Predictions of BuChE Inhibitors Using Support Vector Machine (SVM) and Naive Bayesian Classification Techniques in Drug Discovery.

Article *in* Journal of Chemical Information and Modeling · October 2013

DOI: 10.1021/ci400331p · Source: PubMed

CITATIONS
41

READS
170

7 authors, including:

Jiansong Fang
Guangzhou University of Chinese Medicine
68 PUBLICATIONS **255** CITATIONS

SEE PROFILE

Shengqian Yang
Chinese Academy of Medical Sciences
11 PUBLICATIONS **77** CITATIONS

SEE PROFILE

Ailin Liu
Chinese Academy of Medical Sciences
126 PUBLICATIONS **963** CITATIONS

SEE PROFILE

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Predictions of BuChE Inhibitors Using Support Vector Machine and Naive Bayesian Classification Techniques in Drug Discovery

Jiansong Fang,[†] Ranyao Yang,[†] Li Gao,[†] Dan Zhou,[†] Shengqian Yang,[†] Ai-lin Liu,*[,†,‡,§] and Guan-hua Du*[,†,‡,§]

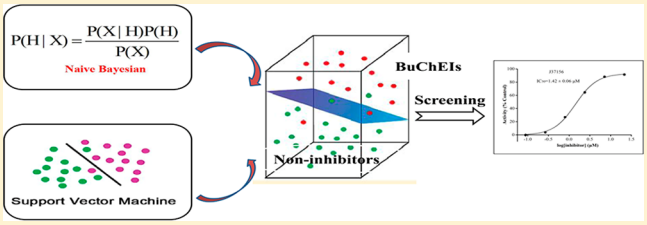
[†]Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, People's Republic of China

[‡]Beijing Key Laboratory of Drug Target Research and Drug Screening, Beijing 100050, People's Republic of China

[§]State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Beijing 100050, People's Republic of China

Supporting Information

ABSTRACT: Butyrylcholinesterase (BuChE, EC 3.1.1.8) is an important pharmacological target for Alzheimer's disease (AD) treatment. However, the currently available BuChE inhibitor screening assays are expensive, labor-intensive, and compound-dependent. It is necessary to develop robust in silico methods to predict the activities of BuChE inhibitors for the lead identification. In this investigation, support vector machine (SVM) models and naive Bayesian models were built to discriminate BuChE inhibitors (BuChEIs) from the noninhibitors. Each molecule was initially represented in 1870 structural descriptors (1235 from ADRIANA.Code, 334 from MOE, and 301 from Discovery studio). Correlation analysis and stepwise variable selection method were applied to figure out activity-related descriptors for prediction models. Additionally, structural fingerprint descriptors were added to improve the predictive ability of models, which were measured by cross-validation, a test set validation with 1001 compounds and an external test set validation with 317 diverse chemicals. The best two models gave Matthews correlation coefficient of 0.9551 and 0.9550 for the test set and 0.9132 and 0.9221 for the external test set. To demonstrate the practical applicability of the models in virtual screening, we screened an in-house data set with 3601 compounds, and 30 compounds were selected for further bioactivity assay. The assay results showed that 10 out of 30 compounds exerted significant BuChE inhibitory activities with IC_{50} values ranging from 0.32 to 22.22 μM, at which three new scaffolds as BuChE inhibitors were identified for the first time. To our best knowledge, this is the first report on BuChE inhibitors using machine learning approaches. The models generated from SVM and naive Bayesian approaches successfully predicted BuChE inhibitors. The study proved the feasibility of a new method for predicting bioactivities of ligands and discovering novel lead compounds.



1. INTRODUCTION

Alzheimer's disease (AD), a progressively degenerative disorder of the brain, is a multifactorial disease caused by dysfunctions of multiple protein targets, contributing to its etiology.[1,2] AD is characterized by widespread loss of central cholinergic function. One proven effective way to treat AD to date is the use of cholinesterase (ChE) inhibitors to augment surviving cholinergic activity.[3] In the central nervous system (CNS), the cholinergic system is important in the regulation of memory and learning processes. There are two cholinesterases that contribute to the degradation of the neurotransmitter acetylcholine (Ach): acetylcholinesterase (E1.1.1.7; AChE) and butyrylcholinesterase (E1.1.1.8; BuChE). In healthy brain, acetylcholinesterase predominates (80%) and butyrylcholinesterase plays a minor role in regulating brain ACh levels. However, in advanced AD pathological condition, AChE activity can be reduced to 55−67% of normal levels in specific brain regions, while BuChE activity increases suddenly.[4,5] In an AD brain, selective BuChE inhibitors (BuChEIs) have demonstrated to have a beneficial effect in vivo, probably by restoring cholinergic activity and/or AChE/BuChE activity ratios to the levels observed in the healthy brain.[6−9] These findings suggest that inhibiting BuChE may provide a new therapeutic strategy for effective augmentation of cholinergic function in affected individuals.

The discovery of BuChEIs by carrying out the binding bioassays on every compound is costly, time-consuming, compound-dependent, and labor-intensive. It is therefore desirable to develop computational methods[10−14] to facilitate the screening of BuChEIs in the early phase of drug discovery. To date, several studies addressing this problem using quantitative structure−activity relationship (QSAR)[15−20] and pharmacophore[21,22] models have been reported. QSAR models define mathematical relationships between a set of descriptors and biological end points or other properties of known ligands

Received: June 6, 2013

   dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

to predict the activities of unknown ligands. Castilho[15] and De Souza[16] made hologram quantitative structure−activity relationship (HQSAR) models using fragment fingerprints as predictive variables of the biological activity variation based on tacrine derivatives and 4-[(diethylamino)methyl]phenol as BuChE inhibitors. Zaheer-ul[17] built three-dimensional quantitative structure−activity relationship (3D-QSAR) models using the comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) for a series of structurally related steroidal alkaloids as BuChE inhibitors. Both CoMFA and CoMSIA yielded significant cross-validated q^2 values of 0.701 and 0.627 and r^2 values of 0.979 and 0.982, respectively. Sakkiah[21] and Abbasi[22] designed the ligand-based pharmacophore models to identify the critical chemical features of BuChE inhibitors, which can be reliable tools to identify structurally diverse compounds with desired biological activity. However, these computational methods still require manual intervention and superimpositions, which limits in silico design and virtual screening of large chemical databases.

Machine learning (ML) approaches are efficacious in the computer-aided drug design. Many statistical learning algorithms have been used to build the models, such as support vector machines (SVMs),[23,24] naive Bayesian,[25] random forest,[26,27] k nearest neighbors (kNNs),[28] and artificial neural networks (ANNs).[29] They have been successfully applied in virtual screening,[25,30] ADEMT prediction,[31,32] and pharmacokinetic studies.[33] To date, several classification models have been built to separate cholinesterase (ChE) inhibitors from the noninhibitors. Chekmarev[34] built SVM classification and regression models for predicting AChE inhibitors with molecular descriptors derived from shape signatures and the molecular operating environment (MOE) application software and proved that the support vector classification algorithm is superior to the classifier based on scores from the molecular docking program GOLD, with the overall prediction accuracies of $Q_{SVC(10CV)}$ = 74% and $Q_{SVC(LNO)}$ = 67% versus Q_{GOLD} = 56%. Besides, Lv and co-workers[35] explored several machine learning methods (support vector machine, k nearest neighbor, and C4.5 decision tree (C4.5DT)) for predicting AChE inhibitors (AChEIs). The SVM prediction accuracy of the best model is up to 88.0% for AChEIs and 79.6% for non-AChEIs. Yan and colleagues[36] built SVM models using a larger number of compounds and data than previous work, and the best model gave a Matthews correlation coefficient of 0.99 and a prediction accuracy (Q) of 99.66% for the test set. However, up to now, there is limited research on classification predictions of the BuChE inhibitors and noninhibitors.

In this study, the workflow of virtually screening BuChE inhibitors is shown in Figure 1. First, a larger number of compounds have been utilized to construct the classification models of BuChE inhibitors. The molecules were represented initially by descriptors derived from ADRIANA.Code (version 2.2.6),[37] MOE (version 2010.10),[38] and Discovery Studio (version 2.5).[39] Correlation analysis and stepwise regression methods were employed for descriptor selections. Classification models were built using support vector machine (SVM), and naive Bayesian learning, to discriminate BuChE inhibitors from nonbinding compounds. Performances were measured by cross-validation, a test set validation, and an external test set validation. In order to guard against the possibility of chance correlation, Y-scrambling was also performed. The two best classification models (NB-d and NB-f) as ligand-based virtual screening tools were used to predict BuChE inhibitors from our in-house database. Finally, selected compounds were validated by bioassay.

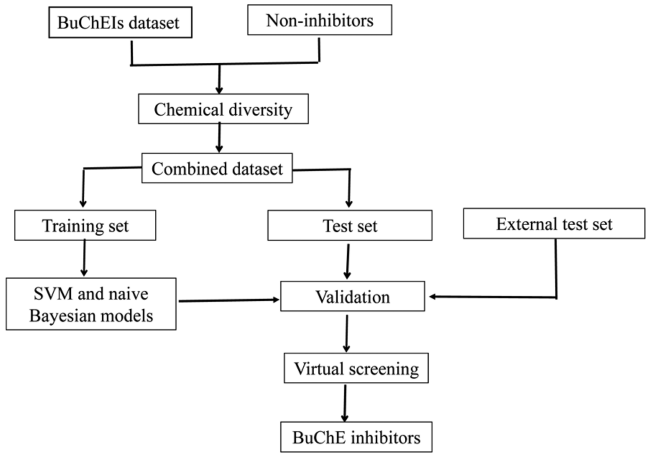


Figure 1. Workflow for classification models building, validation, and virtual screening (VS) as applied to BuChE inhibitors (BuChEIs) and noninhibitor data sets.

2. MATERIALS AND METHODS

2.1. Data Preparation. A total of 1033 BuChE inhibitors (after removing the duplicate molecules), with IC_{50} < 10 μM, were collected from the BindingDB database[40] as positive data. These inhibitors contained a variety of scaffolds and substituents emphasizing their structural diversity. In addition, a data set of 3168 inactive compounds, randomly extracted from the high-throughput screening (HTS) negative results toward BuChE inhibitors in our lab previously, was considered as negative data. Both the inhibitor and noninhibitor data sets were randomly divided into two groups. Finally, the training set was made up of 800 BuChEIs and 2400 inactive compounds, and the test set contained 233 BuChEIs and 768 inactive compounds (detailed information is available in the Supporting Information, see Tables S1 and S2).

All the inorganic counterions were removed prior to calculating descriptors, and the remaining parts were processed through including adding hydrogen atoms, deprotonating strong acids, protonating strong bases, generating stereoisomers, and valid single 3D conformers by means of washing and energy minimizing in MOE. All inhibitors were considered active (class label: "1") in the calculations, while inactive compounds against BuChE were labeled as "−1".

2.2. Molecular Descriptors. A total of 1235 descriptors were calculated using ADRIANA.Code,[37] including 19 global molecular descriptors, 8 size and shape descriptors, 88 2D property autocorrelation descriptors, 96 3D property autocorrelation descriptors, and 1024 3D property-weighted radial distribution functions (property-weighted RDF) descriptors.

Global molecular descriptors represent each chemical structure by a structural, chemical, or physicochemical feature or property of the molecule expressed by a single value. Size and shape descriptors characterize the size and the 3D shape of a molecule, deriving from the 3D structure of a molecule.

The property autocorrelation uses the molecular 2D/3D structure and atom pair properties as a basis to obtain vectorial molecular descriptors.[41] The molecular autocorrelation vectors were calculated based on the following eight atomic properties: identity (Ident), σ charge (SigChg), π charge (PiChg), total charges (TotChg), σ electronegativity (SigEN), π electro-

dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

negativity (PiEN), lone-pair electronegativity (LpEN), and atomic polarizability (Apolariz). For each property, the 2D autocorrelation values for 11 distances (0−10 bonds) and the 3D autocorrelation values for 12 distances (0−11 bonds) were calculated. Thus, for each molecule, 88 2D property autocorrelation descriptors and 96 3D property autocorrelation descriptors can be obtained.

Property RDF uses the 3D structure of a molecule and atom pair properties as the basis to derive vectorial molecular descriptors.[42] Available atom pair properties for property-weighted RDF include RDF_Ident, RDF_SigChg, RDF_PiChg, RDF_TotChg, RDF_SigEN, RDF_PiEN, RDF_LpEN, and RDF_Polariz.

The MOE[33] software can calculate 186 2D descriptors and 148 3D molecular descriptors. Partial charges for each structure were assigned by MOE with the default parameters, and then the energy minimization was carried out in MOE.

Discovery Studio[39] was another software used to calculate the 2D descriptors, which were made up of AlogP, estate keys, molecular properties, molecular property counts, surface area and volume, and topological descriptors. Molecular fingerprints were also calculated with this software, including the SciTegic extended-connectivity fingerprints (FCFP and ECFP) and Daylight-style path-based fingerprints (FPFP and EPFP). Considering that the structural fragments should be neither too small nor too large, two diameters, 4 and 6, were selected for each fingerprint.

2.3. Molecular Descriptor Selection. Pearson correlation analysis[43] can eliminate molecular descriptors that are weakly correlated with activity and highly correlated with each other. In this paper, descriptors whose Pearson correlation coefficients with activity were less than 0.1 were deleted. If the correlation coefficient between two descriptors was higher than 0.9, the descriptor which had a lower correlation coefficient with the activity would be deleted.

In addition, a stepwise variable selection method was performed for the remaining descriptors. First, all the descriptors chosen by correlation analysis were sorted in a descending order according to their correlation coefficient with activity. Linear regression analysis of activity and the first descriptor was performed to get an initial equation. Other descriptors were then added to the regression equation one by one. Then a significance test was done for every new regression equation and each descriptor in the equation. Once the new regression equation was not "statistically significant" after the addition of a new descriptor, the new descriptor would be removed. Descriptors that are not "statistically significant" in the equation were also deleted, followed by the construction of a new regression equation. The process was continued until no descriptor could be added or removed by this method.

According to the above criteria, the descriptors finally chosen from different programs are listed in Table1.

2.4. Methods for Model Building. *2.4.1. Support Vector Machine.* Support vector machine[44,45] was employed to generate classification models as well as naive Bayesian classification technique. SVM was first developed by Vapnik as a general data modeling methodology for pattern recognition, aiming at minimizing structure risk under the frame of VC theory.[46] Its main idea is to map data into a high-dimensional space in which a constrained quadratic programming problem will be solved and a separating hyperplane with the maximal margin will be found. SVM consistently shows

Table 1. Molecular Descriptors Used in This Work

descriptor class	number of descriptors	descriptors
ADRIANA.Code descriptors	25	Dipole, RComplexity, Rgyr, 2DACorr_TotChg_1, 2DACorr_PiEN_7, 2DACorr_LpEN_5, 3DACorr_TotChg_8, RDF_Ident_24, RDF_Ident_28, RDF_Ident_33, RDF_Ident_97, RDF_SigChg_16, RDF_SigChg_40, RDF_SigChg_61, RDF_SigChg_79, RDF_PiChg_14, RDF_PiChg_42, RDF_PiChg_50 RDF_TotChg_108, RDF_PiEN_13, RDF_PiEN_31, RDF_PiEN_49, RDF_PiEN_61, RDF_Polariz_28, RDF_Polariz_33
MOE descriptors	21	b_ar, DCASA, E_ele, E_stb, FASA_H, GCUT_SLOGP_1, logS, opr_nrot, PEOE_RPC+, PEOE_VSA+0, PEOE_VSA-1, pmiX, radius, SlogP_VSA1, SlogP_VSA5, SlogP_VSA7, SMR_VSA1, SMR_VSA5, VAdjMa, vsa_pol, vsurf_HL1
Discovery studio two-dimensional descriptors (DS 2D)	18	Estate_AtomTypes, Estate_AtomValues, FormalCharge, Gasteiger_Charges, LogD, VSA_MR, VSA_PartialCharge, Num_PositiveAtoms, Num_AromaticBonds, Num_Rings, Num_ChainAssemblies, Num_H_Donors_Lipinski, Molecular_PolarSASA, Molecular_FractionalPolarSASA, CHI_V_3_C, CIC, IC, JY

dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

outstanding classification performance, is less penalized by sample redundancy, and has lower risk for overfitting.

SVM algorithm was performed in the LIBSVM 2.9 package developed by Chang and Lin.[47] In our study, a radial basis function (RBF) kernel was used to build models. The two parameters of the SVM (C, γ) for each model were selected using the autosearching program "grid" through a 5-fold cross-validation in LibSVM.

2.4.2. Naive Bayesian Classifiers. The naive Bayesian classification models were developed using Discovery Studio 2.5. The Bayesian approach is a robust classification approach that can discriminate active compounds from inactive 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 details about the method can be found in the references.[48,49] In short, the technique is based on the frequency of occurrence of various descriptors that are found in two or more sets of molecules that well discriminate between these sets. The model learning process generates a large set of Boolean features from the input descriptors and collects the frequency of occurrence of each feature in the active subset and in all data samples. To apply the model to a particular sample, the features of the sample are generated, and a weight is calculated for each feature using a Laplacian-adjusted probability estimate. Finally, the weights are summed to provide a probability estimate, which is a relative predictor of the likelihood of that sample in the active subset.

2.5. Validating the Prediction Accuracies of the SVM and Bayesian Models. The quality of the SVM and Bayesian classifiers 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, which are given by eqs 1−4. TP represents the number of BuChEIs that are predicted as BuChEIs. TN represents the number of inactive compounds that are predicted as inactive compounds. FP stands for the number of inactive compounds that are predicted as BuChEIs, and FN is the number of BuChEIs that are predicted as inactive compounds. SE represents the prediction accuracy for BuChEIs and SP represents the prediction accuracy for noninhibitors.

$$SE = \frac{TP}{TP + FN} \quad (1)$$

$$SP = \frac{TN}{TN + FP} \quad (2)$$

$$Q = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \quad (4)$$

2.6. In Vitro BuChE Inhibitory Assay. The capacity of compounds inhibiting BuChE was assessed using the Ellman's method.[50] 5,5′-Dithiobis(2-nitrobenzoic acid) (Ellman's reagent, DTNB), the substrate (S)-butyrylthiocholine chloride (BuSCh), and ISO-OMPA were purchased from Sigma Aldrich. BuChE was derived from human plasma (purchased from Beijing Red Cross Blood Center), diluted 200 times in 0.05 M phosphate buffer solution. Thirty compounds were dissolved in DMSO and then diluted 100 times in the same buffer for their activity assay.

In the BuChE reaction system, all the assays were carried out in 96-well plates under 0.05 M phosphate buffer solution, using a Spectra Max M5 microplate reader. The assay solution consisted of 10 μL of test compounds, 40 μL of BuChE, 70 μL of 7.5 mM of substrate (BuSCh) from Sigma, and 80 μL of 0.25 mg/mL 5,5′-dithiobis(2-nitrobenzoic acid). Then the reaction was incubated at 37 °C for 60 min. The absorbance value was quantified with the absorbance at a wavelength of 412 nM. The wells without inhibitors were also set for calculating the inhibition of the compounds. The data are expressed as the mean of three independent experiments, and IC_{50} values were determined graphically from log concentration−inhibition curves.

3. RESULTS AND DISCUSSION

3.1. Chemical Space Analysis. The prediction accuracy of machine learning systems is affected by the chemical diversity of samples utilized in the training set and test set. Generally, classification models based on a small group of samples, which only covered a small region of chemical space, limited their applications. Thus, chemical space distribution (defined by MW and ALogP) of the training set, test set, and external test set was performed. As shown in Figure 2, the training set, test set, and external test set distributed in a wide range of chemical space with MW ranging from 100 to 1000, and AlogP ranging from −10 to 15, which indicated that these data sets possess large enough chemical space.

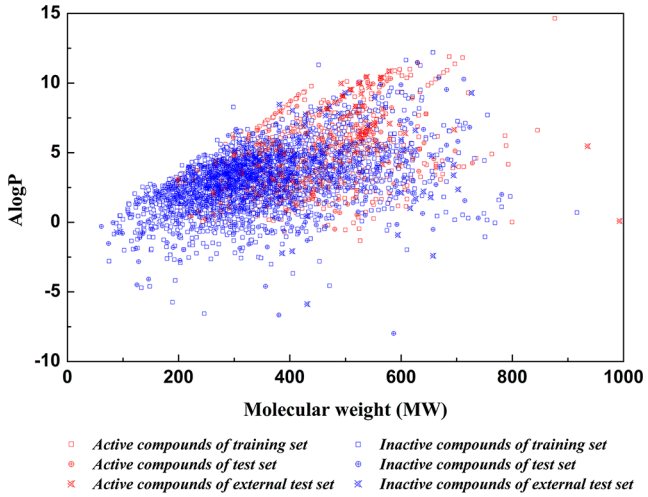


Figure 2. Diversity distribution of the training set (n = 3200 compounds), test set (n = 1001 compounds), and external test set (n = 317 compounds). Chemical space was defined by molecular weight (MW) as X-axis, and ALogP as Y-axis. In the picture, red stands for active compounds, and blue stands for the inactive compounds.

3.2. Model Performance. All the classification models in this study were initially generated using SVM and naive Bayesian classifiers with the descriptors calculated in ADRIANA.Code, MOE, and Discovery Studio (DS). Subsequently, 5-fold, 10-fold (SVM), and leave one out (naive Bayesian) cross-validations were performed. In addition, the models were used to predict a test set comprising 1001 compounds (including 233 inhibitors and 768 inactive

dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. Performance of the Bayesian Classifiers for the Test Sets[a]

descriptors	TP	FN	FP	TN	SE	SP	Q	MCC
DS_2D (MP)	176	57	82	686	0.7554	0.8932	0.8611	0.6267
MP+ECFP_4	217	16	6	762	0.9313	0.9922	0.9780	0.9379
MP+ECFP_6	223	10	6	762	0.9571	0.9922	0.9840	0.9550
MP+EPFP_4	210	23	40	728	0.9013	0.9479	0.9371	0.8291
MP+EPFP_6	212	21	21	747	0.9099	0.9727	0.9580	0.8825
MP+FCFP_4	217	16	27	741	0.9313	0.9648	0.9570	0.8821
MP+FCFP_6	218	15	16	752	0.9356	0.9792	0.9690	0.9134
MP+FPFP_4	208	25	36	732	0.8927	0.9531	0.9391	0.8325
MP+FPFP_6	209	24	16	752	0.8970	0.9792	0.9600	0.8870

[a]TP, true positive; FN, false negative; FP, false positive; TN, true negatives; SE, sensitivity; SP, specificity; *Q*, the overall prediction accuracy; MCC, Matthews correlation coefficient; DS_2D, Discovery studio two-dimensional.

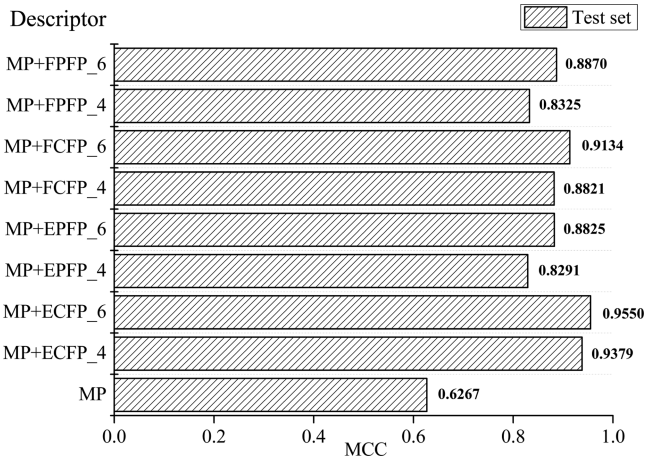


Figure 3. Matthews correlation coefficient of 9 naive Bayesian models based on 18 molecular properties (MP) and 8 sets of fingerprints for the test set.

compounds). At last, the external test set, which was not included in the training and test set, was applied to evaluate the models.

The descriptors calculated in ADRIANA.Code and MOE can depict the whole molecule properties, but they cannot characterize the important substructures or molecular fragments which are important to BuChEIs. Hence, molecular fingerprints, together with property descriptors, were used simultaneously as the descriptors in Bayesian analysis to build new prediction models. The statistical results for these Bayesian classifiers are listed in Table 2 and Figure 3.

As shown in Table 2, the addition of fingerprint can improve the performance of naive Bayesian, because the values of overall accuracy (*Q*) and Matthews correlation coefficient for the test set after considering fingerprints are higher than that (*Q* = 0.8611, MCC = 0.6267) of the model only based on 18 DS 2D descriptors.

Additionally, the models derived from the fingerprints of diameter number 6 (encoded fingerprint names) perform better than those derived from diameter number 4 (Figure 3). For example, the MCC of model A (MP+ECFP_4) is 0.9379, while model B (MP+ECFP_6) is 0.9550. The same explanation applies to the MCC differences (0.8291 and 0.8825) between model C (MP+EPFP_4) and model D (MP+EPFP_6). Among the fingerprints of diameter number 6, ECFP_6 performs better than the others, with the highest MCC of 0.955, the sensitivity of 0.9571, the specificity of 0.9922, and the overall accuracy of 0.9840 (Table 2).

All models constructed by the descriptors from ADRIANA.Code, MOE, and DS with or without ECPF_6 are shown in Table 3. As the results show in Table 3, three models (SVM-a, SVM-b, SVM-c) were generated using SVM, with 25 ADRIANA.Code descriptors, 21 MOE descriptors, and 18 DS 2D descriptors, respectively. In the absence of ECFP_6, with the same descriptors, the performance of SVM models is better than that of Bayesian models. For example, with ADRIANA.Code descriptors, the MCC for test set of SVM-a is 0.8977, which is higher than that (0.7865) of naive Bayesian-a (NB-a). As for MOE descriptors and DS 2D descriptors, the same explanations apply to the MCC differences (0.9468 and 0.6328) between SVM-b model and NB-c model and the MCC

Table 3. Performance of SVM and Bayesian Models and Their Cross-Validation Results

				cross-validation for the training set (%)				
descriptors	model	number of descriptors	accuracy for the training set (%)	5-F	10-F	LOO	accuracy for test set (%)	MCC for test set
ADRIANA.Code	SVM-a	25	99.09	97.38	97.38		96.40	0.8977
MOE	SVM-b	21	99.97	97.97	97.88		98.12	0.9468
DS 2D descriptors	SVM-c	18	99.90	98.53	98.88		90.11	0.7130
ADRIANA.Code	NB-a	25	91.28			97.08	91.90	0.7865
ADRIANA.Code and ECFP_6	NB-b	26	98.72			99.89	98.40	0.9551
MOE	NB-c	21	83.44			90.12	86.41	0.6328
MOE and ECFP_6	NB-d	22	98.84			99.88	98.70	0.9635
DS 2D descriptors	NB-e	18	84.75			90.7	86.11	0.6267
DS 2D descriptors and ECFP_6	NB-f	19	98.88			99.88	98.40	0.9550

dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 4. Performance of SVM and Bayesian Models on the External Test Set

descriptors	model	number of descriptors	sensitivity	specificity	accuracy (Q)	MCC
ADRIANA.Code	SVM-a	25	0.7922	0.9167	0.8864	0.6970
MOE	SVM-b	21	0.7403	0.9833	0.9243	0.7872
DS 2D descriptors	SVM-c	18	0.0343	1.0000	0.5243	0.1331
ADRIANA.Code	NB-a	25	0.8966	0.9258	0.9190	0.7865
ADRIANA.Code and ECFP_6	NB-b	26	0.9091	0.9875	0.9685	0.9132
MOE	NB-c	21	0.7403	0.8292	0.8076	0.5284
MOE and ECFP_6	NB-d	22	0.8571	0.9833	0.9527	0.8689
DS 2D descriptors	NB-e	18	0.7554	0.8932	0.8611	0.6267
DS 2D descriptors and ECFP_6	NB-f	19	0.8961	0.9958	0.9716	0.9221

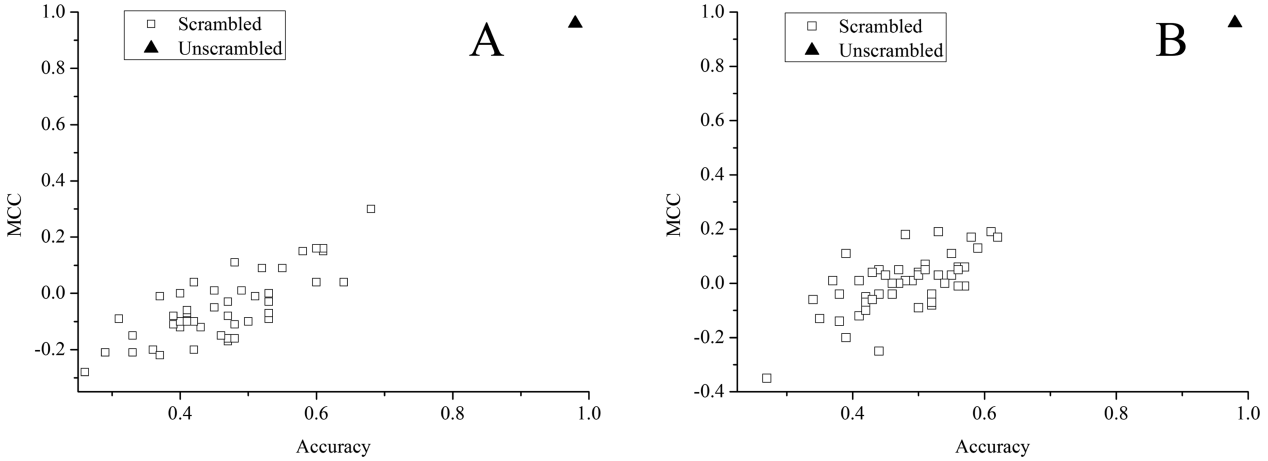


Figure 4. Y-scrambling result of model NB-b (A) and model NB-f (B). The values of NB-b and NB-f are shown as a black triangle, while the values of the scrambled models are shown as white squares. The Matthews correlation coefficient and accuracy (Q) of NB-b and NB-f are significantly higher than any of the Y-scrambled models, which indicates that the models are not results of chance correlation.

differences (0.7130 and 0.6267) between SVM-c model and NB-e model.

Comparing the descriptors calculated by ADRIANA.Code and MOE, both of them have their respective advantage. For the SVM models, SVM-b (with MOE descriptors) performs better than SVM-a (with ADRIANA.Code descriptors). However, for the naive Bayesian models, NB-a (with ADRIANA.Code descriptors) has a higher predictive ability than NB-c (with MOE descriptors). That may mean ADRIANA.Code descriptors are better for building naive Bayesian model, while the descriptors derived from MOE are more suitable to build the SVM model.

Besides, the results in Table 3 show that the three-dimensional (3D) descriptors may generate more accurate results than two-dimensional (2D) descriptors. For example, SVM-a and SVM-b perform better than SVM-c. Similarly, NB-a and NB-c perform better than NB-e. That is because the 3D descriptors can represent not only the plane structure information but also the steric structure information, while the 2D descriptors only contain the plane information.

3.3. External Test Set Validation. A high MCC value of the test set does not necessarily mean the model exhibits a good predictive ability for an external test set. That is because the chemical diversity of the test set may be so similar to that of training set that the models can easily have a good predictive ability for the test set. In order to avoid this problem, a data set containing 77 BuChE inhibitors collected from literature[51−60] with K_i < 500 nM and 240 decoys which were not included in the training and test sets was used as an external test set (shown in Table S3).

In order to make sure that the scaffolds of external test set are diverse from those of the training set, we summarized the scaffolds of 77 active compounds in the external test set (Figure S1). Then six Markush search queries were used to search the 800 active compounds of the training set. Finally, only 18 compounds of 800 can be identified by the scaffold 1. Thus, we conclude that the scaffolds of the external test set are different from those of the training set.

Prediction performances of the nine models are shown in Table 4. NB-b (ADRIANA.Code and ECFP_6) and NB-f (DS 2D descriptors and ECFP_6), which are regarded as the best models in this paper, have the highest MCC (0.9132 and 0.9221, respectively) and prediction accuracy (0.9685 and 0.9716, respectively) in external test set validation. In addition, in the absence of fingerprint, none of the models has a better predictive power than those with ECFP_6 in external test set validation. Therefore, we conclude that bringing fingerprint in the model construction is important in enlarging the predictive range.

3.4. Y-Scrambling. As was mentioned above, NB-b and NB-f were considered to be the best models. Y-scrambling was investigated to prove it was not a result of chance correlation for the best models. The resulting Matthews correlation coefficient and prediction accuracy (Q) for the test set are visualized in Figure 4, from which one can see that all the scrambled models have a MCC less than 0.3 and Q less than 0.7, and the values of MCC and Q of NB-b and NB-f models are significantly greater.

3.5. Analysis of the Important Fragments Given by Naive Bayesian Classifier. One of the advantages of using a

dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

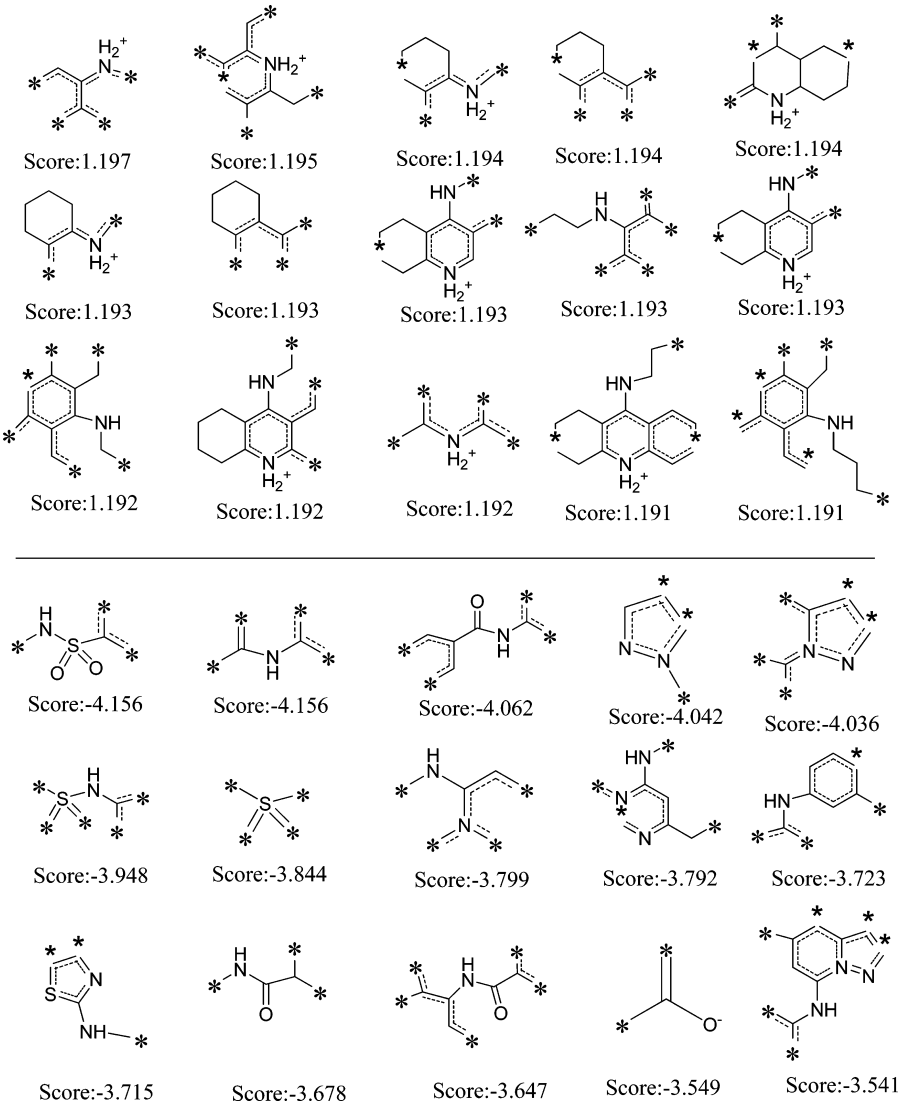


Figure 5. Examples of the top 15 good (top) and bad (bottom) fragments estimated by the NB-b model. The Bayesian score (Score) is given for each fragment.

Table 5. IC_{50} Values (μM) of 10 BuChE Inhibitors

no.	MW	consensus score	IC_{50} (μM/L) ± SD
ISO-OMPA[a]	342.36		2.44 ± 0.17
J39069	491.97	100	22.22 ± 1.71
J39065	487.93	94.96	3.78 ± 0.29
J39067	519.93	93.42	6.09 ± 0.11
J39068	575.99	94.39	8.69 ± 0.56
J27139	380.47	24.55	10.07 ± 1.26
J6487	505.52	23.22	3.57 ± 0.58
J18457	481.50	21.31	2.06 ± 0.44
J18461	425.39	9.79	0.32 ± 0.01
J37156	464.99	6.15	1.42 ± 0.06
J18458	509.47	0	9.34 ± 0.84

[a]ISO-OMPA served as the reference compound for the bioassay.

Bayesian classifier based on structural fingerprints, such as ECFP_6, is that it can identify important fragments or fingerprint features frequently found in two classifying groups. It may be useful for BuChE inhibitor design. The 15 good and 15 bad fragments favorable and unfavorable for BuChE binding were ranked by the Bayesian scores of the NB-b model (Figure 5).

By analyzing the fragments with positive contributions to BuChE binding, it can be observed that most fragments have nitrogen atoms with positive charges. On the one hand, according to the crystal structure of aged human BuChE in complex with MF5 (pdb code: 4BOP),[61] the ring with the positively charged nitrogen atom interacts with Trp82 in the active site by cation−π interactions. On the other hand, the nitrogen atom in the carbon chain in these important fragments may serve as strong hydrogen acceptors and form stable H-bonding interactions with BuChE.

The 15 bad fragments indicate that their existence is unfavorable for BuChE inhibition. Though most of these fragments also include the nitrogen atoms, the nitrogen atoms were not protonated. Therefore, they cannot form the cation−π interactions with Trp82. Besides, most of bad fragments include sulfur atoms, indicating its common occurrence in many noninhibitors.

3.6. Virtual Screening of an In-House Database for BuChE Inhibitors. To illustrate an application of the best models (NB-b and NB-f) in drug discovery research, we

dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

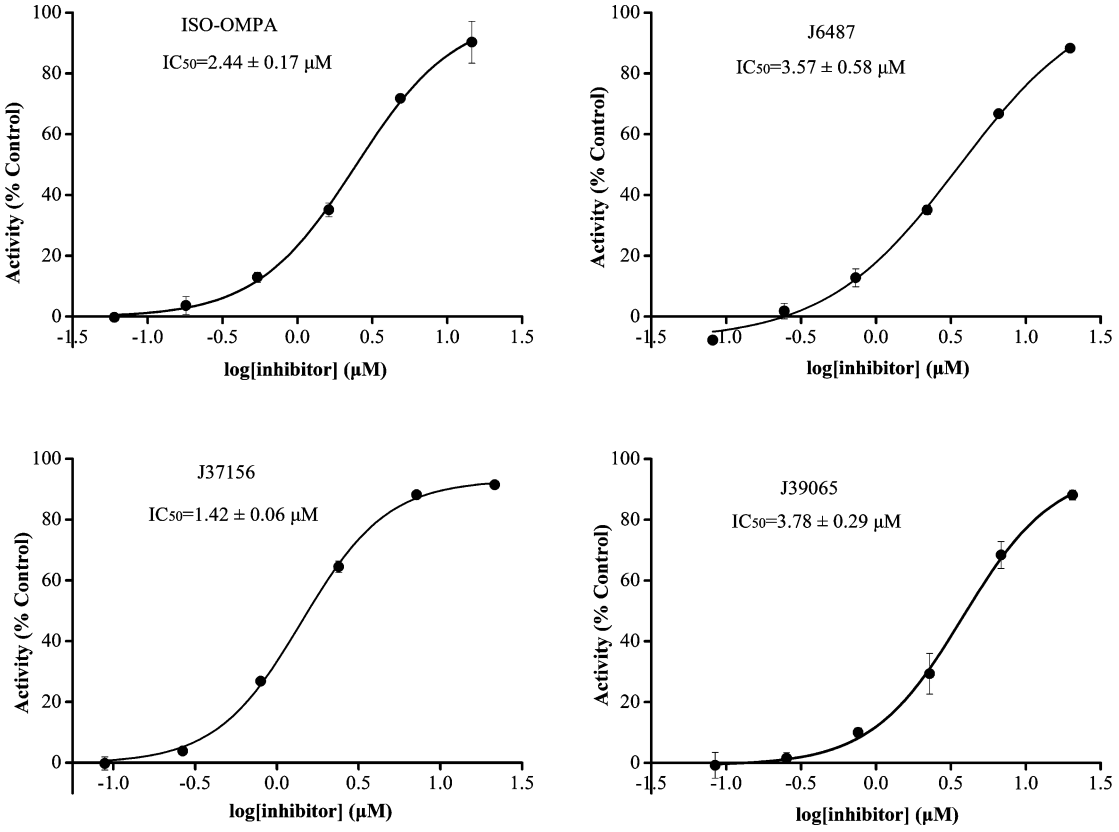


Figure 6. Inhibitory curves of three typical compounds and reference compound ISO-OMPA against BuChE.

performed a virtual screening experiment on our in-house database (including 3601 compounds) to search for BuChE lead ligands. First, each compound was represented by 26 descriptors (25 ADRIANA.Code descriptors and ECFP_6) and 19 descriptors (18 DS 2D descriptors and ECFP_6), respectively, then the two best models (NB-b and NB-f) were used to estimate the probability as BuChE inhibitors for each compound. Using the NB-b model, 163 compounds were predicted as potential BuChE inhibitors, while 110 compounds were predicted as active compounds against BuChE using the NB-f model. Interestingly, it was found that 95 out of them were predicted as BuChE inhibitors with the two models simultaneously. In order to objectively rank the 95 compounds by both the EstPGood score (X) from NB-b model and the EstPGood score (Y) from NB-f model, a consensus score (s) was computed from the normalized X and the normalized Y. The normalized X and Y were computed by the following formulas.

$$x'_i = \left(1 - \frac{x_i - \min(X)}{\max(X) - \min(X)}\right) \times 100 \quad (5)$$

$$y'_i = \left(1 - \frac{y_i - \min(Y)}{\max(Y) - \min(Y)}\right) \times 100 \quad (6)$$

$$s_i = \sqrt{x'_i y'_i} \quad (7)$$

where x' and y' are the normalized EstPGood score from NB-b and the EstPGood score from NB-f, respectively; s_i represents the consensus score. The consensus scores for the 95 compounds are listed in Supporting Information Table S4. The compounds were ranked in the descending order of the consensus score. Finally, 30 compounds (including the top 10 consensus scored compounds and 20 compounds chosen from the other 85 compounds by visual evaluation) were selected for final bioassay validation (shown in Table S4).

3.7. In Vitro BuChE Inhibitory Assay Results. Herein ISO-OMPA served as the reference compound, and its IC_{50} value was 2.44 μM. Out of the 30 candidates tested, 10 compounds were identified to exhibit high or moderate activities against BuChE, with IC_{50} values ranging from 0.32 to 22.22 μM (Table 5). Inhibitory curves of three typical compounds (J6487, J37156, J39065) and reference compound ISO-OMPA toward BuChE are presented in Figure 6. The chemical structures of all the 10 potent BuChE inhibitors are shown in Figure 7. With a success rate of 33% in our virtual screening approach. Among the top 5 consensus scored compounds, 4 out of 5 (80%) were confirmed BuChE inhibitors, which proved that this consensus method improves the hit rate indeed.

Scaffold analyzing results are depicted in Figure 8. Three new BuChE inhibitor scaffolds were discovered from our bioassays. They are *N*-phenylacridin-9-amine (J37156 (IC_{50} = 1.42 μM)), 1,1′-([1,1′-biphenyl]-4,4′-diyl)bis(3-aminopropan-1-one) (J6487 (IC_{50} = 3.57 μM), J18461 (IC_{50} = 0.32 μM), J18458 (IC_{50} = 1.42 μM)), and isoflavone (J27139 (IC_{50} =10.07 μM)), which means these virtual screening models are capable of discovering new BuChE inhibitor scaffolds. The other 4 active compounds, including J39065 that has a IC_{50} value of 3.78 μM, are berberine derivatives, which have been reported as potential dual inhibitors of AChE/BuChE.[62−64] These 4 berberine derivatives, with different substituted groups linked at different positions, exhibit high or moderate activities against BuChE.

dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

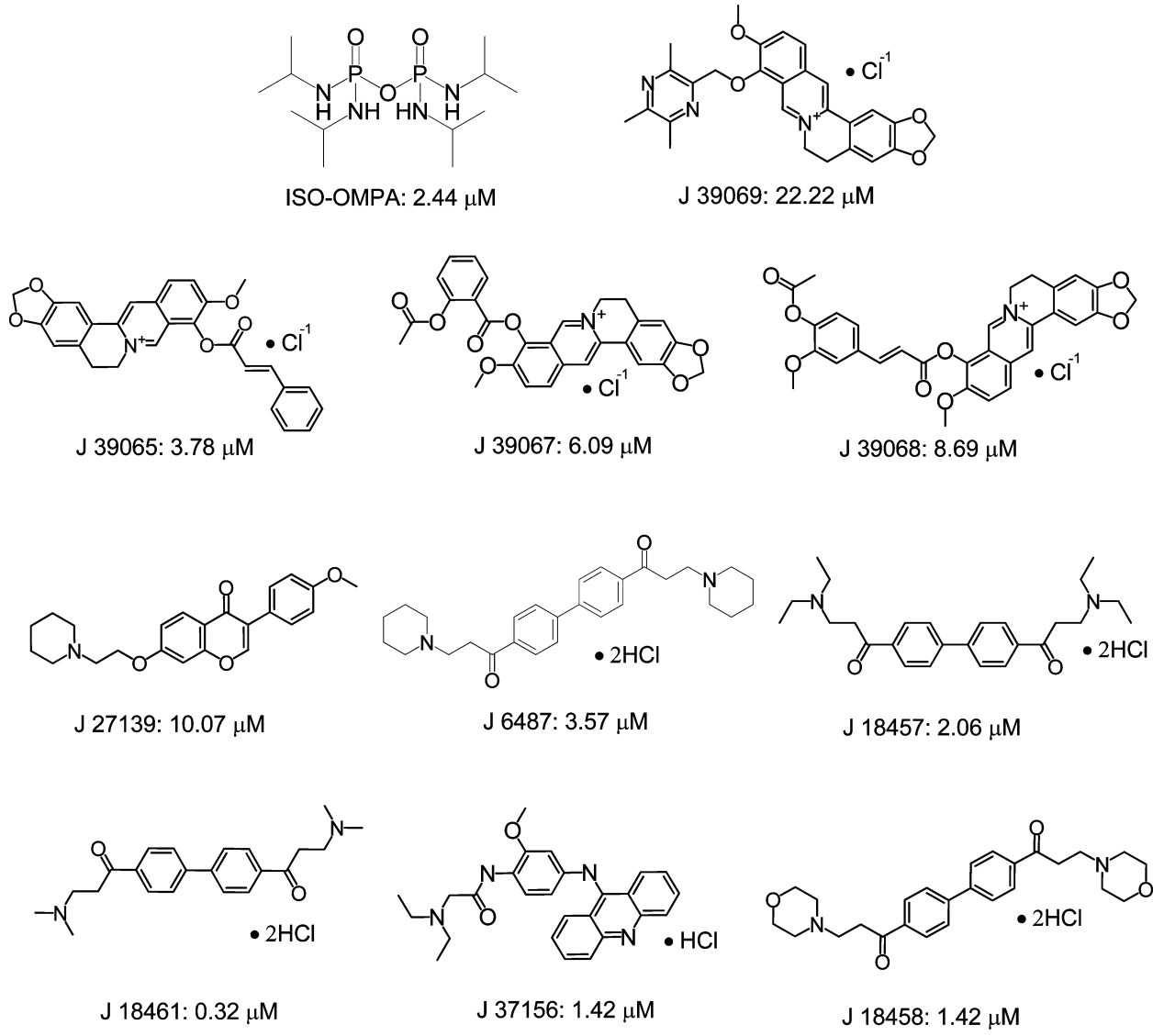


Figure 7. Chemical structures and their IC_{50} values of the 10 potent inhibitors against BuChE.

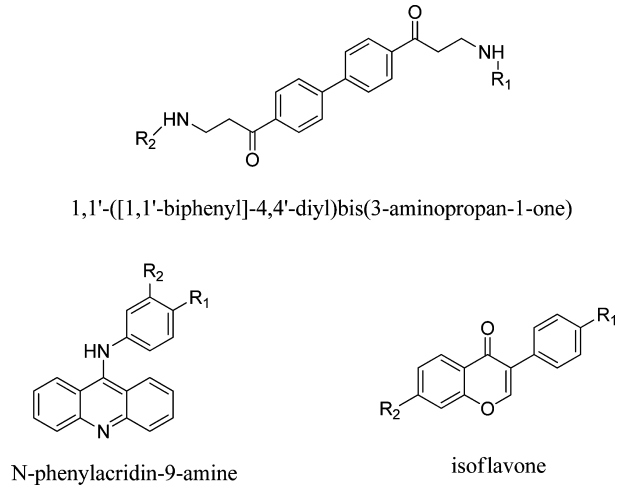


Figure 8. Three new scaffolds of BuChE inhibitors found in this study .

3.8. Comparing with High-Throughput Screening Results. Previously, we developed a high-throughput screening assay for BuChEIs, which screened 76 730 compounds and got 22 hits (the hits rate was 0.029%).[65] In this study, we apply the machine learning approaches (SVM and naive Bayesian) to discover the BuChEIs, and 10 out of 30 (33%) tested compounds exhibit high or moderate activities against BuChE, which means the hit rate is improved for approximately 1150 times at this point. Therefore, the machine learning approaches possess an incomparable advantage to high-throughput screening.

4. CONCLUSION

In this study, several classification models were generated to discriminate BuChE inhibitors (BuChEIs) from noninhibitors by means of SVM and naive Bayesian methods. The molecular descriptors, which were selected by correlation analysis and stepwise variable selection method, played an important role in the construction of the prediction models, while the fingerprint descriptors used in the models could significantly improve their prediction accuracy. The various validations including cross-validation, test set validation, and external test set validation confirmed the prediction reliability of the models; finally, two

dx.doi.org/10.1021/ci400331p | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

best models, NB-b and NB-f, were successfully developed and used in the drug discovery of BuChE inhibitors.

Based on the prediction results, 33.3% of compounds showed BuChE inhibitory activities, which suggested that the prediction models could greatly increase the chance of identifying positive hits and therefore reduce the cost of discovery of a BuChE inhibitor drug. In addition, more importantly, most of the identified active compounds belonging to novel scaffolds support the notable advantage of these models.

In short, this investigation is the first report using machine learning approaches, validated and proven with a successful pilot study of virtual screening, in identifying cholinesterase (ChE) inhibitors. This study suggests a new method of computer-aided drug discovery by screening large structurally diverse databases for predicting biological activities of novel lead candidates of ChE inhibitors.

■ ASSOCIATED CONTENT

Supporting Information

The structures (in SMILE format) of the 3200 compounds of the training set, 1001 compounds of the test set, 317 compounds of the external test set with the corresponding activities (Tables S1–S3), the consensus scores for the 95 compounds (Table S4), six Markush search queries summarized from 77 active compounds of the external test set (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: liuailin@imm.ac.cn. Phone & Fax: 86-10-63165184.
*E-mail: dugh@imm.ac.cn. Phone & Fax: 86-10-63165184.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded in part of the Research Special Fund for Public Welfare Industry of Health (No. 200802041), the National Great Science and Technology Projects (2012ZX09301002-001-001, 2012ZX09301002), and the International Collaboration Project (2011DFR31240). We would like to give special thanks to Jun Xu, professor at Sun Yat-Sen University, for his guidance and help in preparing this paper. We also would like to give special thanks to Simon MingYuen Lee, professor in University of Macau, for his help in polishing English.

■ REFERENCES

(1) Melnikova, I. Therapies for Alzheimer's disease. *Nat. Rev. Drug Discovery* **2007**, *6*, 341–342.
(2) Harel, M.; Sussman, J. L.; Krejci, E.; Bon, S.; Chanal, P.; Massoulie, J.; Silman, I. Conversion of acetylcholinesterase to butyrylcholinesterase: modeling and mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10827–10831.
(3) Greig, N. H.; Utsuki, T.; Yu, Q.; 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.
(4) Greig, N. H.; Lahiri, D. K.; Sambamurti, K. Butyrylcholinesterase: an important new target in Alzheimer's disease therapy. *Int. Psychogeriatr.* **2002**, *14* (Suppl. 1), 77–91.
(5) Wright, C. I.; Geula, C.; Mesulam, M. M. Neurological cholinesterases in the normal brain and in Alzheimer's disease: relationship to plaques, tangles, and patterns of selective vulnerability. *Ann. Neurol.* **1993**, *34*, 373–384.
(6) Giacobini, E. The effect of MF-8622: a selective BuChE inhibitor. *Proc. Soc. Neurosci.* **1996**, *22*, 203.
(7) Yu, Q.; Holloway, H. W.; Utsuki, T.; Brossi, A.; Greig, N. H. Synthesis of novel phenserine-based-selective inhibitors of butyrylcholinesterase for Alzheimer's disease. *J. Med. Chem.* **1999**, *42*, 1855–1861.
(8) Greig, N. H.; De Micheli, E.; Holloway, H. W.; Yu, Q. S.; Utsuki, T.; Perry, T. A.; Brossi, A.; Ingram, D. K.; Deutsch, J.; Lahiri, D. K.; Soncrant, T. T. The experimental Alzheimer drug phenserine: preclinical pharmacokinetics and pharmacodynamics. *Acta Neurol. Scand.* **2000**, *176*, 74–84.
(9) Giacobini, E.; Spiegel, R.; Enz, A.; Veroff, A. E.; Cutler, N. R. Inhibition of acetyl- and butyryl-cholinesterase in the cerebrospinal fluid of patients with Alzheimer's disease by rivastigmine: correlation with cognitive benefit. *J. Neural Transm.* **2002**, *109*, 1053–1065.
(10) Huang, D.; Gu, Q.; Ge, H.; Ye, J.; Salam, N. K; Hagler, A.; Chen, H.; Xu, J. On the value of homology models for virtual screening: discovering hCXCR3 antagonists by pharmacophore-based and structure-based approaches. *J. Chem. Inf. Model.* **2012**, *52*, 1356–1366.
(11) Zhao, W.; Gu, Q.; Wang, L.; Ge, H.; Li, J.; Xu, J. Three-dimensional pharmacophore modeling of liver-X receptor agonists. *J. Chem. Inf. Model.* **2011**, *51*, 2147–2155.
(12) 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.
(13) Wang, L.; Gu, Q.; Zheng, X.; Ye, J.; Liu, Z.; Li, J.; Hu, X.; Hagler, A.; Chen, H.; Xu, J. Discovery of new selective human aldose reductase inhibitors through virtual screening multiple binding pocket conformations. *J. Chem. Inf. Model.* **2013**, *53*, 2409–2422.
(14) Ge, H.; Wang, Y.; Li, C.; Chen, N.; Xie, Y.; Xu, M.; He, Y.; Gu, X.; Wu, R.; Gu, Q.; Zeng, L.; Xu, J. Molecular dynamics-based virtual screening: accelerating drug discovery process by high performance computing. *J. Chem. Inf. Model.* **2013**, *53*, 2757–2764.
(15) Castilho, M. S.; Guido, R. V.; Andricopulo, A. D. Classical and hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors. *Lett. Drug Des. Discovery* **2007**, *4*, 106–113.
(16) De Souza, S. D.; De Souza, A. M.; De Sousa, A. C.; Sodero, A. C.; Cabral, L. M.; Albuquerque, M. G.; Castro, H. C.; Rodrigues, C. R. Hologram QSAR models of 4-[(diethylamino)methyl]-phenol inhibitors of acetyl/butyrylcholinesterase enzymes as potential anti-Alzheimer agents. *Molecules* **2012**, *17*, 9529–9539.
(17) Zaheer-ul, H.; Uddin, R.; Yuan, H.; Petukhov, P. A.; Choudhary, M. I.; Madura, J. D. Receptor-based modeling and 3D-QSAR for a quantitative production of the butyrylcholinesterase inhibitors based on genetic algorithm. *J. Chem. Inf. Model.* **2008**, *48*, 1092–1103.
(18) Takahashi, J.; Hijikuro, I.; Kihara, T.; Murugesh, M. G.; Fuse, S.; Kunimoto, R.; Tsumura, Y.; Akaike, A.; Niidome, T.; Okuno, Y.; Takahashi, T.; Sugimoto, H. Design, synthesis, evaluation and QSAR analysis of N(1)-substituted norcymserine derivatives as selective butyrylcholinesterase inhibitors. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1718–1720.
(19) Lin, G.; Chen, G.; Lu, C.; Yeh, S. C. QSARs for peripheral anionic site of butyrylcholinesterase with inhibitions by 4-acyloxy-biphenyl-4′-*N*-butylcarbamates. *QSAR Comb. Sci.* **2005**, *24*, 943–952.
(20) Fang, J.; Huang, D.; Zhao, W.; Ge, H.; Luo, H. B.; Xu, J. A new protocol for predicting novel GSK-3beta ATP competitive inhibitors. *J. Chem. Inf. Model.* **2011**, *51*, 1431–1438.
(21) Sakkiah, S.; Lee, K. W. Pharmacophore-based virtual screening and density functional theory approach to identifying novel butyrylcholinesterase inhibitors. *Acta Pharmacol. Sin.* **2012**, *33*, 964–978.
(22) Abbasi, S. W.; Kulsoom, S.; Riaz, N. In silico pharmacophore model generation for the identification of novel butyrylcholinesterase inhibitors against Alzheimer's disease. *Med. Chem. Res.* **2012**, *21*, 2716–2722.

(23) Jorissen, R. N.; Gilson, M. Virtual screening of molecular databases using a support vector machine. *J. Chem. Inf. Model.* **2005**, *45*, 549−561.
(24) Liew, C. Y.; Ma, X. H.; Liu, X.; Yap, C. W. SVM model for virtual screening of Lck inhibitors. *J. Chem. Inf. Model.* **2009**, *49*, 877−885.
(25) Abdo, A.; Chen, B.; Mueller, C.; Salim, N.; Willett, P. Ligand-based virtual screening using Bayesian networks. *J. Chem. Inf. Model.* **2010**, *50*, 1012−1020.
(26) Plewczynski, D.; Grotthuss, M.; Rychlewski, L.; Ginalski, K. Virtual high throughput screening using combined random forest and flexible docking. *Comb. Chem. High Throughput Screening* **2009**, *12*, 484−489.
(27) Ehrman, T. M.; Barlow, D. J.; Hylands, P. J. Virtual screening of Chinese herbs with random forest. *J. Chem. Inf. Model.* **2007**, *47*, 264−278.
(28) Miller, D. W. Results of a new classification algorithm combining K nearest neighbors and recursive partitioning. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 168−175.
(29) Molnar, L.; Keseru, G. M. A neural network based virtual screening of cytochrome P450 3A4 inhibitors. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 419−421.
(30) Liu, X. H.; Ma, X. H.; Tan, C. Y.; Jiang, Y. Y.; Go, M. L.; Low, B. C.; Chen, Y. Z. Virtual screening of Abl inhibitors from large compound libraries by support vector machines. *J. Chem. Inf. Model.* **2009**, *49*, 2101−2110.
(31) Cheng, F.; Li, W.; Zhou, Y.; Shen, J.; Wu, Z.; Liu, G.; Lee, P. W.; Tang, Y. admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model.* **2012**, *52*, 3099−3105.
(32) Cheng, F.; Yu, Y.; Shen, J.; Yang, L.; Li, W.; Liu, G.; Lee, P. W.; Tang, Y. Classification of cytochrome P450 inhibitors and non-inhibitors using combined classifiers. *J. Chem. Inf. Model.* **2011**, *51*, 2482−2495.
(33) Moda, T. L.; Torres, L. G.; Carrara, A. E.; Andricopulo, A. D. PK/DB: database for pharmacokinetic properties and predictive in silico ADME models. *Bioinformatics* **2008**, *24*, 2270−2281.
(34) Chekmarev, D.; Kholodovych, V.; Kortagere, S.; Welsh, W. J.; Ekins, S. Predicting inhibitors of acetylcholinesterase by regression and classification machine learning approaches with combinations of molecular descriptors. *Pharm. Res.* **2009**, *26*, 2216−2224.
(35) 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.
(36) Wang, K.; Hu, X.; Wang, Z.; Yan, A. Classification of acetylcholinesterase inhibitors and decoys by a support vector machine. *Comb. Chem. High Throughput Screening* **2012**, *15*, 492−502.
(37) *ADRIANA.Code*, version 2.2.6; Molecular Networks Inc.: Erlangen, Germany, 2011.
(38) *Molecular Operating Environment (MOE)*, version 2010.10; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2010.
(39) *Discovery Studio*, version 2.5; Accelrys Inc.: San Diego, CA, 2009.
(40) 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−D201.
(41) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic Ah receptor activity by neural networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.
(42) Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. The prediction of the 3D structure of organic molecules from their infrared spectra. *Vib. Spectrosc.* **1999**, *19*, 151−164.
(43) Wang, L.; Wang, M.; Yan, A.; Dai, B. Using self-organizing map (SOM) and support vector machine (SVM) for classification of selectivity of ACAT inhibitors. *Mol. Diversity* **2013**, *17*, 85−96.
(44) Byvatov, E.; Schneider, G. Support vector machine applications in bioinformatics. *Appl. Bioinf.* **2003**, *2*, 67−77.
(45) Yang, Z. R. Biological applications of support vector machines. *Briefings Bioinf.* **2004**, *5*, 328−338.
(46) Vapnik, V. N. An overview of statistical learning theory. *IEEE Trans. Neural Networks* **1999**, *10*, 988−999.
(47) Chang, C. C.; Lin, C. J. LIBSVM: a library for support vector machines, 2001; http://www.csie.ntu.edu.tw/~cjlin/libsvm (accessed Sep 8, 2013).
(48) Xia, X.; Maliski, E. G.; Gallant, P.; Rogers, D. Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem.* **2004**, *47*, 4463−4470.
(49) Chen, L.; Li, Y.; Zhao, Q.; Peng, H.; Hou, T. ADME evaluation in drug discovery. 10. Predictions of P-glycoprotein inhibitors using recursive partitioning and naive Bayesian classification techniques. *Mol. Pharmaceutics* **2011**, *8*, 889−900.
(50) Ellman, G. L.; Courtney, K. D.; Andres, V., Jr.; Feather-Stone, R. M. A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem. Pharmacol.* **1961**, *7*, 88−95.
(51) Savini, L.; Campiani, G.; Gaeta, A.; Pellerano, C.; Fattorusso, C.; Chiasserini, L.; Fedorko, J. M.; Saxena, A. Novel and potent tacrine-related hetero- and homobivalent ligands for acetylcholinesterase and butyrylcholinesterase. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1779−1782.
(52) Lin, G.; Tsai, H.; Tsai, Y. Cage amines as the stopper inhibitors of cholinesterases. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2887−2890.
(53) Savini, L.; Gaeta, A.; Fattorusso, C.; Catalanotti, B.; Campiani, G.; Chiasserini, L.; Pellerano, C.; Novellino, E.; McKissic, D.; Saxena, A. Specific targeting of acetylcholinesterase and butyrylcholinesterase recognition sites. Rational design of novel, selective, and highly potent cholinesterase inhibitors. *J. Med. Chem.* **2003**, *46*, 1−4.
(54) Campiani, G.; Fattorusso, C.; Butini, S.; Gaeta, A.; Agnusdei, M.; Gemma, S.; Persico, M.; Catalanotti, B.; Savini, L.; Nacci, V.; Novellino, E.; Holloway, H. W.; Greig, N. H.; Belinskaya, T.; Fedorko, J. M.; Saxena, A. Development of molecular probes for the identification of extra interaction sites in the mid-gorge and peripheral sites of butyrylcholinesterase (BuChE). Rational design of novel, selective, and highly potent BuChE inhibitors. *J. Med. Chem.* **2005**, *48*, 1919−1929.
(55) Gemma, S.; Gabellieri, E.; Huleatt, P.; Fattorusso, C.; Borriello, M.; Catalanotti, B.; Butini, S.; De Angelis, M.; Novellino, E.; Nacci, V.; Belinskaya, T.; Saxena, A.; Campiani, G. Discovery of huperzine A-tacrine hybrids as potent inhibitors of human cholinesterases targeting their midgorge recognition sites. *J. Med. Chem.* **2006**, *49*, 3421−3425.
(56) Darvesh, S.; McDonald, R. S.; Darvesh, K. V.; Mataija, D.; Conrad, S.; Gomez, G.; Walsh, R.; Martin, E. Selective reversible inhibition of human butyrylcholinesterase by aryl amide derivatives of phenothiazine. *Bioorg. Med. Chem.* **2007**, *15*, 6367−6378.
(57) Butini, S.; Campiani, G.; Borriello, M.; Gemma, S.; Panico, A.; Persico, M.; Catalanotti, B.; Ros, S.; Brindisi, M.; Agnusdei, M.; Fiorini, I.; Nacci, V.; Novellino, E.; Belinskaya, T.; Saxena, A.; Fattorusso, C. Exploiting protein fluctuations at the active-site gorge of human cholinesterases: further optimization of the design strategy to develop extremely potent inhibitors. *J. Med. Chem.* **2008**, *51*, 3154−3170.
(58) Butini, S.; Guarino, E.; Campiani, G.; Brindisi, M.; Coccone, S. S.; Fiorini, I.; Novellino, E.; Belinskaya, T.; Saxena, A.; Gemma, S. Tacrine based human cholinesterase inhibitors: synthesis of peptidic-tethered derivatives and their effect on potency and selectivity. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5213−5216.
(59) Darvesh, S.; Pottie, I. R.; Darvesh, K. V.; McDonald, R. S.; Walsh, R.; Conrad, S.; Penwell, A.; Mataija, D.; Martin, E. Differential binding of phenothiazine urea derivatives to wild-type human cholinesterases and butyrylcholinesterase mutants. *Bioorg. Med. Chem.* **2010**, *18*, 2232−2244.
(60) Komloova, M.; Musilek, K.; Horova, A.; Holas, O.; Dohnal, V.; Gunn-Moore, F.; Kuca, K. Preparation, in vitro screening and molecular modelling of symmetrical bis-quinolinium cholinesterase inhibitors—implications for early myasthenia gravis treatment. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 2505−2509.
(61) Wandhammer, M.; Koning, M. D.; Grol, M. V.; Loiodice, M.; Saurel, L.; Noort, D.; Goeldner, M.; Nachon, F. A step toward the

reactivation of aged cholinesterases—crystal structure of ligands binding to aged human butyrylcholinesterase. *Chem. Biol. Interact.* **2013**, *25*, 19−23.
(62) Huang, L.; Luo, Z.; He, F.; Shi, A.; Qin, F.; Li, X. Berberine derivatives, with substituted amino groups linked at the 9-position, as inhibitors of acetylcholinesterase/butyrylcholinesterase. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 6649−6652.
(63) Jiang, H.; Wang, X.; Huang, L.; Luo, Z.; Su, T.; Ding, K.; Li, X. Benzenediol−berberine hybrids: multifunctional agents for Alzheimer's disease. *Bioorg. Med. Chem.* **2011**, *19*, 7228−7235.
(64) Shi, A.; Huang, L.; Lu, C.; He, F.; Li, X. Synthesis, biological evaluation and molecular modeling of novel triazole-containing berberine derivatives as acetylcholinesterase and beta-amyloid aggregation inhibitors. *Bioorg. Med. Chem.* **2011**, *19*, 2298−2305.
(65) Gao, M.; Liu, A.; Du, G. High-throughput screening for butyrylcholinesterase inhibitors. *Chin. J. New Drugs* **2009**, *18*, 1145−1148.