Indian Journal of Biochemistry & Biophysics
Vol. 59, June 2022, pp. 619-631

Identification of potential AChE inhibitors through combined machine-learning and structure-based design approaches

Ankit Ganeshpurkar[1], Ravi Singh[1], Ravi Bhushan Singh[2], Devendra Kumar[3], Ashok Kumar[1] & Sushil Kumar Singh[1]*

[1]Pharmaceutical Chemistry Research Laboratory I, Department of Pharmaceutical Engineering & Technology, Indian Institute of Technology (Banaras Hindu University), Varanasi-221 005, Uttar Pradesh, India

[2]Institute of Pharmacy, Harish Chandra Post Graduate College, Varanasi-221 001, Uttar Pradesh, India

[3]Faculty of Pharmacy, DIT University, Dehradun-248 009, Uttarakhand, India

Received 18 March 2022; *revised 25 June 2022*

Alzheimer's disease (AD) is an irreversible, progressive neurodegenerative disease characterised by dementia.The depletion of acetylcholine (ACh) is involved the synaptic cleft is responsible for dementia due to neuronal loss. The acetylcholinesterase (AChE) enzyme isinvolved in the hydrolytic degradation of ACh and its inhibition is therapeutically beneficial for the treatment in memory loss.The use of machine learning (ML) for the identification of enzyme inhibitors has recently become popular. It identifies important patterns in the reported inhibitors to predict the new molecules. Hence, in this study, a set of support vector classifier-based ML models were developed,validated and employed to predict AChE inhibitors. Further, 247 predicted compounds obtained through PAINS and molecular property filters were docked on the AChE enzyme. The docking study identified compounds AAM132011183, ART21232619 and LMG16204648 as AChE inhibitors with suitable ADME properties. The selected compounds produced stable interactions with enzymes in molecular dynamics studies. The novel inhibitors obtained from the study may be proposed as active leads for AChE inhibition.

Keywords: Alzheimer's disease, Amber, Artificial intelligence, Autodock, Cholinesterase

Acetylcholinesterase (AChE) is a hydrolytic enzyme responsible for the termination of the action of acetylcholine (ACh), a major neurotransmitter of the nervous system[1]. It is widely expressed in vertebrates and has a highly conserved molecular structure. The enzyme is expressed as globular (G_1, G_2, and G_4) and asymmetric (A_4, A_8 and A_{12}) forms[2]. It belongs to a class of α/β protein family withcharacteristic α/β hydrolase folds[3]. AChE consists of eight β-strands connected by α-helices where the almost all the β-strands are arranged parallel to each other except β_2-helix[4]. It has three major sites, *i.e.*, cationic active site (CAS), anionic site (AS) and peripheral anionic site (PAS). The CAS of human AChE consists of a catalytic triad of Ser203, Glu334 and His447 residues. The oxyanion hole (Gly120, Gly121 and Ala204) and acyl pocket (Phe295 and Phe297) are also located with in the active site. Further, the oxyanion hole helps to orient the plane of molecules to facilitate the attack of the nucleophile. Acyl pocket dictates the selectivity of the substrate that could be accommodated in the active site. It is also the site of interaction before the molecule interacts with the active site[5,6].

AS consists of Trp86 residue that interacts with the cationic centre of the choline moiety through the π-cloud. PAS consists of five residues, *i.e.*, Tyr72, Asp74, Tyr124, Trp286 and Tyr341[7]. The mutation of Trp86, Trp286, Phe295, Phe297, Tyr337 and Phe338 residues with alanine resulted in decreased affinity of acetylthiocholine by 660 times, but no decrease in the affinity was observed for 3,3-dimethyl butyl thioacetate, which is a neutral molecule[6]. Further, another study involving a single mutation of Tyr72, Tyr124, Glu285, Trp286 and Tyr341 residues displayed about 10-folds and multiple mutations up to a 400-folds increase in the inhibition constants of the PAS ligands. These findings indicated that the negative charge and π-cloud of the enzyme are the primal features responsible for the movement of the ACh in the gorge[8].

The enzyme initiates its action by nucleophilic attack of serine residue on acyl carbon of ACh. The acid-base catalytic action of histidine residue facilitates abstraction of the proton from the hydroxyl group of

*Correspondence:
E-mail: sksingh.phe@iitbhu.ac.in
Suppl. Data available on respective page of NOPR

the serine. Further, the histidine also transfers proton to choline for completion of the hydrolysis process and release of choline. This is followed by deacylation of serine residue by the attack of water molecule, which is again facilitated by a histidine residue. The glutamate residue present atthe active site restores the proton of the histidine[9].

The classical role of AChE is the termination of the ACh signalling through its metabolism in the synaptic cleft[10,11]. Besides the classical role, it also plays a variety of non-classical roles, including termination of the ACh action in non-synaptic regions. R-splice variant of AChE ($AChE_R$) contains additional 26-residues based C-terminal peptide that modulates haemopoietic differentiation in stress conditions[12-14]. It also promotes neurito genesis in phaeochromocytoma (PC12) cells, neuroblastoma, and primary dorsal root ganglion neurons when transfected with AChE or antisense AChE cRNA[15-18]. The core domain of AChE acts as an adhesive protein similar to neurotactin and displays cell aggregation property[19,20]. Injection of pure AChE in *Xenopus* tadpoles displayed synaptogenic function[21]. Further, AChE also promotes the amyloid fibre assembly. It was found that the inhibition of PAS by edrophonium leads to the inhibition of Aβ aggregation[22].

Alzheimer's Disease (AD) is a multifaceted neurological disorder that progresses into neurodegeneration leading to dementia. Increased oxidative stress due to the deposition of extracellular amyloid β plaques and intracellular hyperphosphorylated tau proteins lead to the contnious reduction in the grey matter and neuronal volume of the brain. This results in deterioration in memory, learning and behaviour in a time-dependent manner. The increased metal ions' concentrations, inflammation, mitochondrial damage and Ca^{+2} signalling dysregulation are also underlying causes for the disease. Davis *et al.* observed that the death of cholinergic neuronsin the brain was correlated tomemory loss[23]. The administration of AChE inhibitors to such patients leads to the improvement in memory function. Further, four AChE inhibitors (AChEI) *viz*. Donepezil. Rivastigmine, Galanthamine and Tacrine are the successful compounds used for the management of the disease. Donepezil is a reversible mixed competitive and non-competitive inhibitor, whereas Rivastigmine is a reversible competitive inhibitor[24]. Galanthamine increases the level of ACh in the synaptic cleft and also causes allosteric modulation of the nicotinic ACh receptor[24]. Further, tacrine was the only AChEI that was withdrawn due to its hepatotoxic effect[25]. Thus, inhibition of AChE is one of the approaches for treating AD[26].

Thousands of AChE inhibitors are reported from different sources with varying AChE inhibition. The vast databases are difficult to analyse in order to identify the features involved in AChE inhibition. Machine learning (ML) is one of the techniques that provides the ability to a machine to make a decision on its own without being explicitly programmed. The decision is made on the basis of the training of available data. The human brain has limitations to perform such a broad set of calculations for the identification of essential or relevant features. ML can deal with a large amount of data and performs the calculations at a much faster pace. The objective of the study is to identify potential AChE inhibitors by combined use of ML and molecular modelling approaches.

Materials and Methods

Machine learning

The ML model was developed by using Python 3.7.6 and Jupyter Notebook within the Anaconda 4.8.2 (64-bit) environment. The other libraries used were Scikit-Learn 0.23.1, Pandas 1.0.5, and Numpy 1.18.

Dataset selection

A human AChE inhibitors dataset of 5761 molecules was prepared using the information obtained from the Binding DB database[27]. The simplified molecular-input line-entry system (SMILES) of the compounds were subjected to substructure filtration by using RDKit in KNIME Analytics Platform (ver. 4.1.2) to obtain *N-benzyl* group-containing compounds. The compounds were assigned labels as active (1) or inactive (0) using pre-decided IC_{50} cut off values. Elite library from Asinex, containing 91473 lead like compounds,was used for the screeningof potential AChE inhibitors.

Molecular descriptor calculation

A total of 101 1D and 2D molecular descriptors were calculated from SMILES using RDKit, excluding molecular descriptors based on functional group count.

Dataset pre-processing

The initial molecular descriptor datasets were further processed by using variance filter, correlation filter and data scaling based on normalisation. Each dataset was divided into training and validation sets in a ratio of 7:3. The validation set of each dataset was used to evaluate the performance ofthe final trained models.

Development of machine learning models

Two different set of models were developed. Initially, the first set of models was developed to predict active and inactive AChE inhibitors with an IC_{50} cut off value of 1000 nM was used (*i.e*. 1 for IC_{50}< = 1000 nM and 0 for IC_{50}> 1000 nM). The second set of models was developed only on active molecules to predict highly potent compounds from the active compounds (*i.e*. 1 for IC_{50}< =100 nM and 0 for IC_{50}> 100 nM). Support vector classifier (SVC) algorithm was used for the development of ML model. Initially, the models were built by using various kernels. *i.e*., linear, polynomial, sigmoidand radial boundary function (RBF). Further, the hyperparameters such as regularisation and gamma values were tuned for the selected kernels obtained after intial screening.

Validation of machine learning

The models were trained on the training datasets using four-fold cross-validation. The trained models were further evaluated on the validation set using a confusion matrix by calculating true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The models were selected based on accuracy, precision, and recall scores[28].

$$Accuracy(A) = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision(P) = \frac{TP}{TP + FP}$$

$$Recall(R) = \frac{TP}{TP + FN}$$

Molecular substructure and properties filters

The identified compounds were subjected to substructure filtration to obtain the compounds with an *N-benzylamine* fragment. The compounds were further examined for the pan assay interference (PAINS) smart patternsusing a KNIME workflow developed by *Saubern et al*[29]. Further, various physicochemical molecular descriptors were calculated by using RDKit, that correlated with the bioavailability and blood-brain permeability of the compounds. The molecular property filter criteria developed by Pajouhesh *et al.* was used for the purpose. It included molecular weight <450, log P <5, H-bond donor <3, H-bond acceptor <7, number of rotatable bonds <8, the total number of H-bonds <8, pKa (neutral or basic) 7.5–10.5 and polar surface area <70 Å[230].

Homology modelling and model validation

Human apo-AChE model (PDB id: 4EY4) was examined through Chimera-1.14rc to identify missing amino acid residues in the protein[31]. The homology modelling was used to fill out missing structures through SWISS-MODEL accessible via the ExPASy web server[32,33]. Further, AChE model was evaluated by using PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK), RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php), Molprobity, GMQE, QMEAN and QMEANDisco. The visualisation of the model was performed with Chimera.

Virtual screening and precision docking

The protein model refinement through energy minimisation was performed as reported in our earlier published studies[34,35]. The protein was minimised by an exhaustive protocol (Table S1 in supporting information (SI)) and monitored for criteria related to protein structure and geometry. SMILES of hit molecules were converted to pdbqt using the reported protocols[34,35]. The non-covalent interactions of three standard AChE inhibitors *viz*. Donepezil, Galanthamine and Huperzine A with the active site of the enzyme was determined using PLIP server (https://plip-tool.biotec.tu-dresden.de). Autogrid 4.0 was used to generate a grid around the selected active site residues for the different autodock atom types *viz*. A, C, HD, NA, N, OA, S, Br, Cl and I, present in the ligands. A grid box of dimensions 62 × 54 × 64 with a 0.375 Å spacing between two consecutive grid pointsand the grid centre was placed at −11.098, −41.511 and 26.758 on the X, Y and Z coordinates, respectively. The validation was performed by redocking a set of AChE inhibitors with diverse chemical structures. The acquired binding energies of the inhibitors were compared with predetermined IC_{50} values. Virtual screening and precision docking were performed using AutoDock 4.2. The conformational search for both the studies was performed using Lamarckian Genetic Algorithm (LGA). The LGA parameters used for virtual screening and precision dockings are summarised in (Table S2 of SI)[36-38]. The virtual screening results were processed through vstools_v0.16, a python script. The binding analysis and visualisation were performed by Discovery Studio visualiser 2020[39,40].

ADMET property

The compounds selected from precision docking were analysed for the absorption, distribution,

metabolism, excretion (ADME) and carcinogenic properties, which were predicted through the Swiss ADME server (http://www.swissadme.ch/)[41,42].

Molecular dynamics

The topology and parameter files of a protein-ligand complex, protein and specific ligands were developed using *tleap* and *antechamber* program in the gas and aqueous phases. Further, the protein-ligand complexes were hydrated as a cubic box using TIP3P water molecules with a cutoff distance of 12 Å between atoms (if any) initially present in the solute and the edge of the periodic box. The residual charges of the system were neutralised by using counter ions. The PME was used to handle long-range electrostatic interactions. The hydrated system was subjected to energy minimisation, heating, equilibration followed by MD run of 50 ns (Table S3 and S4 in SI). The obtained MD simulations were subjected to free energy calculations. The first 50 frames were considered for calculations with a salt concentration of 0.1 mM. The generalised Born implicit solvent model developed by Onufriev *et al*. was used for MM-GBSA calculations[38,43].

Results and Discussion

Machine learning

The substructure filter applied on the dataset obtained from the Bindingdb database yielded 1988 molecules with *N-benzyl amine* fragment similar to donepezil. The molecules were divided into two primary structural datasets, D1 and D2, containing 1988 and 1092 inhibitors with an IC_{50} range of 0.01 nM to 29 μM and 0.01 nM to 100 nM, respectively. Two primary molecular descriptor datasets, *i.e*. D1a and D2a, were obtained with 95 and 90 features, respectively. These datasets were further processed to obtain various subsets. The two primary molecular descriptor datasets were subjected to a variance filter with a criterion of dropping the feature that displayed zero variance. The correlation filter was further applied with the criteria to drop one of the features in any pair of a feature that displayed correlation above +0.8 or below −0.8. It resulted in D1b and D2b datasets. Further, the feature scaling was applied over datasets D1a, D2a, D1b and D2b, based on the normalisation, to obtain four new datasets *viz*. D1c, D2c, D1d and D2d, respectively. The feature scaling helps in the standardisation of independent variable and help to eliminate variance in features. The data pre-processing and feature selection resulted in eight datasets (Table S5 in SI).

SVC is a classification based supervised ML algorithm that classifies elements of a dataset into two or more groups or classes. It identifies a decision boundary that separates the classes based on preassigned labels. The decision boundary could be a line, plane or n-dimensional plane depending upon the dimensionality of the independent variables. The concept of the decision boundary is further enforced by ensuring the maximum margin to separate the classes. The support vectors are data points that lie closest to the hyperplane and dictate the width of the margin. The higher is the margin better is the separation of the two classes. The current work involved the development of two sets of SVC based ML models. The first set of models was developed to predict active compounds with IC_{50} values less than 1000 nM from datasets D1a - D1d. The second set of the models was used to determine the most active compounds with predicted IC_{50} less than 100 nM from datasets D2a – D2d.

The first set of models was initially trained using four available kernels with the default values of remaining hyperparameters. It was found that datasets D1c and D1d outperformed the others. Further, the polynomial and RBF kernels were suitable for model development. It was observed that the removal of correlated and low variance features led to improvement in accuracy, precision and recall for RBF and polynomial kernels but not for the linear kernel. The normalisation of the data further increased the prediction accuracy for the two selected kernels remarkably on both training samples and validation holdouts. Recall, an indicator of correct prediction of positive samples among all the positive samples present in a dataset was increased to more than 80%. Further, the area under the receiver operating characteristic (ROC) was greater than 85%. The results are reported in (Table S6 of SI). The datasets D1c and D1d, along with polynomial and RBF kernels, were selected for tuning the hyperparameters, which resulted in 64 more models. The regularisation parameter (c) is the penalty for the misclassified points and the gamma, which defines how far the influence of a single point reached, were used for hyperparameter tuning. It was observed that the value of 10 was found to be optimum for the regularisation parameter. The scale value, which is equivalent to 1/(number of features × X.var), was found to be better than other values for the gamma parameter. Based on

Tabel 1 — Performance of selected ML models.

					Training Set				Validation Set			
Model	Dataset	Kernel	C	gamma	Accuracy	Precision	Recall	ROC_AUC	Accuracy	Precision	Recall	ROC_AUC
First set of ML models for initial screening												
1.54	D1d	rbf	10	scale	83.97 ± 1.83	85.64 ± 4.63	84.86 ± 4.22	89.96 ± 1.88	83.48	84.65	87.9	83.48
1.22	D1c	rbf	10	scale	83.2 ± 1.65	84.15 ± 4.09	85.52 ± 4.02	89.75 ± 1.83	82.16	83.71	86.43	82.16
1.70	D1d	poly	10	scale	82.6 ± 1.16	84.42 ± 4.29	83.53 ± 5.24	88.7 ± 0.85	81.97	83.47	86.43	81.97
Second set of ML models for identification most active compounds												
2.22	D2d	rbf	10	scale	79.42 ± 5.13	83.14 ± 9.36	83.44 ± 8.44	85.6 ± 1.93	80.42	84	85.27	80.42
2.26	D2d	rbf	100	scale	77.72 ± 7.54	82.05 ± 9.46	80.99 ± 3.75	83.42 ± 3.02	79.15	83	84.26	79.15
2.58	D2d	poly	100	scale	77.16 ± 3.31	80.79 ± 4.79	82.1 ± 4.58	81.53 ± 1.66	78.13	82.08	83.75	78.13
2.70	D2c	rbf	10	scale	79.87 ± 4.94	83.91 ± 8.98	82.77 ± 7.63	86 ± 1.67	81.06	84.5	85.78	81.06
2.74	D2c	rbf	100	scale	78.57 ± 4.41	82.12 ± 6.84	83 ± 4.82	83.77 ± 1.75	81.18	84.84	85.27	81.18

these observations, three out of 64 models, *i.e*. 1.22, 1.54 and 1.70, were selected to carry out the final prediction. The performance of the best and all the models are reported separately in (Table 1 and S7 of SI), respectively.

The second set of models was developed on datasets D2a – D2d in a similar manner as discussed. The initial phase of kernel screening indicated that the model developed on normalised data using polynomial and RBF kernels were better (Table S8 of SI). The hyperparameter tunning indicated that the regularisation penalty of 10, 100 and the scale value for gamma was found to be optimum. The linear kernel had accuracy below 75 % and was not selected (Table S9 of SI). Five out of 80 models developed were used in the final prediction. The details of the performance of the final models are reported in (Table 1). Out of the 91473 compounds of Asinex database, 8913 were found to be active in the prediction and were reduced in number by the second set of the models. Further, 1457 compounds were predicted to have very low IC_{50} values out of 8913 compounds (Fig. S1 in SI).

Molecular substructure and properties filters

Donepezil, a leading drug used in the treatment of AD, bears an *N-benzylamine* group. The group is considered to be responsible for the aqueous solubility and provides selectivity towards AChE. Out of total 1451 compounds, 247 compounds have this group. PAINS are the compounds that may produce false-positive results and off-target effects due to interference with enzymatic assay procedure through cysteine oxidation, chelation, complexation, auto-fluorescence and reactivity towards protein. The physicochemical properties of the compounds are also essential and responsible for the permeation of compounds across the membrane, especially the CNS active drugs that need to cross the complex BBB. The 'Elite Library' or 'Synergy Library', designed by Asinex, screened the properties, and hence the obtained 247 compounds passed both the criteria.

Homology modelling and model validation

The human AChE (PDB id – 4EY4) has the missing residues of peripheral loops 259 – 264 and 495 – 497. The selected PDB was used as a template along with target sequence (access code: P22303) obtained from UniProt (https://www.uniprot.org) to develop a protein model (Fig. S2 in SI). The quality of the model was satisfactory, with more than 99 % of residues falling in Ramachandran allowed and favoured regions (Table S10 in SI). The developed model had Molprobity score of 1.13, which was relatively low and hence the quality of the model was acceptable. The GMQE score of the model was 0.99. The score indicated the expected accuracy of a model built with that template, alignment and coverage of the target sequence. Further, the QMEAN score of the model was found to be 0.86. The QMEAN reflects the quality of the protein model based on the various geometrical parameters, *viz*. the interaction potential between Cβ atoms only, all atoms, the solvation potential and the torsion angle potential. The QMEANDisCo estimated the local quality for the individual residues and was satisfactory for all the residues, except the newly modelled loop.

Virtual screening and precision docking

The first two stages of the energy minimisation of protein resulted in a steep decrease in the potential energy, which subsequently became gradual. There was a gradual decrease in Molprobity and clash scores

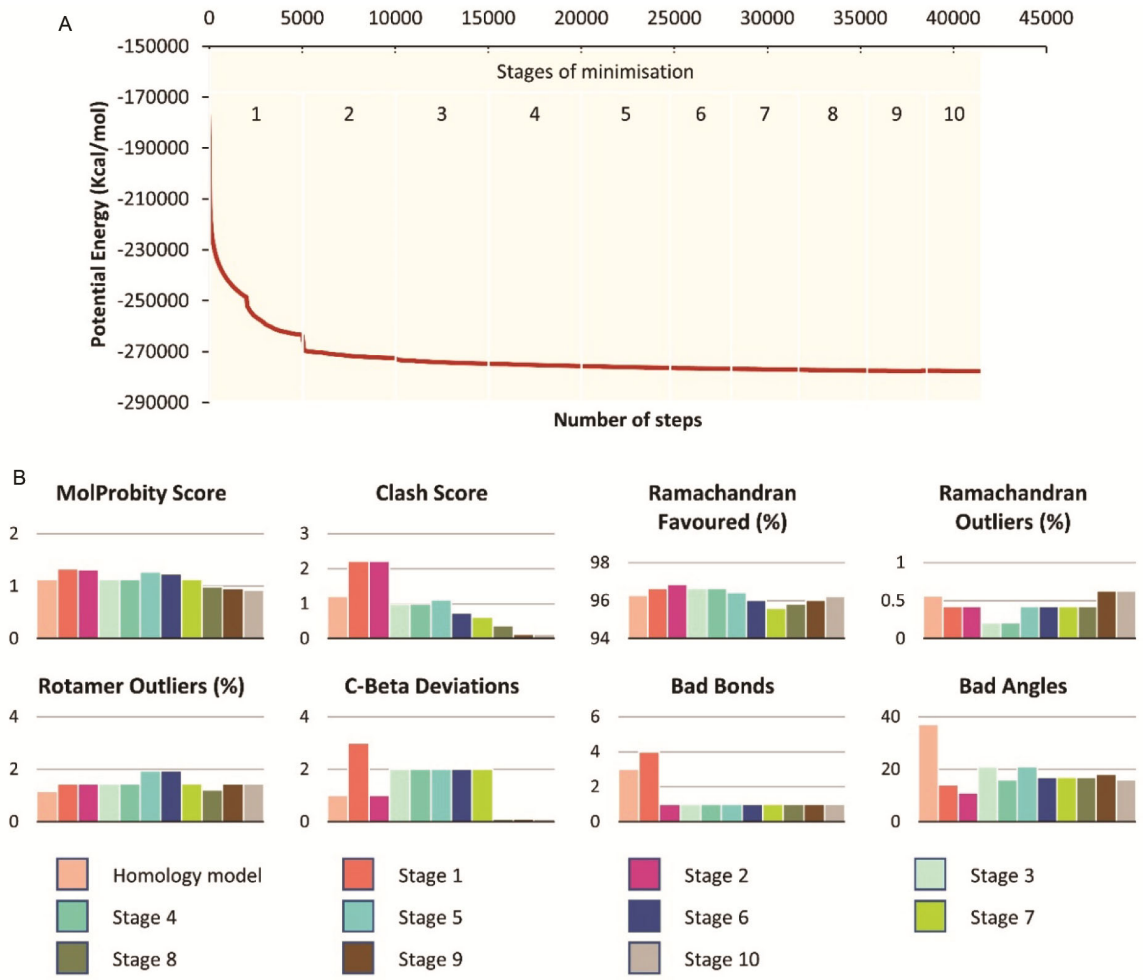

Fig. 1 — (A) Potential energy (Kcal/mol) of the protein models during energyminimisation; and (B) Validation score of protein structures at various energy minimisation stages

too (Fig. 1). The number of Ramachandran favoured and outlier residues remained almost constant, but the rotamer outliers initially remained constant with a slight increase in stages 5 and 6. The inaccurate bond lengths, bond angles and C-β deviations were reduced during the process. The pdb obtained from stage 10 of the energy minimisation was converted to pdbqt using Autodock Tools 1.5.6.

Huperzine A, Galanthamine and Donepezil, cocrystallised with human recombinant AChE, were used for the identification of active site residues for enzyme inhibition. Huperzine A displayed hydrophobic interactions with Trp86 and Tyr337 and hydrogen bonding with Gly120 and Gly121. Tyr337, Tyr133 displayed interactions through the water bridge and Tyr86 through π-π stacking, respectively. Donepezil displayed hydrophobic interactions with Trp86, Tyr337, Phe338 and Tyr341 and hydrogen bonding only with Phe295. It exhibited water bridge mediated interaction with Asp74 and π-π stacking with Trp86, Trp286 and Tyr341. Galanthamine displayed hydrophobic interactions with Trp86, Tyr124, Phe297, Tyr337 and Phe338, water bridge mediated interactions with Gly122, Ser125, Tyr133, Glu202 and Ala204 and hydrogen bonding with Ser203 and Tyr337 (Fig. S3 in SI). Hence, a grid box that could cover all the residues was produced.

A set of 329 chemically diverse AChE inhibitors with predetermined IC_{50} values in the range of 0.006 – 2.09 × 10^7 nM were docked on the enzyme. The compounds were segregated into active and inactive groups based on the twelve IC_{50} cut off values. Various binding energy cutoffs were traded off with IC_{50} values to obtain a confusion matrix (Table S11 in SI). The cut off of 1000 nM displayed the best results among various IC_{50} cut off values. The binding energy cutoff of −8.05 Kcal/mol displayed accuracy and precision values of 73.80 and 77.4 %,

respectively. The selected cutoff returned a high true positive rate of 80.5%. Hence, the binding energy cutoff value of −8.05 Kcal/mol was used to select the active compounds during virtual screening.

The virtual screening resulted in the selection of 247 compounds with binding free energy below the selected cutoff of −8.05 Kcal/mol. The clustering criteria was further used to select the compounds. The compounds displaying cluster size of greater than five were selected. Hence, twenty compounds were finally obtained. The population size of 100 was used in LGA to sample out more conformations for precision docking. Eight ligands, *i.e.*, AOP19760577, ART22341446, AOP19760615, AAM13201183, LMG19707967, ART21232619, ART21295754 and LMG16204648 were found to have cluster sizes greater than 70. The details of binding energiesand interaction patterns are included (Table 2, Fig. 2 and Fig. S4 in SI).

ADMET property

A clinical candidate should have higher potency with low toxicity. Swiss ADME server allows the computation of various parameters *viz*. GI absorption, BBB permeantion and CYP inhibition (CYP1A2, CYP2C19, CYP2C9, CYP2D6 and CYP3A4). Eight compounds obtained from precision docking displayed good absorption profiles and blood-brain barrier permeabilities (Table 3). Further, the compounds were non-inhibitors of most cytochrome P450 enzyme subtypes. The BOILED-Egg model was also used to study the intestinal and brain permeabilities. The model was based on the lipophilicity and polarity of small molecules and used WLOGP and tPSA as molecular descriptors (Fig. S5 in SI). All the selected compounds were found to be in the region with a high probability to permeate the brain[44].

Molecular dynamics

MD is an computational simulation technique that assists to understand the stability protein-ligand complex. In the study, four protein-ligand complexes of AAM13201183, ART21232619, LMG16204648 and Donepezil with AChE were studied through MD simulations. The stability of the protein-ligand complexes was examined by calculating the root mean square deviation (RMSD) for the heavy atoms. The mean RMSDs of AChE and AChE complexed with donepezil was found to be almost identical with

Table 2 — Docking results displaying mean binding energy, cluster size, lowestbinding energy and interactions of ligands against AChE (PDB id 4EY4)

Compound code	Mean binding energy (Kcal/mol) ± SD	Cluster count	Lowest binding energy (Kcal/mol)	Hydrogen bonding	π - interactions	Miscellaneous interaction
AOP19760577	−10.35 ± 0.2014	95	−10.47	Phe295 [a], Arg296 [a], Val294[b]	Trp286 [c, d], Tyr341 [c, d], Phe297[d], Tyr124 [d, f], Asp74 [e]	-
ART22341446	−9.69 ± 0.0168	93	−9.70	Gly121 [a], Gly122 [a], Tyr124 [b], Ser125 [b], His447 [b]	Trp86 [d], Tyr337 [c, d], Phe338 [d], His447 [c]	-
AOP19760615	−9.07 ± 0.0605	83	−9.11	-	Trp86 [c], Tyr124 [f], Phe338 [d], Tyr341 [c, g]	Tyr341 [j]
AAM13201183	−9.02 ± 0.0315	79	−9.13	Tyr72 [b], Asp74 [b], Asn87 [b], Ser125 [b], Tyr337 [b]	Trp86 [c, d, g, h]	-
LMG19707967	−9.66 ± 0.0679	77	−9.69	-	Trp86 [c], Tyr124 [d], Trp286 [c], Tyr337 [c, d], Phe338 [d], Tyr341 [c, f]	His447 [i]
ART21232619	−9.11 ± 0.1267	87	−9.22	Gly126 [a], Ser125 [b], Tyr337 [b], His447 [a]	Trp86 [c, d, g], Tyr337 [c, d], His447 [d]	Glu71 [j], Trp86 [j]
ART21295754	−10.0 ± 0.0400	72	−10.06	-	Trp86 [c, d, g], Tyr337 [d], Phe338 [d], His447 [c, d]	Val73 [h], Pro88 [h]
LMG16204648	−9.56 ± 0.0142	90	−9.60	Tyr72 a, Tyr124 a	Trp86 [d, g], Tyr337 [c], His447 [c, d]	Ile541 [i]
Donepezil	−10.59 ± 0.0807	69	−10.97	Phe295 a, Tyr341 b	Trp86 c, Tyr124 d, Trp286 c, g, Tyr337 c, d, Phe338 d, Tyr341 c	-

[a] : Conventional hydrogen bond, [b] : Carbon hydrogen bond, [c] : π-π interaction, [d] : π-alkyl interaction, [e] : π-anion interaction, [f] : π-lone pair interaction, [g] : π-σ interaction, [h] : π-sulphur interaction, [i] : alkyl interaction. [j] : Halogen interaction

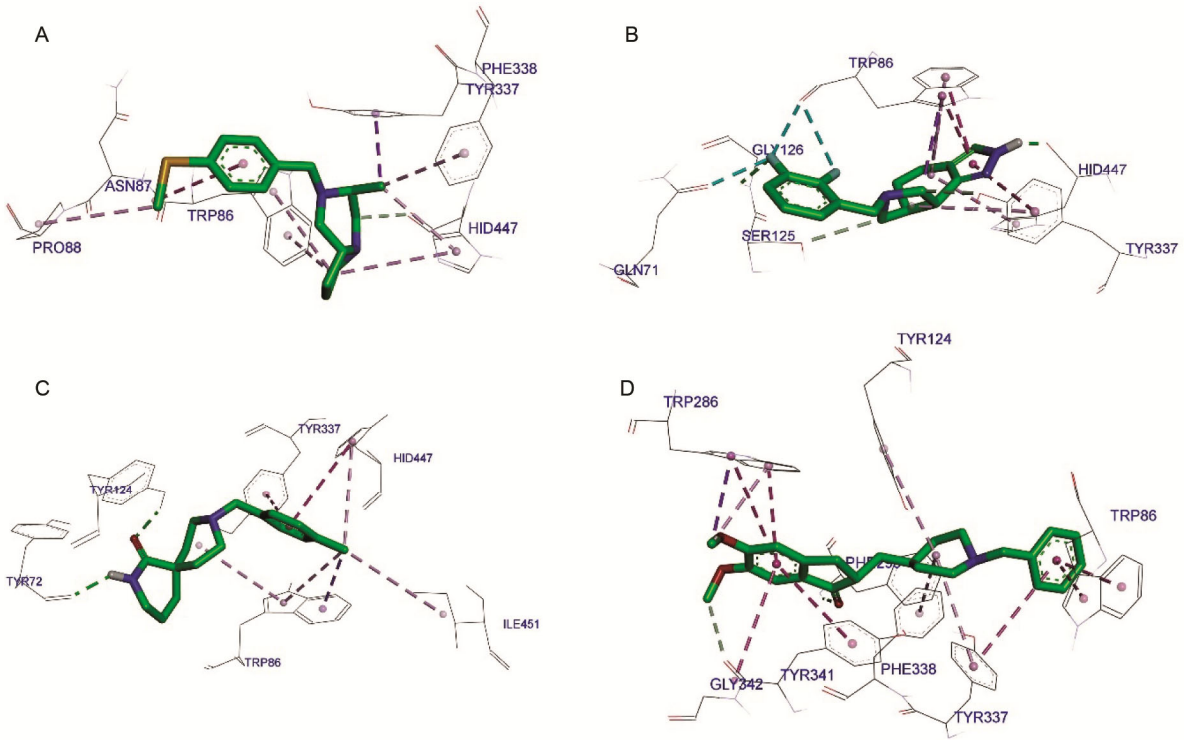

Fig. 2 — 3D interaction diagram of (A) AAM13201183; (B) ART21232619; (C) LMG16204648; and (D) Donepezil docked with AChE

Table 3 — ADME properties of selected compounds

Molecule	GI absorption	BBB permeant	Pgp substrate	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor
AOP19760577	High	Yes	Yes	Yes	Yes	Yes	No	Yes
ART22341446	High	Yes	Yes	No	No	No	Yes	No
AOP19760615	High	Yes	Yes	Yes	Yes	Yes	No	Yes
AAM13201183	High	Yes	No	No	No	No	Yes	No
LMG19707967	High	Yes	Yes	Yes	Yes	No	Yes	Yes
ART21232619	High	Yes	Yes	No	No	No	Yes	No
ART21295754	High	Yes	Yes	No	No	No	Yes	Yes
LMG16204648	High	Yes	Yes	No	No	No	Yes	No

the mean RMSD of 1.7529 ± 0.1632 and 1.6932 ± 0.1566 Å, respectively. The mean RMSDs of AAM132011183 and LMG16204648 complexes were 1.9783 ± 0.1960 and 2.0299 ± 0.2283 Å, respectively, while the ART21232619-AChE complex had the lowest mean RMSD value of 1.6329 ± 0.2283Å. The trajectories of ligands *viz*. AAM13201183, LMG16204648 and Donepezil were found to be stable. However, a much higher fluctuation in RMSD values was observed for ART21232619, which was stabilised after 30thns. Donepezil displayed a mean RMSD of 0.6249 ± 0.1511 Å that was lower than mean RMSD values of AAM132011183, ART21232619 and LMG16204648 of 1.0932 ± 0.2278, 1.1272 ± 0.4989 and 0.9498 ± 0.2542 Å, respectively. The RMSF values of the protein backbone complexed with ligands were found to be similar to the uninhibited enzyme. The active site and tunnel residues displayed RMSF values below 1 Å. The RMSF of ligands displayed relatively higher fluctuations. In the case of donepezil, the solvent-exposed *methoxy* groups displayed higher fluctuation, followed by fluctuations in the *N-benzyl* ring. In AAM132011183, the *octahydropyrrolo [1,2-a]pyrazine* ring displayed minimum RMSF, while the two *methyl* groups displayed higher fluctuations. The fluctuations in ART21232619 was observed in the *fluoro* groups as well asthe fused *pyrazole* ring. The *N-benzyl* moiety of LMG16204648 displayed fluctuations while the remaining structure was relatively stable (Fig. 3).

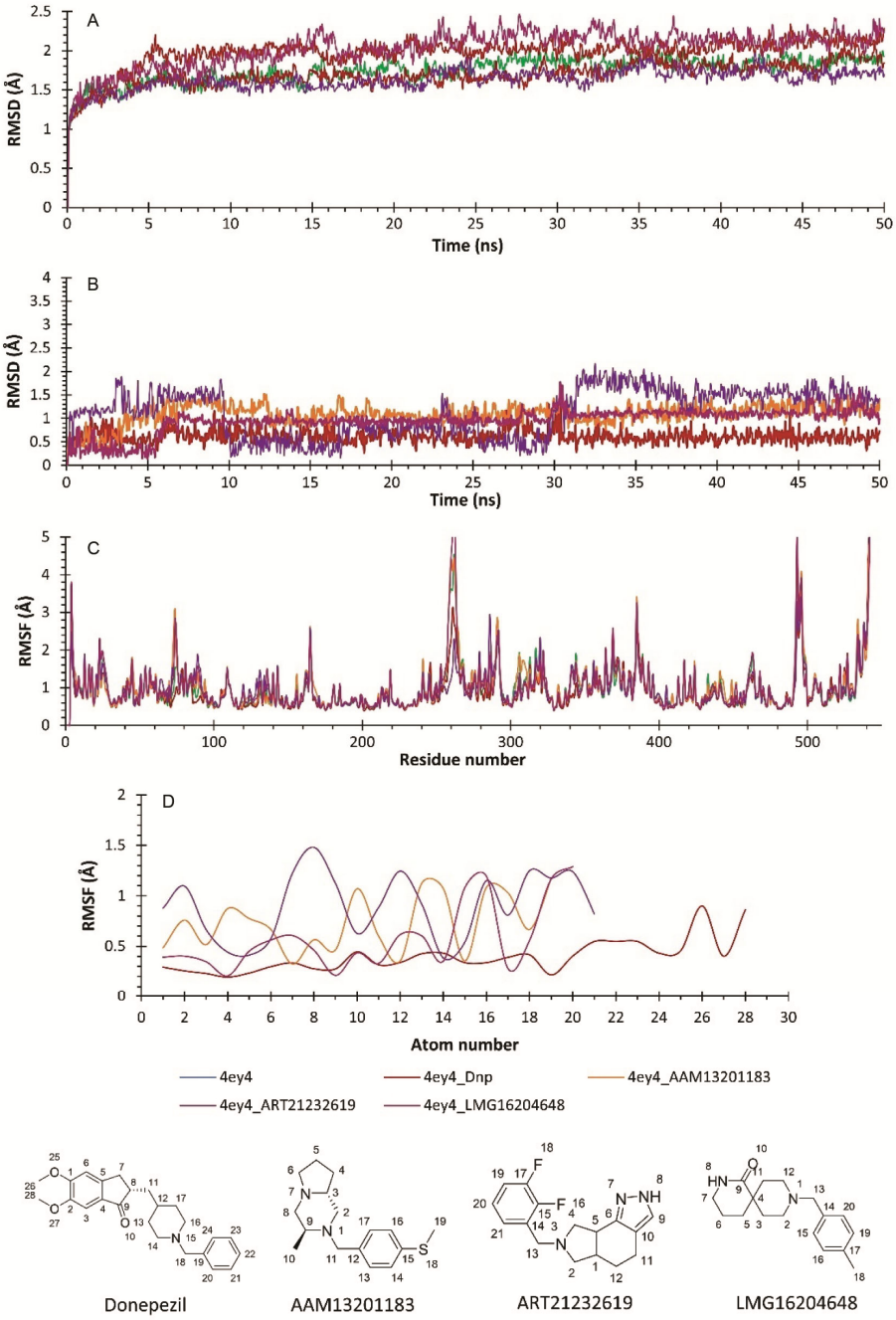

Fig. 3 — (A) RMSD plot of AChE protein backbone; (B) RMSD plots of various ligands bound to AChE; (C) RMSF plot of AChE residues; and (D) RMSF plot of various atoms of ligands

The surface of the enzyme and the tunnel leading to the active site contributed to the solvent-accessible surface area (SASA) of AChE. The mean SASA(s) were 21855.34 ± 314.38, 21253.70 ± 306.70, 21631.65 ± 345.90, 21736.41 ± 357.44, and 21774.03 ± 396.57 $Å^2$ for AChE and AChE complexed donepezil, AAM132011183, ART21232619 and LMG16204648, respectively. A decrease in mean SASA was observed in the case of donepezil due to the two *methoxy* groups displaying intermolecular interactions with the residues present at the opening of the tunnel. In other cases, there was a smaller decrease in the SASA as compared to the uninhibited enzyme. The radius of gyration (RoG) indicates the tightness of the protein

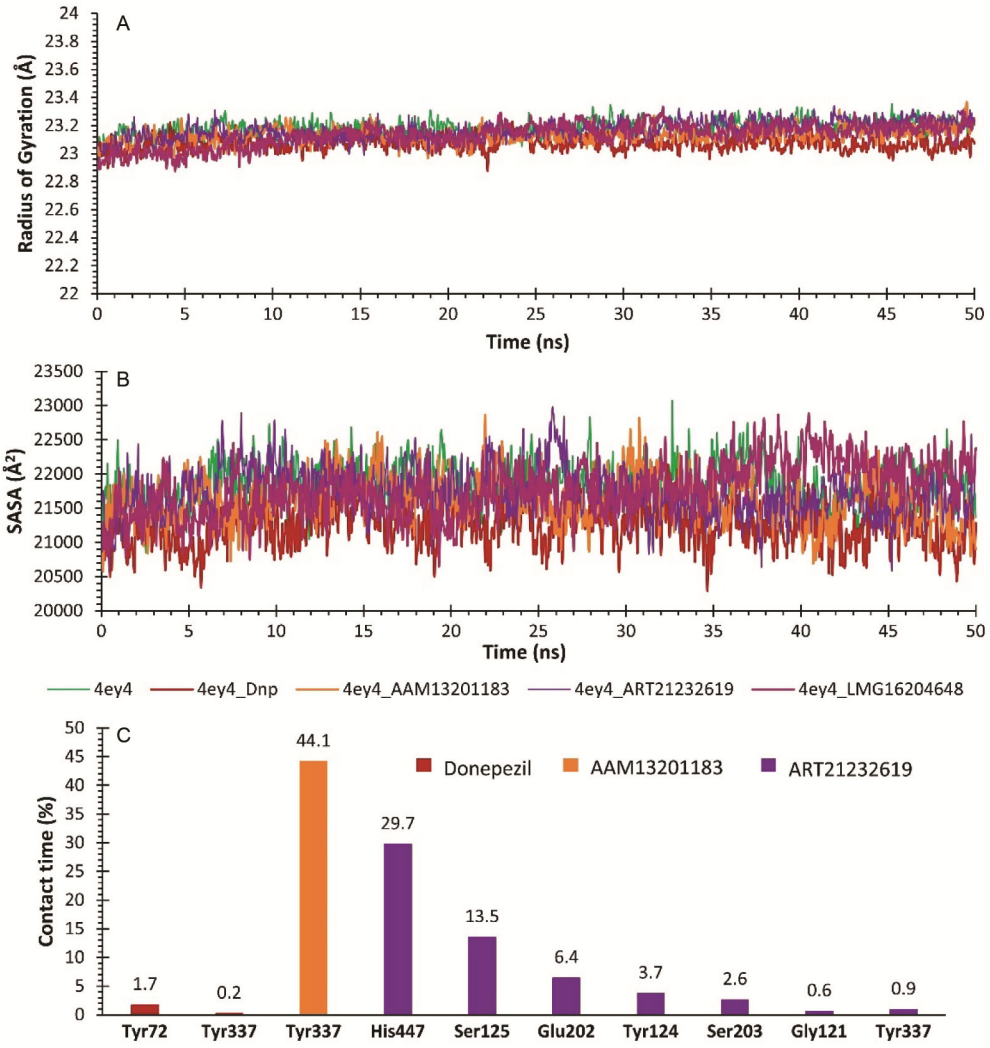

Fig. 4 — (A) Radius of gyration; (B) SASA; and (C) Hydrogen bonding interactions of various protein-ligand complexes

packing. The mean RoG was found to be 23.179 ± 0.056, 23.073 ± 0.05, 23.128 ± 0.054, 23.169 ± 0.064, and 23.132 ± 0.083 Å for AChE and AChE complexed donepezil, AAM132011183, ART21232619 and LMG16204648, respectively. There was no substantial changein RoG, indicating the absence of any alteration in the packing arrangements of the α-chains and β-sheets of the enzyme(Fig. 4).

The contact analysis between the residues of AChE and ligands is tabulated in (Table S12 in SI). Tyr72, Asp74, Thr83, Trp86, Gly120, Gly121, Gly122, Tyr124 Ser125, Gly126, Leu120, Tyr133, Gly202, Ser203, Ala204, Tyr337, Phe338, Tyr341, His447, and Gly448 were prominent residues in contact with the ligands. The residues 287-297 displayed interactions with donepezil but not with other ligands. The hydrogen bond analysis of donepezil displayed hydrogen bonding for only a fraction of time with Tyr72 and Tyr337 residues. However, AAM21232619 displayed hydrogen bonding with Tyr337 residues for 44.1% of the simulation time, while ART212322619 displayed hydrogen bonding with various residues with a prominent contact with His447, Ser125, Glu202, Tyr124 and Ser203. The residues constitute the catalytic site of AChE. LMG16204648 displayed no hydrogen bonding during the simulation, in contrast to molecular docking results.

MM-GBSA is an end-state free energy calculation method that uses multiple conformations obtained from trajectories to determine the stability of the ligand binding. The binding free energy is constituted from combined molecular mechanics and implicit solvent models energy contributions. The binding free energy of AAM13201183 was comparable to

Table 4 — Energy contributions of protein-ligand complexes in MM-GBSA

Compound code	ΔE_{vdw}	ΔE_{ele}	ΔG_{GB}	ΔG_{SA}	ΔG_{MMGBSA}
AAM13201183	−41.804 ± 3.326	−12.936 ± 3.355	28.204 ± 2.415	−5.327 ± 0.253	−31.863 ± 3.019
ART21232619	−35.975 ± 2.704	−14.091 ± 6.043	28.957 ± 5.490	−4.471 ± 0.201	−25.579 ± 3.652
LMG16204648	−37.638 ± 2.814	−13.630 ± 8.327	30.039 ± 5.663	−4.639 ± 0.367	−25.869 ± 4.457
Donepezil	−54.980 ± 2.470	−8.120 ± 3.070	36.038 ± 2.934	−6.253 ± 0.200	−33.327 ± 2.387

donepezil, indicating higher stability (Table 4). The major contribution in donepezil binding was by van der Waal's forces, while in the case of other ligands, it wasdue to electrostatic forces. The per-residue decomposition of the $\Delta G_{MM\text{-}GBSA}$ is presented in (Fig. S6 of SI). Trp86, an anionic site residue, displayed a higher contribution towards the receptor binding. The important residues included Gly120, Gly121, Tyr124 and Tyr337 which were primary interacting residues. The ligands other than donepezil did not display interaction with Trp286. Further, all the ligands also displayed interaction with the cationic site.

Conclusion

The applicationof ML techniques in drug design is an emerging field of drug discovery. It is a convenient tool to identify the features for the target specificity with the availability of inhibitors and the data. Further, the other conventional tools of computational drug design (CDD) provide additional benefits to identify the potential hits. AChE is one of the critical targets in AD, and its inhibition provides symptomatic relief in impaired cognition and memory, which has been proven time and again. The study is an attempt to identify AChE inhibitors by the combined application of ML and conventional CDD approaches. SVC-based ML models predicted 1457 active molecules out of 91473 compounds obtained from the Asinex library. Further, the compounds displayed the absence of PAINS substructures and 257 compounds passed the physicochemical properties filters. The docking study and ADME filter afforded AAM132011183, ART21232619 and LMG16204648 as AChE inhibitors. MD study indicated that the complexes formed by the ligands with the enzyme were stable.

Acknowledgement

The authors would like to acknowledge the financial research support from the Ministry of Education (MoE), New Delhi, India, in the form of teaching assistantships to AG and RS. RBS would like to thank DST, SERB, New Delhi for the award of the TARE grant to him. The authors would also like to extend their gratitude toward Professor David A. Case, Department of Chemistry & Chemical Biology, Rutgers University, New Jersey, the USA for granting a license for Amber 18 and Dr. Stefano Forli, Department of Integrative Structural and Computational Biology, Scripps Research, California Campus for providing python script, vstools_v0.16.

Conflict of interest

All authors declare no conflict of interest.

References

1 Barnard EA, Neuromuscular transmission—enzymatic destruction of acetylcholine, in The peripheral nervous system. Springer: New York. 1974, 201.

2 Bon S, Vigny M & Massoulié J, Asymmetric and globular forms of acetylcholinesterase in mammals and birds. *Proc Natl Acad Sci*, 76 (1979) 2546.

3 Hotelier T, Renault L, Cousin X, Negre V, Marchot P & Chatonnet A, ESTHER, the database of the α/β-hydrolase fold superfamily of proteins. *Nucleic Acids Res*, 32 (2004) D145.

4 Sussman JL & Silman I, Acetylcholinesterase: structure and use as a model for specific cation—protein interactions. *Curr Opin Struct Biol*, 2 (1992) 721.

5 Vellom DC, Radic Z, Li Y, Pickering NA, Camp S & Taylor P, Amino acid residues controlling acetylcholinesterase and butyrylcholinesterase specificity. *Biochemistry*, 32 (1993) 12.

6 Ordentlich A, Barak D, Kronman C, Flashner Y, Leitner M, Segall Y, Ariel N, Cohen S, Velan B & Shafferman A, Dissection of the human acetylcholinesterase active center determinants of substrate specificity. Identification of residues constituting the anionic site, the hydrophobic site, and the acyl pocket. *J Biol Chem*, 268 (1993) 17083.

7 Cavalli A, Bottegoni G, Raco C, De Vivo M & Recanatini M, A computational study of the binding of propidium to the peripheral anionic site of human acetylcholinesterase. *J Med Chem*, 47 (2004) 3991.

8 Barak D, Kronman C, Ordentlich A, Ariel N, Bromberg A, Marcus D, Lazar A, Velan B & Shafferman A, Acetylcholinesterase peripheral anionic site degeneracy conferred by amino acid arrays sharing a common core. *J Biol Chem*, 269 (1994) 6296.

9 Zhou Y, Wang S & Zhang Y, Catalytic reaction mechanism of acetylcholinesterase determined by Born-Oppenheimer ab initio QM/MM molecular dynamics simulations. *J Phys Chem B*, 114 (2010) 8817.

10 Anglister L, Stiles JR & Salpetert MM, Acetylcholinesterase density and turnover number at frog neuromuscular

junctions, with modeling of their role in synaptic function. *Neuron*, 12 (1994) 783.

11 Tai K, Bond SD, MacMillan HR, Baker NA, Holst MJ & McCammon JA, Finite element simulations of acetylcholine diffusion in neuromuscular junctions. *Biophys J*, 84 (2003) 2234.

12 Kaufer D, Friedman A, Seidman S & Soreq H, Acute stress facilitates long-lasting changes in cholinergic gene expression. *Nature*, 393 (1998) 373.

13 Deutsch VR, Pick M, Perry C, Grisaru D, Hemo Y, Golan-Hadari D, Grant A, Eldor A & Soreq H, The stress-associated acetylcholinesterase variant AChE-R is expressed in human CD_{34}^{+} hematopoietic progenitors and its C-terminal peptide ARP promotes their proliferation. *Exp Hematol*, 30 (2002) 1153.

14 Silman I & Sussman JL, Acetylcholinesterase:'classical'and 'non-classical'functions and pharmacology. *Curr Opin Pharmacol*, 5 (2005) 293.

15 Layer PG, Weikert T & Alber R, Cholinesterases regulate neurite growth of chick nerve cells in vitro by means of a non-enzymatic mechanism. *Cell Tissue Res*, 273 (1993) 219.

16 Koenigsberger C, Chiappa S & Brimijoin S, Neurite differentiation is modulated in neuroblastoma cells engineered for altered acetylcholinesterase expression. *J Neurochem*, 69 (1997) 1389.

17 Grifman M, Galyam N, Seidman S & Soreq H, Functional redundancy of acetylcholinesterase and neuroligin in mammalian neuritogenesis. *Proc Natl Acad Sci*, 95 (1998) 13935.

18 Bigbee JW, Sharma KV, Chan EL-P & Bögler O, Evidence for the direct role of acetylcholinesterase in neurite outgrowth in primary dorsal root ganglion neurons. *Brain Res*, 861 (2000) 354.

19 De la Escalera S, Bockamp E, Moya F, Piovant M & Jimenez F, Characterization and gene cloning of neurotactin, a Drosophila transmembrane protein related to cholinesterases. *EMBO J*, 9 (1990) 3593.

20 Darboux I, Barthalay Y, Piovant M & Hipeau-Jacquotte R, The structure-function relationships in Drosophila neurotactin show that cholinesterasic domains may have adhesive properties. *EMBO J*, 15 (1996) 4835.

21 Sternfeld M, Ming G-l, Song H-j, Sela K, Timberg R, Poo M & Soreq H, Acetylcholinesterase enhances neurite growth and synapse development through alternative contributions of its hydrolytic capacity, core protein, and variable C termini. *J Neurosci*, 18 (1998) 1240.

22 Inestrosa NC, Alvarez A, Perez CA, Moreno RD, Vicente M, Linker C, Casanueva OI, Soto C & Garrido J, Acetylcholinesterase accelerates assembly of amyloid-β-peptides into Alzheimer's fibrils: possible role of the peripheral site of the enzyme. *Neuron*, 16 (1996) 881.

23 Davies P, A critical review of the role of the cholinergic system in human memory and cognition. *Ann N Y Acad Sci*, 444 (1985) 212.

24 Zeb MW, Riaz A & Szigeti K, Donepezil: a review of pharmacological characteristics and role in the management of Alzheimer disease. *Clin Med Insights Geriatr*, 10 (2017) 1.

25 Ibach B & Haen E, Acetylcholinesterase inhibition in Alzheimer's Disease. *Curr Pharm Des*, 10 (2004) 231.

26 Dubey S, Ahmad Y & Kohli S, Protective effect of huperzine A on phenytoin induced cognition impairment: Behavioral and biochemical study. *Indian J Biochem Biophys,* 59 (2022) 205.

27 Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L & Chong J, BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res*, 44 (2016) D1045.

28 Kumar N & Shetty N, Machine learning approach for COVID-19 crisis using the clinical data. *Indian J Biochem Biophys*, 57 (2020) 602.

29 Saubern S, Guha R & Baell JB, KNIME Workflow to Assess PAINS Filters in SMARTS Format. Comparison of RDKit and Indigo Cheminformatics Libraries. *Mol Inform*, 30 (2011) 847.

30 Pajouhesh H & Lenz GR, Medicinal chemical properties of successful central nervous system drugs. *NeuroRx*, 2 (2005) 541.

31 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, 25 (2004) 1605.

32 Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L & Schwede T, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*, 42 (2014) W252.

33 Katiyar K, Srivastava RK, Nath R & Singh G, Cryptosporidiosis, a public health challenge: A combined 3D shape-based virtual screening, docking study, and molecular dynamics simulation approach to identify inhibitors with novel scaffolds for the treatment of cryptosporidiosis. *Indian J Biochem Biophys*. 59 (2022) 296.

34 Ganeshpurkar A, Singh R, Kumar D, Gutti G, Sardana D, Shivhare S, Singh RB, Kumar A & Singh SK, Development of homology model, docking protocol and Machine-Learning based scoring functions for identification of Equus caballus's butyrylcholinesterase inhibitors. *J Biomol Struct Dyn*, (2021) 1.

35 Ganeshpurkar A, Singh R, Shivhare S, Kumar D, Gutti G, Singh R, Kumar A & Singh SK, Improved machine learning scoring functions for identification of Electrophorus electricus's acetylcholinesterase inhibitors. *Mol Divers*, 22 (2022) 1455.

36 Ganeshpurkar A, Kumar D & Singh SK, Design, synthesis and collagenase inhibitory activity of some novel phenylglycine derivatives as metalloproteinase inhibitors. *Int J Biol Macromol*, 107 (2018) 1491.

37 Umre R, Ganeshpurkar A, Ganeshpurkar A, Pandey S, Pandey V, Shrivastava A & Dubey N, I*n vitro*, *in vivo* and *in silico* antiulcer activity of ferulic acid. *Future J Pharm Sci*, 4 (2018) 248.

38 Singh R, Ganeshpurkar A, Kumar D, Kumar D, Kumar A & Singh SK, Identifying potential GluN2B subunit containing N-Methyl-D-aspartate receptor inhibitors: an integrative in silico and molecular modeling approach. *J Biomol Struct Dyn*, 38 (2019) 2533.

39 Erdogan T, Computational evaluation of 2-arylbenzofurans for their potential use against SARS-CoV-2: A DFT,

molecular docking, molecular dynamics simulation study. *Indian J Biochem Biophys*, 59 (2022) 59.

40 Choudhury M, Sharma D, Das M & Dutta K, Molecular docking studies of natural and synthetic compounds against human secretory PLA2 in therapeutic intervention of inflammatory diseases and analysis of their pharmacokinetic properties. *Indian J Biochem Biophys*, 59 (2022) 33.

41 Daina A, Michielin O & Zoete V, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*, 7 (2017) 42717.

42 Srivastava S & Pandey A, Computational screening of anticancer drugs targeting miRNA155 synthesis in breast cancer. *Indian J Biochem Biophys*, 57 (2020) 389.

43 Ganeshpurkar A, Singh R, Gore PG, Kumar D, Gutti G, Kumar A & Singh SK, Structure-based screening and molecular dynamics simulation studies for the identification of potential acetylcholinesterase inhibitors. *Mol Simul*, 46 (2020) 169.

44 Daina A & Zoete V, A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem*, 11 (2016) 1117.