Elaborate ligand-based modeling coupled with QSAR analysis and in silico screening reveal new potent acetylcholinesterase inhibitors

Sawsan Abuhamdah · Maha Habash · Mutasem O. Taha

Abstract Inhibition of the enzyme acetylcholinesterase (AChE) has been shown to alleviate neurodegenerative diseases prompting several attempts to discover and optimize new AChE inhibitors. In this direction, we explored the pharmacophoric space of 85 AChE inhibitors to identify high quality pharmacophores. Subsequently, we implemented genetic algorithm-based quantitative structure–activity relationship (QSAR) modeling to select optimal combination of pharmacophoric models and 2D physicochemical descriptors capable of explaining bioactivity variation among training compounds ($r_{68}^2 = 0.94$, F-statistic = 125.8, $r_{LOO}^2 = 0.92$, r_{PRESS}^2 against 17 external test inhibitors = 0.84). Two orthogonal pharmacophores emerged in the QSAR equation suggesting the existence of at least two binding modes accessible to ligands within AChE binding pocket. The successful pharmacophores were comparable with crystallographically resolved AChE binding pocket. We employed the pharmacophoric models and associated QSAR equation to screen the national cancer institute list of compounds. Twenty-four low micromolar AChE inhibitors were identified. The most potent gave IC_{50} value of 1.0 µM.

Keywords Acetylcholinesterase inhibitors · Pharmacophore modeling · Quantitative structure–activity relationship · In silico screening

S. Abuhamdah
Department of Biopharmaceutics and Clinical Pharmacy, Faculty of Pharmacy, The University of Jordan, Amman, Jordan

M. Habash
Department of Pharmaceutical Chemistry and Pharmacognosy, Faculty of Pharmacy, Applied Science University, Amman, Jordan

M. O. Taha (✉)
Drug Discovery Unit, Department of Pharmaceutical Sciences, Faculty of Pharmacy, The University of Jordan, Amman, Jordan
e-mail: mutasem@ju.edu.jo

Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that primarily affects the elderly population and is considered to be responsible for the majority of dementia cases in people aged 65 or older [1]. This disease is characterized by numerous symptoms such as memory and language impairment, cognitive dysfunction and behavioral disturbances (i.e., depression, agitation and psychosis), which become progressively more severe [2]. Currently there is no cure for this disorder, thus there is a real need for novel molecular templates for AD therapy [1–3].

The fact that this neuropathology is associated with central cholinergic deficit [4] suggests that it can be ameliorated by increasing CNS acetylcholine (ACh) concentrations through inhibiting the metabolic enzyme acetylcholinesterase (AChE) [5]. In fact all current pharmacological treatments of AD, i.e., tacrine, donepezil, rivastigmine and galantamine, are cholinesterase inhibitors. However, most of these present many side effects prompting continuous interest in developing new AChE inhibitors [6–8].

Acetylcholinesterase (EC 3.1.1.7) is a serine protease that hydrolyzes the neurotransmitter ACh with high catalytic activity [11, 12]. AChE is found mainly at neuromuscular junctions and cholinergic brain synapses, where its activity serves to terminate synaptic transmission. It

belongs to carboxyl esterase family of enzymes [6–10]. X-ray structures of AChE co-crystallized with various ligands provided insights into the essential structural elements and motifs central to its catalytic mechanism and mode of ACh processing. The active site of AChE comprises 2 subsites—the anionic site and the esteratic subsite. [13–19] The anionic subsite accommodates the positive quaternary amine of ACh as well as other cationic substrates and inhibitors. Interestingly, the cationic moieties of different substrates are not bound by a negatively-charged amino acid in the enzymatic anionic site, but rather by interaction with 14 aromatic residues that line the gorge leading to the active site [15, 17, 20–23].

Inhibition of AChE leads to accumulation of ACh in the synaptic cleft and results in neuroactivation [17]. However, irreversible inhibitors of AChE (e.g., organophosphates) are rather toxic [24], while slowly reversible inhibitors (e.g., carbamates esters) have been successfully used for medical purposes (e.g., physostigmine for the treatment of glaucoma) [25–27].

However, reversible inhibitors seem to be most desirable for treating AD with minimal toxic side effects. They occupy the esteratic site for short periods of time (seconds to minutes) and are used to treat a range of central nervous system diseases. For example, tetrohydroaminoacridine (THA) and donepezil are FDA-approved to improve cognitive function in AD. Rivastigmine is also used to treat Alzheimer's and Lewy body dementia, and pyridostigmine bromide is used to treat myasthenia gravis [25–27].

The continued interest in designing new reversible AChE inhibitors and lack of ligand-based computer-aided drug discovery efforts prompted us to explore the possibility of developing ligand-based three-dimensional (3D) pharmacophore (s) integrated within self-consistent quantitative structure–activity relationship (QSAR) model. The pharmacophore model(s) can be used as 3D search query to discover new AChE inhibitory scaffolds that can be used as new leads for development into anti-AD drugs.

We constructed hundreds of reasonable binding hypotheses for AChE inhibitors by using CATALYST package [28]. Subsequently, QSAR modeling was used as competition arena to select the best orthogonal binding pharmacophores and combine them with other molecular descriptors to yield reasonable predictive QSAR model capable of explaining bioactivity variation within a large collection of AChE inhibitors. The selected pharmacophores were validated by evaluating their abilities to classify a list of compounds as active or inactive through receiver-operating characteristic (ROC) curves. Subsequently, they were employed to screen the national cancer institute (NCI) list of compounds.

We previously reported the use of this innovative approach towards the discovery of new inhibitory leads against glycogen synthase kinase-3 [27], bacterial MurF [29], protein tyrosine phosphatase [30], DPP IV [31], hormone sensitive lipase [32], β-secretase [33], influenza neuraminidase [34], cholesteryl ester transfer protein [35], CDK1 [36], Heat Shock Protein [37], glycogen phosphorylase [38], Rho Kinase [39], nitric oxide synthase (iNOS) inhibitors [40], Ca^{2+}/calmodulin-dependent protein kinase II [41], fungal N-myristoyl transferase [42], renin inhibitory [43], and peroxisome proliferator-activated receptor γ activators [44].

Materials and methods

Molecular modeling

Software and hardware

Pharmacophore and QSAR modeling studies were performed using CATALYST (HYPOGEN module) [28], CERIUS2 [45] and Discovery Studio [46] software suites. Structure drawing was performed employing ChemDraw Ultra 7.0 [47].

The performed modeling workflow involves several subsequent steps: (1) Drawing the chemical structures of the training compounds using ChemDraw software package. (2) Generation of multi-conformations for training compounds using CONFIRM module of CATALYST. (3) Pharmacophore exploration performed using HYPOGEN module of CATALYST. (4) QSAR analysis performed using QSAR and DESCRIPTORS modules within CERIUS software package. (5) Validating the capacity of the QSAR-selected pharmacophores as three-dimensional search queries by plotting their ROC curves. This step was done employing the "Best flexible search" option implemented in CATALYST followed by plotting the output using one of our own scripts written within MATLAB environment. (6) The selected pharmacophores were used as search queries to find promising in silico hits. It remains to be mentioned that we used Ligandfit docking engine within Discovery Studio suit for docking purposes and Discovery Studio environment to visualize and report our findings.

Data Set

The structures of 85 AChE inhibitors (1–85, Table A under supplementary material) were collected from recently published literature [48–51]. Although the inhibitors were gathered from eight separate articles, they were bioassayed employing the same methodology. The bioactivities were expressed as the concentrations of the test compounds that inhibited the activity of AChE by 50 % (IC_{50} in nM). The logarithms of IC_{50} values were used in QSAR and pharmacophores analyses in order to correlate the data linear to the free energy change.

The two-dimensional (2D) chemical structures of the inhibitors were imported into CATALYST, converted into corresponding standard 3D structures and energy minimized to the closest local minimum using the molecular mechanics CHARMm force field in CATALYST. The conformational space of each inhibitor was explored adopting the "best conformer generation" option within CATALYST [28]. A conformational ensemble was generated for each training molecule with an energy threshold of 20 kcal/mol from the local minimized structure, and a maximum limit of 250 conformers per molecule.

Automatic generation and assessment of pharmacophoric hypotheses via catalyst

All 85 molecules with their associated conformational models were rearranged into a worksheet. The biological data of the inhibitors were reported with an "Uncertainty" value of 3, which means that the actual bioactivity of a particular inhibitor is assumed to be within an interval ranging from one-third to three-times the reported bioactivity value of that inhibitor [56, 57]. Subsequently, four structurally diverse training subsets were carefully selected from the collection for pharmacophore modeling: sets **I**, **II**, **III** and **IV** (Table B under supplementary material).

The selected training sets were utilized to conduct 32 modeling runs to explore the pharmacophoric space of AChE inhibitors (Table C under Supplementary Materials). The exploration process included altering interfeature spacing parameter (1 and 3 Å) and the maximum number of allowed features in the resulting pharmacophore hypotheses and in presence or absence of exclusion volumes as in Table C under Supplementary Materials.

Pharmacophore modeling employing CATALYST proceeds through three successive phases: the constructive phase, subtractive phase and optimization phase (see CATALYST Modeling Algorithm under section SM-1 in Supplementary Materials) [28–44, 56, 57]. When generating binding hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: Weight cost, Error cost and Configuration cost (see CATALYST Cost Analysis in Assessment of Generated Binding Hypotheses in section SM-2 under Supplementary Materials).

Additionally, CATALYST cross-validates pharmacophore models using Cat-Scramble algorithm. This validation procedure is based on Fischer's randomization test [61]. In this validation test; we selected a 95 % confidence level that instructs CATALYST to generate 19 random spreadsheets by the Cat-Scramble command. Subsequently, CATALYST-HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses. Success in generating pharmacophores of

comparable cost criteria to those produced by the original unscrambled data reduces the confidence in the training compounds and the unscrambled original pharmacophore models [28, 61, 62].

QSAR modeling

The successful models (259) were clustered into 25 groups utilizing the hierarchical average linkage method available in CATALYST. Subsequently, individual representatives were arbitrary selected to represent their corresponding clusters in subsequent QSAR modeling. Table 1 shows the statistical criteria of representative pharmacophores including their pharmacophoric features, success criteria and differences from corresponding null hypotheses (see CATALYST Cost Analysis in Assessment of Generated Binding Hypotheses in section SM-2 under Supplementary Materials). The table also shows the corresponding Cat. Scramble confidence levels for each representative pharmacophore.

For subsequent QSAR modeling, a subset of 68 compounds from the total list of collected inhibitors (**1–85**, Table A under Supplementary Materials) was utilized as a training set for QSAR. The remaining 17 molecules (ca. 20 % of the dataset) were employed as an external test subset for validating the QSAR models. The test molecules were selected by ranking the collected inhibitors (**1–85**, Table A in Supplementary Materials) according to their IC_{50} values, and then selecting every fifth compound for the test set starting from the high-potency end.

The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformer representations in SD format. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module of CERIUS2. The calculated descriptors included various simple and valence connectivity indices, electro-topological state indices and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, energies of the lowest and highest occupied molecular orbitals, etc.) [45]. Additionally, the training compounds were fitted (using the Best-fit option in CATALYST) against the representative pharmacophores (25 models, Table 1), and their fit values were added as additional descriptors. The fit value for any compound is obtained automatically via equation (D) under SM-1 in Supplementary Materials [28].

Genetic function approximation (GFA) was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, i.e., multiple linear regression modeling (MLR). The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF) [45].

Table 1 Statistical success criteria of representative pharmacophore hypotheses

RUN[a]	Hypotheses[b]	Features	Cost				R[d]	Cat. scramble[e] (%)
			Config.	Total	Null	Residual[c]		
1	6[f]	HBA, Hbic, 2xRingArom, 2xEV[g]	6.2	145.8	252.9	107.1	0.82	90
	7	HBA, Hbic, RingArom, PosIon, EV[g]	6.2	146.9	252.9	106	0.81	90
5	2	HBA, Hbic, RingArom, PosIon	6.2	149.9	252.9	103	0.79	95
	9	HBD, Hbic, RingArom, HBA	6.2	166.8	252.9	86.1	0.73	95
10	3	HBD, Hbic, RingArom, PosIon, 8xEV	16.2	129.9	265.9	136	0.97	95
	6	HBD, Hbic, RingArom, PosIon, 8xEV	16.2	132.4	265.9	133.5	0.96	95
	8	HBD, Hbic, RingArom, HBA, 9xEV	16.2	133.2	265.9	132.7	0.97	95
	9	HBD, Hbic, RingArom, HBA, 8xE	16.2	133.3	265.9	132.6	0.96	95
	10	HBD, Hbic, RingArom, HBA, 4xE	16.2	133.4	265.9	132.5	0.95	95
13	7	HBD, 2xHbic, RingArom	16.8	140.6	265.9	125.3	0.93	95
17	1	HBD, 2xHbic, RingArom, 7xEV	16.3	78.4	167.1	88.7	0.99	95
	3	HBD, 2xHbic, RingArom, 6xEV	16.3	81.6	167.1	85.5	0.97	95
	10	HBA, 2xHbic, RingArom, PosIon	16.3	87.4	167.1	79.7	0.94	95
18	2	HBA, 2xHbic, RingArom, PosIon	15.5	81.1	167.1	86	0.97	95
19	1	HBA, 2xHbic, RingArom, PosIon	13.0	78.4	167.1	88.7	0.977	90
25	3	HBD, 2xHbic, RingArom	17.8	153	360.9	207.9	0.97	95
	5	HBD, 2xHbic, RingArom	17.8	154.2	360.9	206.7	0.96	95
26	7	HBD, 2xHbic, RingArom	16.7	152.9	360.9	208	0.96	95
	8	2xHBD, Hbic, HBA	16.7	153.6	360.9	207.3	0.96	95
	10	HBD, 2xHbic, RingArom	16.7	154.4	360.9	206.5	0.96	95
28	6	HBA, 2xHbic, RingArom, PosIon, 3xEV	13.7	151.6	360.9	209.3	0.96	95
29	9	2xHBD, Hbic, RingArom	17.8	156.4	360.9	204.5	0.96	95
30	8	HBD, 2xHbic, RingArom	16.8	153.3	360.9	207.6	0.96	95
31	2	HBA, HBD, 2xHbic, PosIon	14.6	146.8	360.9	214.1	0.97	95
	6	HBD, 2xHbic, RingArom, PosIon	14.6	151.2	360.9	209.7	0.96	95

[a] Correspond to runs in Table C under supplementary materials

[b] High ranking representative hypotheses (in their corresponding clusters, see "QSAR modeling" section)

[c] Difference between total cost and the cost of the corresponding null hypotheses (see section SM-2 under Supplementary Materials)

[d] Correlation coefficients between pharmacophore-based bioactivity estimates (calculated from equation (C) in SM-1 under Supplementary Materials) and bioactivities of corresponding training compound (subsets in table B under supplementary material)

[e] Fisher confidence level calculated employing the Cat. Scramble methods

[f] Ranking of hypotheses is as generated by CATALYST in each automatic run

[g] Bolded pharmacophores appeared in the best QSAR equations

Diagnostic trials suggested the following optimal GFA parameters: explore linear, quadratic and spline equations at mating and mutation probabilities of 50 %; population size = 500; number of genetic iterations = 30,000 and LOF smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 5 to 25 explanatory terms.

All QSAR models were validated employing leave one-out cross-validation (r_{LOO}^2), bootstrapping (r_{BS}^2) and predictive r^2 (r_{PRESS}^2) calculated from the test subsets. The predictive r_{PRESS}^2 is defined as:

$$r_{PRESS}^2 = SD - PRESS/SD \qquad (1)$$

Where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules, PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.

Receiver operating characteristic (ROC) curve analysis

QSAR-selected pharmacophore models (i.e., Hypo6/1 and Hypo 7/1) were validated by assessing their abilities to

selectively capture diverse AChE inhibitors from a large testing list of actives and decoys.

The testing list was prepared as described by Verdonk and co-workers [64, 65]: Decoys were selected by assessing the distance (D) between any two molecules (e.g., i and j) based on three one-dimensional properties: (1) the number of hydrogen-bond donors (NumHBD); (2) number of hydrogen-bond acceptors (NumHBA) and (3) count of nonpolar atoms (NP, defined as the summation of Cl, F, Br, I, S and C atoms in a particular molecule). For each active compound in the test set, the distance to the nearest other active compound is assessed by their Euclidean Distance (Eq. (2)):

$$D(i,j) = \sqrt{(NumHBD_i - NumHBD_j)^2 + (NumHBA_i - NumHBA_j)^2 + (NP_i - NP_j)^2} \qquad (2)$$

The minimum distances are then averaged over all active compounds (D_{min}). Subsequently, for each active compound in the test set, around 25 decoys were randomly chosen from the ZINC database [66]. The decoys were selected in such a way that they did not exceed D_{min} distance from their corresponding active compound.

To diversify active members in the list, we excluded any active compound having zero distance [D(i, j)] from other active compound(s) in the test set. Active testing compounds were defined as those possessing anti-AChE IC_{50} values ranging from 1.3 nM to 9.5 μM. The test set included 35 active compounds and 868 ZINC decoys.

The testing list (903 compounds) was screened by each particular pharmacophore employing the "Best flexible search" option implemented in CATALYST, while the conformational spaces of the compounds were generated employing the "Fast conformation generation option" implemented in CATALYST. Compounds missing one or more features were discarded from the hit list. In-silico hits were scored employing their fit values as calculated by Eq. (D) in Supplementary Materials.

The ROC curve analysis describes the sensitivity (Se or true positive rate, Eq. (3)) for any possible change in the number of selected compounds (n) as a function of (1-Sp). Sp is defined as specificity or true negative rate (Eq. (4)) [63, 65]

$$Se = \frac{Number\ of\ Selected\ Actives}{Total\ Number\ of\ Actives} = \frac{TP}{TP + FN} \qquad (3)$$

$$Sp = \frac{Number\ of\ Discarded\ Inactives}{Total\ Number\ of\ Inactives} = \frac{TN}{TN + FP} \qquad (4)$$

where, TP is the number of active compounds captured by the virtual screening (VS) method (true positives), FN is the number of active compounds discarded by the VS

method, TN is the number of discarded decoys (presumably inactives), while FP is the number of captured decoys (presumably inactive) [63, 65, 66].

If all molecules scored by a VS protocol with sufficient discriminatory power are ranked according to their score (i.e., fit values), starting with the best-scored molecule and ending with the molecule that got the lowest score, most of the actives will have a higher score than the decoys. Since some of the actives will be scored lower than decoys, an overlap between the distribution of active molecules and decoys will occur, which will lead to the prediction of false positives and false negatives. [63, 65]. The selection of one score value as a threshold strongly influences the ratio of actives to decoys and therefore the validation of a VS method. The ROC curve method avoids the selection of a threshold by considering all Se and Sp pairs for each score threshold [63, 65]. A ROC curve is plotted by setting the score of the active molecule as the first threshold. Afterwards, the number of decoys within this cutoff is counted and the corresponding Se and Sp pair is calculated. This calculation is repeated for the active molecule with the second highest score and so forth, until the scores of all actives are considered as selection thresholds.

The ROC curve representing ideal distributions, where no overlap between the scores of active molecules and decoys exists, proceeds from the origin to the upper-left corner until all the actives are retrieved and Se reaches the value of 1. In contrast to that, the ROC curve for a set of actives and decoys with randomly distributed scores tends towards the Se = 1-Sp line asymptotically with increasing number of actives and decoys [63, 65]. The success of a particular VS workflow can be judged from the following criteria (shown in Table 3):

(1) Area under the ROC curve (AUC) [63, 65, 67]. In an optimal ROC curve an AUC value of 1 is obtained; however, random distributions cause an AUC value of 0.5. VS that performs better than a random discrimination of actives and decoys retrieve an AUC value between 0.5 and 1, whereas an AUC value lower than 0.5 represents the unfavorable case of a VS method that has a higher probability to assign the best scores to decoys than to actives[63, 65].

(2) Overall Accuracy (ACC) describes the percentage of correctly classified molecules by the screening protocol. Testing compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) [63, 65, 67].

(3) Overall specificity (SPC): describes the percentage of discarded inactives by the particular VS workflow. Inactive test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values [63, 65, 67].

(4) Overall True Positive Rate (TPR or overall sensitivity): describes the fraction percentage of captured actives from the total number of actives. Active test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values [63, 65, 67].

(5) Overall False Negative Rate (FNR or overall percentage of discarded actives): describes the fraction percentage of active compounds discarded by the VS method. Discarded active test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values [63, 65, 67].

In silico screening for new AChE inhibitors

Hypo6/1 and Hypo7/1 were employed to screen the NCI structural database. The screening was done employing the "Best Flexible Database Search" within CATALYST. Captured hits were filtered according to Lipinski's [68] and Veber's [69] rules. Remaining hits were fitted against the two pharmacophores using the "best fit" option within CATALYST. The fit values together with the relevant molecular descriptors of each hit were substituted in the optimal QSAR Eq. (5). The highest ranking molecules based on QSAR predictions were acquired and tested in vitro. Table 5 shows active hits and their QSAR-predictions and experimental bioactivities.

In vitro experimental studies

Materials

All of the chemicals were purchased from Sigma–Aldrich (St. Louis, MO, USA) including electric eel AChE (type-VI-S, EC 3.1.1.7), acetylthiocholine iodide (ATCI), 5,5′dithiobis-(2-nitrobenzoic acid) (DTNB), Tris–HCl, bovine serum albumin (BSA), NaCl, MgCl$_2$.6H$_2$O, standard inhibitor galanthamine (G1660), water and dimethyl sulfoxide (DMSO) for bioanalysis. Tested hits were kindly freely provided by the NCI.

Preparation of hit compounds for in vitro assay

The tested compounds were provided as dry powders in variable quantities (5–10 mg). They were initially dissolved in DMSO to give stock solutions of 100 μM.

Subsequently, they were diluted to the required concentrations with Tris buffer (pH 7.4) for enzymatic assay.

Quantification of AChE inhibitory activity in a spectrophotometric assay

Acetylcholinesterase activities were measured through Ellman's colorimetric method with a slight modification [70]. In a typical run, AChE was dissolved in Tris–HCl buffer (50 mM, pH 8.0) to give a final solution of 0.22 U/mL. Subsequently, 1 μL of the enzyme solution was added to each well of a 96-well plate. Thereafter, a predetermined volume of each tested compound solution was added to each well to yield final concentrations of 0.1, 1, 10, 100 μM. Subsequently, Tris–HCl buffer was added (pH 8.0, 50 mM with 0.1 % w/v bovine serum albumin) to each well to reach a total of 40 μL and allowed to stand for 10 min at room temperature. Subsequently, DTNB (50 μL, 0.075 mM) and ATCI (10 μL, 1.5 mM) were added to wells. Color development was measured spectrophotometrically at λ 412 nm using microplate reader (BioTek ELx800, USA) at a rate of one measurement per minute over 15 min period. Positive (galanthamine) and negative (no inhibitors) controls were tested. The reaction rates were compared and the percent inhibition due to the presence of tested compounds was calculated. All samples were assayed in at least duplicate measurements. In general, the amount of DMSO was kept below 1 % in the assay.

Results and discussion

CATALYST-HYPOGEN utilizes a collection of molecules with activities ranging over a number of orders of magnitude for automatic pharmacophore construction. It generates binding hypotheses (pharmacophores) by using the geometric localization of the chemical features present in the molecules to explain the variability of bioactivity. CATALYST-HYPOGEN defines a 3D array of a maximum of five chemical features common to active training molecules that provides relative alignment for each input molecule consistent with binding to certain proposed common binding site. The chemical features can be hydrogen bond donors and acceptors (HBD and HBA), aliphatic and aromatic hydrophobes (Hbic), positive and negative ionizable (PosIon and NegIon) groups and aromatic planes (RingArom). CATALYST pharmacophores have been used as 3D queries for database searching and in 3D-QSAR studies [27, 29–40].

Data mining and conformational coverage

The literature was surveyed to collect many structurally diverse AChE inhibitors (**1–85**, see Table A under

supplementary materials) [48–55]. The conformational space of each inhibitor was extensively sampled utilizing the poling algorithm of CATALYST [55]. Proper conformational exploration is necessary for pharmacophore generation and pharmacophore-based search procedures as both are known for their sensitivity to inadequate conformational sampling [34].

Exploration of AChE pharmacophoric space

The training inhibitors were selected in such a way that they were assayed by the same procedure (compounds **1–85**, Table A in Supplementary Materials). Obviously, statistical consistency necessitates that QSAR and pharmacophore modeling are based on training compounds assayed by a single bioassay procedure [27, 29–40].

The pharmacophoric space of AChE inhibitors was explored through 32 pharmacophore generation automatic runs performed on four carefully selected training subsets:

(A)

(B)
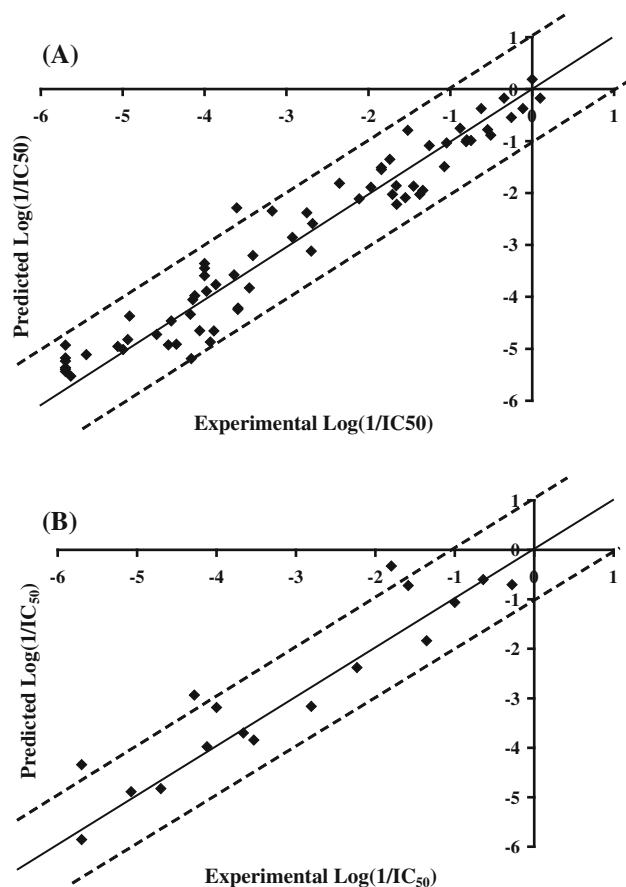
Fig. 1 Experimental versus (**a**) fitted (68 compounds, $r^2_{LOO} = 0.917$), and (**b**) predicted (17 compounds, $r^2_{PRESS} = 0.0841$) bioactivities calculated from the best QSAR model Eq. (5). The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1.0 log error margins

I, **II**, **III** and **IV** (Table B under supplementary Materials). The training compounds were selected to guarantee wide structural diversity with bioactivities extended over more than 3.5 logarithmic cycles. To ensure sufficient molecular diversity within training subsets, member compounds were selected in such a way that each structural cluster of the collected compounds was sampled at least once in each training subset. However, some compounds were repeatedly selected in training subsets because of their critical pharmacophoric features. Training subsets were selected in such a way that differences in AChE inhibitory activities among their member compounds are primarily attributable to the presence or absence of pharmacophoric features [e.g., HBA, HBD, Hbic or ring aromatic (RingArom)] rather than steric shielding and/or bioactivity-enhancing or -reducing auxiliary groups (e.g., electron donating or withdrawing groups). A special emphasis was given to the structural diversity of the most-active compounds in each training subset because of their significant influence on the extent of the evaluated pharmacophoric space during the constructive phase of HYPOGEN algorithm (see CATALYST Modeling Algorithm under section SM-1 in Supplementary Materials) [28, 57–60, 71].

HYPOGEN was instructed to explore only 4- and 5-featured pharmacophores and ignore models of lesser number of features (as shown in Table C in Supplementary Materials). The advantage of this restriction is to narrow the investigated pharmacophoric space while allowing good representation of the feature-rich nature of AChE inhibitors. We previously implemented similar pharmacophore exploration strategies against a multitude of targets [27–44].

Eventually, 259 pharmacophore models resulted from 32 automatic HYPOGEN runs. Fortunately, all generated pharmacophores illustrated ≥90 % Cat.Scramble significance (see "Automatic generation and assessment of pharmacophoric hypotheses via catalyst" section) [28, 61, 62]. These were subsequently clustered and representative models were arbitrary selected from each cluster (total of 25 models, Table 1) to represent their clusters in subsequent QSAR modeling.

QSAR modeling

Clearly from Table 1, representative pharmacophore models shared comparable binding features and excellent statistical criteria. Emergence of numerous statistically comparable pharmacophore hypotheses suggests the ability of AChE ligands to assume multiple binding modes within the binding pocket. Accordingly, it is rather hard to select a particular binding pharmacophore as a single representative of ligand binding. This point combined with the fact

(A)

(B)

(C)

(D)
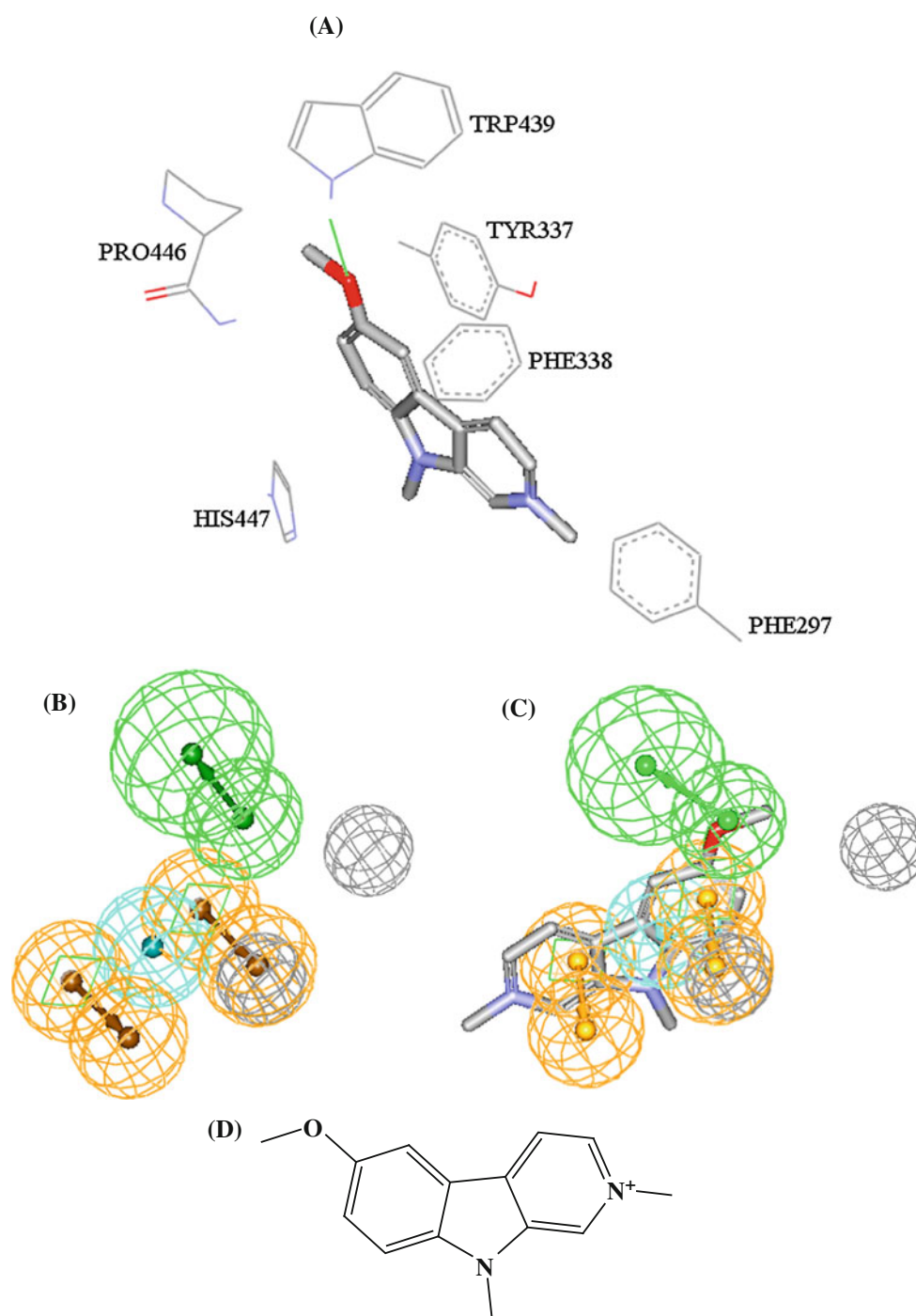
Fig. 2 a Docked structure of training compound **32** ($IC_{50} = 1.3$ nM, Table A under Supplementary Materials) into AChE (PDB code: 1B41, resolution 2.76 Å, the *green li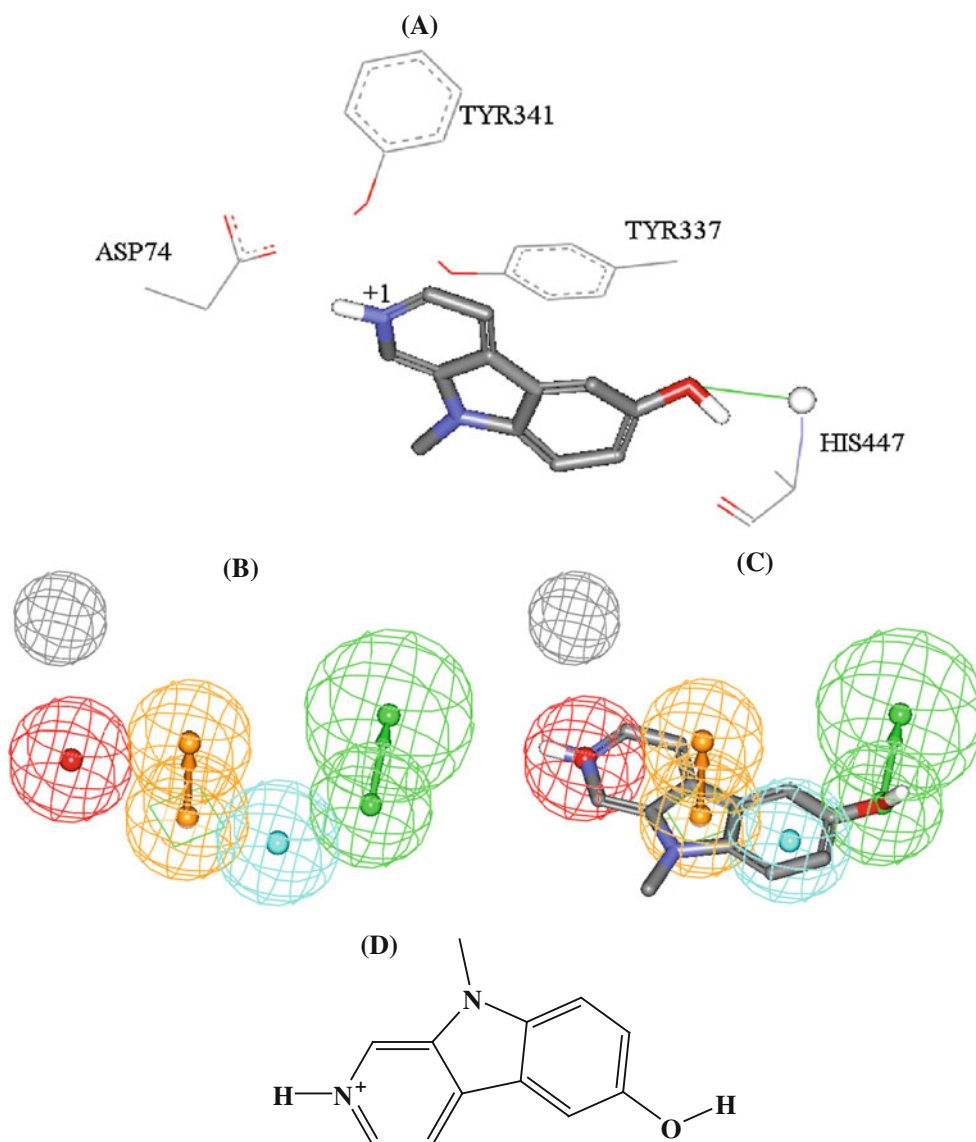ne* represents hydrogen bonding). **b** Pharmacophoric features of Hypo6/1: HBA as *green vectored spheres*, Hbic as *light blue spheres*, RingArom as vectored orange spheres. **c** Hypo6/1 fitted against **32**. **d** Chemical structure of **32**

that pharmacophoric models are limited by steric shielding and bioactivity-enhancing or reducing auxiliary groups [27, 29–44], prompted us to employ classical QSAR analysis to search for the best combination of pharmacophore(s) and other 2D descriptors capable of explaining bioactivity variation across the whole list of collected inhibitors (**1–85**,

Table A under Supplementary Materials). We employed GFA and MLR QSAR (GFA–MLR–QSAR) analysis to search for an optimal QSAR equation(s) [77, 80].

The fit values obtained by mapping representative hypotheses (25 models) against collected AChE inhibitors (**1-25**, Table A under Supplementary Materials) were

Fig. 3 a Docked structure of training compound **35** (IC$_{50}$ = 1.8 nM, Table A under Supplementary Materials) into AChE (PDB code: 1B41, resolution 2.76 Å) binding pocket. **b** Pharmacophoric features of Hypo7/1: HBA as *green vectored spheres*, Hbic as *light blue spheres*, RingArom as *vectored orange spheres*, PosIon as *red spheres*. **c** Hypo1/7 fitted against **35**. **d** Chemical structure of **35**

enrolled, together with around 100 other physicochemical descriptors, as independent variables in GFA-MLR-QSAR analysis (see "QSAR modeling" section) [27–38, 45, 64]. We arbitrarily selected 17 molecules (marked with double asterisks in Table A under Supplementary Materials) as external test molecules for validating the QSAR models. All QSAR models were cross-validated automatically using the leave-one-out cross-validation in CERIUS2 [45, 64].

Equation (5) shows the details of the optimal QSAR model. Figure 1 shows the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing inhibitors.

$$
\begin{aligned}
\text{Log}\ (1/\text{IC}_{50}) = &-3.87 - 0.17(\text{Molecular Solubility}) \\
&+ 0.60(\text{Number of Rotatable Bonds}) \\
&- 1.07(\text{PHI}) + 2.92(\text{JursRNCG}) \\
&+ 0.27(\text{Hypo6/1}) + 0.08(\text{Hypo7/1})
\end{aligned}
$$

$$
r^2_{68} = 0.94,\ r^2_{\text{LOO}} = 0.92,\ \text{F-statistic} = 125.8,
$$
$$
r^2_{\text{PRESS(17)}} = 0.84
\tag{5}
$$

where, r^2_{68} is the correlation coefficient against 68 training compounds and r^2_{PRESS} is the predictive r^2 determined for the 17 test compounds [45, 64]. The different descriptor

Table 2 Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of Hypo6/1 and Hypo7/1

Model	Definition	Chemical features								
		HBA		RingArom		Hbic	RingArom		EV1	EV1
Hypo6/1[a]	Weights	2.70		2.70		2.70	2.70			
	Tolerances	1.60	2.20	1.60	1.60	1.60	1.60	1.60		
	Coordinates									
	X	−4.59	−4.02	1.55	1.62	−0.88	−2.50	−2.42	−2.43	−7.34
	Y	0.19	0.84	0.09	−2.85	−0.12	−0.13	−3.06	−3.04	−0.77
	Z	−1.48	−4.40	−0.23	−0.85	0.42	0.30	0.33	1.88	1.82

Model	Definition	Chemical features						
		HBA	Hbic		RingArom		PosIon	EV
Hypo1/7[b]	Weights	2.40	2.40		2.40		2.40	
	Tolerances	1.60	2.20	1.60	1.60	1.60	1.60	
	Coordinates							
	X	−4.19	−3.68	-2.38	0.04	0.09	3.42	4.65
	Y	−0.83	−3.58	1.00	0.97	2.32	0.26	−1.06
	Z	0.39	1.60	-0.46	-0.49	2.18	-0.20	2.86

[a] Hypo6/1: the 6th pharmacophore hypothesis generated in the 1st HYPOGEN run (Table 1)

[b] Hypo7/1: the 7th pharmacophore hypothesis generated in the 1st HYPOGEN run (Table 1)

coefficients were auto-scaled. **JursRNCG** is the relative negative charge calculated by dividing the charge of most negative atom by the total negative charge [45]. **PHI** is molecular flexibility index [45]. Hypo6/1 and Hypo7/1 represent the fit values of the training compounds against these two pharmacophores (bolded models in Table 1 and shown in Figs. 2, 3) as calculated from equation (D) under section SM-2 in Supplementary Materials.

The contradictory regression slopes associated with the **number of rotatable bonds** and flexibility index (**PHI**) in Eq. (5) suggest certain complex role played by molecular flexibility in ligand-AChE binding. However, the overall influence of molecular flexibility seems to be negative, i.e., on ligand binding, as evident by the larger negative slope associated with **PHI**. The most probable explanation of this trend is related to the entropic cost of binding. Binding of flexible molecules into AChE binding pocket tend to excise higher entropic cost compared to rigid ligands.

The emergence of Molecular Solubility descriptor in combination with negative slope in Eq. (5) suggests that higher water solubility reduces ligand-AChE affinity. This is not unexpected as ligand hydration generally competes with ligand-receptor binding [74–76].

Interestingly, QSAR Eq. (5) shows **JursRNCG** combined with a relatively pronounced positive regression coefficient suggesting significant ligand-AChE affinity promoting effects by focused electrophilic centers. The most probable explanation of this trend is related to the fact that electrophilic heterocycles π-stack efficiently against complementary electron-rich aromatic side chains of amino-acids within the binding pocket of AChE. The catalytic site of AChE includes several electron-rich aromatic rings belonging to Trp439, Tyr337, Phe338 and Phe297 (as in Figs. 2, 3).

On the other hand, emergence of two orthogonal pharmacophoric models, i.e., Hypo6/1 and Hypo7/1 of cross-correlation $r^2 = 0.16$, in Eq. (5) suggests they represent two complementary binding modes accessible to ligands within the binding pocket of AChE, i.e., one of the pharmacophores explains the bioactivities of some training inhibitors while the other explains the remaining inhibitors. Figures 2C and 3C show Hypo6/1 and Hypo7/1 and how they map **32** (IC$_{50}$ = 1.3 nM) and **35** (IC$_{50}$ = 1.8 nM), respectively. The X, Y, and Z coordinates of the two pharmacophores are given in Table 2. Similar conclusions were reached about the binding pockets of other targets based on QSAR analysis [27, 29–44].

To validate the QSAR-selected pharmacophores, we subjected them to ROC curve analysis. In ROC analysis, the ability of a particular pharmacophore model to correctly classify a list of compounds as actives or inactives is indicated by the area under the curve (AUC) of the corresponding ROC together with other parameters: overall accuracy, overall specificity, overall true positive rate and overall false negative rate (see 2.1.5 receiver operating characteristic curve analysis under Experimental for more details) [63–66]. Table 3 and Fig. 4 show the ROC results of our QSAR-selected pharmacophores. Hypo6/1 and

Hypo7/1 illustrated good overall performances with AUC values of 71.5 and 89.7 % respectively.

To further emphasize the validity of our pharmacophore/QSAR modeling approach, we compared the pharmacophoric features of Hypo6/1 and Hypo7/1 and how they map training compounds 32 and 35, respectively, with optimal docked poses of the two compounds. Docking experiments

Table 3 ROC curve analysis criteria for QSAR-selected pharmacophores

Pharmacophore model	ROC–AUC (%)	ACC (%)	SPC (%)	TPR (%)	FNR (%)
Hypo6/1	71.5	96.3	98.7	33.3	1.2
Hypo7/1	88.9	96.3	98.1	50	1.9

ROC receiver operating characteristic curve, *AUC* area under the curve, *ACC* overall accuracy, *SPC* overall specificity, *TPR* overall true positive rate, *FNR* overall false negative rate
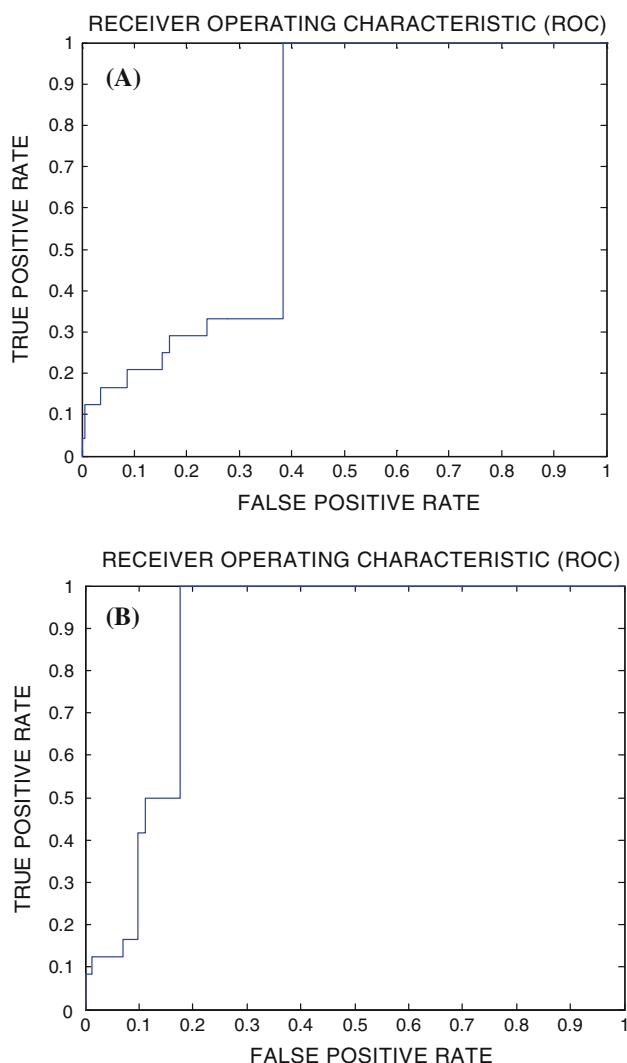
Fig. 4 ROC curves of: **a** Hyp o6/1, **b** Hypo7/1

were conducted employing LigandFit (as implemented in Discovery Studio 2.5) and through default docking settings [71]. Figures 2 and 3 show the pharmacophores, docked poses and corresponding mapped conformers. By comparing the docked pose of 32 (IC_{50} = 1.3 nM) within AChE with the way it fits Hypo6/1 (Fig. 2), one quickly notes the great similarity between the Hypo6/1 and the binding interactions within AChE: Mapping the aromatic methoxy of 32 with hydrogen bond acceptor (HBA) feature in Hypo6/1 (Fig. 2c) corresponds to hydrogen bonding interactions connecting this group with the indole NH of Trp439 (Fig. 2a). Similarly, mapping the pyridinoindole ring system of 32 against two RingArom and a Hbic feature (Fig. 2c) correlates with π-stacking interactions resulting from squeezing the pyridnoindole ring system within the aromatic pouch of the binding pocket comprised from the aormatic side chains of Phe297, Phe338 Tyr337, Trp439 and His447 (Fig. 2a).

Similar analogy can be concluded by comparing the docked pose of 35 (IC_{50} = 1.8 nM) with fitting against Hypo7/1 (Fig. 3): Mapping the protonated pyridinium ion of 35 against PosIon feature in Hypo7/1 (Fig. 3c) corresponds to electrostatic attraction connecting this positive group with the carboxylate anion of Asp74 (Fig. 3a). Likewise, mapping the phenolic OH of 35 against HBA feature in Hypo7/1 (Fig. 3c) seems to correspond to hydrogen bonding interaction connecting the same phenolic group with the amidic NH of His447 (Fig. 3a). Finally, mapping the pyridinoindole electron-deficient ring system of 35 against Hbic and RingArom features in Hypo7/1 (Fig. 3c) correlates with π-stacking interactions against the electron-rich aromatic side chains of Tyr337 and Tyr341 (Fig. 3a).

Clearly from the above discussion, Hypo6/1 and Hypo7/1 represent two valid binding modes assumed by ligands within AChE catalyst site. Interestingly, these pharmacophore models point to limited number of critical interactions required for high ligand-AChE affinity in each of the binding modes. In contrast, docked complexes reveal many bonding interactions without highlighting critical ones. Figures 2a and 3a show only interactions corresponding to pharmacophoric features while other binding interactions were hidden for clarity.

Table 4 Numbers of captured hits by Hypo6/1, Hypo7/1

Pharmacophore models			
3D Database[a]	Post screening filtering[b]	Hypo6/1	Hypo7/1
NCI	Before	46102	14692
	After	11282	4319

[a] NCI: national cancer institute list of available compounds (238,819 structures)

[b] Using Lipinski's and Veber's rules

Table 5 In silico hits, their fit values against (Hypo6/1, Hypo7/1), corresponding QSAR estimates from Eq. (5) and in vitro anti-cholinesterase activity
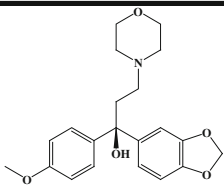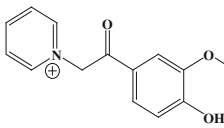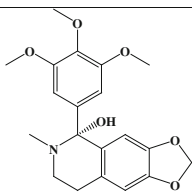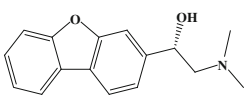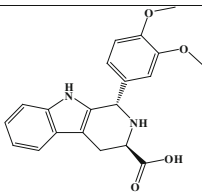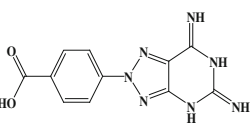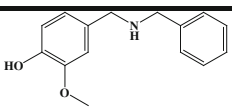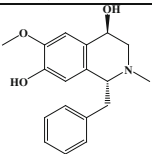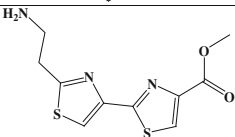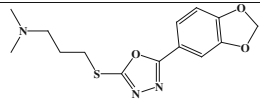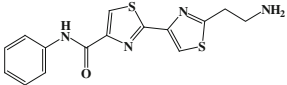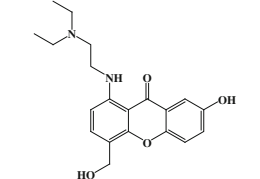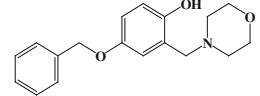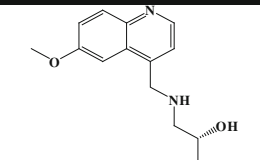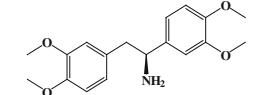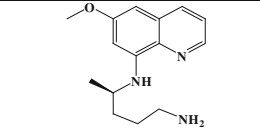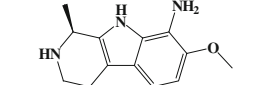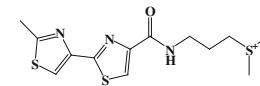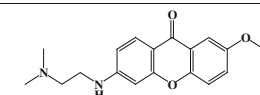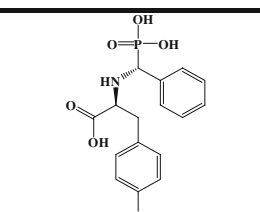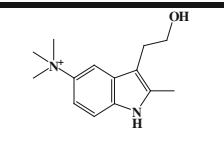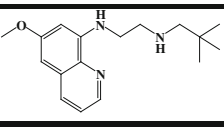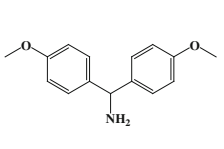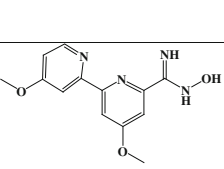
Hit compound	NCI code	Structure	Fit value against		QSAR predictions		Experimental	
			Hypo6/1[a]	Hypo7/1[a]	Log(1/IC$_{50}$)	IC$_{50}$ (μM)	IC$_{50}$ (μM)[b]	Hill slope[c]
86[d]	NSC34198		6.4	6.8	−1.64	43.6	2.00	0.89
87	NSC34674		6.4	7.4	−0.58	3.8	2.51	0.76
88	NSC36355		6.0	6.8	−1.79	62.1	2.51	0.76
89[d]	NSC94759		9.0	7.1	−0.16	1.4	1.00	0.35
90	NSC96603		6.4	6.5	−0.96	9.1	2.51	1.31
91	NSC115883		8.3	6.4	−2.06	114.5	6.34	0.93
92	NSC126255		6.5	7.0	−0.46	2.9	6.30	0.61
93	NSC131665		6.3	7.0	−1.76	58.1	10	0.68
94	NSC143123		8.1	6.3	0.039	0.91	3.98	0.51

Table 5 continued

95	NSC299579		**8.8**	**6.4**	**−1.05**	**11.3**	**10**	**0.60**
96	NSC302667		**8.0**	**6.2**	**−0.50**	**3.2**	**7.9**	**0.60**
97	NSC355357		**6.0**	**7.2**	**−0.91**	**8.1**	**1.77**	**0.97**
98	NSC401623		**6.6**	**6.5**	**−2.32**	**207.7**	**1.77**	**1.34**
99	NSC2457		**0**	**7.8**	**−1.71**	**51.0**	**39.8**	**0.42**
100	NSC26679		**3.2**	**6.4**	**−2.11**	**129.5**	**10**	**0.49**
101	NSC27296		**0**	**7.1**	**−1.34**	**22.0**	**25.11**	**0.62**
102[d]	NSC356217		**0**	**7.4**	**−2.09**	**119.3**	**2.00**	**1.4**
103	NSC359279		**8**	**4.9**	**−1.19**	**15.5**	**2.5**	**1.8**
104[d]	NSC376356		**6.5**	**6.6**	**−2.13**	**134.8**	**2.00**	**1.2**
105	NSC377438		**5.4**	**6.9**	**−2.19**	**155.6**	**2.5**	**1.8**

Table 5 continued

106	NSC405606		0	7.5	−0.31	2.0	3.16	1.0
107	NSC406302		0	7.3	−1.02	10.5	3.98	0.94
108	NSC505712		6.2	5.8	−1.0	10.1	5.62	0.99
109	NSC615148		7.1	6.7	−0.40	2.5	2.51	1.32

[a] Best-fit values calculated by equation (D) under SM-2 in Supplementary Materials

[b] Bioactivity values are the average of at least duplicate measurements. The corresponding dose–response curves are shown in figure I under Supplementary Materials

[c] Hill Slopes were calculated by GraphPad Prism 5.0

[d] The structures of these compounds were validated by proton and 13C NMR spectroscopy as well as mass spectrometry. The corresponding spectrums are shown in figures II to IX in the supplementary materials

In-silico screening and subsequent in vitro evaluation

Hypo6/1 and Hypo7/1 were employed as 3D search queries against the NCIs list of compounds (NCI, 238,819 structures). Subsequently, captured hits were filtered using Lipinski's [68] and Veber's [69] rules such that the remaining hits are more amenable for optimization into promising leads. Table 4 summarizes the numbers of captured hits by each pharmacophore before and after filtration based on Lipinski's and Veber's rules.

The remaining hits were fitted against Hypo6/1 and Hypo7/1 and their fit values, together with other relevant molecular descriptors, were substituted in QSAR equation (5) to predict their anti-AChE bioactivities. The highest-ranking hits were evaluated in vitro against recombinant AChE (Sigma, USA).

Initially, tested hits were screened at 100 μM concentrations, subsequently; compounds showing anti-AChE percentages exceeding 50 % at 100 μM were further assessed to determine their IC$_{50}$ values at 10 and 1 μM concentrations. The resulting dose–response data were fitted using GraphPad Prism. Table 5 shows active hits and their corresponding estimated and experimental anti-AChE bioactivities. The dose–response curves of active hits are depicted in Figure I in the supplementary materials.

Clearly from figure I and Table 5, the dose–response curves of tested inhibitory hits exhibit Hill slope values ≤1.0 and excellent correlation coefficients, which strongly suggest their authenticity (i.e., non-promiscuousity) [78, 79]. To validate our assay conditions and procedure we used the AChE standard inhibitor, galanthamine (G1660), as standard positive control [72]. The chemical structures of the most potent hits, i.e., **86**, **89**, **102** and **104**, were validated by proton and 13C NMR spectroscopy as well as mass spectrometry. The corresponding spectrums are shown in figures **II** to **IX** in the supplementary materials.

Figure 5 shows how active hits **89** and **102** (Table 5) fit pharmacophore models Hypo6/1 and Hypo7/1 (Fig. 5a, d), respectively, and compares their fitted structures with corresponding docked poses into AChE catalytic pocket (Fig. 5b, e). The comparison shows striking resemblance between pharmacophore-fitted **89** and **102** with their docked poses. Moreover, their poses (both pharmacophore-fitted and docked) closely resemble the respective poses generated for training compounds **32** and **35** in Figs. 2 and 3.

Mapping the hydroxyl group of **89** against a HBA feature in Hypo6/1 (Fig. 5a) agrees with hydrogen bonding interaction tying the same hydroxyl with the NH of indole side chain of Trp439 (Fig. 5b). Similarly, mapping the dibenzofuran ring system of **89** against two RingArom

(A)

(B)

(C)

(D)

(E)

(F)
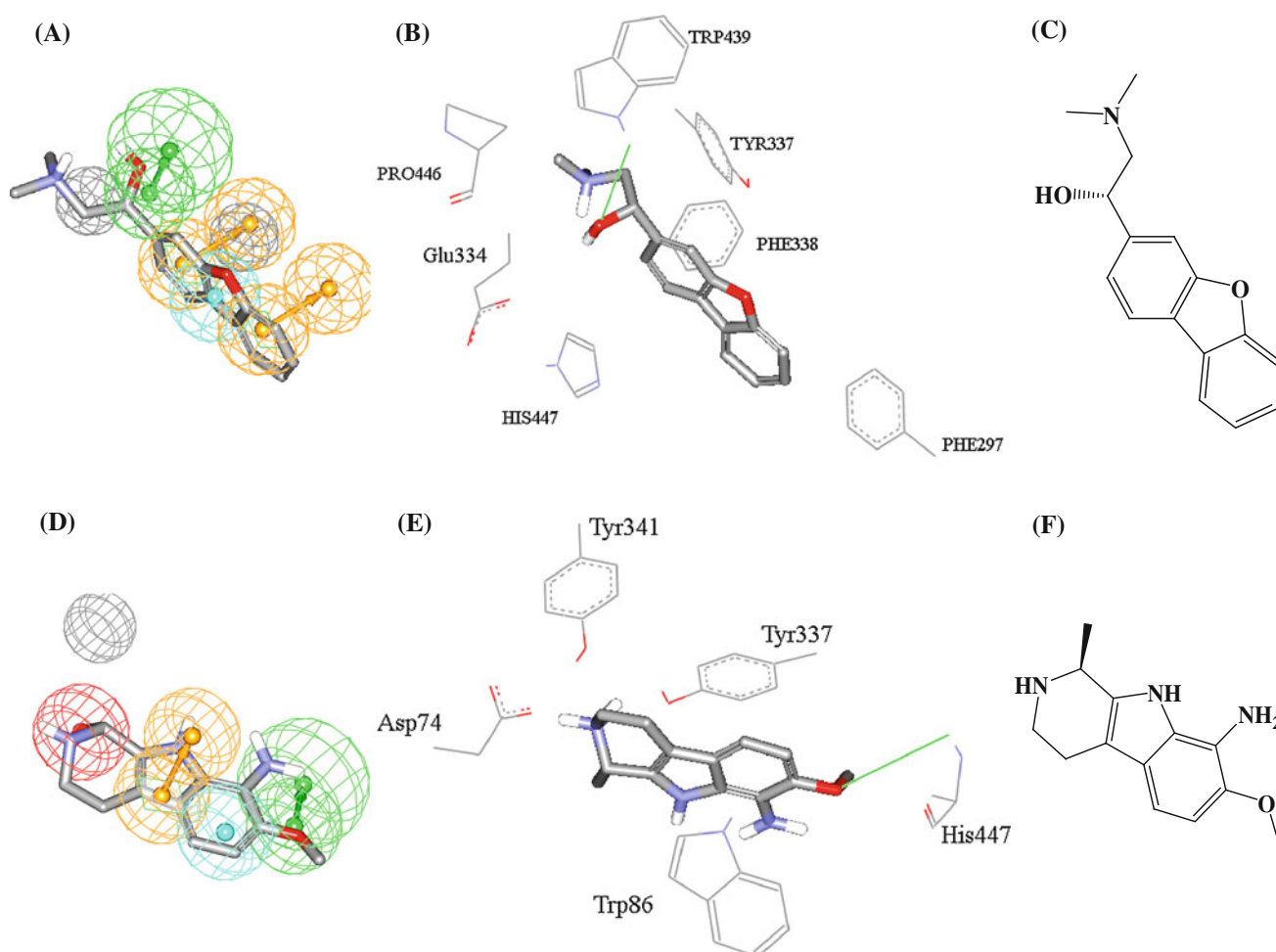
Fig. 5 **a** Hypo6/1 fitted against hit **89** (IC$_{50}$ = 1.0 μM, Table 5). **b** Optimal docked pose of **89** within AChE catalytic site (PDB code: 1B41, resolution 2.76 Å). **c** Chemical structure of **89**. **d** Hypo7/1 fitted against hit **102** (IC$_{50}$ = 2.0 μM, Table 5). **e** Optimal docked pose of **102** within AChE catalytic site. **f** Chemical structure of **102**

features and a Hbic feature in Hypo6/1 (Fig. 5a) corresponds to π-stacking and hydrophobic interactions resulting from squeezing the dibenzofuran with the aromatic pouch of Tyr337, Phe338, Phe297 and His447 (Fig. 5b). All these interactions are reminiscent of those tying **32** within the catalytic pocket of AChE (seen in Fig. 2a, c) albeit **89** shows an additional interaction, namely, electrostatic attraction connecting the dimethyl-ammonium of **89** with the carboxylate of Glu334 (Fig. 5b).

Comparably, mapping the methoxy oxygen and pyrrolidine nitrogen of **102** against HBA and PosIon features in Hypo7/1 (Fig. 5d), agrees with hydrogen bonding and electrostatic attraction interactions connecting the methoxy oxygen and pyrrolidine nitrogen of **102** with the peptidic NH of His447 and carboxylate of Asp74, respectively (Fig. 5e). Similarly, mapping the indole core of **102** against Hbic and RingArom features in Hypo7/1 (Fig. 5d) corresponds to π-stacking and hydrophobic interactions tying this ring system with the aromatic side chains of Tyr337, Tyr341 and Trp86

within the catalytic pocket of AChE (Fig. 5e). All these interactions seem very similar to interactions binding **35** within AChE (Fig. 3) as can be judged from mapping Hypo7/1 against **35** (Fig. 3c) and from the docked pose of **35** with AChE catalytic pocket (Fig. 3a).

Interestingly, upon comparing the scoring values (essentially binding energy estimates based on six scoring functions [71]) of the docked poses of training compounds **32** and **35** with those of hits **89** and **102**, as seen in Table 6, one can quickly notice comparable binding energy readouts indicating comparable binding affinities for these compounds.

It remains to be mentioned that although QSAR prediction was rather accurate with some hits, e.g., **86, 89, 92, 94, 95 96, 101, 106,** and **109**, it deviated significantly from experimental values with other hits (Table 5). We believe these errors are because training compounds used in QSAR and pharmacophore modeling are significantly structurally different from hit molecules, which limits the extrapolatory potential of the QSAR equation.

Table 6 The score values for optimal docked poses of **32**, **35**, **89** and **102** as suggested by LigandFit docking engine. The corresponding docked poses are shown in Figs. 2, 3 and 5

Compound	Scoring functions (kCal/Mol)					
	Ligscore 1	Ligscore 2	−PLP 1	−PLP 2	JAIN	−PMF
32	1.34	2.42	28.73	43.57	6.8	115.12
35	2.71	2.54	22.66	50.51	5.94	99.14
89	3.39	4.84	70.9	70.85	2.57	121.03
102	4.37	5.00	64.32	71.27	4.98	138.92

Similarity analysis between training compounds and active hits

We employed three library comparison methods implemented in Discovery Studio 2.5 to assess the structural similarity/diversity between the modeled compounds (**1–85**, Table A in Supporting Information) (library B) compared with active hits (**86–109**, Table 5, library A), namely, Murcko assemblies, Bayesian model, and global fingerprints. In Murcko assemblies, the algorithm breaks the ligands of each library into unique occurrences of molecular rings, ring assemblies, bridge assemblies, chains, Murcko assemblies, or any combination of these. Murcko assemblies are contiguous ring systems plus chains that link two or more rings [73] The two libraries are compared using a Tanimoto similarity of the assemblies based on the fragments that are common and unique to each library [46] On the other hand, in the Bayesian model approach, two Bayesian models were built, one to learn library A and one to learn library B. Finally, it scores all ligands using both models. A distance is computed as Eq. (6):

$$\text{Distance} = \text{ScoreAA} + \text{ScoreBB} - \text{ScoreAB} - \text{ScoreBA} \tag{6}$$

where ScoreAA is the average score of library A molecules scored by the Bayesian model that learned library A molecules, while ScoreBB is the average score of library B molecules scored by the Bayesian model that learned library B. ScoreAB and ScoreBA are the average scores of libraries A and B molecules scored by the Bayesian models that learned libraries B and A, respectively. The higher the distance, the more dissimilar the libraries are [46]. Finally, the global fingerprint comparison algorithm generates a global fingerprint for all ligands in the training list and all ligands in the hits list and then computes a Tanimoto similarity coefficient between the two libraries [46]. Table 7 shows the results of the three similarity/diversity assessment procedures. Clearly, the three methods suggest minimal structural similarity between modeled AChE inhibitors and our active hits.

Careful analysis of Lipinski's properties of our active hits shows them to be generally more hydrophilic compared to modeled compounds (both training and testing

Table 7 Results of similarity analysis between training compounds and active hits

Murcko assemblies[a]		Bayesian model[b]		Global fingerprints[b,c]	
Number of total assemblies	59	Average LibA score of library A ligands	23.98	Number of total global fingerprint bits	1128
Number of common assemblies	0	Average LibB score of library A ligands	−24.62	Number of common global fingerprint bits	117
Number of assemblies only in library A[d]	19	Average LibA score of library B ligands	−45.53	Number of global fingerprint bits only in library A	429
Number of assemblies only in library B[e]	40	Average LibB score of library B ligands	5.89	Number of global fingerprint bits only in library B	582
Similarity score between the two libraries	0.00	Bayesian distance between the two libraries	100.03	Similarity score between the two libraries	0.10

[a] See "Similarity analysis between training compounds and active hits" section and [73]

[b] See Sect. 3.7 and [46]

[c] Done by implementing the fingerprint descriptor FCFC_6, which correspond to functional-class extended-connectivity fingerprint count up to diameter 6.43

[d] Library A includes active hits (**86–109**, Table 5)

[e] Library B includes all training and testing compounds employed in pharmacophore and QSAR modeling (**1–85**, Figure A under Supplementary Materials)

compounds): The average Lipinsk's parameters of the collected modeled compounds are as follows: LogP = 3.5, molecular weight = 297.39, HBA = 3.31, HBD = 1.03, number of rotable bonds = 4.09, while the same paramters for our active hits are as follows: LogP = 2.00, molecular weight = 295.18, HBA = 4.5, HBD = 1.96, number of rotable bonds = 4.88.

Enhanced hydrophilicity of captured hits mean they are expected to have better pharmacokinetic profiles compared to modeled collected compounds.

Conclusions

AChE inhibitors are currently considered as potential treatments for neurodegenerative disorders such as AD. The pharmacophoric space of AChE inhibitors was explored via four diverse sets of inhibitors and using CATALYST-HYPOGEN to identify high quality binding model(s). Subsequently, genetic algorithm and MLR analysis were employed to achieve optimal QSAR model capable of explaining anti-cholinesterse bioactivity variation across 85 collected inhibitors. Two orthogonal pharmacophoric models emerged in the QSAR equation suggesting the existence of at least two distinct binding modes accessible to ligands within AChE binding pocket. The QSAR equation and the associated pharmacophoric models were experimentally validated through identification of several AChE inhibitors retrieved via in silico screening some of which gave micromolar potencies. Our results suggest that the combination of pharmacophoric exploration and QSAR analyses can be useful tool for finding new diverse AChE inhibitors.

References

1. Clark CM, Karlawish JH (2003) Ann Intern Med 138(5):400–410
2. Chen S, Zhang X-J, Li L, Le W-D (2007) Curr Neuropharmacol 5(2):127–134
3. Terry AV (2003) J Pharmacol Exp Ther 306(3):821–827
4. Bachurin SO (2003) Med Res Rev 23(1):48–88
5. Rollinger JM, Hornick A, Langer T, Stuppner H, Prast H (2004) J Med Chem 47(25):6248–6254
6. Perry EK, Kilford L, Lees AJ, Burn DJ, Perry RH (2003) Ann Neurol 54(2):235–238
7. Mukherjee PK, Satheeshkumar N, Venkatesh P, Venkatesh M (2011) Mini Rev Med Chem 11(3):247–262
8. Lu SH, Wu JW, Liu HL, Zhao JH, Liu KT, Chuang CK, Lin HY, Tsai WB, Ho Y (2011) J Biomed Sci 18:8–18
9. Whittaker VP (1990) Trends Pharmacol Sci 11(1):8–13
10. Purves D, George A, David F, William H, Anthony L, James M, Leonard W (2008) Neuroscience, 4th edn. Sinauer Associates Inc, Sunderland MA, pp 121–122
11. Quinn DM (1987) Chem Rev 87(5):955–979
12. Taylor P, Radic Z (1994) Annu Rev Pharmacol Toxicol 34:281–320
13. Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, Silman I (1991) Science 253(5022):872–879
14. Sussman JL, Harel M, Silman I (1993) Chem Biol Interact 87(1–3):187–197
15. Harel M, Schalk I, Ehret-Sabatier L, Bouet F, Goeldner M, Hirth C, Axelsen PH, Silman I, Sussman JL (1993) PNAS 90(19):9031–9035
16. Silman I, Harel M, Axelsen P, Raves M, Sussman JL (1994) Biochem Soc Trans 22(3):745–749
17. Harel M, Silman I, Quinn DM, Nair HK, Sussman JL (1996) J Am Chem Soc 118(10):2340–2346
18. Greenblatt HMKG, Lewis T, Silman I, Sussman JL (1999) FEBS Lett 463:321–326
19. Kryger G, Silman I, Sussman JL (1998) J Physiol 92(3–4):191–194
20. Inestrosa NC, Dinamarca MC, Alvarez A (2008) FEBS J 275(4):625–632
21. Rees T, Hammond PI, Soreq H, Younkin S, Brimijoin S (2003) Neurobiol Aging 24(6):777–787
22. Rees TM, Berson A, Sklan EH, Younkin L, Younkin S, Brimijoin S, Soreq H (2005) Curr Alzheimer Res 2(3):291–300
23. De Ferrari GV, Canales MA, Shin I, Weiner LM, Silman I, Inestrosa NC (2001) Biochemistry 40(35):10447–10457
24. National Pesticide Information Center-Diazinon Technical Fact Sheet (2012). http://npic.orst.edu/factsheets/diazinontech.pdf
25. Stoelting RK (1999) Anticholinesterase drugs and cholinergic agonists in pharmacology and physiology in anesthetic practice. Lippincott-Raven, Philadelphia
26. Taylor P, Hardman JG, Limbird LE, Molinoff PB., Ruddon RW, Gilman AG (1996) Autonomic pharmacology: cholinergic drugs the pharmacologial basis of therapeutics. The McGraw-Hill Companies, Columbus, Ohio
27. Taha MO, Bustanji Y, Al-Ghussein MAS, Mohammad M, Zalloum H, Al-Masri IM, Atallah N (2008) J Med Chem 51:2062–2077
28. CATALYST 4.11 Users' Manual (2005) Accelrys Software Inc San Diego CA
29. Taha MO, Atallah N, Al-Bakri AG, Paradis-Bleau C, Zalloum H, Younis KS, Levesque RC (2008) Bioorg Med Chem 16(3):1218–1235
30. Taha MO, Bustanji Y, Al-Bakri AG, Yousef A-M, Zalloum WA, Al-Masri IM, Atallah N (2007) J Mol Graph Model 25(6):870–884
31. Al-masri IM, Mohammad MK, Taha MO (2008) ChemMedChem 3(11):1763–1779
32. Taha MO, Dahbiyeh LA, Bustanji Y, Zalloum H, Saleh S (2008) J Med Chem 51:6478–6494
33. Al-Nadaf AS, Taha MO (2010) Bioorg Med Chem 18:3088–3115
34. Abu-Hammad AM, Taha MO (2009) J Chem Inf Model 49:978–996
35. Abu Khalaf R, Abu Sheikha G, Bustanji Y, Taha MO (2010) Eur J Med Chem 45(4):1598–1617
36. Al-Sha'er MA, Taha MO (2010) Eur J Med Chem 45(9):4316–4330
37. Al-Sha'er MA, Taha MO (2010) J Chem Inf Model 50(9):1706–1723
38. Habash M, Taha MO (2011) Bioorg Med Chem 19(16):4746–4771

39. Shahin R, Alqtaishat S, Taha MO (2012) J Comput Aided Mol Des 26(2):249–266
40. Suaifan GARY, Shehadehh M, Al-Ijel H, Taha MO (2012) J Mol Graph Model 37:1–26
41. Shahin R, Taha MO (2012) Bioorg Med Chem 20(1):377–400
42. Taha MO, Qandil AM, Al-Haraznah T, Khalaf RA, Zalloum H, Al-Bakri AG (2011) Chem Biol Drug Des 78(3):391–407
43. Al-Nadaf AH, Taha MO (2011) J Mol Graph Model 29(6): 843–864
44. Al-Najjar BO, Wahab HA, Tengku Muhammad TS, Shu-Chien AC, Ahmad Noruddin NA, Taha MO (2011) Eur J Med Chem 46(6):2513–2529
45. CERIUS2, QSAR Users' Manual, version 4.10, Accelrys Inc., San Diego, CA (2005) 43–88, 221–235, 237–250
46. Discovery Studio 2.5.5 User Guide (2010) Accelrys Inc., San Diego
47. CS ChemDraw Ultra 6.0, Cambridge Soft Corp., USA
48. Fang L, Kraus B, Lehmann J, Heilmann J, Zhang Y, Decker M (2008) Bioorg Med Chem Lett 18(9):2905–2909
49. Fang L, Decker M, Roegler C, Lehmann J, Appenroth D, Fleck C, Kiehntopf M, Deufel T, Peng S, Zhang Y (2008) J Med Chem 51(4):713–716
50. Fang L, Zhang Y, Decker M, Appenroth D, Fleck C, Jumpertz S, Mohr K, Trankle C (2010) J Med Chem 53(5):2094–2103
51. Rook Y, Schmidtke K-U, Gaube F, Schepmann D, Wünsch B, Heilmann J, Lehmann J, Winckler T (2010) J Med Chem 53(9): 3611–3617
52. Schott Y, Decker M, Rommelspacher H, Lehmann J (2006) Bioorg Med Chem Lett 16(22):5840–5843
53. Decker M, Krauth F, Lehmann J (2006) Bioorg Med Chem 14(6):1966–1977
54. Decker M, Kraus B, Heilmann J (2008) Bioorg Med Chem 16(8): 4252–4261
55. Decker M (2006) J Med Chem 49(18):5411–5413
56. Sutter J, Guner O, Hoffmann R, Li H, Waldman M (2000). In: Guner OF (ed) Pharmacophore perception, development,and use in drug design. International University Line, La Jolla, pp 501–511
57. Kurogi Y, Güner OF (2001) Curr Med Chem 8(9):1035–1055
58. Poptodorov K, Luu T, Langer T, Hoffmann R (2006). In: Hoffmann RD (ed) Methods and principles in medicinal chemistry: pharmacophores and pharmacophores searches, vol 2. Wiley-VCH Weinheim, pp 17–47
59. Li H, Sutter J, Hoffmann R (2000) Pharmacophore perception development and use in drug design ed. International University Line, La Jolla, CA, pp 173–178
60. Bersuker I, Bahçeci S, Boggs J (2000) Pharmacophore perception development and use in drug design. International University Line, La Jolla, pp 457–473
61. Fischer R (1966) The principle of experimentation illustrated by a psycho-physical. ExpeHafner Publishing Co, 8th ed. Hafner Publishing,New York Chapter II
62. Krovat EM, Langer T (2003) J Med Chem 46(5):716–726
63. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO (2005) J Med Chem 48(7):2534–2547
64. Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, Watson P (2004). J Chem Inf Comput Sci 44(3)
65. Kirchmair J, Markt P, Distinto S, Wolber G, Langer T (2008). J Comput Aided Mol Des 22(3–4)
66. Irwin JJ, Shoichet BK (2005) J Chem Inf Model 45(1):177–182
67. Jacobsson M, Lidén P, Stjernschantz E, Boström H, Norinder U (2003) J Med Chem 46(26):5781–5789
68. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Adv Drug Deliv Rev 46(1–3):1–3
69. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) J Med Chem 45(12):2615–2623
70. Ellman G, Courtney K, Andresjr V, Featherstone R (1961) Biochem Pharmacol 7(2):88–95
71. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) J Mol Graph Model 21(4):289–307
72. Tamura M, Nakao H, Yoshizaki H, Shiratsuchi M, Shigyo H, Yamada H, Ozawa T, Totsuka J, Hidaka H (2005) Biochim Biophys Acta 1754(1–2):1–2
73. Bemis GW, Murcko MA (1996) J Med Chem 39(15):2887–2893
74. Poornima CS, Dean PM (1995) J Comput Aided Mol Des 9:500–512
75. Poornima CS, Dean PM (1995) J Comput Aided Mol Des 9:513–520
76. Poornima CS, Dean PM (1995) J Comput Aided Mol Des 9:521–531
77. Rogers D, Hopfinger AJ (1994) J Chem Inf Comput Sci 34(4):854–866
78. Walters W, Namchuk M (2003) Nat Rev Drug Discov 2:259–266
79. Brian KS (2006) J Med Chem 49:7274–7277
80. Hasegawa K, Miyashita Y, Funatsu KJ (1997) Chem Inf Comput Sci 37(2):306–310