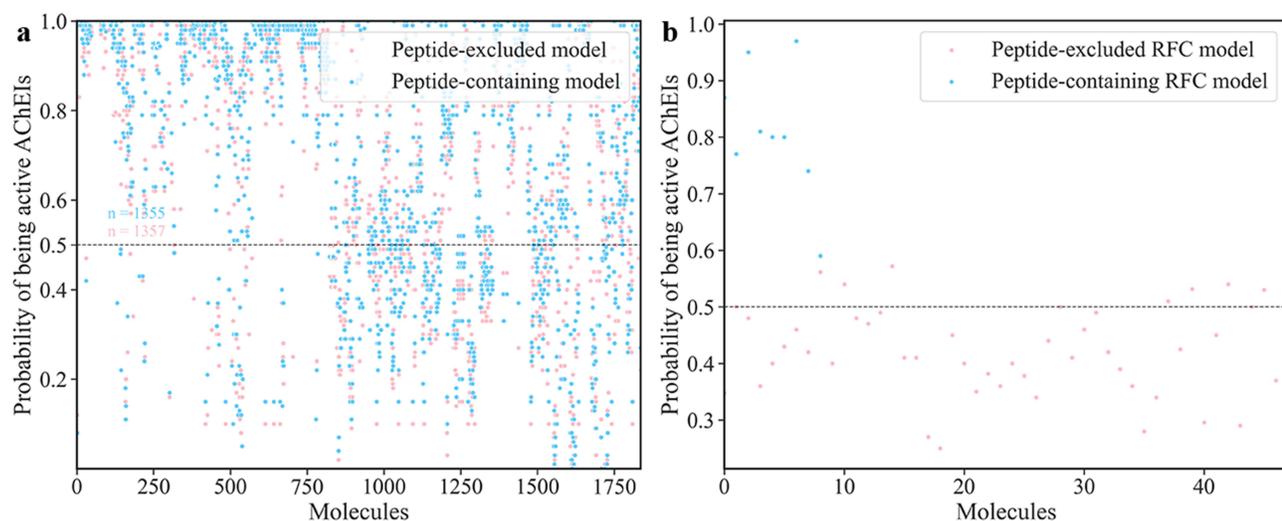


Figure 6 The distributions of prediction probabilities for classifying non-peptide compounds (a) and peptides (b) as active or inactive AChEIs in the external validation datasets. RFC_PC was the peptide-containing Random Forest Classifier model and RFC_PE was the peptide-excluded Random Forest Classifier model. For the development of the RFC_PC model, 38 peptides (representing 80% of the total peptides) were utilized, while the remaining 9 peptides (20% of the total peptides) were reserved as an external validation dataset for peptides. In contrast, for the RFC_PE model, all peptides were used exclusively for external validation purposes. The Y-axis represents the prediction probability of specific small molecules or peptides and X-axis represents the indexes of the small molecules or peptides in the testing dataset. Totally, there are 1836 non-peptides (a) and 47 peptides (b) were predicted. The blue dots represent the predicted active small-molecule AChEIs (a) or peptides (b) by RFC_PC. The pink dots represent the predicted active small-molecule AChEIs (a) or peptides (b) by RFC_PE.

(Table 3). All peptides were predicted to be soluble or moderately soluble, or very soluble in aqueous solution. Two peptides were also predicted to have blood-brain barrier permeability and moderate metabolic stabilities in body fluids.

Anti-AChE Activities of the Leads

The *in vitro* inhibition activities of the six AAPs against AChE were determined. Among them, IFLSMC (Figure 8e) exhibited the highest inhibitory activity, with an IC_{50} value of 7 nM. The IC_{50} values for the other peptides were as follows: WCWIYN (Figure 8a) at 3.4 μ M, WIGCWD (Figure 8b) at 1.9 μ M, LHTMELL (Figure 8c) at 10.6 μ M, WHLCVLF (Figure 8d) at 1.5 μ M, and VWIIGFEHM (Figure 8f) at 3.9 μ M.

Discussion

Alzheimer's disease (AD), the primary cause of dementia, constitutes 60–80% of all dementia cases. By 2050, the global AD patients population is projected to reach 150 million,³³ underscoring the urgent need for novel therapies. Peptides, vital for numerous biological functions, present a promising alternatives to small-molecule drugs in AD treatment and have advantages

Table 3 The Sequences and Affinities of the 6 Peptides

Sequences	Length	Affinity ^a	RMSD	Solubility ^b	Metabolic Stability ^c	BBB ^d	Toxicity ^e
WIGCWD	6	-11.7	0.14±0.004	Soluble	11 ^f	Yes	No
WCWIYN	6	-10.8	0.29±0.006	Soluble	8	Yes	No
WHLCVLF	7	-9.4	0.25±0.008	Moderately soluble	9	No	No
IFLSMC	6	-8.7	0.17±0.004	Very soluble	8	No	No
LHTMELL	7	-8.3	0.24±0.003	Soluble	9	No	No
VWIIGFEHM	9	-8.1	0.14±0.009	Moderately soluble	12	No	No
Huprine X		-10.9	0.17±0.007				

Notes: ^aThe unit of affinity was kcal/mol; ^bThe solubilities were predicted by SwissADME (<http://www.swissadme.ch/index.php>); ^cMetabolic stabilities were predicted by PeptideCutter (https://web.expasy.org/peptide_cutter/); ^dBBB (the brain blood barrier permeabilities) of the peptides were calculated by B3Pred (<https://webs.iitd.edu.in/raghava/b3pred>); ^eThe toxicities were predicted by ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred/index.html>); ^fThe number of enzymes of total 35 enzymes (can be found in ToxinPred) that can cleave the peptide. The H-bonds, π - π stacking, etc., and key residues were included in Supplement 4. AutoDock Vina was used to carry out the docking.

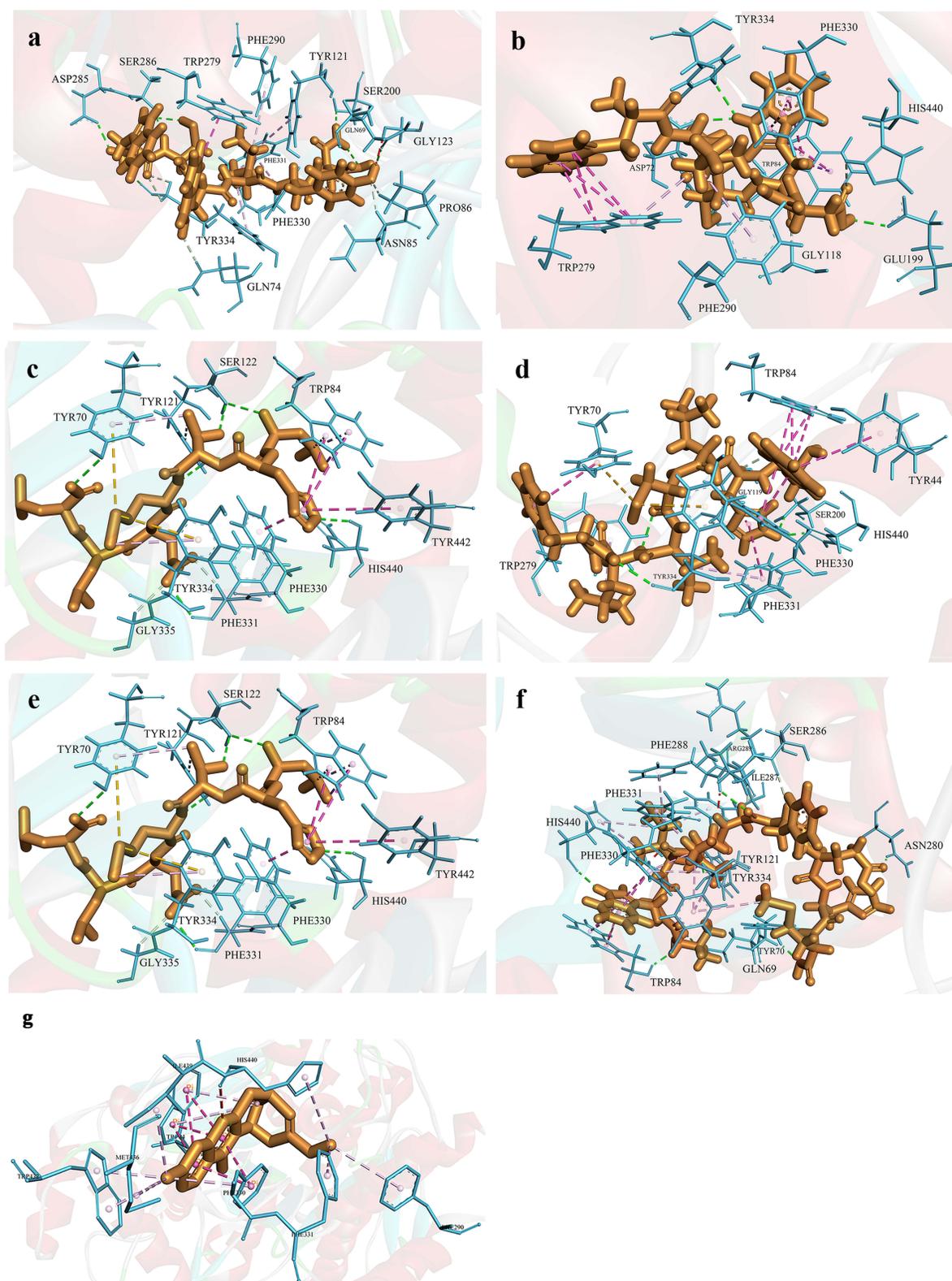


Figure 7 The 3D interactions between AChE (1E66) and WCWIYN (a), WIGCWD (b), LHTMELL (c), WHLCVLF (d), IFLSMC (e), VWIIGFEHM (f), and Huprine X (g), respectively. The light blues represent the amino acid residues of AChE and the thick brown molecules were the corresponding peptides. The interactions were visualized by Discovery Studio Visualizer 2020. Some interactions were not shown and included in [Supplement 4](#). Interaction labels: unfavorable bumps were denoted by red dotted lines; π - π stacked interactions were depicted with hotpink dotted lines; Alkyl/ π -Alkyl were shown as lightpink dotted lines; π -sigma interactions were represented with pink dotted lines; conventional hydrogen and carbon-hydrogen bonds were indicated by green dotted lines; van der Waal were represented with lightgreen dotted lines. The top ranked post of each peptide was chosen to visualized the interactions.

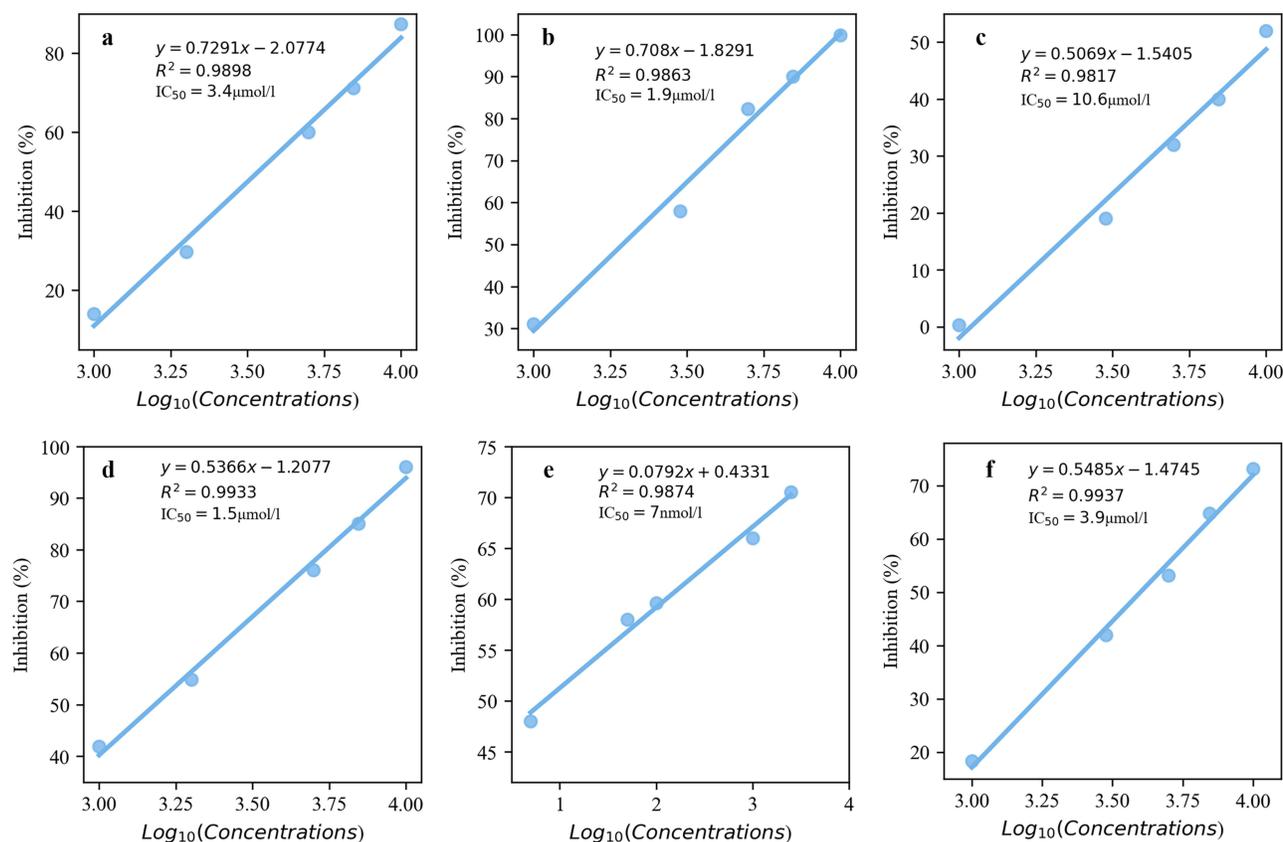


Figure 8 The inhibitions of WCWIYN (a), WIGCWD (b), LHTMELL (c), WHLCVLF (d), IFLSMC (e), and VWIIGFEHM (f) on AChE. The concentrations of the AAPs were presented in nmol/l and then used to construct the linear regressions of Log₁₀(concentrations) vs inhibitions.

over small-molecule medicines.³⁴ Peptide often demonstrate higher efficacy due to their abilities to engage in multiple target interactions, resulting in stronger, more specific binding interactions, and easier cellular penetration. Unlike small molecules, peptides generally possess higher selectivity. Their 3D structures allow selective binding to specific protein targets, reducing off-target effects. This is especially beneficial for treating complex diseases like AD. However, Peptide inhibitors generally have lower stability and more prone to in vivo proteolytic degradation, limiting their half-lives. They may also require particular conditions to maintain activity. In contrast, small-molecule inhibitors can endure harsher physiological conditions and have longer half-lives.

ML offers significant advantages over traditional wet-lab experiments by shortening the drug R&D cycle, elucidating drug pharmacokinetic preclinically, and reducing the development failure rates. Over the past decades, ML has experienced explosive growth in drug discovery⁷ and has been widely applied to identify novel peptides with diverse therapeutic potentials, such as antibacterial,³⁵ anticancer,³⁶ membrane-active,³⁷ human leukocyte antigens,³⁸ cell-penetrating,³⁹ and antihypertensive⁴⁰ peptides. This highlights ML's potential to accelerate the discovery of peptide-based therapeutics for complex diseases like AD.

MD is a structure-based virtual screening technology⁴¹ focusing on receptor-ligand interactions. MD provides critical insights into these interactions' mechanisms and identifies the essential residues required for regulating target activities.⁴² Thus, MD serves as an important complement to ML and has become a vital tool for refining and ranking ML-generated results.^{43,44} The integration of MD and ML has greatly enhanced the drug discovery pipelines' accuracy and reliability.

Due to the significant advantages of ML and MD, their use in the discovering new non-peptide AChEIs has been extensive. However, no ML-based reports on designing AAPs. One reason is that ML relies heavily on high-quality datasets, particularly those derived from wet-lab experiments, which are unfortunately lacking for AAPs, Models based on NPSMs typically exhibit poor predictive power for peptides due to the significant differences between peptides and

NPSMs. For example, although the RFC_PE model performs well on the training set, but sees a marked drop in precision and F1 score on testing set, correctly predicting only 10 out of 47 active peptides. However, it achieved a precision of 0.888 on the external non-peptide validation dataset, surpassing the performance of previously reported RFC^{45,46} and XGBoost⁴⁷ models. By incorporating AAPs into the datasets, the RFC model was significantly enhanced. The resulting RFC_PC model successfully identified all 47 active peptides in the external validation dataset while maintaining similar precision on the non-peptide validation dataset as the RFC_PE model. This improvement highlights the critical role of peptide data in enhancing the RFC model's predictive power. Given that only 38 peptides (0.3% of total instances) were included in the dataset, expanding the peptide dataset could further optimize the RFC_PC model's performance. It also must be stated here that balancing peptide and non-peptide data is crucial for ensuring the model's robustness for both molecule types.

Based on MD results, the predicted AAPs were re-ranked, and six high-affinity peptides (IFLSMC, WCWIYN, WIGCWD, LHTMELL, WHLCVLF, and VWIIGFEHM) were selected. All these leads exhibited docking scores below -8.0 kcal/mol, a threshold distinguishing stable from unstable ligand-receptor complexes. The affinities of WCWIYN and WIGCWD (-10.8 and -11.7 kcal/mol, respectively) were comparable to positive drug of huprine X (-10.8 kcal/mol). Hydrogen bonds are prevalent in all lead-AChE interactions, including conventional hydrogen and carbon hydrogen bonds. The varying residues involved and different bond distances highlight hydrogen bonds' importance in stabilizing peptide-AChE interactions. Hydrophobic interactions are also crucial. Each peptide exhibits diverse hydrophobic interactions with AChE, such as pi-pi stacked and pi-alkyl interactions, particularly with aromatic residues. The specific binding characteristics of each peptide differ in terms of interaction residues and types, reflecting their unique binding modes and specificities. For instance, WCWIYN's multiple hydrogen bonds with SER286 and VWIIGFEHM's several conventional hydrogen bonds with TYR121 and other residues reveal sequence- and structure- dependent binding features. Peptides exhibit various binding patterns with AChE, and the synergistic effect of multiple interactions enables stable peptide-AChE binding. This multi-interaction mode synergy is likely fundamental for peptide bioactivity or function related to AChE and holds great significance for studying peptide-AChE interaction mechanisms and drug design.

These peptides have various solubility from moderately soluble to very soluble. Notably, WIGCWD and WCWIYN could permeate the blood-brain barrier (BBB), and none of the peptides exhibit toxicity. However, four peptides lack simultaneous BBB permeability and metabolic stability, indicating a need for further optimization to balance these properties for effective drug delivery.

The consistency between MD results and experimental data confirms the RFC_PC's success in incorporating active peptides. This approach offers valuable insights for constructing ML models when active peptides are limited, demonstrating that the strategic inclusion of peptide data can significantly enhance model performance and applicability.

In this study, we focused on the discovery of acetylcholinesterase-binding peptides using an integrated approach of ML and MD. The results demonstrated that incorporating active peptide data into ML models significantly enhanced predictive power even with limited data. The identified peptides exhibited favorable solubility and, importantly, two could permeate the BBB, a critical factor for AD therapies. However, four peptides lacked simultaneous BBB permeability and metabolic stability, indicating a need for further optimization.

Data Sharing Statement

The data source was mentioned in the section on materials and methods. The code could be provided through Email to the correspondence author.

Ethical Approval

National Health Commission of the People's Republic of China, Ministry of Education of the People's Republic of China, Ministry of Science and Technology of the People's Republic of China, and National Administration of Traditional Chinese Medicine released the "Notice on Printing and Distributing the Measures for Ethical Review of Biomedical Research Involving Human Beings" on February 18th, 2023 and the full text can be find in https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm. Its Article 32 was as follows:

Article 32: The use of personal information data or biological samples for the following types of life science and medical research involving human beings, which do not cause harm to individuals, do not involve sensitive personal information, or do not involve commercial interests, may be exempted from ethical review in order to reduce unnecessary burdens on researchers and promote the conduct of life science and medical research involving human beings:

1. Research conducted using legally obtained public data, or data generated through observation without interfering with public behavior;
2. Research conducted using anonymized information data.

The data used in our research were from the public ChEMBL database and its data could be freely download and used for research purposes and all of its data were anonymized. Therefore, our research was exempted from approval of Institutional Review Board of our University.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Financial support was received from the Jiangxi Province Department of Education (GJJ160786).

Disclosure

The authors declare they have no financial or non-financial conflicts of interest in this work.

References

1. Medeiros R, Baglietto-Vargas D, LaFerla FM. The role of tau in Alzheimer's disease and related disorders. *CNS Neurosci Ther.* 2011;17:514–524. doi:10.1111/j.1755-5949.2010.00177.x
2. Colovic MB, Krstic DZ, Lazarevic-Pasti TD, Bondzic AM, Vasic VM. Acetylcholinesterase inhibitors: pharmacology and toxicology. *Curr Neuropharmacol.* 2013;11:315–335. doi:10.2174/1570159X11311030006
3. Perry EK, Perry RH, Blessed G, Tomlinson BE. Changes in brain cholinesterases in senile dementia of Alzheimer type. *Neuropath Appl Neuro.* 1978;4:273–277. doi:10.1111/j.1365-2990.1978.tb00545.x
4. Su JQ, Liu HY, Guo K, Chen L, Yang MH, Chen Q. Research advances and detection methodologies for microbe-derived acetylcholinesterase inhibitors: a systemic review. *Molecules.* 2017;22:176–199. doi:10.3390/molecules22010176
5. Munoz-Torrero D. Acetylcholinesterase inhibitors as disease-modifying therapies for Alzheimer's disease. *Curr Med Chem.* 2008;5:2433–2455. doi:10.2174/092986708785909067
6. Ruangritchankul S, Chantharit P, Srisuma S, Gray LC. Adverse drug reactions of acetylcholinesterase inhibitors in older people living with dementia: a comprehensive literature review. *Ther Clin Risk Manag.* 2021;17:927–949. doi:10.2147/TCRM.S323387
7. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov.* 2019;8:463–477. doi:10.1038/s41573-019-0024-5
8. Catacutan DB, Alexander J, Arnold A, et al. Machine learning in preclinical drug discovery. *Nat Chem Biol.* 2022;20:960–973. doi:10.1038/s41589-024-01679-1
9. Venanzi NAE, Basciu A, Vargiu AV, et al. Machine learning integrating protein structure, sequence, and dynamics to predict the enzyme activity of bovine enterokinase variants. *J Chem Inf Model.* 2024;64:2681–2694. doi:10.1021/acs.jcim.3c00999
10. Ahsan MM, Luna SA, Siddique Z. Machine-learning-based disease diagnosis: a comprehensive review. *Healthcare.* 2022;10:541–570. doi:10.3390/healthcare10030541
11. Odusami M, Maskeliūnas R, Damaševičius R, Misra S. Machine learning with multimodal neuroimaging data to classify stages of Alzheimer's disease: a systematic review and meta-analysis. *Cogn Neurodynamics.* 2024;18:775–794. doi:10.1007/s11571-023-09993-5
12. Gunter NB, Gebre RK, Graff-Radford J, et al. Machine learning models of polygenic risk for enhanced prediction of Alzheimer disease endophenotypes. *Neurology.* 2024;10:e200120. doi:10.1212/NXG.0000000000200120
13. García-Gutiérrez F, Alegret M, Marquié M, et al. Unveiling the sound of the cognitive status: machine Learning-based speech analysis in the Alzheimer's disease spectrum. *Alzheimers Res Ther.* 2024;16:26–45. doi:10.1186/s13195-024-01394-y
14. Xu T, Li S, Li AJ, et al. Identification of potent and selective acetylcholinesterase/butyrylcholinesterase inhibitors by virtual screening. *J Chem Inf Model.* 2023;63:2321–2330. doi:10.1021/acs.jcim.3c00230

15. Khan MI, Taehwan P, Cho Y, et al. Discovery of novel acetylcholinesterase inhibitors through integration of machine learning with genetic algorithm based in silico screening approaches. *Front Neurosci-Switz.* 2023;3:16–25. doi:10.3389/fnins.2022.1007389
16. Guo S, Wang J, Wang Q, Wang J, Qin S, Li W. Advances in peptide-based drug delivery systems. *Heliyon.* 2024;10:e26009. doi:10.1016/j.heliyon.2024.e26009
17. Wang L, Wang N, Zhang W, et al. Therapeutic peptides: current applications and future directions. *Signal Transduct Target Ther.* 2022;7:14–40. doi:10.1038/s41392-022-00904-4
18. Othman AM, Lombardi L, Williams DR, Albericio F. Strategies for improving peptide stability and delivery. *Pharmaceuticals-Base.* 2022;15:1283–1315. doi:10.3390/ph15101283
19. Rusu ME, Fizesan I, Pop A, et al. Walnut (*Juglans regia L.*) septum: assessment of bioactive molecules and *in vitro* biological effects. *Molecules.* 2020;25:2187–2207. doi:10.3390/molecules25092187
20. Fan Y, Li X, Ding L, et al. Accelerated solvent extraction of antioxidant compounds from gardeniae fructus and its acetylcholinesterase inhibitory and PC12 cell protective activities. *Foods.* 2021;10:2805–2823. doi:10.3390/foods10112805
21. Liu S, Cao XL, Liu GQ, Zhou T, Yang XL, Ma BX. The in silico and in vivo evaluation of puerarin against Alzheimer's disease. *Food Funct.* 2019;10:799–813. doi:10.1039/C8FO01696H
22. Kandasamy S, Loganathan C, Sakayanathan P, et al. In silico, theoretical biointerface analysis and in vitro kinetic analysis of amine compounds interaction with acetylcholinesterase and butyrylcholinesterase. *Int J Biol Macromol.* 2021;185:750–760. doi:10.1016/j.ijbiomac.2021.06.176
23. Behl T, Kaur I, Sehgal A, et al. AChE as a spark in the Alzheimer's blaze-Antagonizing effect of a cyclized variant. *Ageing Res Rev.* 2023;83:101787. doi:10.1016/j.arr.2022.101787
24. Gaulton A, Hersey A, Nowotka M, et al. The ChEMBL database in 2017. *Nucleic Acids Res.* 2017;45:945–954. doi:10.1093/nar/gkw1074
25. Landrum G. RDKit: open-source cheminformatics from machine learning to chemical registration. 2019. Available from: <https://www.rdkit.org/>. Accessed June 12, 2025.
26. Dvir H, Wong DM, Harel M, et al. 3D structure of Torpedo californica acetylcholinesterase complexed with huprine X at 2.1 Å resolution: kinetic and molecular dynamic correlates. *Biochemistry.* 2002;41:2970–2981. doi:10.1021/bi011652i
27. Trott O, Olson AJ. Software News and Update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31:455–461. doi:10.1002/jcc.21334
28. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Cheminformatics.* 2011;7:33–46. doi:10.1186/1758-2946-3-33
29. Ingkaninan K, Temkithawon P, Chuenchom K, Yuyaem T, Thongnoi W. Screening for acetylcholinesterase inhibitory activity in plants used in Thai traditional rejuvenating and neurotonic remedies. *J Ethnopharmacol.* 2003;89:261–264. doi:10.1016/j.jep.2003.08.008
30. Chang J, Zou SQ, Xu SB, Xiao YW, Zhu D. Screening of inhibitors against idiopathic pulmonary fibrosis: few-shot machine learning and molecule docking based drug repurposing. *Curr Comput-Aid Drug.* 2024;20:134–144. doi:10.2174/1573409919666230417080832
31. Wu JX, Yihao Chen YH, Wu J, et al. Large-scale comparison of machine learning methods for profiling prediction of kinase inhibitors. *J Cheminformatics.* 2024;16:13–25. doi:10.1186/s13321-023-00799-5
32. He SY, Zhao DC, Ling YL, et al. Machine learning enables accurate and rapid prediction of active molecules against breast cancer cells. *Front Pharmacol.* 2021;12:796534.
33. Cummings J, Zhou Y, Lee G, Zhong K, Fonseca J, Cheng F. Alzheimer's disease drug development pipeline: 2023. *Alzh Dement-Dadm.* 2023;9:e12385. doi:10.1002/trc2.12385
34. Henninot A, Collins JC, Nuss JM. The current state of peptide drug discovery: back to the future? *J Med Chem.* 2018;61:1382–1414. doi:10.1021/acs.jmedchem.7b00318
35. Huang J, Xu Y, Xue Y, et al. Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences. *Nat Biomed Eng.* 2023;7:797–8810. doi:10.1038/s41551-022-00991-2
36. Lin YC, Lim YF, Russo E, Schneider P, Bolliger L, Edenharter A. Multidimensional design of anticancer peptides. *Angew Chem Int Edit.* 2015;54:10370–10374. doi:10.1002/anie.201504018
37. Lee EY, Wong GCL, Ferguson AL. Machine learning-enabled discovery and design of membrane-active peptides. *Bioorgan Med Chem.* 2018;26:2708–2718. doi:10.1016/j.bmc.2017.07.012
38. Luo H, Ye H, Ng HW, et al. Machine learning methods for predicting HLA-peptide binding activity. *Bioinform Biol Insig.* 2015;9:21–29. doi:10.4137/BBI.S2946
39. Moran-Torres R, Castillo G, David A, et al. Selective moonlighting cell-penetrating peptides. *Pharmaceutics.* 2021;13:1119–1134. doi:10.3390/pharmaceutics13081119
40. Rauf A, Kiran A, Hassan MT, Mahmood S, Mustafa G, Jeon M. Boosted prediction of antihypertensive peptides using deep learning. *Appl Sci.* 2021;11:2316–2326. doi:10.3390/app11052316
41. Morris GM, Lim-Wilby M. Molecular docking. In: *Molecular Modeling of Proteins, Methods in Molecular Biology.* Vol. 443. Berlin: Springer; 2008:365–382. doi:10.1007/978-1-59745-177-2_19
42. Khan FI, Govender A, Permaul K, et al. Thermostable chitinase II from *Thermomyces lanuginosus* SSBP: cloning, structure prediction and molecular dynamics simulations. *J Theor Biol.* 2015;374:107–114. doi:10.1016/j.jtbi.2015.03.035
43. Yan J, Huang W, Zhang C, Huo H, Chen F. Virtual screening of acetylcholinesterase inhibitors based on machine learning combined with molecule docking methods. *Curr Bioinform.* 2021;16:963–971. doi:10.2174/1574893615999200719234045
44. Yan J, Yan X, Hu S, Zhu H, Yan B. Comprehensive interrogation on acetylcholinesterase inhibition by ionic liquids using machine learning and molecular modeling. *Environ Sci Technol.* 2021;55:14720–14731. doi:10.1021/acs.est.1c02960
45. Sandhu H, Kumar RN, Garg P. Machine learning-based modeling to predict inhibitors of acetylcholinesterase. *Mol Divers.* 2022;26:331–340. doi:10.1007/s11030-021-10223-5
46. Lv W, Xue Y. Prediction of acetylcholinesterase inhibitors and characterization of correlative molecular descriptors by machine learning methods. *Eur J Med Chem.* 2010;45:1167–1172. doi:10.1016/j.ejmech.2009.12.038
47. Vignaux PA, Lane TR, Urbina F, et al. Validation of acetylcholinesterase inhibition machine learning models for multiple species. *Chem Res Toxicol.* 2023;36:188–201. doi:10.1021/acs.chemrestox.2c00283

Drug Design, Development and Therapy

Dovepress
Taylor & Francis Group

Publish your work in this journal

Drug Design, Development and Therapy is an international, peer-reviewed open-access journal that spans the spectrum of drug design and development through to clinical applications. Clinical outcomes, patient safety, and programs for the development and effective, safe, and sustained use of medicines are a feature of the journal, which has also been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/drug-design-development-and-therapy-journal>