

3 Elaborate ligand-based modeling coupled with QSAR analysis 4 and in silico screening reveal new potent acetylcholinesterase 5 inhibitors

6 Sawsan Abuhamdah · Maha Habash ·
7 Mutasem O. Taha

8 Received: 28 August 2013 / Accepted: 3 December 2013
9 © Springer Science+Business Media Dordrecht 2013

10 **Abstract** Inhibition of the enzyme acetylcholinesterase
11 (AChE) has been shown to alleviate neurodegenerative
12 diseases prompting several attempts to discover and opti-
13 mize new AChE inhibitors. In this direction, we explored
14 the pharmacophoric space of 85 AChE inhibitors to iden-
15 tify high quality pharmacophores. Subsequently, we
16 implemented genetic algorithm-based quantitative struc-
17 ture–activity relationship (QSAR) modeling to select
18 optimal combination of pharmacophoric models and 2D
19 physicochemical descriptors capable of explaining bioac-
20 tivity variation among training compounds ($r_{68}^2 = 0.94$,
21 F -statistic = 125.8, $r_{LOO}^2 = 0.92$, r_{PRESS}^2 against 17 external
22 test inhibitors = 0.84). Two orthogonal pharmacophores
23 emerged in the QSAR equation suggesting the existence of
24 at least two binding modes accessible to ligands within
25 AChE binding pocket. The successful pharmacophores
26 were comparable with crystallographically resolved AChE
27 binding pocket. We employed the pharmacophoric models
28 and associated QSAR equation to screen the national

cancer institute list of compounds. Twenty-four low 29
micromolar AChE inhibitors were identified. The most 30
potent gave IC_{50} value of 1.0 μ M. 31

Keywords Acetylcholinesterase inhibitors · 32
Pharmacophore modeling · Quantitative structure–activity 33
relationship · In silico screening 34
35

36 Introduction

37 Alzheimer’s disease (AD) is a progressive neurodegenera-
38 tive disorder that primarily affects the elderly population and
39 is considered to be responsible for the majority of dementia
40 cases in people aged 65 or older [1]. This disease is charac-
41 terized by numerous symptoms such as memory and lan-
42 guage impairment, cognitive dysfunction and behavioral
43 disturbances (i.e., depression, agitation and psychosis),
44 which become progressively more severe [2]. Currently
45 there is no cure for this disorder, thus there is a real need for
46 novel molecular templates for AD therapy [1–3].

47 The fact that this neuropathology is associated with cen-
48 tral cholinergic deficit [4] suggests that it can be ameliorated
49 by increasing CNS acetylcholine (ACh) concentrations
50 through inhibiting the metabolic enzyme acetylcholinester-
51 ase (AChE) [5]. In fact all current pharmacological treat-
52 ments of AD, i.e., tacrine, donepezil, rivastigmine and
53 galantamine, are cholinesterase inhibitors. However, most of
54 these present many side effects prompting continuous
55 interest in developing new AChE inhibitors [6–8].

56 Acetylcholinesterase (EC 3.1.1.7) is a serine protease
57 that hydrolyzes the neurotransmitter ACh with high cata-
58 lytic activity [11, 12]. AChE is found mainly at neuro-
59 muscular junctions and cholinergic brain synapses, where
60 its activity serves to terminate synaptic transmission. It

A1 **Electronic supplementary material** The online version of this
A2 article (doi:10.1007/s10822-013-9699-6) contains supplementary
A3 material, which is available to authorized users.

A4 S. Abuhamdah
A5 Department of Biopharmaceutics and Clinical Pharmacy,
A6 Faculty of Pharmacy, University of Jordan, Amman, Jordan

A7 M. Habash
A8 Department of Pharmaceutical Chemistry and Pharmacognosy,
A9 Faculty of Pharmacy, Applied Science University, Amman,
A10 Jordan

A11 M. O. Taha (✉)
A12 Drug Discovery Unit, Department of Pharmaceutical Sciences,
A13 Faculty of Pharmacy, University of Jordan, Amman, Jordan
A14 e-mail: mutasem@ju.edu.jo

61 belongs to carboxyl esterase family of enzymes [6–10].
 62 X-ray structures of AChE co-crystallized with various
 63 ligands provided insights into the essential structural ele-
 64 ments and motifs central to its catalytic mechanism and
 65 mode of ACh processing. The active site of AChE com-
 66 prises 2 subsites—the anionic site and the esteratic subsite.
 67 [13–19] The anionic subsite accommodates the positive
 68 quaternary amine of ACh as well as other cationic sub-
 69 strates and inhibitors. Interestingly, the cationic moieties of
 70 different substrates are not bound by a negatively-charged
 71 amino acid in the enzymatic anionic site, but rather by
 72 interaction with 14 aromatic residues that line the gorge
 73 leading to the active site [15, 17, 20–23].

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

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

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

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

111 We previously reported the use of this innovative approach
 112 towards the discovery of new inhibitory leads against glyco-
 113 gen 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+} /calmodu-
 lin-dependent protein kinase II [41], fungal N-myristoyl
 transferase [42], renin inhibitory [43], and peroxisome pro-
 liferator-activated receptor γ activators [44].

Materials and methods

Molecular modeling

Software and hardware

Pharmacophore and QSAR modeling studies were per-
 formed using CATALYST (HYPOGEN module) [28],
 CERIOUS2 [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) Pharmaco-
 phore exploration performed using HYPOGEN module of
 CATALYST. (4) QSAR analysis performed using QSAR
 and DESCRIPTORS modules within CERIOUS 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 prom-
 ising 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 phar-
 macophores analyses in order to correlate the data linear to
 the free energy change.

160 The two-dimensional (2D) chemical structures of the
 161 inhibitors were imported into CATALYST, converted into
 162 corresponding standard 3D structures and energy mini-
 163 mized to the closest local minimum using the molecular
 164 mechanics CHARMM force field in CATALYST. The
 165 conformational space of each inhibitor was explored
 166 adopting the “best conformer generation” option within
 167 CATALYST [28]. A conformational ensemble was gen-
 168 erated for each training molecule with an energy threshold
 169 of 20 kcal/mol from the local minimized structure, and a
 170 maximum limit of 250 conformers per molecule.

171 *Automatic generation and assessment of pharmacophoric* 172 *hypotheses via catalyst*

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

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

191 Pharmacophore modeling employing CATALYST pro-
 192 ceeds through three successive phases: the constructive
 193 phase, subtractive phase and optimization phase (see
 194 CATALYST Modeling Algorithm under section SM-1 in
 195 Supplementary Materials) [28–44, 56, 57]. When generat-
 196 ing binding hypotheses, CATALYST attempts to minimize
 197 a cost function consisting of three terms: Weight cost,
 198 Error cost and Configuration cost (see CATALYST Cost
 199 Analysis in Assessment of Generated Binding Hypotheses
 200 in section SM-2 under Supplementary Materials).

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

comparable cost criteria to those produced by the original 211
 unscrambled data reduces the confidence in the training 212
 compounds and the unscrambled original pharmacophore 213
 models [28, 61, 62]. 214

QSAR modeling

 215

The successful models (259) were clustered into 25 groups 216
 utilizing the hierarchical average linkage method available in 217
 CATALYST. Subsequently, individual representatives were 218
 arbitrary selected to represent their corresponding clusters in 219
 subsequent QSAR modeling. Table 1 shows the statistical 220
 criteria of representative pharmacophores including their 221
 pharmacophoric features, success criteria and differences 222
 from corresponding null hypotheses (see CATALYST Cost 223
 Analysis in Assessment of Generated Binding Hypotheses in 224
 section SM-2 under Supplementary Materials). The table also 225
 shows the corresponding Cat. Scramble confidence levels for 226
 each representative pharmacophore. 227

For subsequent QSAR modeling, a subset of 68 com- 228
 pounds from the total list of collected inhibitors (**1–85**, 229
 Table A under Supplementary Materials) was utilized as a 230
 training set for QSAR. The remaining 17 molecules (ca. 231
 20 % of the dataset) were employed as an external test 232
 subset for validating the QSAR models. The test molecules 233
 were selected by ranking the collected inhibitors (**1–85**, 234
 Table A in Supplementary Materials) according to their 235
 IC₅₀ values, and then selecting every fifth compound for 236
 the test set starting from the high-potency end. 237

The chemical structures of the inhibitors were imported 238
 into CERIU2 as standard 3D single conformer represen- 239
 tations in SD format. Subsequently, different descriptor 240
 groups were calculated for each compound employing the 241
 C2.DESRIPTOR module of CERIU2. The calculated 242
 descriptors included various simple and valence connec- 243
 tivity indices, electro-topological state indices and other 244
 molecular descriptors (e.g., logarithm of partition coeffi- 245
 cient, polarizability, dipole moment, molecular volume, 246
 molecular weight, molecular surface area, energies of the 247
 lowest and highest occupied molecular orbitals, etc.) [45]. 248
 Additionally, the training compounds were fitted (using the 249
 Best-fit option in CATALYST) against the representative 250
 pharmacophores (25 models, Table 1), and their fit values 251
 were added as additional descriptors. The fit value for any 252
 compound is obtained automatically via equation (D) under 253
 SM-1 in Supplementary Materials [28]. 254

Genetic function approximation (GFA) was employed to 255
 search for the best possible QSAR regression equation 256
 capable of correlating the variations in biological activities 257
 of the training compounds with variations in the generated 258
 descriptors, i.e., multiple linear regression modeling 259
 (MLR). The fitness function employed herein is based on 260
 Friedman’s ‘lack-of-fit’ (LOF) [45]. 261

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
13	10	HBD, Hbic, RingArom, HBA, 4xE	16.2	133.4	265.9	132.5	0.95	95
	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

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

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

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

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

Receiver operating characteristic (ROC) curve analysis 280

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

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

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

$$D(i,j) = \sqrt{(\text{NumHBD}_i - \text{NumHBD}_j)^2 + (\text{NumHBA}_i - \text{NumHBA}_j)^2 + (\text{NP}_i - \text{NP}_j)^2} \quad (2)$$

296 The minimum distances are then averaged over all active
297 compounds (D_{\min}). Subsequently, for each active com-
298 pound in the test set, around 25 decoys were randomly
299 chosen from the ZINC database [66]. The decoys were
300 selected in such a way that they did not exceed D_{\min} dis-
301 tance from their corresponding active compound.

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

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

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

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

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

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

method, TN is the number of discarded decoys (presum- 327
ably inactive), while FP is the number of captured decoys 328
(presumably inactive) [63, 65, 66]. 329

330 If all molecules scored by a VS protocol with sufficient
331 discriminatory power are ranked according to their score (i.e.,
332 fit values), starting with the best-scored molecule and ending
333 with the molecule that got the lowest score, most of the actives
334 will have a higher score than the decoys. Since some of the
335 actives will be scored lower than decoys, an overlap between
336 the distribution of active molecules and decoys will occur,
337 which will lead to the prediction of false positives and false
338 negatives. [63, 65]. The selection of one score value as a
339 threshold strongly influences the ratio of actives to decoys and

340 therefore the validation of a VS method. The ROC curve
341 method avoids the selection of a threshold by considering all
342 Se and Sp pairs for each score threshold [63, 65]. A ROC curve
343 is plotted by setting the score of the active molecule as the first
344 threshold. Afterwards, the number of decoys within this cutoff
345 is counted and the corresponding Se and Sp pair is calculated.
346 This calculation is repeated for the active molecule with the
347 second highest score and so forth, until the scores of all actives
348 are considered as selection thresholds.

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

- 358 (1) Area under the ROC curve (AUC) [63, 65, 67]. In an
359 optimal ROC curve an AUC value of 1 is obtained;
360 however, random distributions cause an AUC value
361 of 0.5. VS that performs better than a random
362 discrimination of actives and decoys retrieve an
363 AUC value between 0.5 and 1, whereas an AUC
364 value lower than 0.5 represents the unfavorable case
365 of a VS method that has a higher probability to assign
366 the best scores to decoys than to actives [63, 65].
- 367 (2) Overall Accuracy (ACC) describes the percentage of
368 correctly classified molecules by the screening pro-
369 tocol. Testing compounds are assigned a binary score
370 value of zero (compound not captured) or one
371 (compound captured) [63, 65, 67].

372	(3) Overall specificity (SPC): describes the percentage of	Subsequently, they were diluted to the required concen-	417
373	discarded inactives by the particular VS workflow.	trations with Tris buffer (pH 7.4) for enzymatic assay.	418
374	Inactive test compounds are assigned a binary score		
375	value of zero (compound not captured) or one	<i>Quantification of AChE inhibitory activity</i>	419
376	(compound captured) regardless to their individual	<i>in a spectrophotometric assay</i>	420
377	fit values [63, 65, 67].		
378	(4) Overall True Positive Rate (TPR or overall sensitiv-	Acetylcholinesterase activities were measured through	421
379	ity): describes the fraction percentage of captured	Ellman's colorimetric method with a slight modification	422
380	actives from the total number of actives. Active test	[70]. In a typical run, AChE was dissolved in Tris-HCl	423
381	compounds are assigned a binary score value of zero	buffer (50 mM, pH 8.0) to give a final solution of 0.22 U/	424
382	(compound not captured) or one (compound captured)	mL. Subsequently, 1 µL of the enzyme solution was added	425
383	regardless to their individual fit values [63, 65, 67].	to each well of a 96-well plate. Thereafter, a predetermined	426
384	(5) Overall False Negative Rate (FNR or overall per-	volume of each tested compound solution was added to	427
385	centage of discarded actives): describes the fraction	each well to yield final concentrations of 0.1, 1, 10,	428
386	percentage of active compounds discarded by the VS	100 µM. Subsequently, Tris-HCl buffer was added (pH	429
387	method. Discarded active test compounds are	8.0, 50 mM with 0.1 % w/v bovine serum albumin) to each	430
388	assigned a binary score value of zero (compound	well to reach a total of 40 µL and allowed to stand for	431
389	not captured) or one (compound captured) regardless	10 min at room temperature. Subsequently, DTNB (50 µL,	432
390	to their individual fit values [63, 65, 67].	0.075 mM) and ATCI (10 µL, 1.5 mM) were added to	433
		wells. Color development was measured spectrophoto-	434
		metrically at λ 412 nm using microplate reader (BioTek	435
391	<i>In silico screening for new AChE inhibitors</i>	ELx800, USA) at a rate of one measurement per minute	436
		over 15 min period. Positive (galanthamine) and negative	437
392	Hypo6/1 and Hypo7/1 were employed to screen the NCI	(no inhibitors) controls were tested. The reaction rates were	438
393	structural database. The screening was done employing the	compared and the percent inhibition due to the presence of	439
394	"Best Flexible Database Search" within CATALYST.	tested compounds was calculated. All samples were	440
395	Captured hits were filtered according to Lipinski's [68] and	assayed in at least duplicate measurements. In general, the	441
396	Veber's [69] rules. Remaining hits were fitted against the	amount of DMSO was kept below 1 % in the assay.	442
397	two pharmacophores using the "best fit" option within		
398	CATALYST. The fit values together with the relevant	Results and discussion	443
399	molecular descriptors of each hit were substituted in the		
400	optimal QSAR Eq. (5). The highest ranking molecules	CATALYST-HYPOGEN utilizes a collection of molecules	444
401	based on QSAR predictions were acquired and tested	with activities ranging over a number of orders of magnitude	445
402	in vitro. Table 5 shows active hits and their QSAR-pre-	for automatic pharmacophore construction. It generates	446
403	dictions and experimental bioactivities.	binding hypotheses (pharmacophores) by using the geo-	447
		metric localization of the chemical features present in the	448
404	<i>In vitro experimental studies</i>	molecules to explain the variability of bioactivity. CATA-	449
		LYST-HYPOGEN defines a 3D array of a maximum of five	450
405	<i>Materials</i>	chemical features common to active training molecules that	451
		provides relative alignment for each input molecule consis-	452
406	All of the chemicals were purchased from Sigma-Aldrich (St.	tent with binding to certain proposed common binding site.	453
407	Louis, MO, USA) including electric eel AChE (type-VI-S, EC	The chemical features can be hydrogen bond donors and	454
408	3.1.1.7), acetylthiocholine iodide (ATCI), 5,5'-dithiobis-(2-	acceptors (HBD and HBA), aliphatic and aromatic hydro-	455
409	nitrobenzoic acid) (DTNB), Tris-HCl, bovine serum albumin	phobes (Hbic), positive and negative ionizable (PosIon and	456
410	(BSA), NaCl, MgCl ₂ ·6H ₂ O, standard inhibitor galanthamine	NegIon) groups and aromatic planes (RingArom). CATA-	457
411	(G1660), water and dimethyl sulfoxide (DMSO) for bioanal-	LYST pharmacophores have been used as 3D queries for	458
412	ysis. Tested hits were kindly freely provided by the NCI.	database searching and in 3D-QSAR studies [27, 29–40].	459
413	<i>Preparation of hit compounds for in vitro assay</i>	Data mining and conformational coverage	460
414	The tested compounds were provided as dry powders in	The literature was surveyed to collect many structurally	461
415	variable quantities (5–10 mg). They were initially dis-	diverse AChE inhibitors (1–85, see Table A under	462
416	solved in DMSO to give stock solutions of 100 µM.		

463 supplementary materials) [48–55]. The conformational
 464 space of each inhibitor was extensively sampled utilizing
 465 the poling algorithm of CATALYST [55]. Proper confor-
 466 mational exploration is necessary for pharmacophore gen-
 467 eration and pharmacophore-based search procedures as
 468 both are known for their sensitivity to inadequate confor-
 469 mational sampling [34].

470 Exploration of AChE pharmacophoric space

471 The training inhibitors were selected in such a way that
 472 they were assayed by the same procedure (compounds 1–
 473 85, Table A in Supplementary Materials). Obviously, sta-
 474 tistical consistency necessitates that QSAR and pharma-
 475 cophore modeling are based on training compounds
 476 assayed by a single bioassay procedure [27, 29–40].

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

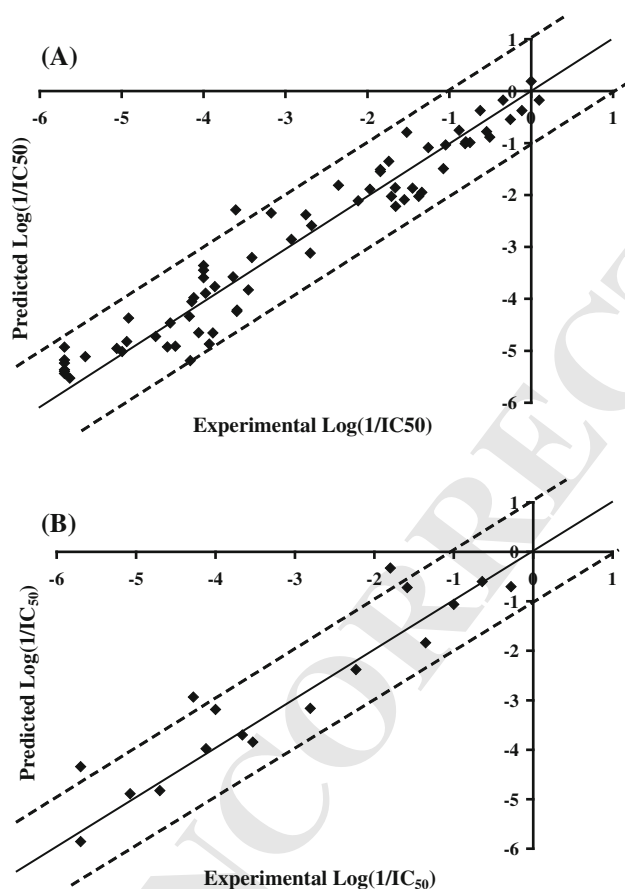


Fig. 1 Experimental versus (a) fitted (68 compounds, $r_{\text{LOO}}^2 = 0.917$), and (b) predicted (17 compounds, $r_{\text{PRESS}}^2 = 0.0.841$) bioactivities calculated from the best QSAR model Eq. (5). The solid lines are the regression lines for the fitted and predicted bioactivities of training 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).
 480 The training compounds were selected to guarantee wide
 481 structural diversity with bioactivities extended over more
 482 than 3.5 logarithmic cycles. To ensure sufficient molecular
 483 diversity within training subsets, member compounds were
 484 selected in such a way that each structural cluster of the
 485 collected compounds was sampled at least once in each
 486 training subset. However, some compounds were repeat-
 487 edly selected in training subsets because of their critical
 488 pharmacophoric features. Training subsets were selected in
 489 such a way that differences in AChE inhibitory activities
 490 among their member compounds are primarily attributable
 491 to the presence or absence of pharmacophoric features
 492 [e.g., HBA, HBD, Hbic or ring aromatic (RingArom)]
 493 rather than steric shielding and/or bioactivity-enhancing or
 494 -reducing auxiliary groups (e.g., electron donating or
 495 withdrawing groups). A special emphasis was given to the
 496 structural diversity of the most-active compounds in each
 497 training subset because of their significant influence on the
 498 extent of the evaluated pharmacophoric space during the
 499 constructive phase of HYPOGEN algorithm (see CATA-
 500 LYST Modeling Algorithm under section SM-1 in Sup-
 501 plementary Materials) [28, 57–60, 71].

502 HYPOGEN was instructed to explore only 4- and
 503 5-featured pharmacophores and ignore models of lesser
 504 number of features (as shown in Table C in Supplementary
 505 Materials). The advantage of this restriction is to narrow
 506 the investigated pharmacophoric space while allowing
 507 good representation of the feature-rich nature of AChE
 508 inhibitors. We previously implemented similar pharma-
 509 cophore exploration strategies against a multitude of targets
 510 [27–44].

511 Eventually, 259 pharmacophore models resulted from
 512 32 automatic HYPOGEN runs. Fortunately, all generated
 513 pharmacophores illustrated $\geq 90\%$ Cat.Scramble signifi-
 514 cance (see “Automatic generation and assessment of
 515 pharmacophoric hypotheses via catalyst” section) [28, 61,
 516 62]. These were subsequently clustered and representative
 517 models were arbitrary selected from each cluster (total of
 518 25 models, Table 1) to represent their clusters in sub-
 519 sequent QSAR modeling.

520 QSAR modeling

521 Clearly from Table 1, representative pharmacophore
 522 models shared comparable binding features and excellent
 523 statistical criteria. Emergence of numerous statistically
 524 comparable pharmacophore hypotheses suggests the ability
 525 of AChE ligands to assume multiple binding modes within
 526 the binding pocket. Accordingly, it is rather hard to select
 527 a particular binding pharmacophore as a single representa-
 528 tive of ligand binding. This point combined with the fact
 529

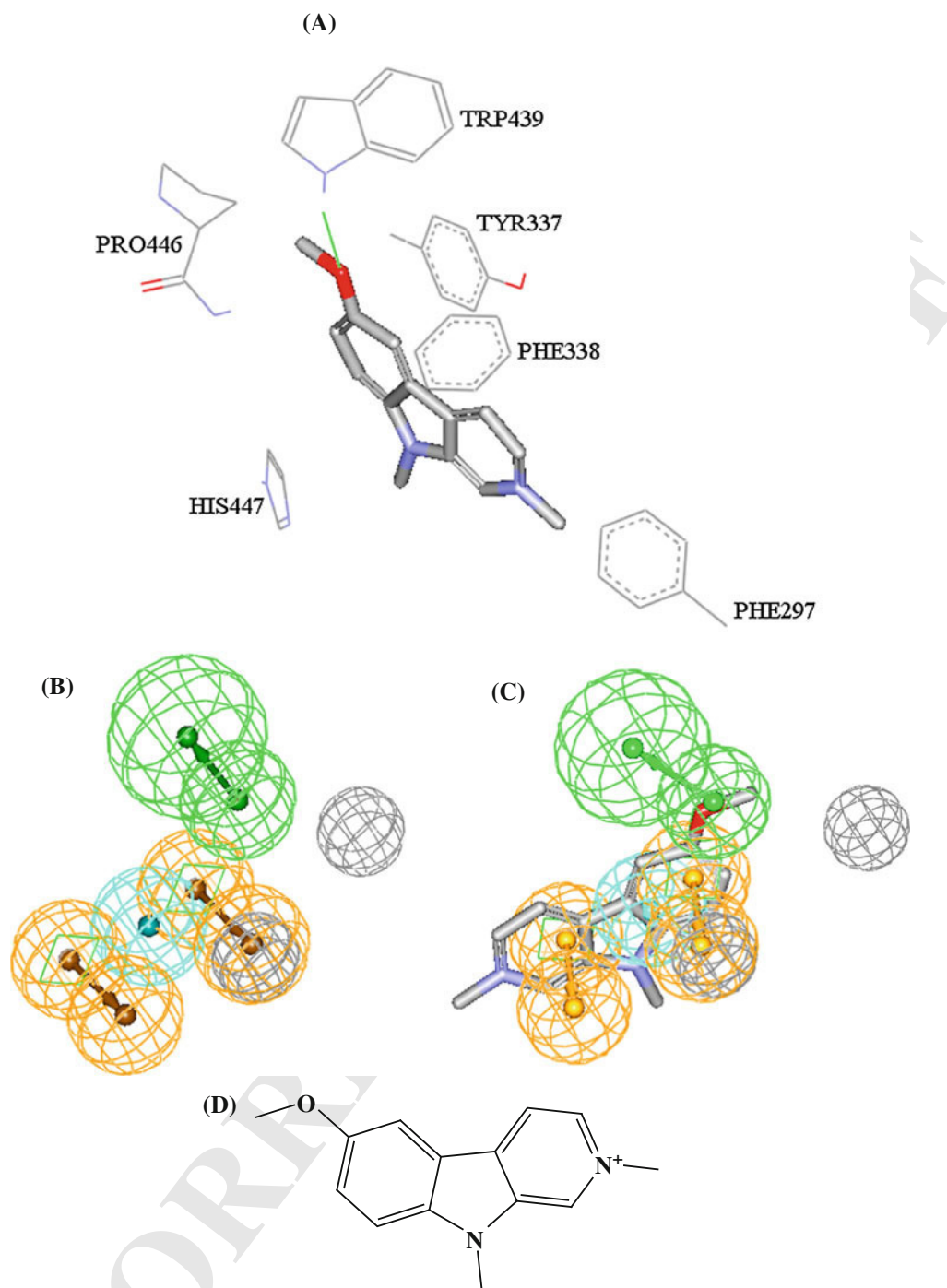


Fig. 2 a Docked structure of training compound **32** ($IC_{50} = 1.3$ nM, Table A under Supplementary Materials) into AChE (PDB code: 1B41, resolution 2.76 Å, the green line represents hydrogen bonding).

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**

530 that pharmacophoric models are limited by steric shielding
531 and bioactivity-enhancing or reducing auxiliary groups [27,
532 29–44], prompted us to employ classical QSAR analysis to
533 search for the best combination of pharmacophore(s) and
534 other 2D descriptors capable of explaining bioactivity
535 variation across the whole list of collected inhibitors (1–85,

536 Table A under Supplementary Materials). We employed 536
537 GFA and MLR QSAR (GFA–MLR–QSAR) analysis to 537
538 search for an optimal QSAR equation(s) [77, 80]. 538

539 The fit values obtained by mapping representative 539
540 hypotheses (25 models) against collected AChE inhibitors 540
541 (1–25, Table A under Supplementary Materials) were 541

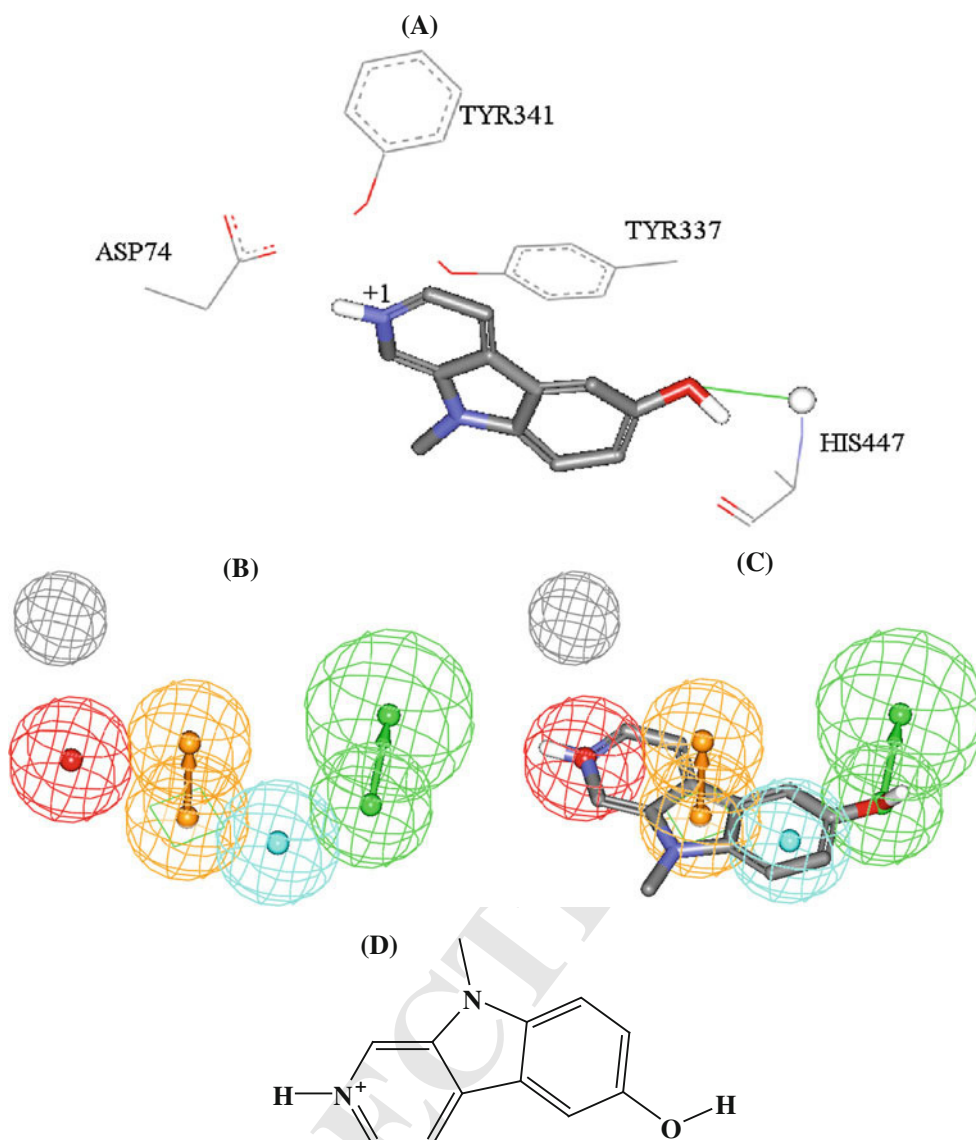


Fig. 3 **a** Docked structure of training compound **35** ($IC_{50} = 1.8$ nM, Table A under Supplementary Materials) into AChE (PDB code: 1B41, resolution 2.76 Å) binding pocket. **b** Pharmacophoric features

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**

542 enrolled, together with around 100 other physicochemical
543 descriptors, as independent variables in GFA-MLR-QSAR
544 analysis (see “QSAR modeling” section) [27–38, 45, 64].
545 We arbitrarily selected 17 molecules (marked with double
546 asterisks in Table A under Supplementary Materials) as
547 external test molecules for validating the QSAR models. All
548 QSAR models were cross-validated automatically using the
549 leave-one-out cross-validation in CERIU2 [45, 64].

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

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

$$\begin{aligned} r_{68}^2 = 0.94, \quad r_{\text{LOO}}^2 = 0.92, \quad F\text{-statistic} = 125.8, \\ r_{\text{PRESS}(17)}^2 = 0.84 \end{aligned} \quad (5) \quad 555$$

557 where, r_{68}^2 is the correlation coefficient against 68 training
558 compounds and r_{PRESS}^2 is the predictive r^2 determined for
559 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	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	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)

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

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

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

583 Interestingly, QSAR Eq. (5) shows **JursRNCG** com-
584 bined with a relatively pronounced positive regression
585 coefficient suggesting significant ligand-AChE affinity
586 promoting effects by focused electrophilic centers. The
587 most probable explanation of this trend is related to the fact
588 that electrophilic heterocycles π -stack efficiently against

589 complementary electron-rich aromatic side chains of
590 amino-acids within the binding pocket of AChE. The cat-
591 alytic site of AChE includes several electron-rich aromatic
592 rings belonging to Trp439, Tyr337, Phe338 and Phe297 (as
593 in Figs. 2, 3).

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

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

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

620 To further emphasize the validity of our pharmacophore/
621 QSAR modeling approach, we compared the pharmaco-
622 phoric features of Hypo6/1 and Hypo7/1 and how they map
623 training compounds **32** and **35**, respectively, with optimal
624 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

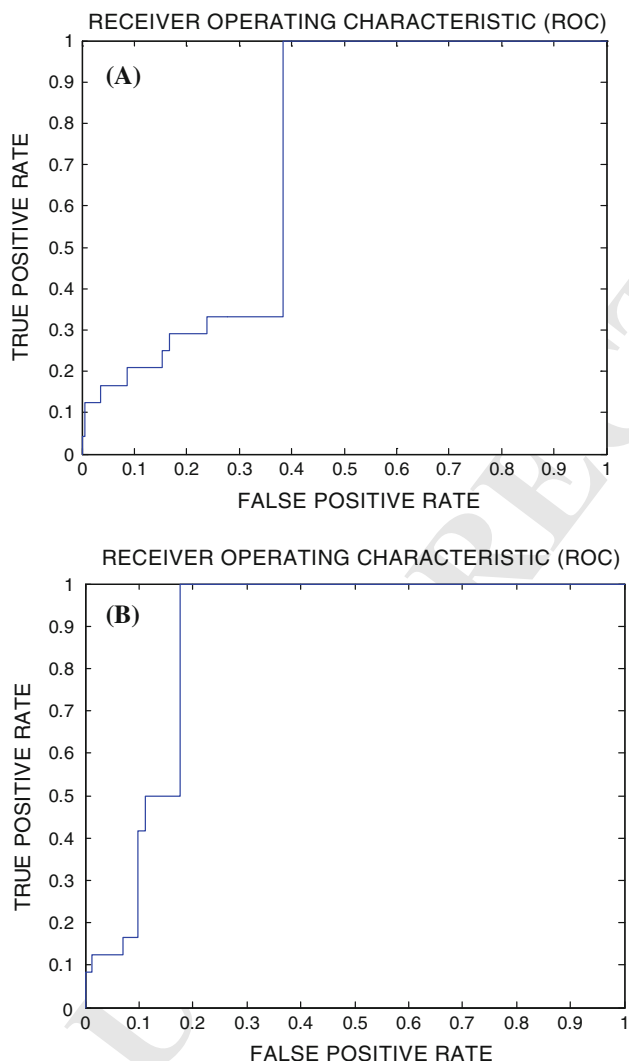


Fig. 4 ROC curves of: **a** Hypo6/1, **b** Hypo7/1

625 were conducted employing LigandFit (as implemented in
626 Discovery Studio 2.5) and through default docking settings
627 [71]. Figures 2 and 3 show the pharmacophores, docked
628 poses and corresponding mapped conformers. By compar-
629 ing the docked pose of **32** ($IC_{50} = 1.3$ nM) within
630 AChE with the way it fits Hypo6/1 (Fig. 2), one quickly
631 notes the great similarity between the Hypo6/1 and the
632 binding interactions within AChE: Mapping the aromatic
633 methoxy of **32** with hydrogen bond acceptor (HBA) feature
634 in Hypo6/1 (Fig. 2c) corresponds to hydrogen bonding
635 interactions connecting this group with the indole NH of
636 Trp439 (Fig. 2a). Similarly, mapping the pyridinoindole
637 ring system of **32** against two RingArom and a Hbic feature
638 (Fig. 2c) correlates with π -stacking interactions result-
639 ing from squeezing the pyridinoindole ring system within the
640 aromatic pouch of the binding pocket comprised from the
641 aromatic side chains of Phe297, Phe338 Tyr337, Trp439
642 and His447 (Fig. 2a).

643 Similar analogy can be concluded by comparing the
644 docked pose of **35** ($IC_{50} = 1.8$ nM) with fitting against
645 Hypo7/1 (Fig. 3): Mapping the protonated pyridinium ion of
646 **35** against PosIon feature in Hypo7/1 (Fig. 3c) corresponds
647 to electrostatic attraction connecting this positive group with
648 the carboxylate anion of Asp74 (Fig. 3a). Likewise, mapping
649 the phenolic OH of **35** against HBA feature in Hypo7/1
650 (Fig. 3c) seems to correspond to hydrogen bonding inter-
651 action connecting the same phenolic group with the amidic
652 NH of His447 (Fig. 3a). Finally, mapping the pyridinoindole
653 electron-deficient ring system of **35** against Hbic and Rin-
654 gArom features in Hypo7/1 (Fig. 3c) correlates with π -
655 stacking interactions against the electron-rich aromatic side
656 chains of Tyr337 and Tyr341 (Fig. 3a).

657 Clearly from the above discussion, Hypo6/1 and Hypo7/
658 1 represent two valid binding modes assumed by ligands
659 within AChE catalyst site. Interestingly, these pharmaco-
660 phore models point to limited number of critical interac-
661 tions required for high ligand-AChE affinity in each of the
662 binding modes. In contrast, docked complexes reveal many
663 bonding interactions without highlighting critical ones.
664 Figures 2a and 3a show only interactions corresponding to
665 pharmacophoric features while other binding interactions
666 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

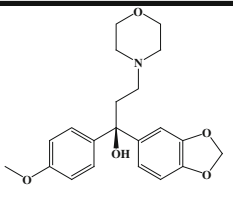
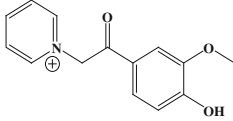
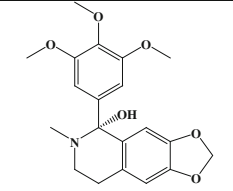
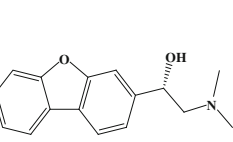
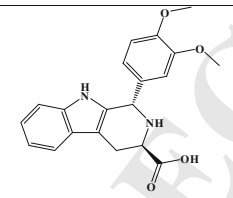
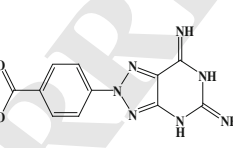
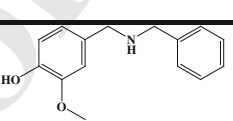
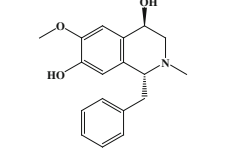
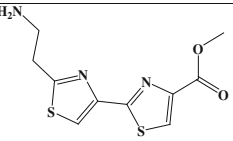
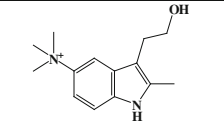
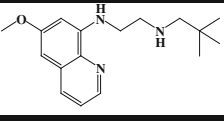
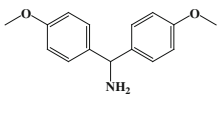
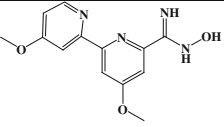
Hit compound	NCI code	Structure	Fit value against		QSAR predictions		Experimental	
			Hypo6/1 ^a	Hypo7/1 ^a	Log(1/IC ₅₀)	IC ₅₀ (μM)	IC ₅₀ (μM) ^b	Hill slope ^c
86 ^d	NSC34198		6.4	6.8	-1.64	43.6	2.00	0.89
87	NSC34674		6.4	7.4	-0.58	3.8	2.51	0.76
88	NSC36355		6.0	6.8	-1.79	62.1	2.51	0.76
89 ^d	NSC94759		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

667 In-silico screening and subsequent in vitro evaluation

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

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

682 Initially, tested hits were screened at 100 μ M concen-
683 trations, subsequently; compounds showing anti-AChE
684 percentages exceeding 50 % at 100 μ M were further
685 assessed to determine their IC₅₀ values at 10 and 1 μ M
686 concentrations. The resulting dose–response data were fit-
687 ted using GraphPad Prism. Table 5 shows active hits and
688 their corresponding estimated and experimental anti-AChE
689 bioactivities. The dose–response curves of active hits are
690 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 cor-
responding 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 fea-
ture 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

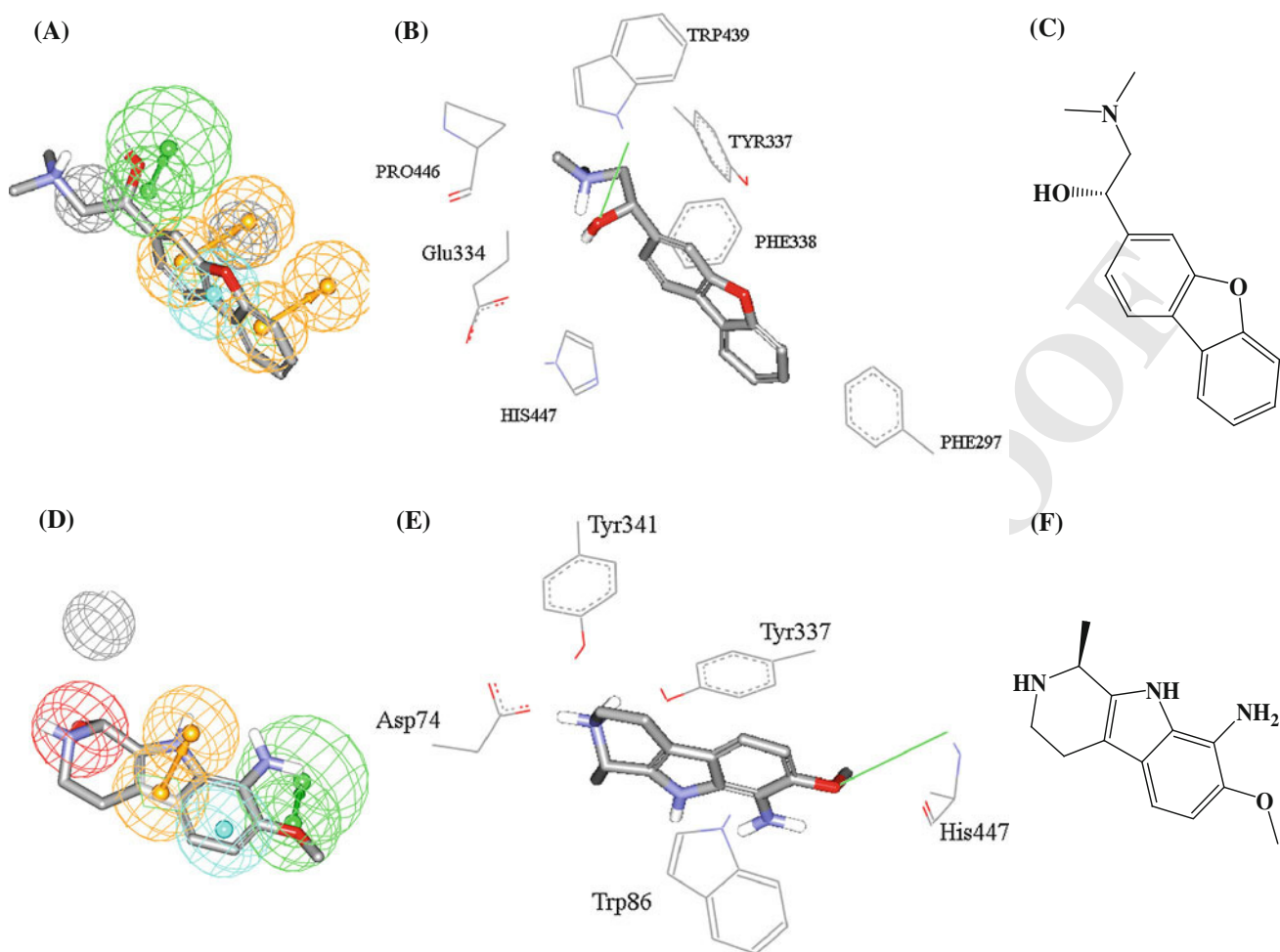


Fig. 5 a Hypo6/1 fitted against hit **89** ($IC_{50} = 1.0 \mu M$, Table 5). b Optimal docked pose of **89** within AChE catalytic site (PDB code: 1B41, resolution 2.76 Å). c Chemical structure of **89**. d Hypo7/1

fitted against hit **102** ($IC_{50} = 2.0 \mu M$, Table 5). e Optimal docked pose of **102** within AChE catalytic site. f Chemical structure of **102**

716 features and a Hbic feature in Hypo6/1 (Fig. 5a) corre-
 717 sponds to π -stacking and hydrophobic interactions result-
 718 ing from squeezing the dibenzofuran with the aromatic
 719 pouch of Tyr337, Phe338, Phe297 and His447 (Fig. 5b).
 720 All these interactions are reminiscent of those tying **32**
 721 within the catalytic pocket of AChE (seen in Fig. 2a, c)
 722 albeit **89** shows an additional interaction, namely, elec-
 723 trostatic attraction connecting the dimethyl-ammonium of
 724 **89** with the carboxylate of Glu334 (Fig. 5b).

725 Comparably, mapping the methoxy oxygen and pyrroli-
 726 dine nitrogen of **102** against HBA and PosIon features in
 727 Hypo7/1, respectively (Fig. 5d), agrees with hydrogen
 728 bonding and electrostatic attraction interactions connecting
 729 the methoxy oxygen and pyrrolidine nitrogen of **102** with the
 730 peptidic NH of His447 and carboxylate of Asp74, respec-
 731 tively (Fig. 5e). Similarly, mapping the indole core of **102**
 732 against Hbic and RingArom features in Hypo7/1 (Fig. 5d)
 733 corresponds to π -stacking and hydrophobic interactions
 734 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).

735 Interestingly, upon comparing the scoring values (essen-
 736 tially binding energy estimates based on six scoring func-
 737 tions [71]) of the docked poses of training compounds **32** and
 738 **35** with those of hits **89** and **102**, as seen in Table 6, one can
 739 quickly notice comparable binding energy readouts indicat-
 740 ing comparable binding affinities for these compounds.

741 It remains to be mentioned that although QSAR pre-
 742 diction was rather accurate with some hits, e.g., **86**, **89**, **92**,
 743 **94**, **95**, **96**, **101**, **106**, and **109**, it deviated significantly from
 744 experimental values with other hits (Table 5). We believe
 745 these errors are because training compounds used in QSAR
 746 and pharmacophore modeling are significantly structurally
 747 different from hit molecules, which limits the extrapolatory
 748 potential of the QSAR equation.
 749
 750
 751
 752
 753

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

754 Similarity analysis between training compounds
755 and active hits

756 We employed three library comparison methods imple-
757 mented in Discovery Studio 2.5 to assess the structural
758 similarity/diversity between the modeled compounds (**1–**
759 **85**, Table A in Supporting Information) (library B) com-
760 pared with active hits (**86–109**, Table 5, library A),
761 namely, Murcko assemblies, Bayesian model, and global
762 fingerprints. In Murcko assemblies, the algorithm breaks
763 the ligands of each library into unique occurrences of
764 molecular rings, ring assemblies, bridge assemblies, chains,
765 Murcko assemblies, or any combination of these. Murcko
766 assemblies are contiguous ring systems plus chains that
767 link two or more rings [73] The two libraries are compared
768 using a Tanimoto similarity of the assemblies based on the
769 fragments that are common and unique to each library [46]
770 On the other hand, in the Bayesian model approach, two
771 Bayesian models were built, one to learn library A and one
772 to learn library B. Finally, it scores all ligands using both
773 models. A distance is computed as Eq. (6):

$$\text{Distance} = \text{ScoreAA} + \text{ScoreBB} - \text{ScoreAB} - \text{ScoreBA} \quad (6)$$

775 where ScoreAA is the average score of library A molecules
776 scored by the Bayesian model that learned library A mole-
777 cules, while ScoreBB is the average score of library B mole-
778 cules scored by the Bayesian model that learned library B.
779 ScoreAB and ScoreBA are the average scores of libraries A
780 and B molecules scored by the Bayesian models that learned
781 libraries B and A, respectively. The higher the distance, the
782 more dissimilar the libraries are [46]. Finally, the global
783 fingerprint comparison algorithm generates a global finger-
784 print for all ligands in the training list and all ligands in the hits
785 list and then computes a Tanimoto similarity coefficient
786 between the two libraries [46]. Table 7 shows the results of
787 the three similarity/diversity assessment procedures. Clearly,
788 the three methods suggest minimal structural similarity
789 between modeled AChE inhibitors and our active hits.

790 Careful analysis of Lipinski's properties of our active
791 hits shows them to be generally more hydrophilic com-
792 pared 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 common assemblies	0	Average LibB score of library A ligands -24.62
Number of assemblies only in library A ^d	19	Average LibA score of library B ligands -45.53
Number of assemblies only in library B ^e	40	Average LibB score of library B ligands 5.89
Similarity score between the two libraries	0.00	Bayesian distance between the two libraries 100.03
		Number of total global fingerprint bits 1128
		Number of common global fingerprint bits 117
		Number of global fingerprint bits only in library A 429
		Number of global fingerprint bits only in library B 582
		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)

793 compounds): The average Lipinski's parameters of the
794 collected modeled compounds are as follows: $\text{LogP} = 3.5$,
795 molecular weight = 297.39, HBA = 3.31, HBD = 1.03,
796 number of rotatable bonds = 4.09, while the same parameters
797 for our active hits are as follows: $\text{LogP} = 2.00$, molecular
798 weight = 295.18, HBA = 4.5, HBD = 1.96, number of
799 rotatable bonds = 4.88.

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

803 Conclusions

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

823 **Acknowledgments** This project was sponsored by the Deanship of
824 Scientific Research at the University of Jordan. The authors wish to
825 thank the National Cancer Institute for freely providing hit compounds
826 for experimental validation.

827

828 References

- 829 1. Clark CM, Karlawish JH (2003) *Ann Intern Med* 138(5):400–410
830 2. Chen S, Zhang X-J, Li L, Le W-D (2007) *Curr Neuropharmacol*
831 5(2):127–134
832 3. Terry AV (2003) *J Pharmacol Exp Ther* 306(3):821–827
833 4. Bachurin SO (2003) *Med Res Rev* 23(1):48–88
834 5. Rollinger JM, Hornick A, Langer T, Stuppner H, Prast H (2004) *J*
835 *Med Chem* 47(25):6248–6254
836 6. Perry EK, Kilford L, Lees AJ, Burn DJ, Perry RH (2003) *Ann*
837 *Neurol* 54(2):235–238
838 7. Mukherjee PK, Satheeshkumar N, Venkatesh P, Venkatesh M
839 (2011) *Mini Rev Med Chem* 11(3):247–262
840 8. Lu SH, Wu JW, Liu HL, Zhao JH, Liu KT, Chuang CK, Lin HY,
841 Tsai WB, Ho Y (2011) *J Biomed Sci* 18:8–18

9. Whittaker VP (1990) *Trends Pharmacol Sci* 11(1):8–13 842
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 121–122 843
11. Quinn DM (1987) *Chem Rev* 87(5):955–979 844
12. Taylor P, Radic Z (1994) *Annu Rev Pharmacol Toxicol* 34:281–320 845
13. Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L,
Silman I (1991) *Science* 253(5022):872–879 846
14. Sussman JL, Harel M, Silman I (1993) *Chem Biol Interact*
87(1–3):187–197 847
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 848
16. Silman I, Harel M, Axelsen P, Raves M, Sussman JL (1994)
Biochem Soc Trans 22(3):745–749 849
17. Harel M, Silman I, Quinn DM, Nair HK, Sussman JL (1996) *J*
Am Chem Soc 118(10):2340–2346 850
18. Greenblatt HMKG, Lewis T, Silman I, Sussman JL (1999) *FEBS*
Lett 463:321–326 851
19. Kryger G, Silman I, Sussman JL (1998) *J Physiol* 92(3–4):191–194 852
20. Inestrosa NC, Dinamarca MC, Alvarez A (2008) *FEBS J* 275(4):
625–632 853
21. Rees T, Hammond PI, Soreq H, Younkin S, Brimijoin S (2003)
Neurobiol Aging 24(6):777–787 854
22. Rees TM, Berson A, Sklan EH, Younkin L, Younkin S, Brimijoin
S, Soreq H (2005) *Curr Alzheimer Res* 2(3):291–300 855
23. De Ferrari GV, Canales MA, Shin I, Weiner LM, Silman I,
Inestrosa NC (2001) *Biochemistry* 40(35):10447–10457 856
24. National Pesticide Information Center-Diazinon Technical Fact
Sheet (2012). <http://npic.orst.edu/factsheets/diazinontech.pdf> 857
25. Stoelting RK (1999) *Anticholinesterase drugs and cholinergic*
agonists in pharmacology and physiology in anesthetic practice.
Lippincott-Raven, Philadelphia 858
26. Taylor P, Hardman JG, Limbird LE, Molinoff PB., Ruddon RW,
Gilman AG (1996) *Autonomic pharmacology: cholinergic drugs*
the pharmacological basis of therapeutics. The McGraw-Hill
Companies 859
27. Taha MO, Bustanji Y, Al-Ghoussein MAS, Mohammad M,
Zalloum H, Al-Masri IM, Atallah N (2008) *J Med Chem*
51:2062–2077 860
28. CATALYST 4.11 Users' Manual (2005) Accelrys Software Inc
San Diego CA 861
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 862
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):
870–884 863
31. Al-masri IM, Mohammad MK, Taha MO (2008) *ChemMedChem*
3(11):1763–1779 864
32. Taha MO, Dahbiyeh LA, Bustanji Y, Zalloum H, Saleh S (2008)
J Med Chem 51:6478–6494 865
33. Al-Nadaf AS, Taha MO (2010) *Bioorg Med Chem* 18:3088–3115 866
34. Abu-Hammad AM, Taha MO (2009) *J Chem Inf Model* 49:
978–996 867
35. Abu Khalaf R, Abu Sheikha G, Bustanji Y, Taha MO (2010) *Eur*
J Med Chem 45(4):1598–1617 868
36. Al-Sha'er MA, Taha MO (2010) *Eur J Med Chem* 45(9):4316–4330 869
37. Al-Sha'er MA, Taha MO (2010) *J Chem Inf Model* 50(9):1706–1723 870
38. Habash M, Taha MO (2011) *Bioorg Med Chem* 19(16):4746–4771 871
39. Shahin R, Alqtaishat S, Taha MO (2012) *J Comput Aided Mol*
Des 26(2):249–266 872
40. Suaifan GARY, Shehadeh M, Al-Ijel H, Taha MO (2012) *J Mol*
Graph Model 37:1–26 873
41. Shahin R, Taha MO (2012) *Bioorg Med Chem* 20(1):377–400 874
42. Taha MO, Qandil AM, Al-Haraznah T, Khalaf RA, Zalloum H,
Al-Bakri AG (2011) *Chem Biol Drug Des* 78(3):391–407 875

- 908 43. Al-Nadaf AH, Taha MO (2011) *J Mol Graph Model* 29(6): 843–864
- 909
- 910 44. Al-Najjar BO, Wahab HA, Tengku Muhammad TS, Shu-Chien
- 911 AC, Ahmad Noruddin NA, Taha MO (2011) *Eur J Med Chem*
- 912 46(6):2513–2529
- 913 45. CERIUS2, QSAR Users' Manual, version 4.10, Accelrys Inc.,
- 914 San Diego, CA (2005) 43–88, 221–235, 237–250
- 915 46. Discovery Studio 2.5.5 User Guide (2010) Accelrys Inc., San
- 916 Diego
- 917 47. CS ChemDraw Ultra 6.0, Cambridge Soft Corp., USA
- 918 48. Fang L, Kraus B, Lehmann J, Heilmann J, Zhang Y, Decker M
- 919 (2008) *Bioorg Med Chem Lett* 18(9):2905–2909
- 920 49. Fang L, Decker M, Roegler C, Lehmann J, Appenroth D, Fleck C,
- 921 Kiehtopf M, Deufel T, Peng S, Zhang Y (2008) *J Med Chem*
- 922 51(4):713–716
- 923 50. Fang L, Zhang Y, Decker M, Appenroth D, Fleck C, Jumpertz S,
- 924 Mohr K, Trankle C (2010) *J Med Chem* 53(5):2094–2103
- 925 51. Rook Y, Schmidtko K-U, Gaube F, Schepmann D, Wünsch B,
- 926 Heilmann J, Lehmann J, Winckler T (2010) *J Med Chem* 53(9):
- 927 3611–3617
- 928 52. Schott Y, Decker M, Rommelspacher H, Lehmann J (2006)
- 929 *Bioorg Med Chem Lett* 16(22):5840–5843
- 930 53. Decker M, Krauth F, Lehmann J (2006) *Bioorg Med Chem* 14(6):
- 931 1966–1977
- 932 54. Decker M, Kraus B, Heilmann J (2008) *Bioorg Med Chem*
- 933 16(8):4252–4261
- 934 55. Decker M (2006) *J Med Chem* 49(18):5411–5413
- 935 56. Sutter J, Guner O, Hoffmann R, Li H, Waldman M (2000). In:
- 936 Guner OF (ed) *Pharmacophore perception, development, and use in*
- 937 **AQ3** *drug design*. International University Line, La Jolla, pp 501–511
- 938 57. Kurogi Y, Güner OF (2001) *Curr Med Chem* 8(9):1035–1055
- 939 58. Poptodorov K, Luu T, Langer T, Hoffmann R (2006). In: Hoff-
- 940 mann RD (ed) *Methods and principles in medicinal chemistry: phar-*
- 941 *macophores and pharmacophores searches*, vol 2. Wiley-
- 942 VCH Weinheim, pp 17–47
- 943 59. Li H, Sutter J, Hoffmann R (2000) *Pharmacophore perception*
- 944 *development and use in drug design* ed. International University
- 945 Line, La Jolla, CA, pp 173–178
- 946 60. Bersuker I, Bahçeci S, Boggs J (2000) *Pharmacophore perception*
- 947 *development and use in drug design*. International University
- 948 Line, La Jolla, pp 457–473
- 949 61. Fischer R (1966) *The principle of experimentation illustrated by a*
- 950 *psycho-physical*. ExpeHafner Publishing Co, 8th ed. Hafner
- 951 Publishing, New York Chapter II
62. Krovat EM, Langer T (2003) *J Med Chem* 46(5):716–726 952
63. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO (2005) *J*
- 953 *Med Chem* 48(7):2534–2547 954
64. Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW,
- 955 Taylor RD, Watson P (2004). *J Chem Inf Comput Sci* 44(3) 956
65. Kirchmair J, Markt P, Distinto S, Wolber G, Langer T (2008).
- 957 *J Comput Aided Mol Des* 22(3–4) 958
66. Irwin JJ, Shoichet BK (2005) *J Chem Inf Model* 45(1):177–182 959
67. Jacobsson M, Lidén P, Stjernschantz E, Boström H, Norinder U
- 960 (2003) *J Med Chem* 46(26):5781–5789 961
68. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) *Adv*
- 962 *Drug Deliv Rev* 46(1–3):1–3 963
69. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW,
- 964 Kopple KD (2002) *J Med Chem* 45(12):2615–2623 965
70. Ellman G, Courtney K, Andresjr V, Featherstone R (1961) *Bio-*
- 966 *chem Pharmacol* 7(2):88–95 967
71. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) *J*
- 968 *Mol Graph Model* 21(4):289–307 969
72. Tamura M, Nakao H, Yoshizaki H, Shiratsuchi M, Shigyo H,
- 970 Yamada H, Ozawa T, Totsuka J, Hidaka H (2005) *Biochim*
- 971 *Biophys Acta* 1754(1–2):1–2 972
73. Bemis GW, Murcko MA (1996) *J Med Chem* 39(15):2887–2893 973
74. Poornima CS, Dean PM (1995) *J Comput Aided Mol Des* 9:
- 974 500–512 975
75. Poornima CS, Dean PM (1995) *J Comput Aided Mol Des* 9:
- 976 513–520 977
76. Poornima CS, Dean PM (1995) *J Comput Aided Mol Des*
- 978 9:521–531 979
77. Rogers D, Hopfinger AJ (1994) *J Chem Inf Comput Sci* 34(4):
- 980 854–866 981
78. Walters W, Namchuk M (2003) *Nat Rev Drug Discov* 2:259–266 982
79. Brian KS (2006) *J Med Chem* 49:7274–7277 983
80. Hasegawa K, Miyashita Y, Funatsu KJ (1997) *Chem Inf Comput*
- 984 *Sci* 37(2):306–310 985

Journal : 10822



Article : 9699

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ1	Kindly check and confirm the edit made in abstract.	
AQ2	Please supply the publisher location for Ref [26].	
AQ3	Please check and confirm the author names and initials are correct for references [26, 56, 58, 59].	