Struct Chem (2015) 26:467–476
DOI 10.1007/s11224-014-0494-3

ORIGINAL RESEARCH

Computational design of novel flavonoid analogues as potential AChE inhibitors: analysis using group-based QSAR, molecular docking and molecular dynamics simulations

Chakshu Vats · Jaspreet Kaur Dhanjal · Sukriti Goyal · Navneeta Bharadvaja · Abhinav Grover

Received: 28 March 2014 / Accepted: 11 August 2014 / Published online: 11 September 2014

Abstract Acetyl cholinesterase (AChE) is an enzyme associated with the loss of cholinergic neurones in Alzheimer's disease. Acetylcholine is an important neurotransmitter found in the brain and the levels of which decrease significantly in Alzheimer's patients due to increased expression of AChE. In this study, a novel fragment-based QSAR model has been developed using twenty-seven flavonoid-derived compounds exhibiting inhibitory activity against AChE. This fragment-based method gives the advantage of studying the effect of individual fragments on the biological activity of the compound by evaluating the descriptors. The compounds were divided into training and test sets, where the test set was used for cross-validation of the model. The QSAR model exhibited good statistical values for the training set ($r^2 = 0.8070$, $q^2 = 0.7088$, F-ratio = 31.3616) and test set (pred_$r^2 = 0.8131$). The regression equation obtained had three descriptors describing effect of substitutions in terms of quantitative values. Evaluation of the model implied that electronegative substitution at R1 position lowers the inhibitory activity, while the presence of hydroxyl group improves the same. The presence of rings increased the activity of the compounds. The model thus generated was used to obtain six combinatorial libraries and predicts the activity of these compounds. These compounds were selected for docking and molecular dynamics simulation studies and two leads were identified against AChE.

Keywords QSAR · Acetyl cholinesterase · Alzheimer's disease · Inhibitor · Docking · Molecular dynamics simulations

Introduction

Alzheimer's is one of the leading causes of dementia affecting around 37 million people worldwide. It is also the seventh biggest cause of deaths in United States [1]. The estimated figures are expected to double every 20 years till 2040 due to rapidly ageing population around the globe. Alzheimer's is a progressive neurodegenerative disorder characterized by loss of cognition and impaired intellectual ability and functionality. The two important pathological features that characterize the disease include formation of senile plaques due to aggregation of the peptides of β-amyloid which are derived from amyloid precursor protein and neurofibrillary tangles composed of intra-neuronal cluster of paired helical filaments [2]. Other features of the disease include synaptic degeneration of acetylcholine neurotransmitter, accumulation of lysosomes and glia and mitochondria-mediated inflammation. However, none of the above mentioned factors completely describe the clinical features associated with the disorder. The actual mechanism by which amyloid-beta plaques give rise to Alzheimer's is still unknown but it has been assumed that

Electronic supplementary material The online version of this article (doi:10.1007/s11224-014-0494-3) contains supplementary material, which is available to authorized users.

C. Vats · N. Bharadvaja
Department of Biotechnology, Delhi Technological University, New Delhi 110042, India

J. K. Dhanjal · A. Grover (✉)
School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India
e-mail: abhinavgr@gmail.com; agrover@mail.jnu.ac.in

S. Goyal
Apaji Institute of Mathematics & Applied Computer Technology, Banasthali University, Tonk 304022, Rajasthan, India

perturbance in the protein might be the initiating factor [3]. This hypothesis illustrates the triggering of neuron degeneration on accumulation of peptides of amyloid-beta. It has been shown that the toxic aggregates of the protein disrupt the calcium ion channel and initiate apoptosis [4].

Cholinergic deficit or the loss of cholinergic neurons associated with AD leads to the decreased activity of acetylcholine, a neurotransmitter responsible for memory and function [5]. This led to the development of acetyl cholinesterase inhibitors (AChEIs) as potent drugs for the treatment of AD. AChEIs prevent the degradation of AChE by inhibiting the hydrolysis of acetylcholine into choline and acetyl group thereby improving the cognition ability and memory loss [6]. Subsequently, AChEIs were introduced as the preliminary treatment against the disorder. AChEIs and memantine are being used for treatment of AD since last 15 years. All the FDA-approved Alzheimer's drugs, donepezil, tacrine, galantamine, rivastigmine and memantine, belong to these two categories only [1]. These drugs, however, are mildly efficacious and provide only symptomatic relief [7].

The paucity of these drugs and their adverse effects [8] demand major breakthrough in the advancement of new drugs preferably from natural sources. Various compounds of plant origin have been identified to be effective against AD [9]. Anisodamine from *Anisodus tanguticus*, a Chinese herb, has been found to mimic the function of choline in mouse models. Experiments also suggest the use of alkaloids like Withanolide A as probable ligands for inhibition of acetylcholinesterase [10]. Flavonoids, a group of natural compounds found in a variety of fruits and vegetables have also established a presence around the globe due to the important pharmacological functions they possess. Literature provides evidence for its free radical scavenging activity, neuroprotective role, acetyl cholinesterase inhibitory activity and anti-amyloid-beta fibril activity [11, 12]. Thus, the identification of natural products with anti-AD properties is gaining keen interest among the researchers these days.

Traditional drug development method based on random screening, chance discovery is a lengthy, expensive and intellectually inefficient method. Computer-assisted drug designing methods are fast and a viable option for screening of potential drug-like candidates. These methods are low cost and have high success rate. One such method is known as QSAR or quantitative structure activity relationship. QSAR has been long used in scientific community around the world for identification of structure–activity relationships. Previously, QSAR experiments have been performed for inhibitors targeting AChE enzyme. These experiments, however, were based on a 3-D QSAR involving CoMFA and CoMSIA studies [13–16]. With high correlation values, in the range of 0.7–0.9, these models were quite suitable for the prediction of activity of newly synthesized inhibitors based on their 3-D conformation. Although these models could predict the activity of new compounds, they could not signify the importance of the substitution of a particular group at a particular site. In order to identify the contribution of a particular group at a specific site, a novel fragment-based or group-based QSAR method has been developed. This ligand-based drug designing method is advantageous over 2-D and 3-D QSAR. Since it is a fragment-based method, descriptors can be calculated for different fragments instead of a whole molecule. This method can be applied to congeneric as well as non-congeneric inhibitors. Another advantage of this method lies in the fact that the knowledge of the effect of a particular fragment on a substitution site can be used to generate a series of compounds known as combinatorial library. This library can thus be screened for the prediction of more potent drug-like candidates. From the crystallographic structure of AChE, two different ligand binding sites have been identified, i.e. a catalytic active site (CAS) and peripheral cationic site (PAS) [17]. Therefore, in order to completely inhibit AChE, inhibition of both the sites has been advocated. A 3-D QSAR study using flavonoid inhibitors of AChE has been reported previously as well. However, the model was generated using fewer compounds and the correlation values were also low [18]. This study presents an improved model, generated using G-QSAR method and a well-trained system.

A flavonoid scaffold, developed by Li et al., possessing terminal amine groups attached with carbon spacers has been derived to fulfil the aim of designing dual binding site inhibitors of AChEIs. Varied length carbon spacers were used to enhance the dual binding inhibition by the compound. This spacer resides in the mid-gorge and allows the terminal amine groups to occupy CAS site via cation-π interaction and aromatic stacking interactions allows binding of flavonoids to PAS site [19]. In this study, a congeneric series of flavonoid analogues designed and evaluated by Li et al. were used to develop a fragment-based G-QSAR model. These derivatives having basic side chains of different lengths ranging from two to six alkanes depicted metal chelation, AChE inhibition and anti-amyloid-beta aggregation properties [20]. A viable model was obtained which predicted relationship between physiochemical properties of these compounds and their anti-AD properties. The model was then used to recognize important molecular sites and their properties to aid in the development of novel molecule using the approach of virtual combinatorial chemistry. In addition to building the QSAR model, an attempt has been made to provide detailed insights into the molecular mechanism of action of this class of compounds.

Materials and methods

Data set preparation

The structures of 27 flavonoid derivatives [20] were drawn using Marvin sketch 5.12 (Supplementary data), a ChemAxom Ltd. product. These 2-dimensional structures were converted into 3-dimensional form followed by optimization using force field batch minimization, a Vlife engine platform. The minimization was carried out to allow molecules to acquire a stable conformation. Merck molecular force field and Gasteiger charges were the preferred options for the optimization procedure. A common template, representative of all the derivatives, was prepared with the presence of a dummy atom (X) at the substitution site. The reported inhibitory activity (IC_{50}) value of these inhibitors [20] was converted into logarithmic pIC_{50} value to be used for G-QSAR model building. The study was performed using Vlife MDS, version 4.3 provided by Vlife Sciences, Pune, India on Intel® Xeon(R) CPU E31230 @ 3.20 GHz with 8.00 GB RAM [21].

Calculation of descriptors

The common scaffold (Fig. 1) prepared above was used as a template for the fragment-based QSAR model. The G-QSAR module from the VLife MDS was used for the model building. The optimized molecules were imported and their activity values were manually inserted into the worksheet. Activity data can be stored in a .qsr file with molecule names and corresponding activities in each row in this file.

Calculation of 2-D descriptors is one of the most inevitable steps. Molecular descriptors are a numerical representation of chemical information encoded by a molecule. These descriptors are obtained by certain mathematical and logical operations based on the equation and mathematical formulas for different properties. Various physico-chemical descriptors were calculated for the groups present at the substitution site in each of the molecule. The descriptor names adopt a predefined nomenclature such that the name clearly defines the substitution site and properties associated with it. For example, descriptors were named as R1-Hosoya Index, where R1 is the substitution site and Hosoya Index denotes the topological index or number of ways to arrange the edges of bonds of a graph such that no two bonds are placed together. All the columns with a constant value of the molecular descriptor were removed to eliminate the physiochemical properties which do not correlate with the biological activity.

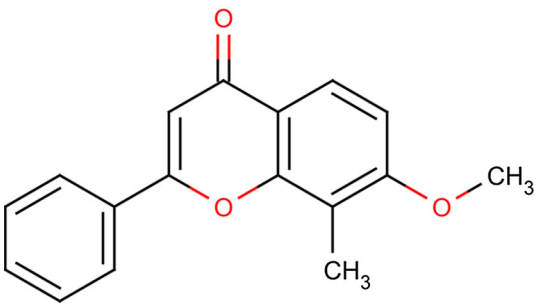

Fig. 1 Common scaffold used for preparing flavonoid derivatives

Data selection and building G-QSAR model

The dataset of the 27 derivatives was divided into test and training sets such that there was a uniform distribution of molecules with respect to activity values (pIC_{50}). This selection was done manually. 9 molecules (7d, 8b, 8d, 9b, 9c, 11c, 11d and 15d out of 27 molecules) were selected for test set and the remaining 18 formed the training set. Unicolumn statistics were calculated for both test and training sets. This is done to observe whether the test set is derived within the max–min range of training set. The mean and standard deviation data, for test and training sets give the relative point density distribution along the mean. Partial Least Square cross-validation method was selected to build G-QSAR model. This method works by removing or adding a predictor variable thereby improving the previous model. The process continues until all the significant variables were included in the model.

From variable selection and model building wizard, simulated annealing algorithm was selected. Simulated annealing was the algorithm of choice for variable (descriptor) selection. The G-QSAR model was built based upon the chosen descriptors. Simulating annealing is a probabilistic method where a temperature variable is kept for simulation of the heating process. This method derives its name from the process of annealing of metals. The temperature variable is initially set high and then decreased as the algorithm progresses. At this stage, algorithm is allowed to accept worst solutions than the current solution, with greater frequency. This ensures that the algorithm, in its initial stages, is not trapped in any local optimum. As the temperature is reduced, the acceptance of solutions becomes more stringent. This allows it to focus on the space where optimum solution can likely be found. This process of gradual cooling makes the algorithm effective enough to obtain an optimal solution in case of large number of local optimum solutions. The value of cross correlation, maximum temperature, number of iterations and variance cutoff were set as 0.5, 1000, 10 and 0.0, respectively.

Validation of the developed G-QSAR model

Certain statistical parameters have to be considered in order to establish a G-QSAR model. These include r^2, q^2, pred_r^2, F-test and standard error [22]. The r^2, coefficient of determination, is a statistical measure of how close the regression line spans the real data points. An F-test is a statistical method of comparing two different models, to identify the best fit. For the model to be robust, the value of these parameters should be above the threshold i.e. $r^2 > 0.5$, $q^2 > 0.5$ and pred_$r^2 > 0.5$. High value of F-test and low values pred_r^{2}se, q^2_se and r^2_se are desirable for a good model [23].

Cross-validation

The model was validated both internally and externally. For internal validation, leave-one-out (q^2) method was adopted. In this method, molecules in the training set were removed consecutively one by one and using the same descriptors, the model was refit. This accounted for prediction of the biological value of the removed molecule. This method is based on the formula:

$$q^2 = 1 - \left(\frac{\sum(y_i - y)}{\sum(y_i - y_{\text{mean}})}\right),$$

where, y_i is the actual and y is the predicted activity of the i^{th} molecule in the training set, and y_{mean} represents the average activity of all molecules of the training set. Model generated from the training set was then used to perform external validation using the test set compounds. The value of pred_r^2 was calculated using the following formula:

$$\text{pred\_}r^2 = 1 - \left(\frac{\sum(y_i - y)}{\sum(y_i - y_{\text{mean}})}\right),$$

where, y_{i} is the actual and y is the predicted activity of the i^{th} molecule in the test set and y_{mean} represents the average activity of all molecules in the training set. Y randomisation tests were used for testing the robustness of the model by comparing it to those derived from random data sets, obtained by shuffling molecules to form new training and test sets. Z score was calculated by the formula:

$$Z \text{ score} = \frac{(h - \mu)}{\sigma}.$$

It is used for calculating the significance of models, by comparing individual scores with the mean score of the entire data set. In this case, h represents the q^2 value calculated for the actual data set; μ is the average q^2 value and σ denotes standard deviation calculated for various iterations using models built by different data sets chosen at random.

Generation of combinatorial library

Six different combinatorial libraries were created on the basis of six different templates. The templates were differing in the length of the spacers, varying from 2 to 6. The lead grow module of VLifeMDS was used to create the libraries. Since there was only one substitution site in the template, each library consisted of 233 molecules. The prediction of activity, however, is done using the generic prediction step in the G-QSAR module. The validated G-QSAR model was used to predict the activity values of the molecules in these six libraries. A highly variable library was obtained by this method.

Preparation of protein and ligand for docking

The protein crystal structure of AChE was obtained from Protein Data Bank [PDB ID: 4M0E]. The CAS site is formed by the residues Trp84, Tyr130, Gly199, His441 and His444 and one conserved residue Phe330, is also involved in the recognition of ligands. The PAS site consists of Tyr70, Asp72, Tyr121, Trp279 and Tyr334 amino acid residues. The water molecules and non-bonded heteroatoms were removed using Accelyrs Viewerlite 5.0 [24]. In order to perform docking, protein was prepared further using Schrodinger's protein preparation wizard [25]. In this preparation process, hydrogens were added, bond lengths were optimized, disulphide bonds were created, terminal residues were capped and selenomethionine was converted to methionine. The compounds obtained from the combinatorial library with higher predicted pIC_{50} value were also prepared using LigPrep. Different chiral, stereochemical and ionization variants of these compounds were generated by this method.

In order to perform docking, a grid was created around the active site of the protein molecule using Glide module of Schrodinger [26–28]. All the small molecules were then docked against the active cleft of the protein using extra precision docking protocol of Glide. The top two complexes, ranked on the basis of their binding energies, were examined for hydrogen bonds and hydrophobic interactions using Ligplot program [29].

Molecular dynamics simulations of the docked complexes

In order to investigate the in vivo stability of the docked complexes, molecular dynamic study was performed in the presence of an explicit solvent on a fully hydrated model using explicit triclinic boundary with harmonic restraints. The simulations were performed using Desmond molecular dynamics module of Schrodinger, with optimized potentials for liquid simulations all-atom force field 2005 [30–32]. The

Table 1 Uni-column statistics

Data	Average	Maximum	Minimum	Standard deviation	Sum
Training	3.0134	3.88	1.59	0.62	54.24
Test	2.99	3.66	1.70	0.63	26.99

complexes were prepared by addition of hydrogens followed by optimization, removal of water molecules, capping of end terminals and generation of disulphide bonds using the protein preparation wizard. Prepared protein–ligand complexes were then solvated with SPC water model in a triclinic periodic boundary box. To avoid direct interaction of the protein complex with its own periodic image, the distance between the complex and the box wall was kept 10 Å. Energy of the prepared systems was minimized to 5,000 steps using steepest descent method or until a gradient threshold of 25 kcal/mol/Å was reached. It was followed by low-memory Broyden-Fletcher-Goldfarb Shanno quasi-Newtonian minimiser until a convergence threshold of 1 kcal/mol/Å was met. For system equilibration, the default parameters in Desmond were applied. The equilibrated systems were then used for simulations at a temperature of 300 K and a constant pressure of 1 atm, with a time step of 2 fs. For handling long-range electrostatic interactions, smooth particle mesh Ewald method was used whereas cutoff method was selected to define the short-range electrostatic interactions. A cut-off of 9 Å radius was used.

Results and discussion

QSAR molecular modelling

The prerequisite for generating a QSAR model is a set of congeneric series with adequate variability in the activity (pIC_{50}) values. The activity values have been derived experimentally [20]. The structures of all the 27 ligands present in the series are provided along with their activity values (Online resource 1). A QSAR model in general is generated considering three important factors: steric, electrostatic and hydrophobic interactions. The descriptors for the three energies were calculated at each lattice point around the grid after the optimization of the ligands. 67 % of the total compounds i.e. 18 were selected as training set and rest 9 as test set. Uni-column statistics (Table 1) were calculated and the two sets were found to adhere to the rule which states that maximum of the test set should be less than the maximum of training set, and minimum of test set should be greater than training set.

Interpretation of G-QSAR model

The model obtained was statistically significant with value of $r^2 = 0.80$, $q^2 = 0.70$ and pred_$r^2 = 0.81$. The values of other statistical parameters are also given in Table 2. The predicted activity data were complying with the actual data with small variations for both test and training sets, which were demonstrated in radar plots (Fig. 2a, b). The data points on the fitness plot lie near the regression line

Table 2 Statistical parameters obtained in G-QSAR model

r^2	q^2	*f*-test	r^2_se	q^2_se	pred_r^2	pred_r^2_se	Y-randomization Avg r^2	Y-randomization Avg q^2
0.8070	0.7088	31.36	0.2918	0.3584	0.8131	0.2764	0.1765	−0.447

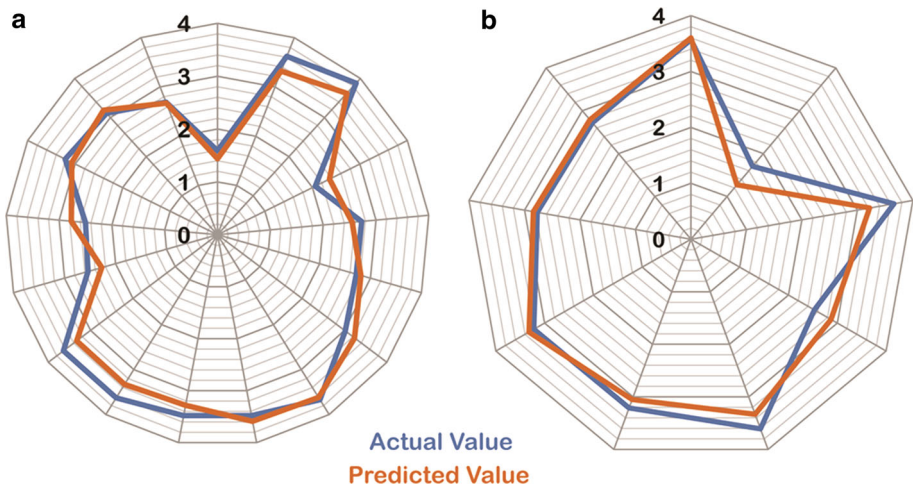

Fig. 2 *Plots* representing actual and predicted activity values of **a** training set, **b** test set

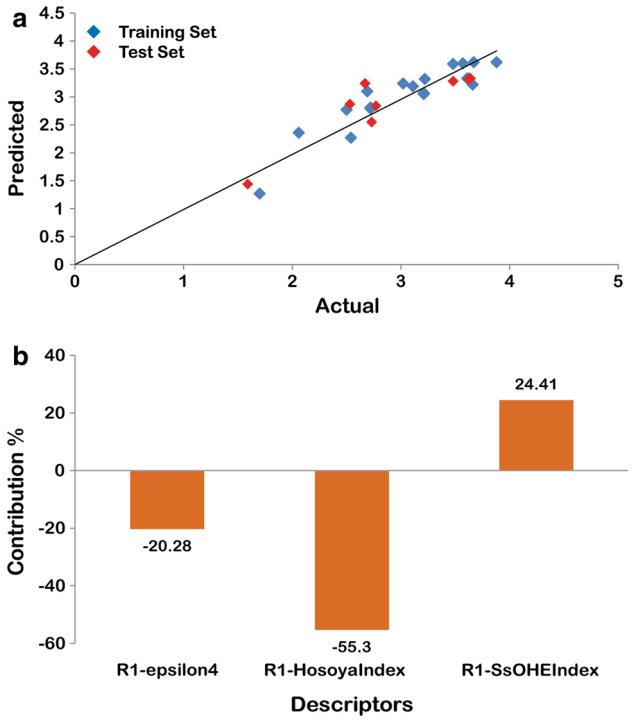

Fig. 3 **a** Fitness plot for test and training sets, **b** molecular descriptors for the G-QSAR model

indicating that the model is acceptable (Fig. 3a). Three important descriptors were calculated for the compounds namely, R1-SsOHE, R1-Hosoy Index and R1-Epsilon4 (Fig. 3b).

The model had good internal and external prediction. The model can be given by the equation:

$$\begin{aligned} \text{pIC50} = {} & (-39.014 * \text{R1} - \text{Epsilon4}) \\ & + (-0.0009 * \text{R1} - \text{Hosoya Index}) \\ & + (0.074 * \text{R1} - \text{SsOHE} - \text{Index}) + 20.82 \end{aligned} ,$$

with $n = 18$, Degree of freedom $= 15$, Z Score $R^2 = 7.95408$, Z Score $Q^2 = 3.24697$ and where n = no. of compounds in training set. The equation obtained had three physico-chemical descriptors which are explained below:

Hosoya Index

This descriptor signifies that the topological index or Z index of a graph is the total number of matching in it plus 1 ("plus 1" accounts for the number of matchings with 0 edges).

$$z = \sum\nolimits_k p(G,k),$$

where, $p(G,k)$ = Number of ways in which K edges from all bonds of a graph G may be chosen so that no two of them are adjacent. The descriptor has negative contribution of −55.30 indicating that less branching or smaller number of edges in the molecule graph is more favourable.

SsOHE-index

This Electro-topological descriptor gives indices for number of –OH group connected with one single bond. It had a positive value of 24.41, indicating that the presence of –OH group increases the activity of the inhibitors.

Epsilon4

Epsilon4 is the electronegativity index of the saturated carbon skeleton i.e. it indicates the electronegativity when all bonds are saturated and all atoms have made carbon. This descriptor has a negative contribution of −20.28 on the inhibitory activity of the inhibitors. This indicates that lower electronegativity is more favourable.

Y-randomization

In order to check the robustness of the model, Y-randomization test was performed using Y-randomization MLR tool of DTClab (http://dtclab.webs.com/software-tools). The test was done using multiple linear regression, which served the purpose of validation of the QSAR model as well. The model values obtained by Y-randomization were similar to the values obtained using PLS. The values for r^2 and q^2 were found to be 0.813 and 0.727, respectively, which were in accordance with the developed QSAR model. To check the effectiveness, 200 random models were generated by shuffling the activity values and keeping the descriptors constant. For a model to be robust, the averaged value of r^2 and q^2 obtained after shuffling should be low and cRp^2 should be high. Our model fulfilled these criteria with the average r^2 as 0.1765, average q^2 as −0.4476 and cRp^2 as 0.7310. This result indicated that the QSAR model was not just a chance prediction and hence is capable of predicting the activity of new compounds.

Applicability domain analysis

Applicability domain (AD) is the physico-chemical, structural or biological space on which the training set of the model has been developed. It gives information about the subspace of the chemical space in which the model can give reliable prediction. Thus, model will be applicable only to compounds falling within this domain. AD for our model was calculated using an online tool available at http://dtclab.webs.com/software-tools. AD calculation was based on distance scores and was obtained using Euclidean

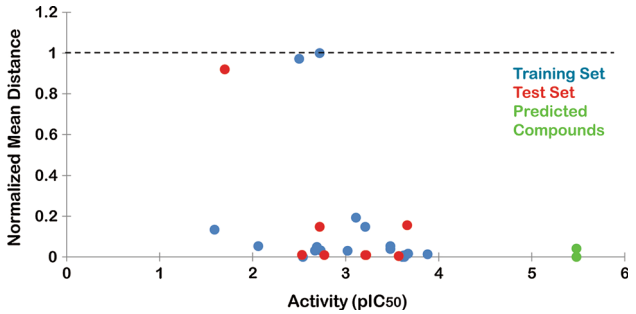

Fig. 4 Scatter plot representing applicability domain values for test set and predicted compounds

distance norms. In this method, the descriptor values were considered and normalized distance was calculated for all the compounds comprising the training and the test sets. The values for training set compounds lied between 0 and 1, where 0 represented the least diverse compound and 1 was for the most diverse compound. In case of test and external set compounds, the value should be within 0 and 1 to consider them within the AD. All the compounds were observed to be within this domain (Fig. 4).

Combinatorial library analysis

Six combinatorial libraries were generated based on the above generated model and their activities were predicted. The six different libraries were generated on different templates with different spacer sequences varying from $n = 2$ to $n = 6$. The substitution was made by different alkanes, atoms, aromatic compounds and rings. Each library consisted of 233 molecules. 40 % of the molecules had predicted values above the highest value in the congeneric series.

Docking and molecular dynamics studies

The top scoring compounds from the combinatorial library with pIC_{50} values above 3.88 were selected. In order to choose the top scoring compound, the extrapolation values were also taken into account. The compounds with extrapolation values above 1.5 were not included. This filtering gave 35 top scoring compounds from all the 6 libraries. In order to identify the potential of these compounds, molecular docking studies were performed using Glide.

Extra precision docking was performed for all the 35 compounds with AChE protein. Results were analysed and the two top scoring compounds were selected to study the interaction pattern with the 3D structure of AChE. The AD value for these two compounds was also calculated as discussed above (Fig. 4). The first compound with activity value of 5.48 contained a spacer sequence with $n = 5$ and had substitution at R1 with cyclohexane (Fig. 6a). The

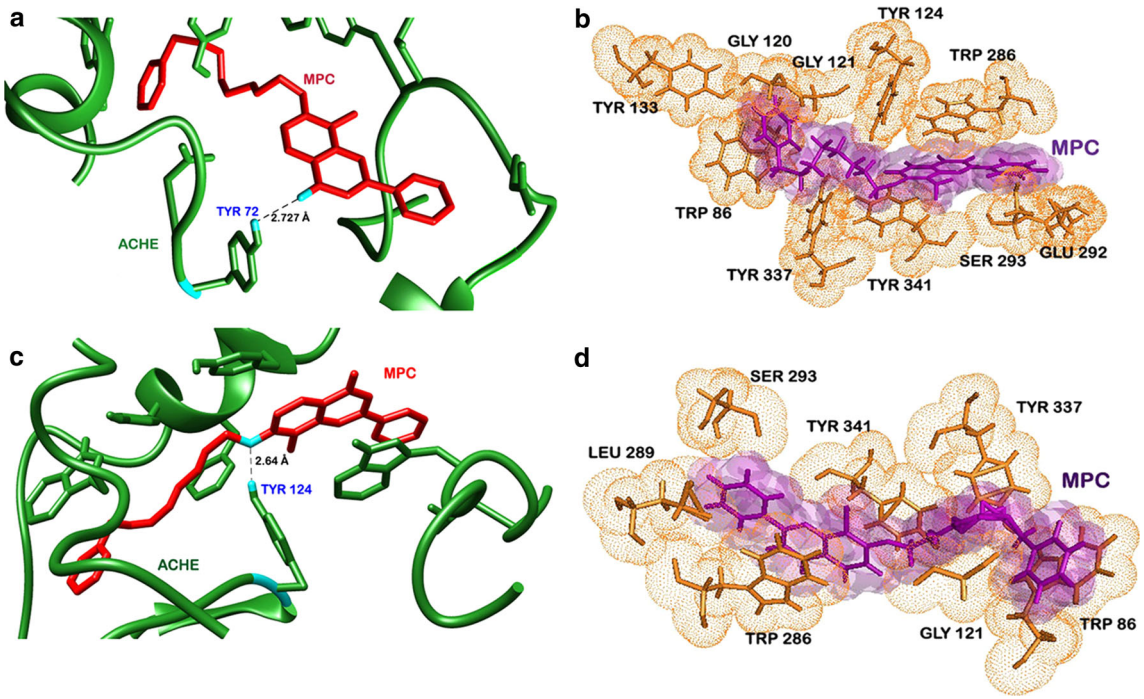

Fig. 5 **a** Pre-MD hydrogen bonds of MPC, **b** pre-MD hydrophobic interactions of MPC, **c** post-MD hydrogen bonds of MPC, **d** post-MD hydrophobic interactions of MPC

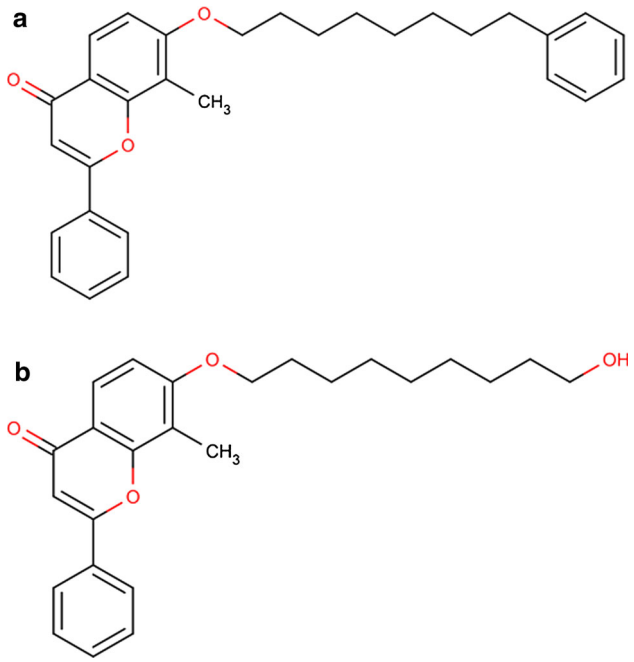

Fig. 6 Structures of the two selected compounds from the library **a** MPC, **b** MPC2

IUPAC name of the compound is 8-methyl-2-phenyl-7-[(8-phenyloctyl)oxy]-4H-chromen-4-one. For the sake of convenience, this compound would be further referred to as MPC. The glide score or the binding energy for the compound was found to be −9.82 kcal/mol. Another important energy value, i.e. van der Waal's energy was also calculated and was found to be −52 kcal/mol. These values suggested strong binding between ligand and protein. Two other important interactions namely hydrogen and hydrophobic interactions were studied using interaction networks plotted using ligplot. MPC made one hydrogen bond with Tyr 72 of AChE (Fig. 5a), one of the active site residues in PAS site, and the bond length was 2.72 Å. Many other strong hydrophobic bonds, with active site residues Gly 120, Gly 121, Leu 289, Tyr 337, Phe 338, Trp 286 and several others (Fig. 5b) were also formed. MPC was found deeply engraved in the active site of AChE.

The second compound (Fig. 6b) with $n = 2$ and activity value of 5.48 had a substitution at R1 with cyclohexane. The IUPAC name of the compound is 8-methyl-2-phenyl-7-(4-phenylbutoxy)-4H-chromen-4-one (further referred to as MPC2). The glide score or binding energy of the compound was calculated as −9.38 kcal/mol and the van der Waal's energy was observed to be −42.91 kcal/mol. In order to explore further the interaction network, ligplot was generated for the complex. MPC2 was making one hydrogen bond with Phe 295 (Fig. 7a) and its bond length was found to be 3.15 Å. Hydrophobic interactions with Trp 286, Leu 289, Val 294,

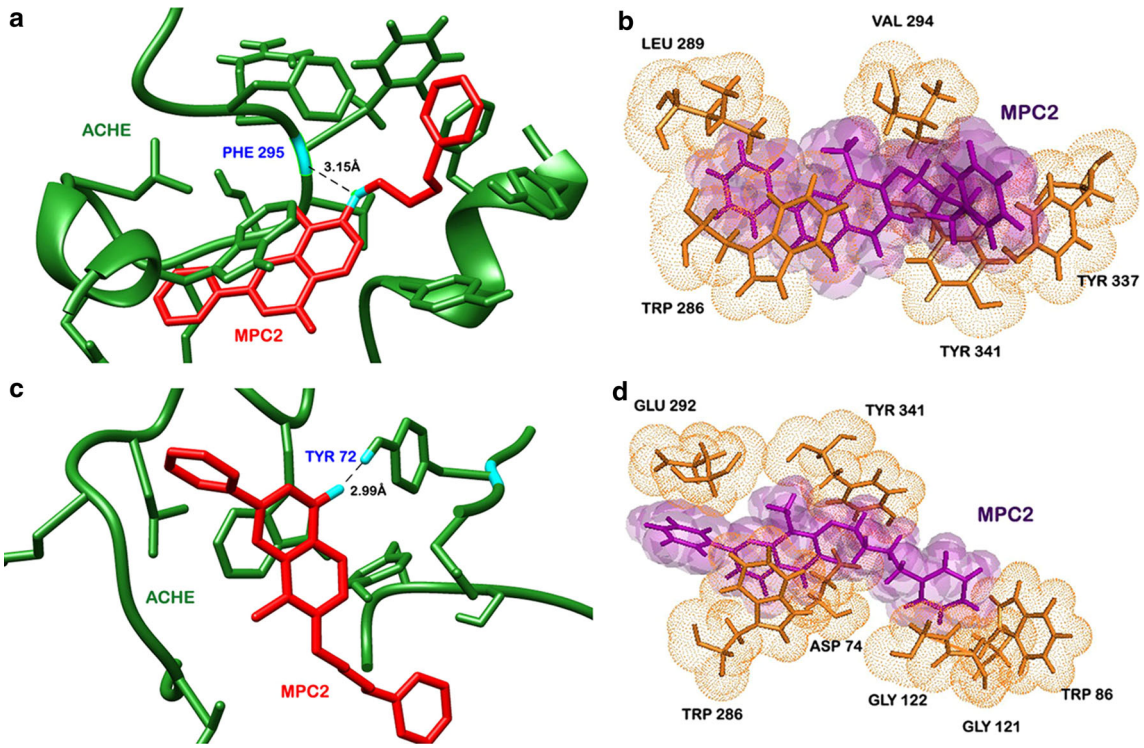

Fig. 7 **a** Pre-MD hydrogen bonds of MPC2, **b** pre-MD hydrophobic interactions of MPC2, **c** post-MD hydrogen bonds of MPC2, **d** post-MD hydrophobic interactions of MPC2

Tyr 337 and Tyr 341 were also (Fig. 7b) observed. This network of strong hydrogen and hydrophobic bonds was stably holding the ligand in place. However, in vivo conditions are different and behaviour of the ligands often changes. Therefore, to further investigate the dynamic behaviour of the ligand–protein complexes, molecular dynamics simulation studies were performed.

Both the complexes were simulated in water box for around 15 ns. A slightly different interaction pattern was observed. After MD, MPC was making one strong hydrogen bond with Tyr 124 (Fig. 5c) instead of Tyr 72 and bond length was found to be 2.64 Å. MPC was now involved in hydrophobic interactions with Trp 86, Gly 121, Trp 286, Ser 293, Tyr 337 and Tyr 341 (Fig. 5d). However, these changes did not affect the stability of the complex. The RMSD plot (Fig. 8a) showed that the complex was stable throughout with fluctuations in the range of 1–2.5 Å. Higher stability was observed from 2 to 6 ns. However, standard deviation was less than 0.8, which indicated that the complex did not deviate much from its initiation conformation. MPC2 also underwent significant changes and was now making a hydrogen bond with Tyr 72 (Fig. 7c) with bond length of 2.99 Å. Tyr 72 is an important residue of PAS site. Hydrophobic bonds with Asp 74, Trp 86, Gly 121, Gly 122, Trp 286, Glu 292 and Tyr 34 (Fig. 7d) were also formed. The changes were significant but did not interfere with the stability of the complex. RMSD of all frames in reference to the first frame was plotted. The docked complex was found to be quite stable with minor deviations (ranging between 1 and 2.5 Å) in the conformation of backbone of the protein (Fig. 8b). Thus, we can strongly suggest these two compounds to be good inhibitors of AChE.

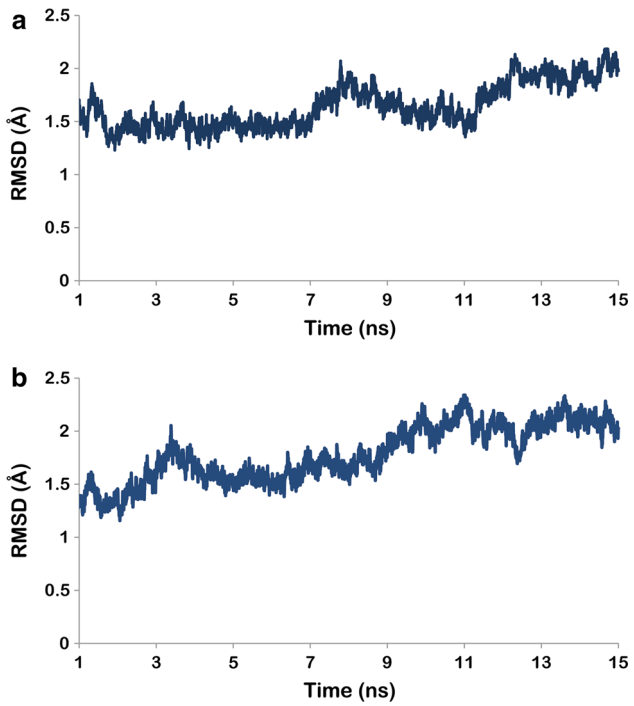

Fig. 8 RMSD trajectory for **a** MPC, **b** MPC2

Conclusion

In this study, a fragment-based QSAR model was developed based on 27 flavanoid molecules with known anti-AChE activity. The activity data and structures were obtained from the literature. The compounds were divided into training and test sets, and model was generated using PLS coupled with simulated annealing method. The statistical parameters obtained were namely r^2, q^2, *F*-test and standard error for the training set and the pred_r^2 for the test set fulfilled the conditions for a model to be considered predictive. The model equation contained three physicochemical descriptors R1-hosoyaindex, R1-SsOHE and R1-epsilon4. The second descriptor was having a positive contribution and the rest two displayed negative contribution in determining the activity of the compounds. Based on the analysis of these descriptors, six combinatorial libraries were created and their activities were predicted using the developed G-QSAR model. The compounds with an activity above 3.88 and in a valid extrapolation range were selected for further studies. Docking studies were performed for these compounds and good binding energies were obtained. This suggested that the binding between the compounds and the protein was favourable. Molecular dynamics studies were performed for the top two compounds obtained after docking to identify the stability of the complex in in vivo conditions. The RMSD reported depicts stable conformation of the protein–ligand complex. Thus, the fragment-based QSAR model developed in this study could be a useful tool in identification and development of lead molecules by taking into account specifically the properties of the substituents. The detailed analysis carried out in this study provides a substantial basis for MPC and MPC2 to be prospective lead molecules against AChE.

Acknowledgments AG would like to thank University Grants Commission, India for the Faculty Recharge position. AG is also thankful to Jawaharlal Nehru University for usage of all computational facilities.

References

1. Melnikova I (2007) Therapies for Alzheimer's disease. Nat Rev Drug Discov 6(5):341–342. doi:10.1038/nrd2314
2. Mayeux R (2006) Genetic epidemiology of Alzheimer disease. Alzheimer Dis Assoc Disord 20(3 Suppl 2):S58–S62

3. Huang Y, Mucke L (2012) Alzheimer mechanisms and therapeutic strategies. Cell 148(6):1204–1222. doi:10.1016/j.cell.2012.02.040
4. Yankner BA, Duffy LK, Kirschner DA (1990) Neurotrophic and neurotoxic effects of amyloid beta protein: reversal by tachykinin neuropeptides. Science 250(4978):279–282
5. Francis PT (2005) The interplay of neurotransmitters in Alzheimer's disease. CNS Spectr 10(11 Suppl 18):6–9
6. Stahl SM (2000) The new cholinesterase inhibitors for Alzheimer's disease, Part 2: illustrating their mechanisms of action. J Clin Psychiatry 61(11):813–814
7. Cutler NR, Sramek JJ (2001) Review of the next generation of Alzheimer's disease therapeutics: challenges for drug development. Prog Neuropsychopharmacol Biol Psychiatry 25(1):27–57
8. Allain H, Bentue-Ferrer D, Tribut O, Gauthier S, Michel BF, Drieu-La Rochelle C (2003) Alzheimer's disease: the pharmacological pathway. Fundam Clin Pharmacol 17(4):419–428
9. Kim H, Park BS, Lee KG, Choi CY, Jang SS, Kim YH, Lee SE (2005) Effects of naturally occurring compounds on fibril formation and oxidative stress of beta-amyloid. J Agric Food Chem 53(22):8537–8541. doi:10.1021/jf051985c
10. Grover A, Shandilya A, Agrawal V, Bisaria VS, Sundar D (2012) Computational evidence to inhibition of human acetyl cholinesterase by withanolide a for Alzheimer treatment. J Biomol Struct Dyn 29(4):651–662. doi:10.1080/07391102.2012.10507408
11. Schroeter H, Spencer JP, Rice-Evans C, Williams RJ (2001) Flavonoids protect neurons from oxidized low-density-lipoprotein-induced apoptosis involving c-Jun N-terminal kinase (JNK), c-Jun and caspase-3. Biochem J 358(Pt 3):547–557
12. Lou H, Fan P, Perez RG (2011) Neuroprotective effects of linarin through activation of the PI3K/Akt pathway in amyloid-beta-induced neuronal cell death. Bioorg Med Chem 19(13):4021–4027. doi:10.1016/j.bmc.2011.05.021
13. Shen LL, Liu GX, Tang Y (2007) Molecular docking and 3D-QSAR studies of 2-substituted 1-indanone derivatives as acetylcholinesterase inhibitors. Acta Pharmacol Sin 28(12):2053–2063. doi:10.1111/j.1745-7254.2007.00664.x
14. Li YP, Weng X, Ning FX, Ou JB, Hou JQ, Luo HB, Li D, Huang ZS, Huang SL, Gu LQ (2013) 3D-QSAR studies of azaoxoisoaporphine, oxoaporphine, and oxoisoaporphine derivatives as anti-AChE and anti-AD agents by the CoMFA method. J Mol Graph Model 41:61–67. doi:10.1016/j.jmgm.2013.02.003
15. Akula N, Lecanu L, Greeson J, Papadopoulos V (2006) 3D QSAR studies of AChE inhibitors based on molecular docking scores and CoMFA. Bioorg Med Chem Lett 16(24):6277–6280. doi:10.1016/j.bmcl.2006.09.030
16. Gharaghani S, Khayamian T, Ebrahimi M (2013) Molecular dynamics simulation study and molecular docking descriptors in structure-based QSAR on acetylcholinesterase (AChE) inhibitors. SAR QSAR Environ Res 24(9):773–794. doi:10.1080/1062936X.2013.792877
17. Sheng R, Lin X, Zhang J, Chol KS, Huang W, Yang B, He Q, Hu Y (2009) Design, synthesis and evaluation of flavonoid derivatives as potent AChE inhibitors. Bioorg Med Chem 17(18):6692–6698. doi:10.1016/j.bmc.2009.07.072
18. Goyal M, Grover S, Dhanjal J, Goyal S, Tyagi C, Grover A (2014) Molecular modelling studies on flavonoid derivatives as dual site inhibitors of human acetyl cholinesterase using 3D-QSAR, pharmacophore and high throughput screening approaches. Med Chem Res 23(4):2122–2132. doi:10.1007/s00044-013-0810-2
19. Fernandez-Bachiller MI, Perez C, Monjas L, Rademann J, Rodriguez-Franco MI (2012) New tacrine-4-oxo-4H-chromene hybrids as multifunctional agents for the treatment of Alzheimer's disease, with cholinergic, antioxidant, and beta-amyloid-reducing properties. J Med Chem 55(3):1303–1317. doi:10.1021/jm201460y
20. Li RS, Wang XB, Hu XJ, Kong LY (2013) Design, synthesis and evaluation of flavonoid derivatives as potential multifunctional acetylcholinesterase inhibitors against Alzheimer's disease. Bioorg Med Chem Lett 23(9):2636–2641. doi:10.1016/j.bmcl.2013.02.095
21. VLifeMDS: Molecular Design Suite. In., 3.0 edn; 2004: Vlife Sciences Technologies Pvt. Ltd., Pune, India
22. Gonzalez MP, Teran C, Saiz-Urra L, Teijeira M (2008) Variable selection methods in QSAR: an overview. Curr Top Med Chem 8(18):1606–1627
23. Golbraikh A, Tropsha A (2002) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J Comput Aided Mol Des 16(5–6):357–369
24. Viewerlite 5.0: Discovery Studio Visualizer. 5.0 edn
25. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27(3):221–234. doi:10.1007/s10822-013-9644-8
26. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47(7):1739–1749
27. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49(21):6177–6196
28. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 47(7):1750–1759
29. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8(2):127–134
30. Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: SC 2006 Conference, Proceedings of the ACM/IEEE. IEEE, pp 43–43
31. Guo Z, Mohanty U, Noehre J, Sawyer TK, Sherman W, Krilov G (2010) Probing the α-helical structural stability of stapled p53 peptides: molecular dynamics simulations and analysis. Chem Biol Drug Des 75(4):348–359
32. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J Chem Theory Comput 6(5):1509–1519