JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Discovery of Multi-target-directed Ligands against Alzheimer's Disease through Systematic Prediction of Chemical-protein Interactions

Jiansong Fang, Yongjie Li, Rui Liu, Xiaocong Pang, Chao Li, Ranyao Yang, Yangyang He, Wenwen Lian, Ai-lin Liu, and Guan-hua Du

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/ci500574n • Publication Date (Web): 22 Dec 2014

Just Accepted

Journal of Chemical Information and Modeling is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Discovery of Multi-target-directed Ligands against Alzheimer's Disease through Systematic Prediction of Chemical-protein Interactions

Jiansong Fang[1], Yongjie Li[1], Rui Liu[1], Xiaocong Pang[1], Chao Li[1], Ranyao Yang[1], Yangyang He[1], Wenwen Lian[1], Ai-lin Liu[1,2,3], and Guan-hua Du[1,2,3*]*

[1]Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, PR China

[2]Beijing Key Laboratory of Drug Target and Screening Research, Beijing 100050, PR China

[3]State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Beijing 100050, PR China

Corresponding authors:

Ai-Lin Liu* (Email: liuailin@imm.ac.cn); Tel & Fax: 86-10-83150885; Address: 1 Xian Nong Tan Street, Beijing 100050 China

Guan-Hua Du* (Email: dugh@imm.ac.cn); Tel & Fax: 86-10-63165184; Address: 1 Xian Nong Tan Street, Beijing 100050 China

ABSTRACT

To determine chemical-protein interactions (CPI) is costly, time-consuming, and labor-intensive. *In silico* prediction of CPI can facilitate the target identification and drug discovery. Though many *in silico* target prediction tools have been developed, few of them could predict active molecules against multi-target for a single disease. In this investigation, naive Bayesian (NB) and recursive partitioning (RP) algorithms were applied to construct classifiers for predicting the active molecules against 25 key targets toward Alzheimer's disease (AD) using multitarget-quantitative structure-activity relationships (mt-QSAR) method. Each molecule was initially represented with two kinds of fingerprint descriptors (ECFP6 and MACCS). 100 classifiers were constructed and their performance was evaluated and verified with internally five-fold cross-validation and external test set validation. The range of the area under the receiver operating characteristic curve (ROC) for the test sets was from 0.741 to 1.0, with an average of 0.965. In addition, the important fragments for multi-target against AD given by NB classifiers were also analyzed. Finally, the validated models were employed to systematically predict the potential targets for 6 approved anti-AD drugs, and 19 known active compounds related to AD. The prediction results were confirmed by reported bioactivity data and our *in vitro* experimental validation, resulting in several multi-target-directed ligands (MTDLs) against AD, including 7 acetylcholinesterase (AChE) inhibitors ranging from 0.442 to 72.26 μM and 4 histamine receptor 3 (H_3R) antagonists ranging from 0.308 to 58.6 μM. To be exciting, the best MTDL DL0410 was identified as an dual cholinesterase inhibitor with IC_{50} values of 0.442 μM (AChE) and 3.57 μM (BuChE) as well as a H_3R antagonist with an IC_{50} of 0.308 μM. This investigation is the first report using mt-QASR approach to predict chemical-protein interaction

for a single disease and discovering highly potent MTDLs. This protocol may be useful for *in silico* multi-target prediction of other diseases.

1. INTRODUCTION

Alzheimer's disease (AD) is a devastating condition leading to progressive cognitive decline, functional impairment and loss of independence. AD is characterized by a loss of basal forebrain neurons and reduced cortical and hippocampal levels of acetylcholine (ACh). It incurs an enormous personal cost to those affected and there is an urgent need to develop more effective therapies to treat and delay the onset of the disease.[1] To confront AD, an innovative strategy is to design single chemical entities able to simultaneously modulate more than one target, which called "multi-target-directed ligands (MTDLs)".[2] MTDLs are, in principle, effective in treating complex diseases because of their ability to interact with multiple targets supposed to be responsible for the pathogenesis.

Polypharmacology has emerged as a new theme in drug discovery, especially for complex diseases such as AD that involve functional modulation of multiple proteins such as AD.[3-5] Polypharmacology focuses on the fact that one drug can hit multiple targets. Prediction of polypharmacology for known drugs is highly useful for finding new indication and explaining the molecular mechanism of action.

The essence of predicting polypharmacology is to identify the interactions between drugs and target proteins. However, the determination of chemical-protein interaction (CPI) remains very challenging and time-consuming at the experiment level. Thus, many *in silico* target prediction tools have been developed, which were summarized by a recent review.[6] The most widely used methods are ligand-based target prediction (LBTP), and structured-based target prediction (SBTP) approaches, such as similarity search,[7-9] pharmacophore modeling,[10, 11] and inverse docking.[12, 13]

Most of the tools are available for predicting the targets covering many diseases, which results in poor accuracy and time wasting when one compound is known to be active toward specific disease such as AD. Until recently, Xie et al assembled TargetHunter,[14] and HTDocking to construct an Alzheimer's chemogenomics knowledgebase (AlzPlatform) for target identification and polypharmacology analysis for AD.[15] However, the basic principle of TargetHunter program is based on the concept that structurally similar compounds may have similar biological profiles, which make it difficult to identify ligands with novel structural scaffolds that differ from the reference, and HTDocking is constrained by the available crystallographic structure of target and computational speed, thus limiting its application. Therefore, it is still necessary to develop computational methods to predict chemical-protein interactions toward AD.

Nowadays several ligand-based methods apply data mining methods in order to identify unknown drug-target interactions, such as quantitative structure-activity relationships (QSAR) and computational chemogenomics. Conventional QSAR models are unspecific or only consider a series of ligands against a single target.[16] To address this problem, Vina et al. exploited a multitarget-QSAR (mt-QSAR) classification method with an accuracy of 72% for the training set and 72% in cross-validation.[17] Computational chemogenomic methods have been developed to predict the interactions between compounds and proteins. This protocol aims at exploiting the whole chemical space, which corresponds to not only the space of the small molecules but also of drug targets interacting with the molecules.[18-21] Wang et al. constructed a model for predicting CPI based only on the primary sequence of proteins and the structural features of small molecules, and used it to identify novel ligands for four targets (i.e., GPR40, SIRT1, p38, and GSK3β) validated by experimental assays.[18] The major advantage of chemogenomic model is that it can predict CPI by a single binary model. However, the increasing of target descriptor

variables increases the risk of "overfitting" and computational complexity. Cheng et al. compared the two methods (mt-QSAR and computational chemogenomics) for CPI, and found the performance of mt-QSAR method was better than that of the chemogenomic for the external validation set.[21]

In this study, we have applied mt-QSAR method to predict the chemical-protein interactions for 25 key targets related to AD. The workflow of mt-QSAR is shown in Figure 1. Based on two machine learning tools (naive Bayesian and recursive partitioning), one hundred binary classifiers were constructed by integrating the chemical and pharmacological information derived from the BindingDB database. All developed models were validated by 5-fold cross-validation and test set validation. To test the applicability of this paradigm, the developed tools were used to predict polypharmacology for 6 approved AD drugs and 19 known active compounds related to AD. The predictions were then confirmed by reported bioactivity data and our *in vitro* experimental validation, which indicated the potential of this strategy in target prediction of compounds and MTDLs discovery.

2. MATERIALS AND METHODS

2.1 Data Collection and Preparation. Thomson Reuters Integrity Database[18] and Therapeutic Target Database (TTD)[22] were applied to explore the important targets for AD. The retrieval condition was limited to the key word 'Alzheimer's disease', and they were further filtered to keep the targets with drug candidates that had entered into Phase I clinical trial at least. After that, 25 targets related to AD were obtained (Figure 2), then the corresponding ligands together with bioactivity data were collected from the Binding Database (http://www.bindingdb.org, accessed March 2014).[23]

In this study, the following criteria was adopted to refine the data sets: (1) the duplicate molecules were removed; (2) Salts were converted to the corresponding acids or bases and water molecules were removed from hydrates; (3) the compound was defined as positive (designated as +1) if its K_i, EC_{50} or $IC_{50} \leq 10$ μM. After filtering, 18,741 active ligands were got. The decoy compounds (designated as -1) were generated through three ways: (1) experimentally-validated non-inhibitors, such as non-inhibitors of butyrylcholinesterase (BuChE),[24] (2) extracted from DUD subsets, such as non-inhibitors of acetylcholinesterase (AChE), monoamine oxidase B (MAO-B) and beta-secretase 1(BCAE1); (3) generated in DUD online database with known active compounds for the other 21 targets.[25] The ratio of inactive compounds versus active compounds is 3. Both the active and decoy compounds were randomly divided into two groups (training set and test set at a ratio of 3).

2.2 Chemical Descriptors Calculation. Two kinds of fingerprint descriptors (ECFP6 and MACCS) were used for small molecular description. Extended connectivity fingerprints (ECFP) are circular fingerprints using a variant of the Morgan algorithm.[26] The advantage of circular fingerprints was that they could rapidly calculated, might present stereochemical information and could also be interpreted as chemical substructures. In this study, the ECFP_6 fingerprints were calculated by Discovery Studio software.[27]

Each molecule was also represented as a binary string using the MACCS Keys, which contained 166 most common substructure patterns and was free available from PaDEL-Descriptor (version 2.18, http://padel.nus.edu.sg, Access Date: April, 2014).[28]

2.3 mt-QSAR Method. The objective of mt-QSAR is to solve the multi-label classification problems. In this study, multi-label problem was decomposed into multiple binary classification problems. Herein, 100 mt-QSAR models were built for 25 AD important targets using two

fingerprints (ECFP_6 and MACCS) and two machine learning algorithms (naive Bayesian and recursive partitioning). For each target, there were four classifiers (NB_ECFP6, NB_MACCS, RP_ECFP6 and RP_MACCS) to predict the activity of compounds.

2.3.1 Naive Bayesian. The Bayesian approach is a robust classification approach that can distinguish between active and decoy compounds. It considers the likelihood of a model but takes the complexity of the model into consideration. As a result, it automatically picks the simplest model that can explain the observed data to prevent overfitting. A more detailed introduction can be found in the following references.[29] In general, the technique is based on the frequency of occurrence of various descriptors that are found in two or more sets of molecules that can discriminate best between these sets. For naive Bayesian classifiers, it can generate the posterior probabilities based on the core of the function, which are given by Equation 1.

$$P(+|A_1, ..., A_n) = \frac{P(A_1,..., A_n|+)P(+)}{P(A_1,..., A_n)} \qquad (1)$$

$P(A_1,...,A_n|+)$ is the conditional probability of a particular molecule being classified as active; P(+) is the prior probability, a probability induced from a set of compounds in the training set; $P(A_1,...,An)$ is the marginal probability of the given descriptors that will occur in the training set.

2.3.2 Recursive partition. Recursive partitioning (RP) in Discovery Studio 4.0 was used to develop decision trees to classify the dataset into active compounds and decoys. RP is a classification method used to decipher the relationship between a dependent property Y (activity class: 1 or -1) to a set of independent molecular descriptors X. Models are constructed by successively splitting a dataset into increasingly homogeneous subsets until it is infeasible to continue, based on a set of "stopping rules".[30]

The result of a RP model can be featured by a "decision tree" or "graph", which is used to make predictions about novel data. When making a prediction with the model, each sample is assigned to a certain node of the tree. The class predicted for that sample is the class to which a plurality of the training samples in the node belongs. In this study, 5-fold cross-validation was used to determine the degree of pruning required for the best predictive performance. The minimum number of samples at each node, maximum knots per property, and the maximum tree depth was set at 10, 20, and 20, respectively.

2.4 Performance Evaluation of Models. The internal 5-fold cross-validation and external test set validation were used to evaluate all models. In a 5-fold cross-validation, the entire data set was equally divided into five cross-validation splits. The model was trained on a set of four cross validation splits together, and the fifth sub-sample set was used as an internal validation set (test set).

All developed models was measured by the quantity of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), sensitivity (SE), specificity (SP), the overall prediction accuracy (Q), and Matthews correlation coefficient (MCC), which are given by eqs 2−5.

$$SE = \frac{TP}{TP + FN} \quad (2)$$

$$SP = \frac{TN}{TN + FP} \quad (3)$$

$$Q = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \quad (5)$$

The value of MCC falls in the range of $-1 \leq MCC \leq +1$. A perfect classification gives a MCC value of 1. In addition, the receiver operating characteristic (ROC) curve was plotted. The ROC curve was used to graphically present the model behavior of true positive rate against false positive rate in a visual way.[31] Performance was also measured by the area under the ROC curve (AUC). A perfect classifier gives AUC = 1, whereas random performance gives AUC = 0.5.

2.5 Experimental Validation.

In vitro AChE Inhibitory Assay

AChE (E.C. 3.1.1.7) was extracted from rat cortex homogenate. 5,5'-dithiobis-(2-nitrobenzoic acid) (Ellman's reagent, DTNB), acetylthiocholine chloride (ATC), and donepezil hydrochloride were supplied by Sigma Aldrich. The AChE activity was measured by detecting the hydrolysis product of the substrate following the method of Ellman.[32,33] 19 compounds and denepezil were detected on their inhibition of AChE activity.

In the AChE reaction system, the assay was performed in 96-well plates using a Spectra Max M5 microplate reader (Molecular Devices, Sunnyvale, CA, USA). Donepezil hydrochloride was selected as the reference compound. The assay solution consisted of 10 μL test compounds, 30 μL 0.05 M PBS, 20 μl AChE, 60 μL 3.75 mM of ATC and 80 μL 0.25 mg/ml DTNB. Then the assay mixture was incubated at 37 °C for 60 min and then the absorbance value was quantified at the absorbance wavelength 412 nm. The wells without test compounds were also set for calculating the inhibition of the test compounds. Each concentration was analyzed in triplicate independently, and IC_{50} values were determined graphically from log concentration–inhibition curves.

In vitro Assay on Histamine Receptor 3 (H_3R)

The cell-based histamine receptor 3 (H_3R) assay was detected based on β-lactamase complementation technology.[34] The H_3-bla U2OS cells (invitrogen, USA) stably express two fusion proteins as well as a β-lactamase (bla) reporter gene under the control of a UAS response element. The first fusion protein is human H_3R linked to a Gal4-VP16 transcription factor via TEV protease site, and the other is β-arrestin/TEV protease fusion protein.

H_3-bla U2OS cells (32μl, 7,500 cells/well) were plated in a 384-well format with FreeStyle™ Expression Medium (Invitrogen, USA) and incubated for 18 hours. Cells were exposed to 4 μl test compounds and the control compound thioperamide (Sigma-Aldrich, USA) for 30 mins, and then stimulated with 4 μl methylhistamine at 400 nM (Sigma-Aldrich, USA) for 5 hours. Cells were then loaded with 8 μl LiveBLAzer™-FRET B/G Substrate (Invitrogen, USA) for 2 hours before detection. Fluorescence emission values at 460 nm and 530 nm were obtained using a Spectra Max M5 microplate reader (Molecular Devices, Sunnyvale, CA, USA) and the inhibition percentage was plotted against the indicated concentrations of compounds. Each concentration was analyzed in triplicate, and IC_{50} values were determined graphically from log concentration–inhibition curves.

In Vitro CDK5/p35 Inhibitory Assays

The CDK5/p35 inhibitory activity of the 10 selected compounds was tested using the Kinase-Glo luminescent technique, which is a safe non-radioactive assay.[35] The assay was performed as previously described.[36]

Kinase-Glo assays were performed in opaque white-walled 384-well microplates. The reaction system included 2.5 μl of test compounds, 5 μl of cdk5/p35 kinase (50 ng, Promega), and 5 μl of a complex of Histone H1 peptide (Apeptide Co., Ltd (Shanghai, China)) and adenosine triphosphate (ATP, Roche). After 1 h of incubation at 37 °C, the enzymatic reaction was stopped

by the addition of 12.5 µl of Kinase-Glo Reagent (catalogue number V6713, Promega). The plate was mixed, incubated for 1 h at 37 °C, and read with a SpectraMax M5 (Molecular Devices, Sunnyvale, CA, USA) after 10 min. The wells without inhibitors were also set for calculating the inhibition of the test compounds. The data are expressed as the mean of three independent experiments.

3. RESULTS AND DISCUSSION

3.1 Data Set Analysis. The prediction accuracy of binary classifiers is influenced by the chemical diversity of samples utilized in the training set and test set. Generally, classification models that only covered a small region of chemical space, limited their applications. Tanimoto similarity index is a classic method to explore the diversity of compounds within a chemical data set. Smaller tanimoto similarity index means that compounds within the data set have better diversity. Thus the tanimoto similarity analysis was performed with the fingerprint of ECFP_2. As shown in Table 1, the tanimoto indexes range from 0.107 to 0.194 for training sets and 0.117 to 0.191 for test sets, which suggests that the entire data set is diverse enough.

The distributions of targets and ligands space in the entire data set are given in Figure 3. As presented in Figure 3A, the target space (n = 25) can be divided into four subfamilies, namely neurotransmitter receptors (n = 14), enzymes (n = 6), kinases (n = 2), and targets related to amyloid beta precursor protein (APP) metabolism (n = 3), and the number of corresponding active compounds for four subfamilies was 8192, 6232, 1402, and 2915, respectively (Figure 3B). The results above show that the entire data set has diverse ligand and target coverage.

3.2 Model Evaluation and Comparison. All the classification models in this study were initially developed using NB and RP classifiers with the two kinds of fingerprints (ECFP_6 and

MACCS). Subsequently, 5-fold cross-validations were performed. In addition, the models were used to predict the respective test sets.

The performance of the 5-fold cross-validation for training set is given in Table 2. Among the one hundred models, 79 models out of 100 (79%) give the MCC value higher than 0.8, whereas 93 models out of 100 (93%) give the AUC value higher than 0.9. At the same time, the values of MCC range from 0.463 to 1, with an average of 0.826, whereas the values of AUC range from 0.725 to 1, with an average of 0.954. These data indicates that the overall predictive accuracies of the mt-QSAR models are high. The detailed performance of the training sets is given in Table S1. It is interesting to see that 93 out of 100 models (93%) give both of the SE and SP values higher than 0.8 simultaneously. The average values of SE and SP for all the models are 0.946 and 0.940, respectively.

The internal cross validation for training set cannot represent the true predictive ability of the models. Thus, the external test set validation is extremely important for controlling the quality of computational model. As given in Table 3, a high performance is also yielded for the test sets of 100 mt-QSAR models. The range of the MCC value is 0.285 to 1, with an average value of 0.826. The range of AUC value is 0.741 to 1.0, with an average of 0.965. The four models toward GABA-A performed worst, with the average MCC value of 0.519 and AUC value of 0.841. The reason was that the performance of the SP (average = 0.724) was considerably lower than that of the SE (average = 0.864), which resulted in the low MCC and AUC values. The detailed performance of the test sets are given in Table S2.

To compare the performance of models from different fingerprints, the corresponding MCC values from ECFP6 and MACCS on the test sets were calculated. For the models built by naive Bayesian (Figure 4A), the performance of models from ECFP6, is superior to that from MACCS.

For example, the MCC values from ECFP6 range from 0.763 to 1 with an average of 0.975, whereas those from MACCS range from 0.285 to 0.951 with an average of 0.702. However, for the models constructed by recursive partitioning (Figure 4B), the MCC values from ECFP6 and MACCS show no significant difference, since most of the models share the approximate MCC values.

Additionally, the performance of models from different algorithms on the test sets was also compared. As given in Figure 5A, with the same fingerprint ECFP6, the models built by naive Bayesian performed better than those constructed by recursive partitioning. For example, the MCC values from naive Bayesian ranged from 0.763 to 1 with an average of 0.975, whereas those from recursive partitioning ranged from 0.530 to 0.988 with an average of 0.834. However, with the same fingerprint MACCS, the models from naive Bayesian (Figure 5B) performed worse than those built by recursive partitioning. For example, the MCC values from naive Bayesian ranged from 0.285 to 0.951 with an average of 0.702, whereas those from recursive partitioning ranged from 0.497 to 0.957 with an average of 0.794. In general, both of algorithms have their respective advantage.

As discussed above, it is difficult to judge that which fingerprint or algorithm is better than the other one. On the one hand, after the introduction of different fingerprints (ECFP6 and MACCS), each molecule can be represented as binary strings in different ways, therefore, the active compounds predicted by the models based on different fingerprints may be different. On the other hand, each algorithm has its respective advantage and limitation. For example, the models based on naive Bayesian algorithm perform better than those based on recursive partitioning when ECFP6 fingerprint is used, whereas the results are opposite when MACCS fingerprint is used. Because of this, it is necessary to combine the results of single classifiers to predict

chemical-protein interactions. In the following cases, chemical-protein interaction is defined as a potential interaction if the compound was predicted active by at least two out of four single classifiers within one target.

3.3 Analysis of the Important Fragments for Multi-target against AD Given by Naive Bayesian Classifier. One of the advantages of using a Bayesian classifier based on structural fingerprints is that it can identify important fragments frequently found in two classifying groups. Here we analyzed the important fragments for 21 targets against AD based on ECFP_6 fingerprint, of which the number of ligands was higher than 100. For each target, 20 good fragments favorable for binding and corresponding Bayesian scores were collected, then 420 fragments were obtained and their frequencies occurred in different targets were analyzed. This may be useful for MTDL design against AD. As given in Table 4, 24 fragments favorable for binding at least two targets toward AD are presented, which can be classified into three types according to their action mechanism. They are the fragments that can favorable bind to cholinesterase (ID = 1 to 13), targets related to APP (ID = 14 to 15), and neurotransmitter receptors (ID = 16 to 24). 8 fragments out of 24 are found to be favorable for binding to at least three targets of AD. For example, the fragment N, 2-dimethylpentan-3-amine (ID = 1) can make positive contribution to the binding of AChE, BuChE, BCAE1, and gamma-seretase.

3.4 Case 1. The Prediction of Polypharmacology for Six Known AD Drugs. In order to explore the polypharmacology of known AD drugs, we used all the 100 classifiers based on 25 key targets to predict potential targets for six approved AD drugs. Among them, five AD drugs (tacrine, rivastigmine, galantamine, donepezil, and huperizine A) are cholinesterase inhibitors and the last one (memantine) is an N-methyl-Daspartate (NMDA) receptor antagonist. The prediction results of polypharmacology for six approved AD drugs are given in Table S3. Here

chemical-protein interaction was defined as a positive interaction if the drug was predicted as "+1" by at least two out of the four single classifiers. Then 44 chemical-protein interaction pairs (Table S4) were obtained. In order to validate the predicted results, the PubChem bioassay database[37] was used to identify the known targets validated by experiments for each drug. The retrieve results are presented in Table 5 and Figure 6. Table 5 lists the predicted targets validated by experiments and corresponding binding affinities. There are 17 chemical-protein interaction pairs validated by the references.[38-52] For example, tacrine can target muscarnic M1 receptor with an IC_{50} value of 2 μM, and donepezil can target histamine receptor 3 (IC_{50} = 0.35 μM). Figure 6 shows the number of predicted targets versus that of identified targets for six approved drugs. A success rate of 38.6% (17/44) indicates the reliability with this method. Besides, the remaining predicted targets could be the new targets for the known drugs that merit further validation by experiments.

The six drugs and their predicted targets (44 CPI pairs) were compiled to build a polypharmacological interaction network with Cytoscape 2.8 (Figure 7). Octahedrons and circles correspond to drug nodes and protein nodes, respectively. A red edge is placed between a drug node and a protein node if the protein is a known target of that drug validated by experiments, while the other edges are colored black. As shown in Figure 7, all of the 6 drugs are predicted active toward at least four targets for each drug. Among them, donepezil is predicted positive toward 16 targets and 6 out of them have been validated by experiments, which supports the hot theme "one drug can hit multiple targets".

3.5 Case 2. The Target Prediction and Validation for 10 BuChE Inhibitors and 9 cdk5 inhibitors. In an AD brain, BuChE inhibitors (BuChEIs) have demonstrated to have a beneficial effect *in vivo*.[53] Previously, we built support vector machine (SVM) models and naive Bayesian

models to discriminate BuChE inhibitors (BuChEIs) from the non-inhibitors, resulting in 10 compounds exerting significant BuChE inhibitory activities with IC_{50} values ranging from 0.32 to 22.22 μM.[54] Cyclin-dependent kinase 5 (cdk5) has emerged as a principal therapeutic target for Alzheimer's disease.[55, 56] Recently, we developed two types of consensus models (CC-ANN and consensus prediction) to predict the inhibitory effects of a compound toward cdk5 activity. The assay results showed that 9 out of 40 compounds exerted cdk5/p35 inhibitory activities with IC_{50} values ranging from 9.23 to 95.57 μM.[36] Histamine receptor 3 (H_3R) blockade results in increased release of glutamate, ACh and dopamine in the prefrontal cortex, which helps to improve dementia symptoms, specifically in AD.[57] Preclinical studies in animal models indicate that H_3R antagonism may enhance memory and cognition.[58] Therefore, H_3R antagonists have been studied as drug candidates for the treatment of AD.

Now we are interested in whether these 19 compounds (10 BuChE inhibitors and 9 cdk5 inhibitors) can be correctly predicted by the new models in this study and whether they can also target H_3R as well as the other cholinesterase, AChE. To answer this question, the generated classifiers were used to predict the 19 compounds toward BuChE, cdk5, AChE, and H_3R. The predicted results are given in Table 6. Similarly, chemical-protein interaction is defined as a positive interaction if the compound was predicted active by at least two out of four single classifiers. Unsurprisingly, the results show that 10 out of 19 compounds are predicted active toward BuChE, and they are precisely the 10 BuChE inhibitors we have reported. At the same time, 7 out of 19 compounds are predicted as cdk5 inhibitors, and all of the 7 compounds belong to the 9 known cdk5 inhibitors. This means 17 chemical-protein interactions out of 19 (10+9) are predicted correctly toward BuChE and cdk5, and a success rate of 89% (17/19) is obtained,

which indicates the reliability of the models. Interestingly, there are 12 out of 19 compounds predicted as AChE inhibitors, and 14 out of 19 compounds are predicted as H_3R ligands.

In order to see whether our models miss the false negative compounds, all the 19 compounds was further evaluated by *in vitro* assays on AchE, cdk5 and H_3R, and the assay results were presented in Table 6. The assay results showed that none of the compounds predicted inactive by our models can inhibit or antagonize one of these three targets, which meant our models could exclude true negative compounds very well.

Herein donepezil served as the reference compound on AChE, and its IC_{50} value was 251 nM. Out of the 12 compounds predicted as AChE inhibitors, 7 compounds (all of them are known BuChE inhibitors) were identified to exhibit high or moderate activities against AChE, with the IC_{50} values ranging from 0.442 to 72.26 μM (Table 7). Inhibitory curves of three typical compounds (DL0410, J27139 and J18461) and reference compound donepezil toward AChE are presented in Figure 8.

On H_3R assay *in vitro*, thioperamide was selected as the reference compound with an IC_{50} value of 305 nM, which was consistent with the IC_{50} value in the reference.[59] Out of the 14 compounds predicted as H_3R ligands, 4 compounds were identified as H_3R antagonists. Their IC_{50} values ranged from 0.308 to 58.6 μM (Table 7). Inhibitory curves of three typical compounds (DL0410, J18461 and J18911) and reference compound thioperamide on H_3R are shown in Figure 9.

The chemical structures of all the 7 dual cholinesterase inhibitors and 4 H_3R antagonists are shown in Figure 10. Among the 7 dual cholinesterase inhibitors, the two best active compounds, DL0410 (IC_{50} = 0.442 μM) and J27139 (IC_{50} = 0.57 μM), share the same order of activities toward AChE with donepezil (IC_{50} = 0.251 μM), which shows a promising prospect on AD.

Among the 4 H_3R antagonists, 3 out of them (DL0410, J18457, and J18461) belong to dual cholinesterase inhibitors, which means all of the 3 compounds can target 3 proteins (AChE, BuChE, and H_3R) related to AD at least. They are typical multi-target-directed ligands (MTDLs) against AD. J18911, known as cdk5 inhibitor (9.23 μM) in our previous study, was also found as a H_3R antagonist (29.23 μM) here. To be exciting, DL0410, the most potent MTDL against AD, was identified as a dual cholinesterase inhibitor with IC_{50} values of 0.442 μM (AChE) and 3.57 μM (BuChE) as well as a H_3R antagonist with an IC_{50} of 0.308 μM, which showed promising to further develop as a drug candidate against AD.

In order to check whether the 8 MTDLs in this study are structurally diverse to known ligands, fingerprint similarity was performed with the fingerprint of FCFC6 to find the closest known active molecule (CKAM) to the MTDLs. Four databases of ligands (including 3469 AChE inhibitors, 1033 BuChE inhibitors, 2932 H3R ligands, and 462 cdk5 inhibitors, respectively) were prepared as input database, and each MTDL is set as reference ligand. The results are given in Table S5. Except for berberine derivatives J39068 and J39065, all the fingerprint similarities for 15 ligand-target pairs are below 0.82, which shows the MTDLs in this study are structurally diverse to known ligands.

To compare machine learning models with TargetHunter, the 8 MTDLs were further used as query ligands to explore the possible targets with TargetHunter, and the predicted top 10 targets versus the validated AD targets for each query were given in Table S6. The results showed that most of the predicted top 10 targets for each query were not related to AD, and TargetHunter could not identify the validated AD targets among the predicted top 10 targets except for J39068 and J18458. To sum up, machine learning models are more suitable for identifying possible targets of AD than structural similarity tool (TargetHunter) in this study.

3.6 Predicting ADEM Properties of the 19 Active Compounds.

The ADEM descriptors module available in Discovery Studio (DS) 4.0[27] was used to predict the 19 active compounds. The following properties, including human intestinal absorption (HIA), aqueous solubility, blood-brain-barrier penetration (BBB), cytochrome P450 (CYP450) 2D6 inhibition, plasma protein binding (PPB) and polar surface area (PSA), were predicted., and the resulted data were listed in Table S7. It showed most of the compounds have good intestinal absorption, good or optimal aqueous solubility, and medium or low ability to cross the blood-brain-barrier (BBB). Especially for DL0410, all of the ADEM properties were in the required druggability ranges. The detailed results and comparisons can be found in Table S5 (Supporting Information).

3.7 Limitations and Appropriate Application of mt-QSAR

Compared with structural similarity and docking methods, mt-QSAR has some distinct advantages for predictive accuracy and scaffold hopping. However, there are still some pitfalls and disadvantages which limit its application. First, it is clear that the high quality of negative data should be chosen from experimentally-validated inactive compounds, but this kind of data is quite limited. In most cases, DUD online database or compounds randomly extracted from commercial database are used for decoy generation, which may make some noise compounds involved. Secondly, a good classification model depends on large and diverse chemical space in the training set and test set. For the targets which only have few ligands or no ligand (such as orphan receptors), structural similarity or inverse docking will be better choices for target fishing. For example, for the targets such as 5HT1A, COMT, and GABA-B in this study, all of their corresponding ligands in training set are lower than 50, therefore, the models based on these dataset limit their application. Finally, compared with docking method, machine learning models

and structural similarity can not directly interpret receptor-ligand interaction which aids in understanding the mechanism of action and rational structural modification. Thus, appropriate application for each method or combination of different methods provide a new perspective to overcome their own shortages.

4. CONCLUCSION

Traditional computational approaches have focused on considering a series of compounds with a single target. However, identification of compounds that interact with multiple targets in a particular disease network such as Alzheimer's disease may provide unique insight into drug discovery. Here the mt-QSAR technique was adopted based on the "one-versus-the-rest" approach.

In this study, we have built 100 binary classifiers to predict the chemical-protein interactions for 25 key targets related to AD using mt-QSAR method. The various validations including cross-validation and test set validation confirmed the prediction reliability of the models. Moreover, the important fragments were also identified to characterize MTDLs against AD based on ECFP_6 fingerprint.

To demonstrate the application of the validated models, two cases were illustrated to systematically predict the polypharmacology for 6 approved AD drugs (case 1), and predict potential novel targets for 19 known AD active compounds (case 2). The prediction were confirmed by reported bioactivity data with a success rate of 38.6% for case 1 and 89% toward BuChE and cdk5 for case 2. To our surprise, several MTDLs against AD, including 7 dual cholinesterase inhibitors and 4 H_3R antagonists, were discovered by our *in vitro* experimental validation, and some of them showed nanomolar order of activities toward AChE or H_3R.

In short, this investigation is the first report using mt-QSAR approach, validated and proven with a successful pilot study on exploring polypharmacology for anti-AD drugs and discovering MTDLs. The methodology has potential application in identifying targets related to AD for small molecules, drug repurposing, and virtual screening for MTDLs. It also can provide a methodological reference for exploring polypharmacology of other complex diseases.

ASSOCIATED CONTENT

Supporting Information. The detailed performance of the 5-Fold cross-validation and the test set validation for 25 targets towards Alzheimer disease using NB and RP classifiers (Tables S1-S2), the prediction results of polypharmacology for six approved AD drugs (Table S3), and 44 chemical-protein interaction pairs predicted by 100 classifiers (Table S4). This material is available, free of charge, via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel & Fax: 86-10-83150885. E-mail: liuailin@imm.ac.cn or dugh@imm.ac.cn.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was funded in part of the Research Special Fund for the National Great Science and Technology Projects (2012ZX09301002-001-001), the International Collaboration Project (2011DFR31240), Peking Union Medical College graduate student innovation fund (2013-1007-18), and Beijing New-star Plan of Science and Technology (xx2013065). We also would like to

give special thanks to Hanzhong Ke, a PhD candidate in University of Illinois Urbana Champaign, for his help in polishing English.

REFERENCES

(1) Wimo, A.; Jönsson, L.; Bond, J.; Prince, M.; Winblad, B. The Worldwide Economic Impact of Dementia 2010. *Alzheimers Dement.* **2013**, 9, 1-11.

(2) Cavalli, A.; Bolognesi, M. L.; Minarini, A.; Rosini, M.; Tumiatti, V.; Recanatini, M.; Melchiorre, C. Multi-Target-Directed Ligands to Combat Neurodegenerative Diseases. *J. Med. Chem.* **2008**, 51, 347-372.

(3) Merino, A.; Bronowska, A. K.; Jackson, D. B.; Cahill, D. J. Drug Profiling: Knowing Where It Hits. *Drug Discovery Today* **2010**, 15, 749-756.

(4) Hopkins, A. L. Network Pharmacology: The Next Paradigm in Drug Discovery. *Nat. Chem. Biol.* **2008**, 4, 682-690.

(5) Ashburn, T. T.; Thor, K. B. Drug Repositioning: Identifying and Developing New Uses for Existing Drugs. *Nat. Rev. Drug Discovery* **2004**, 3, 673-683.

(6) Jenkins, J. L.; Bender, A.; Davies, J. W. *In silico* Target Fishing: Predicting Biological Targets from Chemical Structure. *Drug Discovery Today: Technol.* **2007**, 3, 413-421.

(7) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating Protein Pharmacology by Ligand Chemistry. *Nat. Biotechnol.* **2007**, 25, 197-206.

(8) Dunkel, M.; Günther, S.; Ahmed, J.; Wittig, B.; Preissner, R. SuperPred: Drug Classification and Target Prediction. *Nucleic Acids Res.* **2008**, 36, W55-W59.

(9) Yan, X.; Li, J.; Liu, Z.; Zheng, M.; Ge, H.; Xu, J. Enhancing Molecular Shape Comparison by Weighted Gaussian Functions. *J. Chem. Inf. Model.* **2013**, 53, 1967-1978.

(10) Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper Server: a Web Server for Potential Drug Target Identification Using Pharmacophore Mapping Approach. *Nucleic Acids Res.* **2010**, 38, W609-W614.

(11) Gong, J.; Cai, C.; Liu, X.; Ku, X.; Jiang, H.; Gao, D.; Li, H. ChemMapper: A Versatile Web Server for Exploring Pharmacology and Chemical Structure Association Based on Molecular 3D Similarity Method. *Bioinformatics* **2013**, 29, 1827-1829.

(12) Luo, H.; Chen, J.; Shi, L.; Mikailov, M.; Zhu, H.; Wang, K.; He, L.; Yang, L. DRAR-CPI: A Server for Identifying Drug Repositioning Potential and Adverse Drug Reactions Via the Chemical–Protein Interactome. *Nucleic Acids Res.* **2011**, 39, W492-W498.

(13) Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J. TarFisDock: A Web Server for Identifying Drug Targets with Docking Approach. *Nucleic Acids Res.* **2006**, 34, W219-W224.

(14) Wang, L.; Ma, C.; Wipf, P.; Liu, H.; Su, W.; Xie, X. Q. TargetHunter: An In Silico Target Identification Tool for Predicting Therapeutic Potential of Small Organic Molecules Based on Chemogenomic Database. *AAPS J.* **2013**, 15, 395-406.

(15) Liu, H.; Wang, L.; Lv, M.; Pei, R.; Li, P.; Pei, Z.; Wang, Y.; Su, W.; Xie, X. Q. AlzPlatform: An Alzheimer's Disease Domain-Specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research. *J. Chem. Inf. Model.* **2014**, 54, 1050-1060.

(16) Fang, J.; Huang, D.; Zhao, W.; Ge, H.; Luo, H. B.; Xu, J. A New Protocol for Predicting Novel GSK-3β ATP Competitive Inhibitors. *J. Chem. Inf. Model.* **2011**, 51, 1431-1438.

(17) Vina, D.; Uriarte, E.; Orallo, F.; González-Díaz, H. Alignment-Free Prediction of a Drug-Target Complex Network Based on Parameters of Drug Connectivity and Protein Sequence of Receptors. *Mol. Pharmaceutics* **2009**, 6, 825-835.

(18) Wang, F.; Liu, D.; Wang, H.; Luo, C.; Zheng, M.; Liu, H.; Zhu, W.; Luo, X.; Zhang, J.; Jiang, H. Computational Screening for Active Compounds Targeting Protein Sequences: Methodology and Experimental Validation. *J. Chem. Inf. Model.* **2011**, 51, 2821-2828.

(19) Yu, H.; Chen, J.; Xu, X.; Li, Y.; Zhao, H.; Fang, Y.; Li, X.; Zhou, W.; Wang, W.; Wang, Y. A Systematic Prediction of Multiple Drug-Target Interactions from Chemical, Genomic, and Pharmacological Data. *PLoS One* **2012**, 7, e37608.

(20) Cao, D. S.; Liu, S.; Xu, Q. S.; Lu, H. M.; Huang, J. H.; Hu, Q. N.; Liang, Y. Z. Large-Scale Prediction of Drug–Target Interactions Using Protein Sequences and Drug Topological Structures. *Anal. Chim. Acta* **2012**, 752, 1-10.

(21) Cheng, F.; Zhou, Y.; Li, J.; Li, W.; Liu, G.; Tang, Y. Prediction of Chemical-Protein Interactions: Multitarget-QSAR Versus Computational Chemogenomic Methods. *Mol. Biosyst.* **2012**, 8, 2373-2384.

(22) Zhu, F.; Shi, Z.; Qin, C.; Tao, L.; Liu, X.; Xu, F.; Zhang, L.; Song, Y.; Liu, X.; Zhang, J.; Han, B.; Zhang, P.; Chen, Y. Therapeutic Target Database Update 2012: A Resource for Facilitating Target-Oriented Drug Discovery. *Nucleic Acids Res.* **2012**, 40, D1128-1136.

(23) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: A Web-Accessible Database of Experimentally Determined Protein-Ligand Binding Affinities. *Nucleic Acids Res.* **2007**, 35, D198-201.

(24) Gao, M.; Liu, A. L.; Du, G. H. High-Throughput Screening for Butyrylcholinesterase Inhibitors. *Chin. J. New Drugs* **2009**, 12, 021.

(25) 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. *J. Med. Chem.* **2012**, 55, 6582-6594.

(26) Morgan, H. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, 5, 107-113.

(27) *Discovery Studio, version 4.0*, Accelrys Inc.: San Diego, CA: 2014.

(28) Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, 32, 1466-1474.

(29) Xia, X.; Maliski, E. G.; Gallant, P.; Rogers, D. Classification of Kinase Inhibitors Using a Bayesian Model. *J. Med. Chem.* **2004**, 47, 4463-4470.

(30) Glick, M.; Jenkins, J. L.; Nettles, J. H.; Hitchings, H.; Davies, J. W. Enrichment of High-Throughput Screening Data with Increasing Levels of Noise Using Support Vector Machines, Recursive Partitioning, and Laplacian-Modified Naive Bayesian Classifiers. *J. Chem. Inf. Model.* **2006**, 46, 193-200.

(31) Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. Assessing the Accuracy of Prediction Algorithms for Classification: An Overview. *Bioinformatics* **2000**, 16, 412-424.

(32) Ellman, G. L.; Courtney, K. D.; Featherstone, R. M. A New and Rapid Colorimetric Determination of Acetylcholinesterase Activity. *Biochem. Pharmacol.* **1961**, 7, 88-95.

(33) Fang, J.; Wu, P.; Yang, R.; Gao, L.; Li, C.; Wang, D.; Wu, S.; Liu, A. L.; Du, G. L. Inhibition of Acetylcholinesterase by Two Genistein Derivatives: Kinetic Analysis, Molecular Docking and Molecular Dynamics Simulation. *Acta Pharmaceutica Sinica B* **2014**, DOI: 10.1016/j.apsb.2014.10.002

(34) Galarneau, A.; Primeau, M.; Trudeau, L.E.; Michnick, S.W. Beta-Lactamase Protein Fragment Complementation Assays as in Vivo and in Vitro Sensors of Protein Protein Interactions. *Nat. Biotechnol.* **2002**, 20, 619–622.

(35) Koresawa, M; Okabe, T. High-Throughput Screening with Quantitation of ATP Consumption: A Universal Non-Radioisotope, Homogeneous Assay for Protein Kinase. *Assay Drug Dev. Technol.* **2004**, 2: 153-160.

(36) Fang, J.; Yang, R.; Gao, L.; Yang, S.; Pang, X; Li, C; He, Y; Liu, A. L.; Du, G. H. Consensus Models for CDK5 Inhibitors in Silico and Their Application to Inhibitor Discovery. DOI: 10.1007/s11030-014-9561-3

(37) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B. A. PubChem's BioAssay database. *Nucleic Acids Res.* **2012**, 40, D400-D412.

(38) Khorana, N.; Changwichit, K.; Ingkaninan, K.; Utsintong, M. Prospective Acetylcholinesterase Inhibitory Activity of Indole and its Analogs. *Bioorg. Med. Chem. Lett.* **2012**, 22, 2885-2888.

(39) Schott, Y.; Decker, M.; Rommelspacher, H.; Lehmann, J. 6-Hydroxy-and 6-methoxy-β-carbolines as Acetyl-and Butyrylcholinesterase Inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, 16, 5840-5843.

(40) Mohamed, T.; Yeung, J. C.; Vasefi, M. S.; Beazely, M. A.; Rao, P. P. Development and Evaluation of Multifunctional Agents for Potential Treatment of Alzheimer's Disease: Application to a Pyrimidine-2, 4-diamine Template. *Bioorg. Med. Chem. Lett.* **2012**, 22, 4707-4712.

(41) Takayama, H.; Yaegashi, Y.; Kitajima, M.; Han, X.; Nishimura, K.; Okuyama, S.; Igarashi, K. Design, Synthesis, and Biological Evaluation of Tricyclic Heterocycle-tetraamine Conjugates as Potent NMDA Channel Blockers. *Bioorg. Med. Chem. Lett.* **2007**, 17, 4729-4732.

(42) Fernández-Bachiller, M. I.; Pérez, C. n.; Monjas, L.; Rademann, J. r.; Rodríguez-Franco, M. I. New Tacrine–4-Oxo-4 H-chromene Hybrids as Multifunctional Agents for the Treatment of Alzheimer's Disease, with Cholinergic, Antioxidant, and β-Amyloid-Reducing Properties. *J. Med. Chem.* **2012**, 55, 1303-1317.

(43) Sowell, J. W., Sr.; Tang, Y.; Valli, M. J.; Chapman, J. M., Jr.; Usher, L. A.; Vaughan, C. M.; Kosh, J. W. Synthesis and Cholinergic Properties of bis[[(dimethylamino)methyl]furanyl] Analogues of Ranitidine. *J. Med. Chem.* **1992**, 35, 1102-1108.

(44) Cavalli, A.; Bolognesi, M. L.; Minarini, A.; Rosini, M.; Tumiatti, V.; Recanatini, M.; Melchiorre, C. Multi-Target-Directed Ligands to Combat Neurodegenerative Diseases. *J. Med. Chem.* **2008**, 51, 347-372.

(45) Bolognesi, M. L.; Bartolini, M.; Cavalli, A.; Andrisano, V.; Rosini, M.; Minarini, A.; Melchiorre, C. Design, Synthesis, and Biological Evaluation of Conformationally Restricted Rivastigmine Analogues. *J. Med. Chem.* **2004**, 47, 5945-5952.

(46) Samadi, A.; Estrada, M.; Pérez, C.; Rodríguez-Franco, M. I.; Iriepa, I.; Moraleda, I.; Chioua, M.; Marco-Contelles, J. < i> Pyridonepezils</i>, New Dual AChE Inhibitors as Potential Drugs for the Treatment of Alzheimer's Disease: Synthesis, Biological Assessment, and Molecular Modeling. *Eur. J. Med. Chem.* **2012**, 57, 296-301.

(47) Camps, P.; Formosa, X.; Galdeano, C.; Gómez, T.; Munoz-Torrero, D.; Scarpellini, M.; Viayna, E.; Badia, A.; Clos, M. V.; Camins, A. Novel Donepezil-Based Inhibitors of Acetyl-and Butyrylcholinesterase and Acetylcholinesterase-Induced β-amyloid Aggregation. *J. Med. Chem.* **2008**, 51, 3588-3598.

(48) Kwon, Y. E.; Park, J. Y.; No, K. T.; Shin, J. H.; Lee, S. K.; Eun, J. S.; Yang, J. H.; Shin, T. Y.; Kim, D. K.; Chae, B. S. Synthesis, in Vitro Assay, and Molecular Modeling of New Piperidine Derivatives Having Dual Inhibitory Potency Against Acetylcholinesterase and Aβ< sub> 1–42</sub> Aggregation for Alzheimer's Disease Therapeutics. *Bioorg. Med. Chem.* **2007**, 15, 6596-6607.

(49) Bembenek, S. D.; Keith, J. M.; Letavic, M. A.; Apodaca, R.; Barbier, A. J.; Dvorak, L.; Aluisio, L.; Miller, K. L.; Lovenberg, T. W.; Carruthers, N. I. Lead Identification of Acetylcholinesterase Inhibitors–Histamine H< sub> 3</sub> Receptor Antagonists from Molecular Modeling. *Bioorg. Med. Chem.* **2008**, 16, 2968-2973.

(50) Bolea, I.; Juárez-Jiménez, J.; de los Ríos, C. b.; Chioua, M.; Pouplana, R. n.; Luque, F. J.; Unzeta, M.; Marco-Contelles, J.; Samadi, A. Synthesis, Biological Evaluation, and Molecular Modeling of Donepezil and N-[(5-(benzyloxy)-1-methyl-1 H-indol-2-yl) methyl]-N-methylprop-2-yn-1-amine Hybrids as New Multipotent Cholinesterase/monoamine Oxidase Inhibitors for the Treatment of Alzheimer's Disease. *J. Med. Chem.* **2011**, 54, 8251-8270.

(51) Yan, J.; Sun, L.; Wu, G.; Yi, P.; Yang, F.; Zhou, L.; Zhang, X.; Li, Z.; Yang, X.; Luo, H.; Qiu, M. Rational Design and Synthesis of Highly Potent Anti-Acetylcholinesterase Activity Huperzine A Derivatives. *Bioorg. Med. Chem.* **2009**, 17, 6937-6941.

(52) Camps, P.; Gómez, E.; Muñoz-Torrero, D.; Badia, A.; Clos, M. V.; Curutchet, C.; Muñoz-Muriedas, J.; Luque, F. J. Binding of 13-amidohuprines to acetylcholinesterase: Exploring the Ligand-Induced Conformational Change of the gly117-gly118 Peptide Bond in the Oxyanion Hole. *J. Med. Chem.* **2006**, 49, 6833-6840.

(53) Greig, N. H.; Utsuki, T.; Yu, Q.-s.; Zhu, X.; Holloway, H. W.; Perry, T.; Lee, B.; Ingram, D. K.; Lahiri, D. K. A New Therapeutic Target in Alzheimer's Disease Treatment: Attention to Butyrylcholinesterase. *Curr. Med. Res. Opin.* **2001**, 17, 159-165.

(54) Fang, J.; Yang, R.; Gao, L.; Zhou, D.; Yang, S.; Liu, A. L.; Du, G. L. Predictions of BuChE Inhibitors Using Support Vector Machine and Naive Bayesian Classification Techniques in Drug Discovery. *J. Chem. Inf. Model.* **2013**, 53, 3009-3020.

(55) Goedert, M. Tau Protein and the Neurofibrillary Pathology of Alzheimer's Disease. *Trends Neurosci.* **1993**, 16, 460-465.

(56) Lau, L. F.; Seymour, P. A.; Sanner, M. A.; Schachter, J. B. Cdk5 as a Drug Target for the Treatment of Alzheimer's Disease. *J. Mol. Neurosci.* **2002**, 19, 267-273.

(57) Schneider, E.H.; Neumann, D.; Seifert, R. Modulation of Behavior by the Histaminergic System: Lessons from HDC-, H3R- and H4R-Deficient Mice. *Neurosci. Biobehav. R.* **2014**, 47, 101-121.

(58) Panula, P.; Nuutinen, S. The Histaminergic Network in the Brain: Basic Organization and Role in Disease. *Nat. Rev. Neurosci.* **2013**, 14, 472–487.

(59) Tang, L.; Zhao, L.; Hong, L.; Yang, F.; Sheng, R.; Chen, J.; Shi, Y.; Zhou, N.; Hu, Y. Design and Synthesis of Novel 3-Substituted-Indole Derivatives as Selective H3 Receptor Antagonists and Potent Free Radical Scavengers. *Bioorg. Med. Chem.* **2013**, 21, 5936-5944.

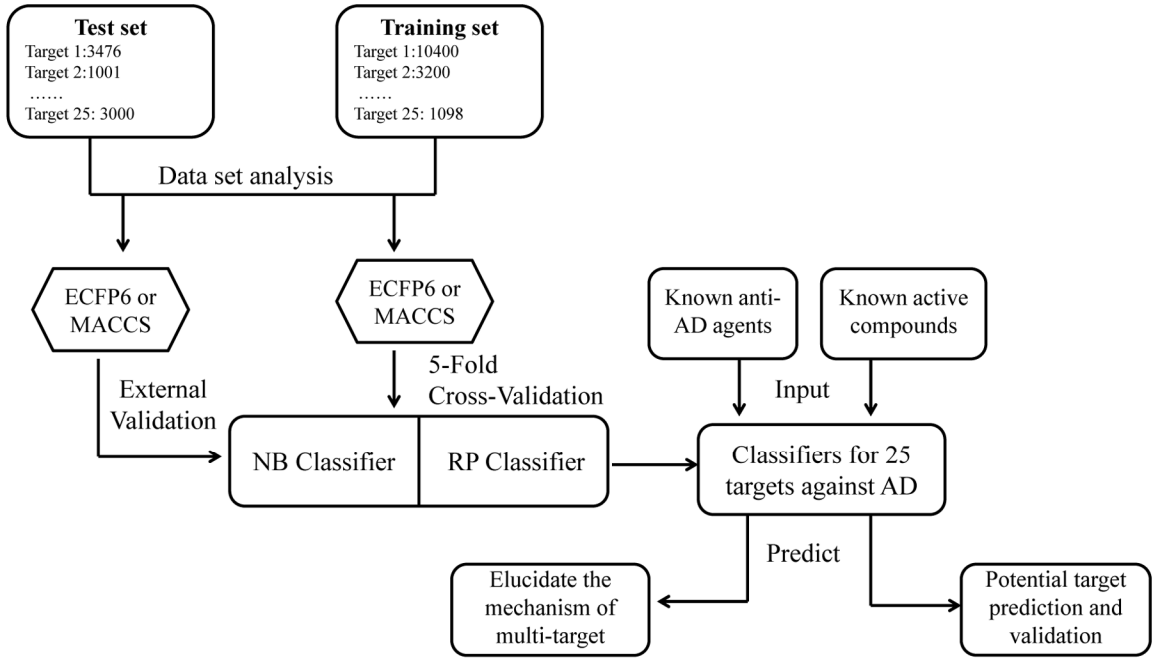
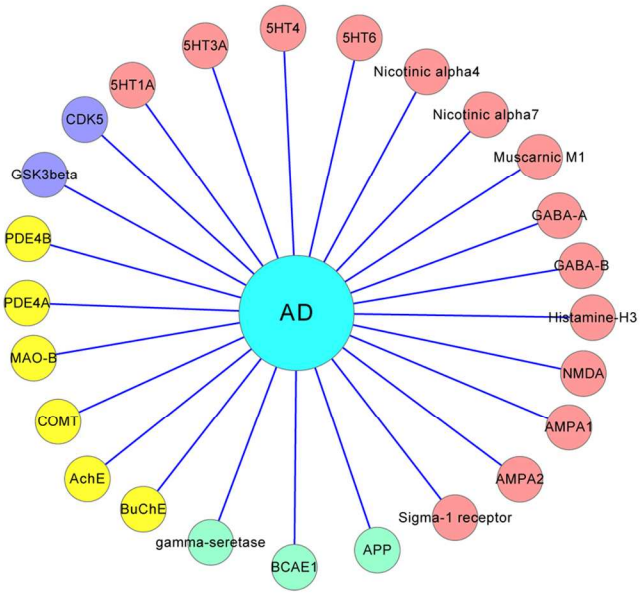


Figure 1. The workflow of multitarget quantitative structure-activity relationships (mt-QSAR) toward Alzheimer's disease.



Key target

AchE	Acetylcholinesterase
AMPA1	Glutamate receptor ionotropic AMPA1 receptor
AMPA2	Glutamate receptor ionotropic AMPA2 receptor
APP	Amyloid beta A4 precursor binding protein
BACE1	Beta-secretase 1
BuChE	Butyrylcholinesterase
CDK5	Cyclin-dependent kinase 5
COMT	Catechol O-methyltransferase
GABA-A	Gamma-Aminobutyric acid A receptor
GABA-B	Gamma-Aminobutyric acid B receptor
GSK3beta	Glycogen synthase kinase 3 beta
MAO-B	Monoamine Oxidase B
NMDA	Glutamate (NMDA) receptor
PDE4A	Phosphodiesterase Type 4A
PDE4B	Phosphodiesterase Type 4B
5HT1A	5-hydroxytryptamine 1A
5HT3A	5-hydroxytryptamine 3A
5HT4	5-hydroxytryptamine 4
5HT6	5-hydroxytryptamine 6

Figure 2. Summary of 25 key targets related to Alzheimer's disease.

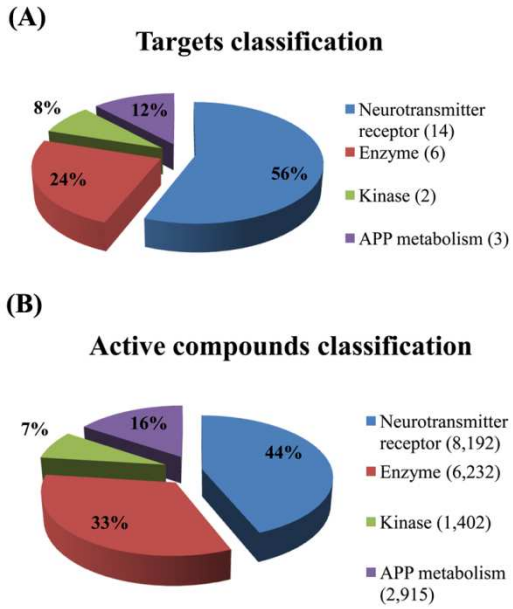


Figure 3. Targets and active compounds classification within the entire data set

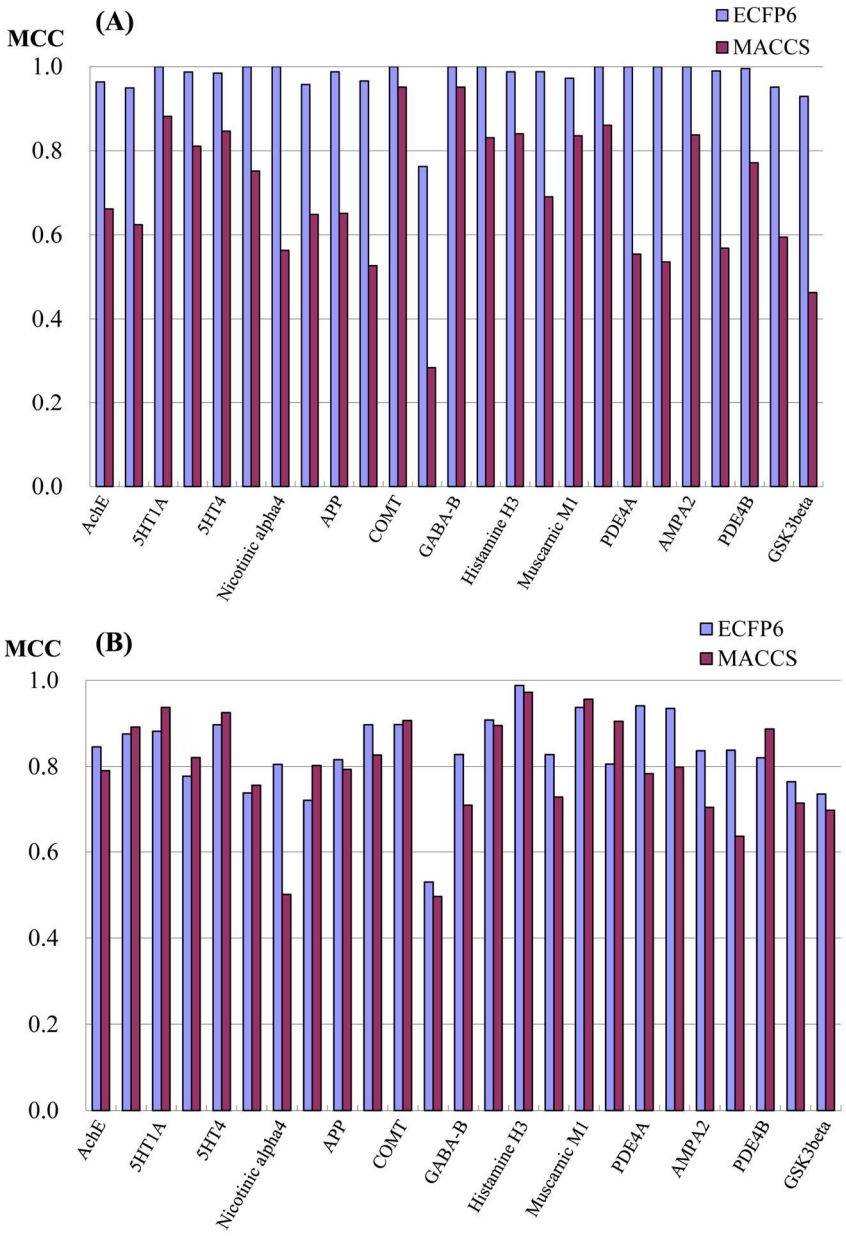


Figure 4. The performance comparisons of classifiers using naive Bayesian (A) and recursive partition (B) with different fingerprints on the test set.

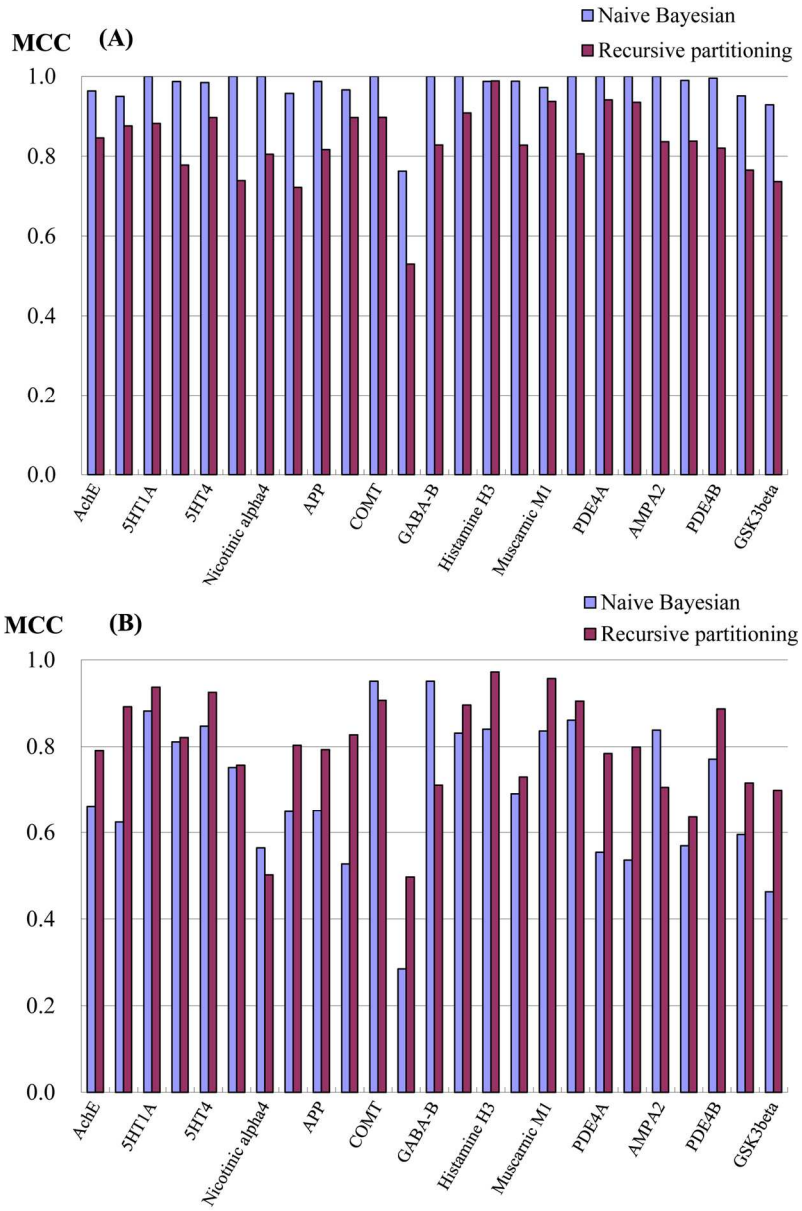


Figure 5. The performance comparisons of classifiers using different algorithms with ECFP6 (A) and MACCS (B) fingerprints on the test sets.

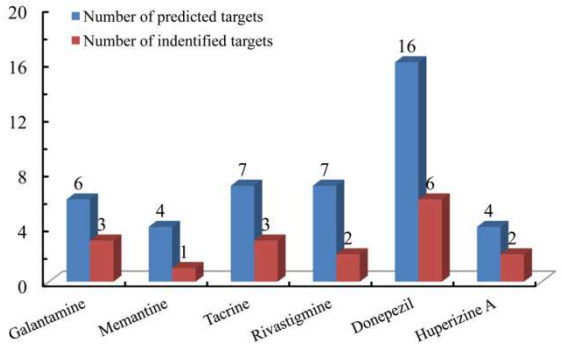


Figure 6. The number of predicted targets (cuboids in dark blue) versus that of identified targets (cuboids in red) for six approved drugs

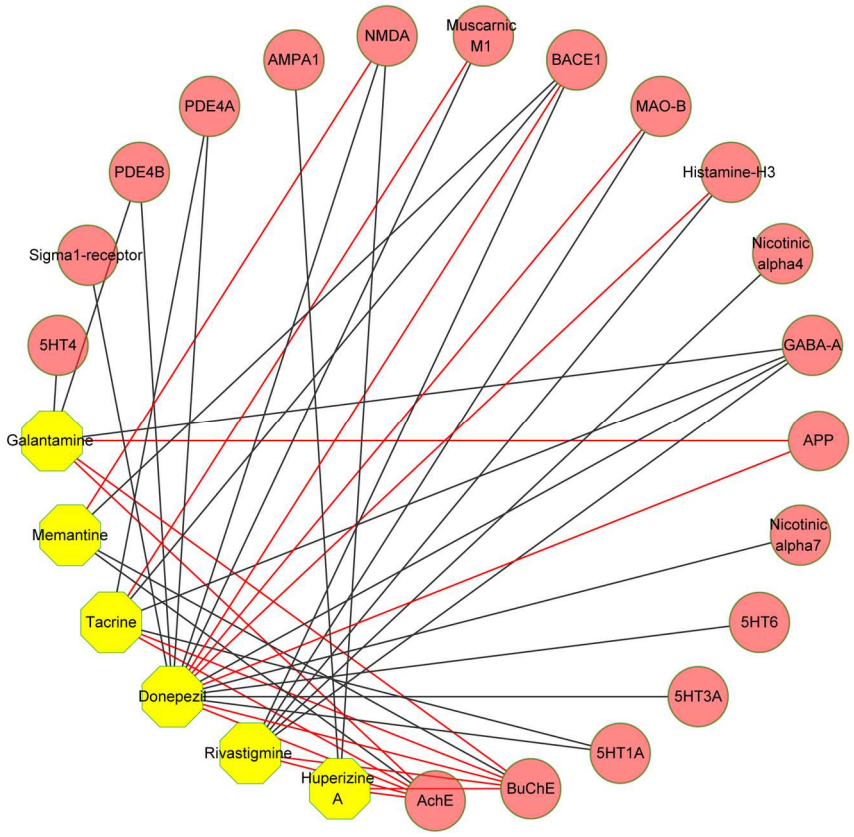


Figure 7. The polypharmacology analysis of 6 approved AD drugs based on the 100 classifiers in this study. Octahedrons and circles represent drug nodes and protein nodes, respectively. The red edge stands for the chemical-protein interaction validated by experiments, and the black edge stands for the remaining predicted interactions which merit further validation by experiments.

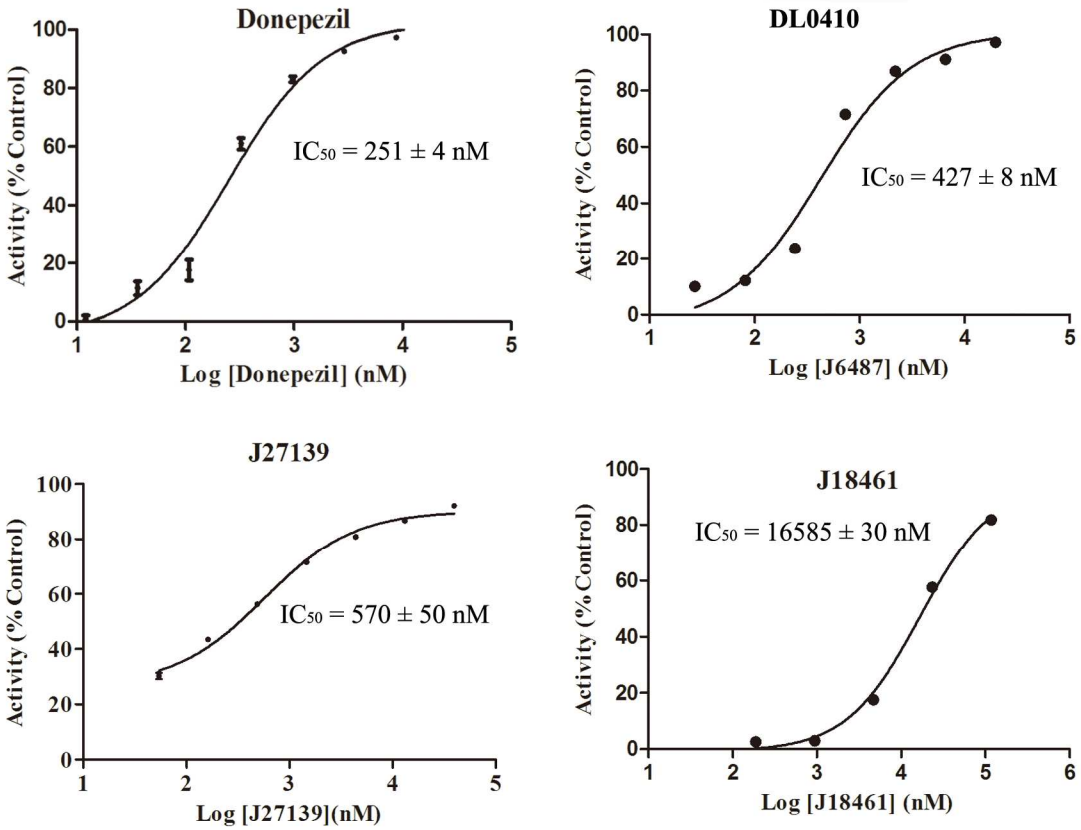


Figure 8. Inhibitory curves of three typical compounds and reference compound donepezil against AChE.

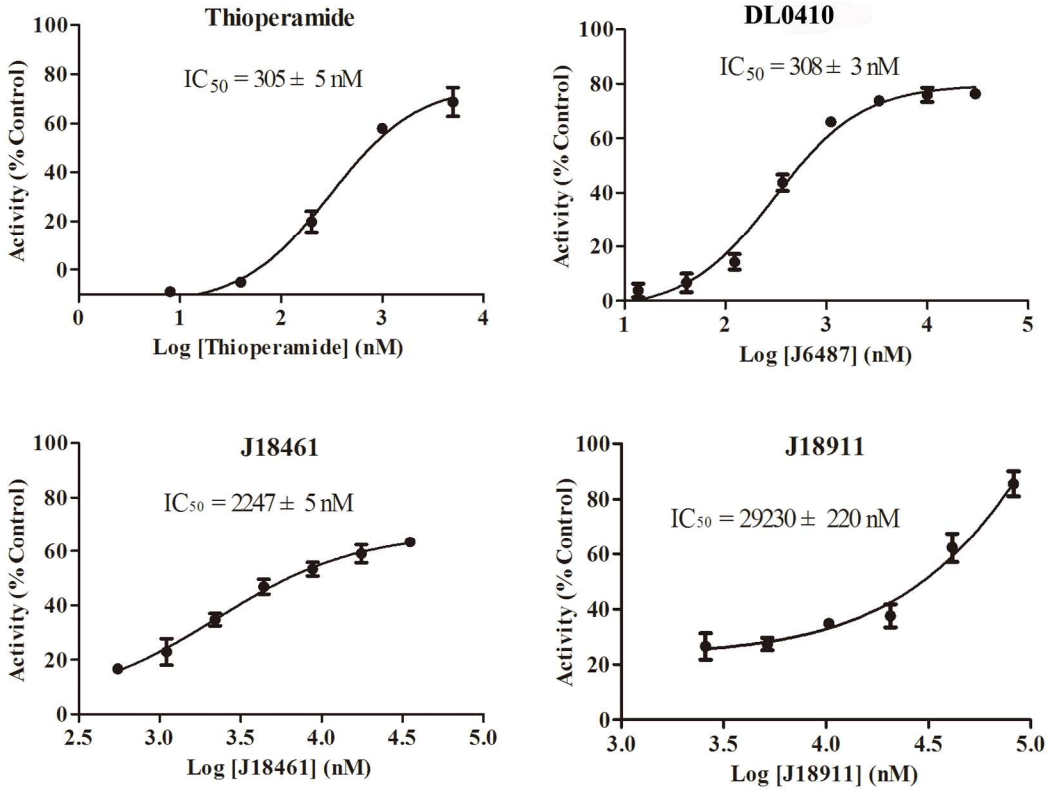


Figure 9. Inhibitory curves of three typical compounds and reference compound thioperamide on histamine receptor 3 (H_3R).

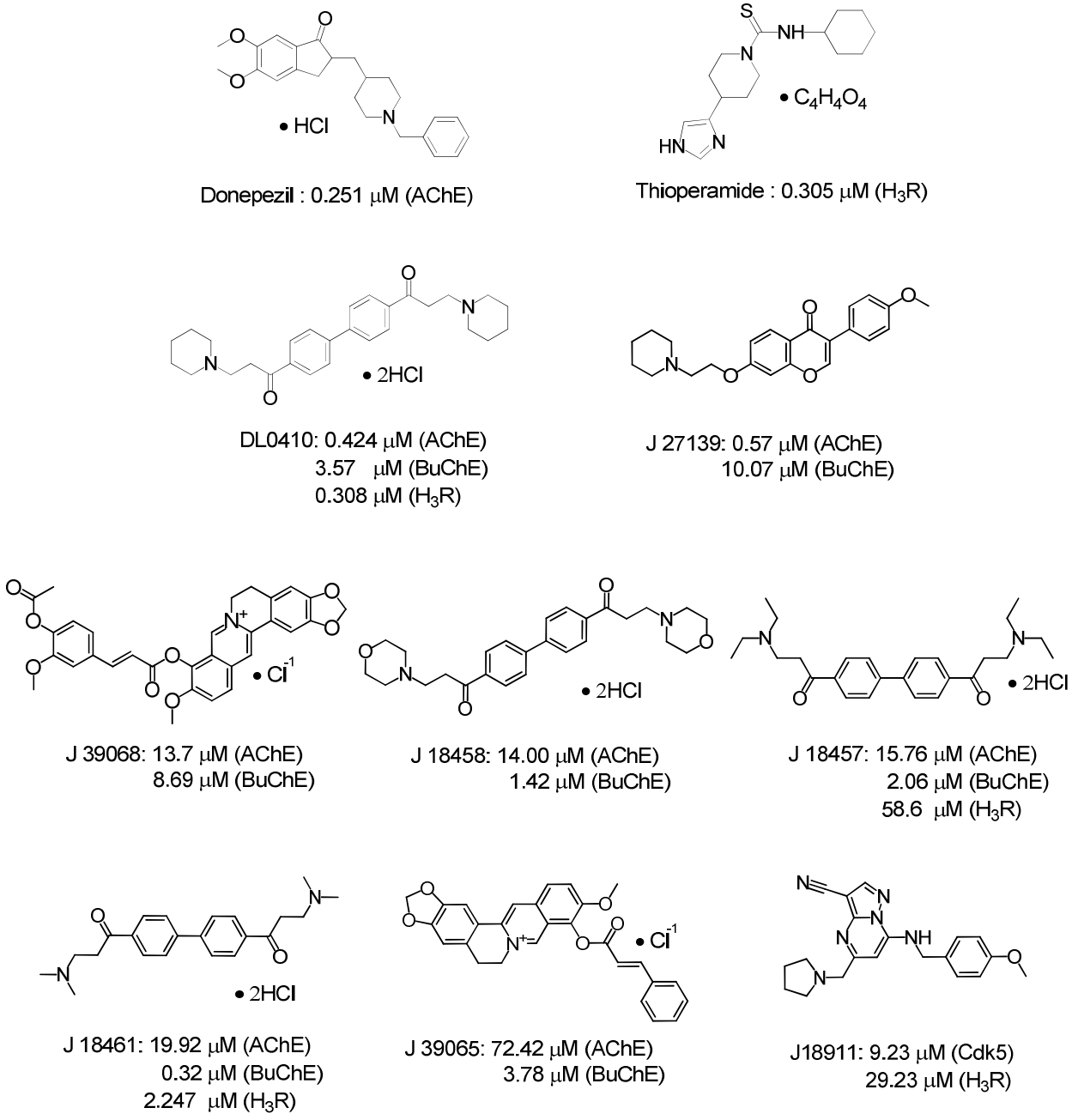


Figure 10. Chemical structures and their IC_{50} values of the 7 dual cholinesterase inhibitors and 4 histamine receptor 3 (H_3R) antagonists.

Table 1 Detailed Statistical Description of the Entire Data Set Based on the Multi-label Classification Strategy.

Target	Training set (ECFP2)				Test set (ECFP2)			
	Inhibitors	decoys	Total	Tanimoto index	Inhibitors	decoys	Total	Tanimoto index
AchE	2600	7800	10400	0.157	869	2607	3476	0.171
BuChE	800	2400	3200	0.194	233	768	1001	0.191
5HT1A	35	105	140	0.146	10	30	40	0.171
5HT3A	300	900	1200	0.138	102	306	408	0.154
5HT4	375	1125	1500	0.164	126	378	504	0.176
5HT6	80	240	320	0.133	31	93	124	0.149
Nicotinic alpha4	165	495	660	0.127	57	171	228	0.130
Nicotinic alpha7	700	2000	2700	0.151	257	749	1006	0.158
APP	155	465	620	0.140	53	159	212	0.157
BCAE1	1780	5340	7120	0.182	591	1773	2364	0.182
COMT	40	120	160	0.133	13	39	52	0.147
GABA-A	105	315	420	0.107	35	105	140	0.117
GABA-B	30	90	120	0.107	13	39	52	0.121
gamma-seretase	250	750	1000	0.163	86	258	344	0.176
H_3R	2200	6600	8800	0.158	732	2196	2928	0.168
MAO-B	490	1470	1960	0.133	161	483	644	0.146
Muscarnic M1	930	2790	3720	0.146	309	927	1236	0.159
NMDA	410	1230	1640	0.146	134	402	536	0.158
PDE4A	330	990	1320	0.160	110	330	440	0.170

PDE4B	440	1320	1760	0.159	146	438	584	0.175
AMPA1	110	330	440	0.142	40	120	160	0.160
AMPA2	100	300	400	0.142	31	93	124	0.161
Sigma-1 receptor	580	1740	2320	0.159	195	585	780	0.172
CDK5	360	1080	1440	0.161	102	420	522	0.157
GSK3beta	750	2250	3000	0.158	190	908	1098	0.158

Table 2 Performance of the 5-Fold Cross-Validation for 25 targets towards Alzheimer disease Using NB and RP Classifiers

Modeling Methods	ECFP6				MACCS			
	NB		RP		NB		RP	
	MCC	AUC	MCC	AUC	MCC	AUC	MCC	AUC
AchE	0.976	0.995	0.836	0.942	0.553	0.874	0.817	0.928
BuChE	0.982	0.998	0.892	0.975	0.652	0.917	0.918	0.971
5HT1A	0.981	0.998	0.728	0.907	0.892	0.987	0.812	0.896
5HT3A	0.993	0.995	0.866	0.941	0.841	0.957	0.88	0.942
5HT4	0.986	0.998	0.892	0.975	0.872	0.984	0.925	0.974
5HT6	0.983	0.999	0.778	0.908	0.828	0.967	0.852	0.933
α4-nAChR	0.996	0.995	0.934	0.976	0.759	0.971	0.858	0.973
α7-nAChR	0.978	0.994	0.845	0.955	0.655	0.921	0.861	0.948
APP	0.991	0.996	0.903	0.943	0.649	0.918	0.829	0.907
BCAE1	0.966	0.998	0.93	0.976	0.575	0.914	0.894	0.965
COMT	1	1	1	1	0.967	0.999	0.952	0.979
GABA-A	0.91	0.929	0.59	0.756	0.514	0.814	0.664	0.725
GABA-B	1	0.996	0.833	0.949	0.914	0.992	0.888	0.854
γ-seretase	1	1	0.929	0.981	0.79	0.969	0.914	0.965
H_3R	0.989	1	0.993	0.997	0.852	0.986	0.975	0.992
MAO-B	0.977	0.995	0.825	0.949	0.668	0.892	0.779	0.923
Muscarnic M1	0.984	0.995	0.951	0.987	0.823	0.978	0.955	0.989
NMDA	0.993	1	0.894	0.972	0.82	0.944	0.878	0.949

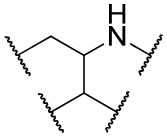
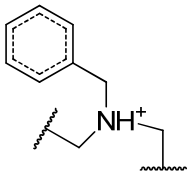
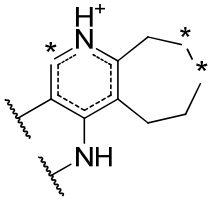
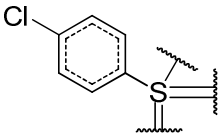
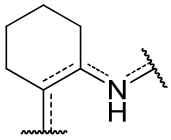
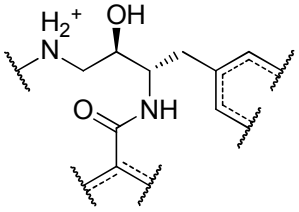
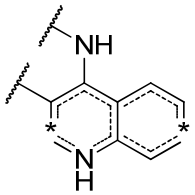
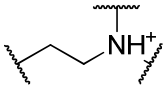
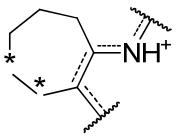
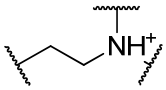
PDE4A	0.994	0.998	0.947	0.97	0.65	0.912	0.861	0.954
PDE4B	0.983	0.991	0.853	0.971	0.586	0.895	0.81	0.934
AMPA1	0.97	0.976	0.873	0.969	0.786	0.908	0.825	0.916
AMPA2	1	0.994	0.852	0.968	0.548	0.875	0.763	0.919
Sigma1	0.994	0.999	0.891	0.981	0.829	0.977	0.879	0.962
CDK5	0.989	0.999	0.864	0.962	0.67	0.934	0.838	0.965
GSK3beta	0.977	0.997	0.868	0.949	0.463	0.863	0.729	0.872

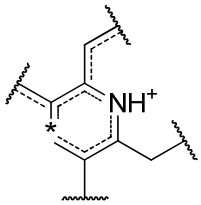
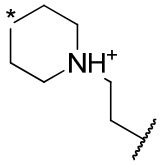
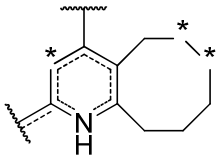
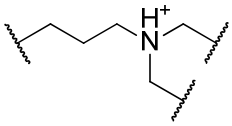
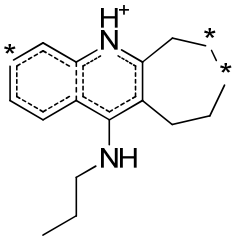
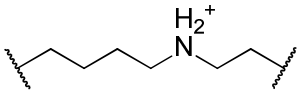
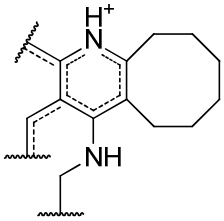
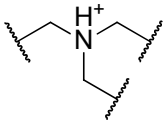
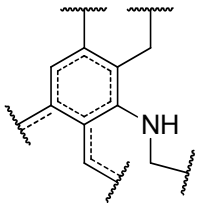
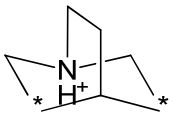
Table 3 Performance of the test set Validation for 25 targets towards Alzheimer's disease Using NB and RP Classifiers

Modeling	ECFP6				MACCS			
Methods	NB		RP		NB		RP	
	MCC	AUC	MCC	AUC	MCC	AUC	MCC	AUC
AChE	0.964	0.986	0.846	0.979	0.662	0.929	0.791	0.981
BuChE	0.95	0.982	0.876	0.978	0.624	0.919	0.892	0.972
5HT1A	1	1	0.882	1	0.882	1	0.937	0.983
5HT3A	0.987	0.995	0.778	0.971	0.811	0.985	0.821	0.974
5HT4	0.984	0.994	0.897	0.981	0.847	0.995	0.925	0.978
5HT6	1	1	0.739	0.942	0.752	0.99	0.757	0.954
α4-nAChR	1	1	0.805	0.917	0.564	0.874	0.502	0.861
α7-nAChR	0.957	0.983	0.722	0.968	0.649	0.947	0.803	0.977
APP	0.988	1	0.816	0.967	0.651	0.93	0.793	0.977
BCAE1	0.966	0.999	0.897	0.988	0.527	0.916	0.827	0.986
COMT	1	1	0.897	0.949	0.951	1	0.907	0.974
GABA-A	0.763	0.886	0.53	0.89	0.285	0.741	0.497	0.846
GABA-B	1	1	0.828	0.972	0.951	1	0.711	0.945
γ-secretase	1	1	0.908	0.995	0.831	0.95	0.896	0.985
H_3R	0.987	0.995	0.988	1	0.84	0.987	0.972	0.998
MAO-B	0.988	0.995	0.828	0.988	0.691	0.894	0.73	0.958
Muscarnic M1	0.972	0.989	0.937	0.988	0.836	0.979	0.957	0.999
NMDA	1	1	0.806	0.975	0.861	0.975	0.905	0.994
PDE4A	1	1	0.941	0.993	0.554	0.878	0.784	0.976
PDE4B	1	0.998	0.935	0.969	0.536	0.907	0.799	0.978
AMPA1	1	1	0.837	0.986	0.838	0.991	0.706	0.965

AMPA2	0.99	1	0.838	1	0.569	0.895	0.636	0.955
Sigma1	0.995	1	0.821	0.969	0.771	0.99	0.887	0.993
CDK5	0.951	0.985	0.765	0.955	0.595	0.925	0.716	0.949
GSK3beta	0.929	0.977	0.736	0.957	0.463	0.863	0.699	0.868

Table 4 Analysis of the Important Fragments for multi-target against AD Given by Naive Bayesian Classifier.

ID	Structure	Targets and Bayesian score	ID	Structure	Targets and Bayesian score
1	 1334840514	AChE (1.355); BuChE (1.196); BCAE1 (1.294); gamma-seretase (1.226)	13	 485677615	AChE (1.359); Sigma-1 receptor (1.335)
2	 -1830436798	AChE (1.359); BuChE (1.193);	14	 -979993137	APP (1.234); gamma-seretase (1.259)
3	 -1441803809	AChE (1.36); BuChE (1.193)	15	 945431654	AMPA2 (1.254); gamma-seretase (1.232)
4	 -1108879003	AChE (1.359); BuChE (1.193)	16	 -1794005192	GABA-A (1.008); H_3R (1.279); Muscarnic M1 (1.244)
5	 -1056448394	AChE 1.361 BuChE 1.194	17	 -244159614	H_3R (1.28); Muscarnic M1 (1.264)

6	 -128679049	AChE (1.361); BuChE (1.195)	18	 368983122	5HT3A (1.143); GABA-A (1.008); H3R (1.277); Muscarnic M1 (1.235);
7	 440744344	AChE (1.358); BuChE (1.194)	19	 681865297	5HT3A (1.127); GABA-A (0.912); H3R (1.276); Sigma-1 receptor (1.32)
8	 552399862	AChE (1.357); BuChE (1.191)	20	 1063370408	AMPA1 (1.175); PDE4A (1.182); PDE4B (1.219)
9	 805978089	AChE (1.357); BuChE (1.192)	21	 1133499173	GABA-A (0.96); H_3R (1.279); Muscarnic M1 (1.257)
10	 1334101934	AChE (1.357); BuChE (1.192)	22	 1960253356	5HT3A (1.142); Muscarnic M1 (1.244); Nicotinic alpha7 (1.322)

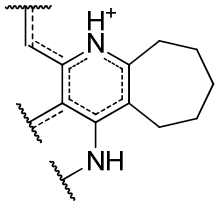
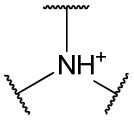
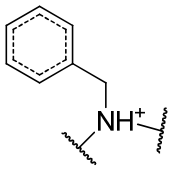
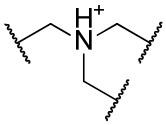
11	 1973629459	AChE (1.358); BuChE (1.193)	23	 1976330679	H_3R (1.28); Muscarnic M1 (1.265)
12	 -1811250099	AChE (1.356) ; Sigma-1 receptor (1.34)	24	 2069906330	5HT3A (1.127); Muscarnic M1 (1.247); Nicotinic alpha7 (1.315)

Table 5. Verification of Predicted Targets by Experiments for Six Approved AD Drugs

Drug	Target	Experimental Potency	Ref
Galantamine	AChE	IC_{50} = 300 nM	[38]
Galantamine	BuChE	IC_{50} = 8.4 μM	[39]
Galantamine	APP	Inhibition self-induced Aβ aggregation of 48 % at 100 μM	[40]
Memantine	NMDA	IC_{50} = 940 nM	[41]
Tacrine	AChE	IC_{50} = 350 nM	[42]
Tacrine	BuChE	IC_{50} = 40 nM	[42]
Tacrine	Muscarnic M1	IC_{50} = 2 μM	[43]
Rivastigmine	AChE	IC_{50} = 0.92 μM	[44]
Rivastigmine	BuChE	IC_{50} = 0.3 μM	[45]
Donepezil	AChE	IC_{50} = 10 nM	[46]
Donepezil	BuChE	IC_{50} = 0.93 μM	[47]
Donepezil	APP	IC_{50} = 86.5 μM	[48]
Donepezil	H_3R	IC_{50} = 0.35 μM	[49]
Donepezil	MAO-B	IC_{50} = 15 μM	[50]
Donepezil	BACE1	IC_{50} = 3.2 μM	[40]
Huperizine A	AchE	IC_{50} = 10 nM	[51]
Huperizine A	BuChE	IC_{50} = 10 μM	[52]

Table 6. The predicted and experimental results for 19 known active compounds towards AChE, BuChE, cdk5, and H_3R using the generated classifiers.

Name	BuChE/Pred[a]	BuChE/Exp[b]	CDK5/Pred	CDK5/Exp	AChE/Pred	AChE/Exp	H_3R/Pred	H_3R/Exp
DL0410	True	Active	False	NA[d]	True	Active	True	Active
J14683	False	-[c]	False	Active	True	NA	False	NA
J18457	True	Active	False	NA	True	Active	True	Active
J18458	True	Active	False	NA	True	Active	True	NA
J18461	True	Active	False	NA	True	Active	True	Active
J18803	False	-	True	Active	False	NA	True	NA
J18811	False	-	True	Active	False	NA	True	NA
J18836	False	-	True	Active	False	NA	True	NA
J18842	False	-	True	Active	False	NA	True	NA
J18848	False	-	True	Active	False	NA	True	NA
J18854	False	-	True	Active	False	NA	True	NA
J18879	False	-	False	Active	False	NA	True	NA
J18911	False	-	True	Active	True	NA	True	Active
J27139	True	Active	False	NA	True	Active	True	NA
J37156	True	Active	False	NA	True	NA	True	NA
J39065	True	Active	False	NA	True	Active	False	NA
J39067	True	Active	False	NA	True	NA	False	NA
J39068	True	Active	False	NA	True	Active	False	NA
J39069	True	Active	False	NA	True		False	NA

"Target/Pred[a]" represents the predicted result for a compound against one target; "Target/Exp[b]" represents the experimental result for a compound against one target; "-[c]" means that the predicted result has not been confirmed by BuChE assay yet; "NA[d]" represents that the experimental result has proved the compound to be inactive against one target

Table 7 IC_{50} Values (μM) of Active Compounds found in this Study

No.	MW	AChE IC_{50} (μM)±SD	BuChE IC_{50} (μM)±SD	CDK5 IC_{50} (μM)±SD	H_3R IC_{50} (μM)±SD
Donepezil [a]	415.95	0.251 ±0.004	-	-	-
Thioperamide [a]	408.52	-	-	-	0.305±0.005
DL0410	505.52	0.424 ± 0.008	3.57 ± 0.58	NA[b]	0.308±0.003
J27139	380.47	0.57 ± 0.05	10.07 ± 1.26	NA	NA
J39068	575.99	13.7 ± 0.438	8.69 ± 0.56	NA	NA
J18458	509.47	14.00 ± 0.506	9.34 ± 0.84	NA	NA
J18457	481.50	15.76 ±0.130	2.06 ± 0.44	NA	58.6±2.1
J18461	425.39	16.585 ±0.03	0.32 ± 0.01	NA	2.247±0.005
J39065	487.93	72.42 ±2.678	3.78 ± 0.29	NA	NA
J18911	363.5	NA	-	9.23±2.33	29.23±0.22

[a]Donepezil and thioperamide served as the reference compounds for the bioassay. NA[b] represents that the experimental result has proved the compound to be inactive against one target.

"For Table of Contents Use Only"

Discovery of Multi-target-directed Ligands against Alzheimer's Disease through Systematic Prediction of Chemical-protein Interactions Jiansong Fang, Yongjie Li, Rui Liu, Xiaocong Pang, Chao Li, Ranyao Yang, Yangyang He, Wenwen Lian, Ai-lin Liu*, and Guan-hua Du*