4D-QSAR Molecular Modeling and Analysis of Flavonoid Derivatives as Acetylcholinesterase Inhibitors

Yanyu Wang, a, b, # Yanping Zhao, a, b Chaochun Wei, a Nana Tian, a, b and Hong Yan* a

a Beijing Key Laboratory of Environmental and Viral Oncology, Faculty of Environment and Life, Beijing University of Technology; Beijing 100124, China; and b Beijing Tide Pharmaceutical Co., Ltd.; No. 8 East Rongjing Street, Beijing Economic Technological Development Area (BDA), Beijing 100176, China.

Received March 30, 2021; accepted April 28, 2021

Flavonoids are potential strikingly natural compounds with antioxidant activity and acetylcholinesterase (AChE) inhibitory activity for treating Alzheimer's disease (AD). In present study, in line with our interests in flavonoid derivatives as AChE inhibitors, a four-dimensional quantitative structure–activity relationship (4D-QSAR) molecular model was proposed. The data required to perform 4D-QSAR analysis includes 52 compounds reported in the literature, usually analogs, and their measured biological activities in a common assay. The model was generated by a complete set of 4D-QSAR program which was written by our group. The best model was found after trying multiple experiments. It had a good predictive ability with the cross-validation correlation coefficient $Q^2 = 0.77$, the internal validation correlation coefficient $R^2 = 0.954$, and the external validation correlation coefficient $R^2_{\text{pred}} = 0.715$. The molecular docking analysis was also carried out to understand exceedingly the interactions between flavonoids and the AChE targets, which was in good agreement with the 4D-QSAR model. Based on the information provided by the 4D-QSAR model and molecular docking analysis, the idea for optimizing the structures of flavonoids as AChE inhibitors was put forward which maybe provide theoretical guidance for the research and development of new AChE inhibitors.

Key words Alzheimer's disease; flavonoid; acetylcholinesterase inhibitor; four-dimensional quantitative structure–activity relationship (4D-QSAR); docking study

INTRODUCTION

Alzheimer's disease (AD) is the most common cause of senile dementia by a progressive neurodegenerative brain disorder. Acetylcholinesterase (AChE) inhibitors are the most promising drugs for symptomatic treatment of AD via increasing the neurotransmitter acetylcholine levels at cerebral cortex synapses. Many AChE inhibitors have been reported, such as Tacrine, Donepezil, Rivastigmine, and Galantamine have been approved by the European and U.S. regulatory authorities. These inhibitors have beneficial effects on the symptoms of AD and diminish the overall clinical applicability of the accumulative side effects. Therefore, the development of efficient AChE inhibitors with low toxicity is still a critical goal in AD therapy. The natural products of plant origin with efficient AChE inhibitors with low to toxicity is still a critical goal in AD therapy.

The four-dimensional (4D)-QSAR was originally proposed by Recanatini et al. in 1997. It can make up for the deficiency of 3D-QSAR without manually adjust and align the conformation of compounds by performing molecular state ensemble averaging. Laboratório de Quimiometria Teórica e Aplicada (LQTA)-QSAR was one of the 4D-QSAR methods proposed by Martins et al. in 2009. Based on the generation of a Conformational Ensemble Profile (CEP) for each compound, it was not only one conformation and the calculation of Lennard–Jones (LJ) potential and Coulomb interaction (C) descriptors for compounds. To filter the descriptor and reduce the dimensionality, a complete set of the 4D-QSAR program in LQTA-QSAR was written by us with Python and Linux shells. This 4D-QSAR model was constructed by using GROMACS dynamics simulation to generate descriptors for LJ and C. Through the selection of descriptors, this model had achieved satisfactory results which can be downloaded from https://github.com/masgils/QSAR_KING.

To our knowledge, the QASAR studies of the AChE inhibitors mainly involve synthetic compounds, no natural products like flavonoids. In this paper, we focused on the 4D-QSAR modeling and analysis of flavonoids as AChE inhibitors. The 4D-QSAR model was established by the data set of 52 synthesized flavonoids with AChE inhibitory activity. By using the established 4D-QSAR program, the model of flavonoids as AChE inhibitors were obtained and analyzed in details. The molecular docking analysis was also carried out to understand one of the molecular mechanics-based methods to predict the biological activity of unknown compounds. As an important part of computer-aided drug design, it is a model based on chemical structure characteristics and biological activity data of compounds. The four-dimensional (4D)-QSAR was originally proposed by Recanatini et al. in 1997. It can make up for the deficiency of 3D-QSAR without manually adjust and align the conformation of compounds by performing molecular state ensemble averaging. Laboratório de Quimiometria Teórica e Aplicada (LQTA)-QSAR was one of the 4D-QSAR methods proposed by Martins et al. in 2009. Based on the generation of a Conformational Ensemble Profile (CEP) for each compound, it was not only one conformation and the calculation of Lennard–Jones (LJ) potential and Coulomb interaction (C) descriptors for compounds. To filter the descriptor and reduce the dimensionality, a complete set of the 4D-QSAR program in LQTA-QSAR was written by us with Python and Linux shells. This 4D-QSAR model was constructed by using GROMACS dynamics simulation to generate descriptors for LJ and C. Through the selection of descriptors, this model had achieved satisfactory results which can be downloaded from https://github.com/masgils/QSAR_KING.

© 2021 The Pharmaceutical Society of Japan
exceedingly the interactions between flavonoids and the AChE targets. Based on the information provided by the 4D-QSAR model and molecular docking analysis, the idea for optimizing the structures of flavonoid derivatives as AChE inhibitors was put forward. The results are the first report on the 4D-QSAR modeling of flavonoid derivatives as AChE inhibitors and can provide theoretical guidance for the research and development of new AChE inhibitors.

DATASET AND METHODS

Dataset The dataset contains 52 flavonoid derivatives as AChE inhibitors reported by Luo et al. and Singh et al.8–10) The inhibitory activities expressed as $IC_{50}$ values were converted into $pIC_{50}$ values, which were used as dependent variables in the following QSAR analysis (Table 1). The data set was randomly divided into two sets with 41 compounds in the training set and 11 compounds in the test set, respectively.

Computer Hardware and Software Computational work was done by CentOS 7.0 operating system. The 4D-QSAR model was built by GaussView 6.0.16,16) Gaussian 09 D01,17) AmberTools 20,18,19) GROMACS, Acpype 2018-9-20, LQTA grid, and VMD 1.9.3. The other software were AutoDock 4.2 with MGLTools 1.5.6, AutoDock Vina, and PyMOL 2.5. Data analysis and regression used a Python3 script written by our research group and downloaded from https://github.com/masgilis/QSAR_KING.

Generating Topology Files The schematic diagram of the 4D-QSAR method was shown in Fig. 2. GaussView 6.0.16 was used to draw the structures of flavonoids. Geometry optimization was carried out in Gaussian 09D on density functional theory (DFT) with B3LYP/cc-pVDZ level. GD3 dispersion correction was added and a Merz–Kollmann population analysis was carried out.15) AmberTools 20 was applied to fit the Restrained Electrostatic Potential (RESP) charge and generated the Generalized Amber Force Field (GAFF) files. GROMACS molecule topology files (*.gro and *.top) were transformed from Amber files by Acpype.20,21)

Molecular Dynamics Simulations All selected flavonoids were placed in the center of a rhombic dodecahedron box. The minimum distance from the molecular and the box border was set to 10 Å. The box was filled by water molecules with an extended Single Point Charge (SPC/E).22) By adding counter ions, the electrostatic charge of the system was zero. The Particle Mesh Ewald (PME) method was used to calculate Van der Waals (VDW) interaction energies with a cut-off radius of 10 Å and the long-range electrostatic. The molecular and solvent water were added to the constant–pressure and constant–temperature ensemble (NPT). The pressure and temperature of the system remain constant controlled by Parrinello Rahman coupling and Berendsen’s thermostat, respectively.23,24) The steepest descent and conjugate gradient method were used to optimizeatomic position. A stepwise heating method was following: 50, 100, 200, and 350 K for 10 picoseconds (ps) simulation time with 1-femtosecond (fs) step size. Then the system backed to 300 K for 500 ps simulation and the trajectory file output every 10 ps. The conformational ensemble profile (CEP) of all compounds was assembled by considering their conformations recorded from 10 to 500 ps.

Generating LQTA Grid Matrix GROMACS was used to achieve molecular alignment. Compound 2 was selected as the reference of alignment for it’s the most active one among all compounds. In the light of the index number of common atoms, all conformations of compounds were superimposed to the reference. The minimum root-mean-square of the distances (RMSD) between the corresponding atoms was calculated by the least-squares method.

The aligned CEPs were referred to the LQTA-grid program. A grid box was defined as $24 \times 25 \times 19$ Å, which was large enough to accommodate the CEPs and the aligned conformation set file. A matrix of interaction energy descriptors was generated by hypothetical N-terminal of protein as $-\text{NH}_3^+$ probe which was calculated once per 1 Å. The Lennard–Jones potential descriptors ($LJ$ descriptors), represented the molecular interaction between the probe and the compound structure, were calculated by Eq. (1):

$$V(r) = 4\varepsilon \left[ \frac{\sigma}{r} \right]^12 - \left[ \frac{\sigma}{r} \right]^6$$

where $\varepsilon$ is the depth of the potential well, $\sigma$ is the infinite distance with zero potential between particles, $r$ is the particles’ distance.

The Coulomb interaction descriptors ($C$), represented the molecular electrostatic interaction between the probe and the compound structure, were calculated by Formula (2):
Table 1. Structures of the 52 Flavonoids as AChE Inhibitors and IC$_{50}$ Values

| Compound | $R_1$ | $R_2$ | $R_3$ | $n$ | IC$_{50}$ (nM) | pH IC$_{50}$ |
|----------|-------|-------|-------|-----|----------------|--------------|
| 1        |       |       |       | 9   | 1.09           | 5.963        |
| 2        |       |       |       | 9   | 0.64           | 6.194        |
| 3        |       |       |       | 9   | 0.67           | 6.174        |
| 4        |       |       |       | 9   | 1.7            | 5.770        |
| 5        |       |       |       | 9   | 0.83           | 6.081        |
| 6        |       |       |       | 9   | 1.08           | 5.967        |
| 7        |       |       |       | 9   | 2.01           | 5.997        |
| 8        |       |       |       | 9   | 3.07           | 5.513        |
| 9*       |       |       |       | 9   | 1.58           | 5.804        |
| 10       |       |       |       | 9   | 1.12           | 5.951        |
| 11*      |       |       |       | 9   | 1.9            | 5.721        |
| 12       |       |       |       | 9   | 1.21           | 5.917        |
| 13       |       |       |       | 9   | 1.25           | 5.903        |
| 14       |       |       |       | 9   | 1.12           | 5.951        |
| 15       |       |       |       | 9   | 2.15           | 5.668        |
| 16       |       |       |       | 9   | 1.38           | 5.860        |
| 17       |       |       |       | 9   | 0.72           | 6.143        |
| 18       |       |       |       | 9   | 4.8            | 5.319        |
| 19       |       |       |       | 9   | 1.99           | 5.701        |
| 20       |       |       |       | 4   | 2.24           | 5.650        |
| 21*      |       |       |       | 4   | 2.77           | 5.558        |
| 22       |       |       |       | 4   | 3.04           | 5.517        |
| 23       |       |       |       | 4   | 8.20           | 5.081        |
| 24       |       |       |       | 6   | 3.11           | 5.507        |
| 25       |       |       |       | 6   | 9.18           | 5.037        |
| 26       |       |       |       | 6   | 6.73           | 5.172        |

| Compound | $R_1$ | $R_2$ | $R_3$ | $n$ | IC$_{50}$ (nM) | pH IC$_{50}$ |
|----------|-------|-------|-------|-----|----------------|--------------|
| 27       |       |       |       | 6   | 7.85           | 5.105        |
| 28*      |       |       |       | 8   | 1.83           | 5.738        |
| 29       |       |       |       | 8   | 1.99           | 5.701        |
| 30       |       |       |       | 8   | 7.88           | 5.103        |
| 31       |       |       |       | 8   | 23.99          | 4.620        |
| 32       |       |       |       | 10  | 3.74           | 5.427        |
| 33       |       |       |       | 10  | 2.26           | 5.646        |
| 34       |       |       |       | 10  | 11.4           | 4.963        |
| 35       |       |       |       | 12  | 33.2           | 4.479        |
| 36*      |       |       |       | -   | 35             | 4.456        |
| 37*      |       |       |       | -   | 37.3           | 4.428        |
| 38       |       |       |       | -   | 44.2           | 4.355        |
| 39       |       |       |       | 5   | 7.7            | 5.114        |
| 40*      |       |       |       | 5   | 8.25           | 5.084        |
| 41       |       |       |       | 3   | 6.77           | 5.169        |
| 42       |       |       |       | 3   | 5.87           | 5.231        |
| 43*      |       |       |       | 5   | 6.48           | 5.188        |
| 44*      |       |       |       | 5   | 6.12           | 5.213        |
| 45       |       |       |       | 6   | 13.53          | 4.869        |
| 46       |       |       |       | 6   | 16.5           | 4.783        |
| 47*      |       |       |       | 7   | 15.6           | 4.807        |
| 48*      |       |       |       | 7   | 17.3           | 4.762        |
| 49       |       |       |       | 8   | 48.12          | 4.318        |
| 50       |       |       |       | 8   | 65.7           | 4.182        |
| 51       |       |       |       | 10  | 57.2           | 4.243        |
| 52       |       |       |       | 10  | 67.2           | 4.173        |

*The marker compound was selected as test sets.
was used to evaluate the model. The descriptors selected by Comparative Distribution Detection Algorithm (CDDA) to remove the descriptors that were extremely inconsistent with the model. Secondly, the X matrix was divided into a training set and a test set. There were 41 compounds in the training set and 11 compounds in the test set. Thirdly, all descriptors with a variance less than 0.01 were removed for their low variance generated by grid points far away from the ligand. The Pearson correlation coefficient between the descriptor and the bioactive label y was calculated automatically. Using the Comparative Distribution Detection Algorithm (CDDA) to move the descriptors with a huge value would damage the model. If the number of descriptors used in modeling was too large, the overfitting will occur. And it also uses the K-fold cross-validation (K = 10) to evaluate these models on the training set, so as to select the model with the lowest the root mean square error of the cross-validation (RMSECV). The results of grid search was shown in Fig. 3. The lower or higher m will increase the RMSECV. If m was in lower value, few descriptors tend to underfit the model. If m was in higher value, the overfitting will occur. Under comprehensive consideration, the m was selected as 0.18, that means taking the top 18% descriptors of Pearson’s correlation coefficient to build a model. If n was with the lowest RMSECV of 0.36. So, the 604 descriptors and 6 components of 41 flavonoids were used to build the 4D-QSAR model by the PLS algorithm.

After the model was established, the fitting effect of the model was observed by the learning curve. The abscissa is the number of training samples, the ordinate is the standard error root mean square error (RMSE). The learning curve was made by split the training set into a smaller training set (80%) and a validation set (20%) randomly. And then constantly adding the sample to training model of the small training set, the effect of the model was tested with a validation set. The results was plotted as a learning curve (Fig. 4). With the increase of training samples, the two curves gradually approach with a small gap in the end, which indicates that the model has a good variance and bias tradeoff without over-fitting or under-fitting.

The predictive ability of the model was evaluated by the linear relationship between the experimental value and the predicted value of the biological activity. The abscissa is the experimental value of the biological activity, the ordinate is the predicted value. The values of each compound and test compounds were in a linear relationship diagram (Fig. 5). The best PLS model was selected by K-fold (K = 10) cross. Its correlation coefficient validation (R^2) was 0.954, cross-validation correlation coefficient (Q^2) was 0.77 and external validation correlation coefficient (R^2_ex) was 0.715, and external validation standard error (RMSEP) was 0.257. The experimental and predictive pIC_{50} values of the 11 test sets are shown in Table 2. The maximum residual error of the pIC_{50} is 0.174, the minimum residual error is 0.004, and the average residual error is 0.012. All the results display that the established model has strong stability and good external prediction ability.

**4D-QSAR Analysis**

The descriptors selected by the

\[ V_q = f \times \frac{q_i q_j}{r} \]  

(2)

f is an electric conversion factor, and the value is 138.935485 kJ·mol⁻¹, qi and qj are the signed magnitudes of the charges, and r is the distance between the charges.

Through the above formula calculation, a data matrix (52 × 121524) dimensions was generated with the row as a sample and column as a descriptor.

**4D-QSAR Modeling**

The QSAR-KING program written in this article can be downloaded from https://github.com/masgils/QSAR_KING. The main steps of the QSAR-KING program are as follows. Firstly, the LJ descriptor and the C descriptor needed to be truncated, otherwise, a calculated descriptor with a huge value would damage the model. Secondly, the X matrix was divided into a training set and a test set. There were 41 compounds in the training set and 11 compounds in the test set. Thirdly, all descriptors with a variance less than 0.01 were removed for their low variance generated by grid points far away from the ligand. The Pearson correlation coefficient between the descriptor and the bioactive label y was calculated automatically. Using the Comparative Distribution Detection Algorithm (CDDA) to remove the descriptors that were extremely inconsistent with the active label y distribution. 3356 descriptors were obtained and arranged in descending order of Pearson’s correlation coefficient. And using the K-fold cross-validation (K = 10) grid search to find the hyperparameter m and n combination with the smallest Mean Square Error (MSE). Finally, the model was constructed by partial least squares (PLS) using two hyperparameters m and n. After the model was built, the test set was used to evaluate the model. The descriptors selected by the model were visualized by VMD 1.9.3.

**Docking Process**

The docking studies of all 52 flavonoids in the whole data set were conducted to further visualize the interaction mechanism and binding conformation. Docking studies were carried out using AutoDock Vina. The crystal structure of the AChE (code ID: 1ACJ) was obtained in the Protein Data. During docking, polar hydrogen atoms were added and Kollman charges were assigned to the structure. The grid box of size 40 × 40 × 40 was used spaced at 0.375 Å and with x, y, z center 4.226, 69.901, 65.807, respectively. PyMOL software was used to visualize the molecular interactions of the stable complex.

**RESULTS AND DISCUSSION**

The QSAR-KING program written by our research group was used to process the obtained X matrix. QSAR-KING is an automatic and efficient 4D-QSAR modeling program that can help us complete data processing and model automatically. When the PLS were used to construct the model, two hyperparameters m and n needed to be determined. The hyperparameter m (ranging from 0.05 to 1, with a step size of 0.01) was used to adjust the number of descriptors used in modeling. The top m descriptors of Pearson’s correlation coefficient was take to build a model. The hyperparameter n represented the principal component in the PLS algorithm. As reported in the literature of 4D-QSAR, the number of principal components were selected using the ordered predictors’ selection (OPS) method. For the OPS algorithm can only find the local minimum, the grid search was used to find the global minimum. The grid search used all hyperparameter combinations to construct a PLS model in a loop. And it also uses the K-fold cross-validation (K = 10) to evaluate these models on the training set, so as to select the model with the lowest the root mean square error of the cross-validation (RMSECV). The results of grid search was shown in Fig. 3. The lower or higher m will increase the RMSECV. If m was in lower value, few descriptors tend to underfit the model. If m was in higher value, the overfitting will occur. Under comprehensive consideration, the m was selected as 0.18, that means taking the top 18% descriptors of Pearson’s correlation coefficient to build a model. If n was with the lowest RMSECV of 0.36. So, the 604 descriptors and 6 components of 41 flavonoids were used to build the 4D-QSAR model by the PLS algorithm.

After the model was established, the fitting effect of the model was observed by the learning curve. The abscissa is the number of training samples, the ordinate is the standard error root mean square error (RMSE). The learning curve was made by split the training set into a smaller training set (80%) and a validation set (20%) randomly. And then constantly adding the sample to training model of the small training set, the effect of the model was tested with a validation set. The results was plotted as a learning curve (Fig. 4). With the increase of training samples, the two curves gradually approach with a small gap in the end, which indicates that the model has a good variance and bias tradeoff without over-fitting or under-fitting.

The predictive ability of the model was evaluated by the linear relationship between the experimental value and the predicted value of the biological activity. The abscissa is the experimental value of the biological activity, the ordinate is the predicted value. The values of train compounds and test compounds were in a linear relationship diagram (Fig. 5). The best PLS model was selected by K-fold (K = 10) cross. Its correlation coefficient validation (R^2) was 0.954, cross-validation correlation coefficient (Q^2) was 0.77 and external validation correlation coefficient (R^2_ex) was 0.715, and external validation standard error (RMSEP) was 0.257. The experimental and predictive pIC_{50} values of the 11 test sets are shown in Table 2. The maximum residual error of the pIC_{50} is 0.174, the minimum residual error is 0.004, and the average residual error is 0.012. All the results display that the established model has strong stability and good external prediction ability.

**4D-QSAR Analysis**

The descriptors selected by the
model are visualized in Fig. 6. It is similar to the contour map analysis of Comparative Molecular Field Analysis (CoMFA). Green and yellow are electrostatic field descriptors (C), red and blue are three-dimensional field descriptors (LJ). The green regions indicate that the addition of positive charge groups can improve the activity of the compounds, while the yellow region favors negatively charged groups. Red regions denote steric interactions corresponding to positive coefficients, while blue regions represent steric descriptors related to negative coefficients.

The relationship between activity and substituents properties can be seen in Fig. 6. There are 3LJ+ and 3C+ regions at the near distance of $R_2$ substituent, indicating that the introduction of positive charge groups with larger volume in the near distance of $R_2$ is beneficial to increase the inhibitory activity of the compound. But there are 4LJ- at the far dis-

---

**Table 2.** Experimental and Predictive $pIC_{50}$ Values of the Test Set Compounds

| Compound | $pIC_{50}$         | Residuals |
|----------|--------------------|-----------|
|          | Experimental       | Predicted |         |
| 9        | 5.801              | 5.825     | 0.024   |
| 11       | 5.721              | 5.748     | 0.027   |
| 21       | 5.558              | 5.564     | 0.006   |
| 28       | 5.738              | 5.742     | 0.004   |
| 36       | 4.456              | 4.375     | -0.081  |
| 37       | 4.428              | 4.457     | 0.029   |
| 40       | 5.084              | 4.808     | -0.276  |
| 43       | 5.188              | 5.151     | -0.037  |
| 44       | 5.213              | 5.201     | -0.012  |
| 47       | 4.807              | 4.981     | 0.174   |
| 48       | 4.762              | 4.916     | 0.154   |
tance of \( R_2 \) substituent, meaning that the introduction of large groups far away from \( R_2 \) is not conducive to inhibitory activity of the compound. There are \( 2LJ \) and \( 2C^+ \) regions on the upper left of the \( R_1 \) substituent, indicating that small groups are favored at the \( R_1 \), and the introduction of large groups is not conducive to the improvement of the activity. There are \( 1LJ^+ \), \( 1C^+ \) regions at the lower right of the \( R_1 \) substituent, indicating that the introduction of a large group with positive charged groups at the lower right of \( R_1 \) can improve the inhibitory activity. Those results were consistent with the \( pIC_{50} \) values of compounds. Compounds 36 and 37 introduce a hydroxyl group at the bottom right of the \( R_1 \) substituent, which significantly reduces the inhibitory activity. Compared with compound 29, compound 28 has a higher inhibitory activity by introducing small steric groups at the \( R_1 \), which proves that \( R_1 \) substituent favors small groups. Compared to compound 47, compound 48 has a lower activity by introducing hydroxyethyl groups, which also proves that the \( R_1 \) substituent should avoid introducing negative charge groups. Compared with compound 1, compound 17 has a higher inhibitory activity by

"Fig. 7. Docking Models of Compound 2 and Two-Dimensional Interaction Force Expansion Diagram"
introducing large steric groups in the near distance of $R_2$. But compound 9 and 11 introduction of large groups in the far distance of $R_2$, which causes these compound have lower activity than others. Those proves that near the $R_1$ substituent favors large groups and far the $R_2$ substituent favors small groups.

Docking Studies Molecular docking is an effective tool for verifying the accuracy and stability of the 4D-QSAR model. It can help us to better predict the possible binding conformation between candidate molecules and target proteins with the binding power and biological activity. The docking studies of all 52 molecules in the whole data set were conducted to further visualize the interaction mechanism and binding conformation. The molecular docking results of 52 molecules with AChE (PDB code: 1ACJ) showed that compound 2 was the most active compound with high docking score ($\Delta G = -7.88$ kcal/mol and $K_i = 1.66 \mu M$).

As is known to all, electrostatic hydrophobic interaction may constitutes steric LJ potential contributions and the VDW interaction may constitutes Coulombic contributions in 4D-QSAR descriptors.\(^{20}\) From the three-dimensional molecular docking Fig. 7a, compound 2 is completely wrapped in the docking pocket. And there are hydrogen bonds between compound 2 and Tyr121 and Tyr334, which makes compound 2 more active. It is consistent with three regions of the descriptor visualization map analyzed by the conformation. From the two-dimensional interaction force expansion diagram Fig. 7b, the $R_2$ substituent is mainly wrapped by Ser81, Tyr121, Tyr70, Asp72, and Ser122 at a distance of 5 Å. The Asp72 is a negative charge amino acid, which means that the introduction of a positive charge group at the $R_2$ substituent is beneficial to the activity. The aromatic $R_1$ of compound 2 interacts with Asp72 to form anion–π interaction, which constitutes the LJ+ contribution around $R_1$ site in 4D-QSAR model. The $R_2$ substituent is mainly wrapped by Gly117, Gly123, Tyr116, Tyr442, Ile439, Trp84, Tyr130, Leu127, and Ser124. They form VDW force at the $R_1$ position of compound 2 and in line with the $C^+$ contribution in 4D-QSAR model. So, the introduction of a positive charge at $R_1$ substituent of the compound 2 will increase biological activity, but the large groups is not conducive to improving the biological activity for the hydrophobic Leu127 and Ile439. There are Gly441, His440, Ghu199, and Ser200 forms VDW force at the lower left of the $R_1$ substituent, which constitutes the $C^+$ contribution in 4D-QSAR model. And the steric hindrance here is relatively small from the three-dimensional molecular docking Fig. 7a. So the introduction of large groups of $R_1$ substituent is beneficial to improve the activity of compound 2. Those results are in good agreement with that of the 4D-QSAR model.

CONCLUSION

In the present study, the 4D-QSAR model of AChE inhibitors was established by the data set of 52 synthesized flavonoids reported in the literature. The data set was randomly divided into two sets with 41 compounds in the training set and 11 compounds in the test set, respectively. The QSAR model was generated by a complete set of the 4D-QSAR programs from a training set of 41 compounds. The 4D-QSAR model had a good predictive ability with the cross-validation correlation coefficient $Q^2 = 0.77$, the internal validation correlation coefficient $R^2 = 0.954$, and the external validation correlation coefficient $R^2_{pred} = 0.715$. Comparing the data of 11 test set compounds with the experimental data, the maximum residual error of the pIC$_{50}$ was 0.174, the minimum residual error was 0.004, and the average residual error was 0.012, which further illustrated the accuracy of the model prediction. Through the visualization of descriptors, the introduction of a positive charge on the $R_1$ substituent was conducive to the increase of the activity of the compound, but the introduction of a large group was not conducive to the increase of the activity. The positive charge groups and large groups were introduced at the near distance of $R_2$, substituent will increase the inhibitory activity. But the introduction of large groups far away from $R_2$ is not conducive to inhibitory activity of the compound. And At the same time, the conformational analysis of molecular docking verified the accuracy of the model. According to the information provided by the model and docking analysis, the modification at $R_1$ position and $R_2$ position of flavonoid derivatives will be the best choices for the development of new AChE inhibitors.

Acknowledgments This work was supported by the Beijing Natural Science Foundation [No. 2192004].

Conflict of Interest The authors declare no conflict of interest.

REFERENCES

1) McGleenon BM, Dynan KB, Passmore AP. Acetylcholinesterase inhibitors in Alzheimer’s disease. Br. J. Clin. Pharmacol., 48, 471–480 (1999).
2) Watkins PB, Zimmerman HJ, Knapp MJ, Gracon SL, Lewis KW. Hepatotoxic effects of tacrine administration in patients with Alzheimer’s disease. JAMA. JAMA, 271, 992–998 (1994).
3) Houghton PJ, Ren Y, Howes M. Acetylcholinesterase inhibitors from plants and fungi. Nat. Prod. Rep., 23, 181–199 (2006).
4) Brühlmann C, Marston A, Hostettmann K, Carrupt PA, Testa B. Screening of non-alkaloidal natural compounds as acetylcholinesterase inhibitors. Chem. Biodivers., 1, 819–829 (2004).
5) Urban A, Marston A, Grilo LS, Bravo J, Pare A, Purevasuren B, Batsuren D, Reist M, Carrupt PA, Hostettmann K. Xanthones from gentianella amarella ssp. acuta with acetylcholinesterase and monoamine oxidase inhibitory activities. J. Nat. Prod., 71, 895–897 (2008).
6) Williams P, Sorribas A, Howes MR. Natural products as a source of Alzheimer’s drug leads. Nat. Prod. Rep., 28, 48–77 (2011).
7) Ji H, Zhang H. Theoretical evaluation of flavonoids as multipotent agents to combat Alzheimer’s disease. J. Mol. Struct. THEOCHEM, 676, 3–9 (2006).
8) Luo W, Chen Y, Wang T, Hong C, Chang L, Chang C, Yang Y, Xie S, Wang C. Design, synthesis and evaluation of novel 7-aminoalkyl-substituted flavonoid derivatives with improved cholinesterase inhibitory activities. Bioorg. Med. Chem., 24, 672–680 (2016).
9) Luo W, Su Y, Hong C, Tian R, Su L, Wang Y, Li Y, Yue J, Wang C. Design, synthesis and evaluation of novel 4-dimethylamino flavonoid derivatives as potential multi-functional anti-Alzheimer agents. Bioorg. Med. Chem., 21, 7275–7282 (2013).
10) Singh M, Kaur M, Singh N, Silakari O. Exploration of multi-target potential of chromen-4-one based compounds in Alzheimer’s disease: design, synthesis and biological evaluations. Bioorg. Med. Chem., 25, 6273–6285 (2017).
11) Karelson M, Lobanov VS, Katritzky AR. Quantum-chemical descriptors in QSAR/QSPR studies. Chem. Rev., 96, 1027–1044 (1996).
12) Recanatini M, Cavalli A, Hansch C. A comparative QSAR analysis of acetylcholinesterase inhibitors currently studied for the treatment of Alzheimer's disease. *Chem. Biol. Interact.*, **105**, 199–228 (1997).

13) Shim J, Mackrell JAD Jr. Computational ligand-based rational design: role of conformational sampling and force fields in model development. *MedChemComm*, **2**, 356–370 (2011).

14) Martins JPA, Barbosa EG, Pasqualoto KFM, Ferreira MMC. LQTA-QSAR: a new 4D-QSAR methodology. *J. Chem. Inf. Model.*, **49**, 1428–1436 (2009).

15) Ma W, Wang Y, Chu D, Yan H. 4D-QSAR and MIA-QSAR study on the Bruton’s tyrosine kinase (Btk) inhibitors. *J. Mol. Graph. Model.*, **92**, 357–362 (2019).

16) A RDTA. GaussView, Version 6. 0.16. 2016.

17) Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian 09 Rev. D.01. 2009.

18) Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174 (2004).

19) Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, **25**, 247–260 (2006).

20) Batista PR, Wilter A, Durham EHAB, Pascutti PG. Molecular dynamics simulations applied to the study of subtypes of HIV-1 protease common to Brazil, Africa, and Asia. *Cell Biochem. Biophys.*, **44**, 395–404 (2006).

21) Sousa da Silva AW, Vranken WF. ACYPPE-Antechamber PYthon parser interface. *BMC Res. Notes*, **5**, 367 (2012).

22) Svishchev IM, Kusalik PG. Structure in liquid water: a study of spatial distribution functions. *J. Chem. Phys.*, **99**, 3049–3058 (1993).

23) Parrinello M, Rahman A. Crystal structure and pair potentials: a molecular-dynamics study. *Phys. Rev. Lett.*, **45**, 1196–1199 (1980).

24) Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690 (1984).

25) Barbosa EG, Ferreira MMC. Digital filters for molecular interaction field descriptors. *Mol. Inform.*, **31**, 75–84 (2012).

26) Yoshida T, Hirono S. A 3D-QSAR analysis of CDK2 inhibitors using FMO calculations and PLS regression. *Chem. Pharm. Bull.*, **67**, 546–555 (2019).

27) Chen F, Yin X, Wang Y, Lv Y, Sheng S, Ouyang S, Zhong Y. Pharmacokinetics, tissue distribution, and druggability prediction of the natural antitumor active compound cytisine N-isoflavones combined with computer simulation. *Biol. Pharm. Bull.*, **43**, 976–984 (2020).

28) Patil RB, Sawant SD. 4D-QSAR studies of coumarin derivatives as HIV-1 integrase 3′-processing inhibitors. *Med. Chem. Res.*, **24**, 3062–3076 (2015).