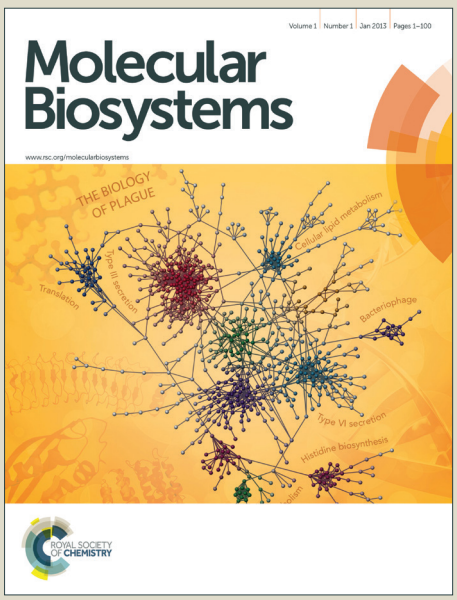
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Chemometric design to explore pharmacophore features of BACE inhibitors for controlling alzheimer's disease

Tabassum Hossain,[a] Arup Mukherjee,[a] and Achintya Saha*[a]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

The β-amyloid precursor protein cleavage enzyme (BACE) has been conceived to be an attractive therapeutic target to control alzheimer's disease (AD). Validated ligand-based pharmacophore mapping was combined with 3D QSAR modeling approaches that include CoMFA, CoMSIA and HQSAR techniques to identify structural and physico-chemical requirements for potential BACE inhibitor using a database containing 980 structuraly diverse compounds, assembled from different literatures. Structure-based docking technique was also used to validate the features obtained from the ligand-based models which were further used to screen the database of compounds designed through *de novo* approach. Contour maps of 3D QSAR models, CoMFA (R^2 = 0.880, *se* = 0.402, Q^2 = 0.596, R^2_{pred} = 0.713,) and CoMSIA (R^2 = 0.903, *se* = 0.362, Q^2 = 0.578, R^2_{pred} = 0.715), and pharmacophore space model (R^2 = 0.833, *rmsd* = 1.578, Q^2 = 0.845, R^2_{pred} = 0. 764) depict the models are robust and provide explanation of the important features, like steric, electrostatic, hydrophobic, positive ionization, hydrogen bond acceptor and donor, which play important role for interaction with the receptor site cavity. The HQSAR study (R^2 = 0.823, *se* = 0.488, Q^2 = 0.823, R^2_{pred} = 0.768) and *de novo* design which generate new fragments, illustrating the important molecular fingerprints for inhibition. The docking study elucidates the important interactions between the amino acid residues (Gly11, Thr72, Asp228, Gly230, Thr231, Arg235) at the catalytic site of the receptor and the ligand, indicating the structural requirements of the inhibitors. The *de novo* designed molecules were further screened for ADMET properties, and ligand-receptor interaction of top hits was analysed by molecular docking to explore pharmacophore features of BACE inhibitors.

Introduction

Alzheimer's disease (AD) is the degeneration of nervous system in brain, which worsens as it progresses. It affects especially old people, causing gradual loss of memory, thinking and other mental abilities, often leads to abnormal behavior. It is fundamentally caused by progressive neurodegenerative disorder, the most common form of dementia that affects about 6% of the population aged over 65 years[1]. In the past decade, massive research efforts have been directed toward understanding of β-amyloid precursor protein cleavage enzyme (BACE) as a critical target for AD therapy. The β–secretase is identified as the membrane–anchored aspartyl protease of BACE and found to be attractive therapeutic target of patients with sporadic AD. Currently there is no specific therapy available for treatment of this disease[2]. The etiology of AD remains unidentified until case-control studies demonstrated involvement of several risk factors, including age, familiar influence, depression, hypertension, diabetes, high cholesterol levels and atrial fibrillation as well as low physical and cognitive activity[3].

Cerebral deposition of amyloid β peptide (AβP) postulates an early and vital characteristic of AD[1]. The insoluble plaques are the principal collections of AβPs of 39–43 amino acids, formed via the sequential cleavage of β-amyloid precursor protein (APP) by aspartyl proteases, β- and γ-secretases[1]. Hence inhibition of one of these key proteases, β-secretase (BACE, β-site APP cleaving enzyme) may represent a modifying treatment for AD by blocking Aβ production[4].

In this study, structurally diverse compounds are considered for molecular modeling studies in order to explore the structural and physicochemical requirements to exhibit potential BACE inhibitors through 3D quantitative structure activity relationships (QSARs), hologram QSAR (HQSAR) and pharmacophore mapping studies. The popular 3D QSAR methods, comparative molecular field analysis (CoMFA)[5] and similarity analysis (CoMSIA)[6, 7] involve in the generation of a common 3D lattice around a set of molecules and calculate the steric and electrostatic interaction energies as well as similarity functions (Gaussian). The HQSAR study identifies important sub-structural features in the molecule that are significant for biological activity[8]. Pharmacophore map is the 3D spatial arrangement of the different chemical features of ligand for potential binding to the active site of the receptor. It can also be used as a query in 3D database search to identify new structural classes of potential lead compounds[9]. The *de novo* drug design, which can produce novel molecular entities without structural limitations, is also considered for modeling study. It produces highly efficient scaffolds with the required pharmacological profiles[10]. The

analyses are further correlated with the receptor-ligand interaction study at the active site cavity of BACE.

Materials and Methods

In order to generate suitable model, a desirable set comprising 980 structurally diverse compounds (Supp. file, Table S1) of BACE inhibitors (IC_{50}, 1 nM – 2.8 M) has been considered. The compounds were divided into training and multiple test sets by the k-means cluster analysis (k-MCA) method[11]. The k-MCA technique separates different descriptors into groups which arrange in clusters according to their Euclidian distances in multidimensional space based on both biological activity and descriptors. The reliability of the splitting method may be justified if the test set compounds are well distributed in all clusters. In the present work, k-MCA method divided the dataset into four clusters containing different members of training and test sets. From there, 423 compounds were selected randomly from each cluster for designing of the test set, remaining were considered in the training set for QSAR model developement. The selective BACE inhibitory concentration (IC_{50}), expressed in terms of pIC_{50} ($\log_{10}10^6$/ IC50) has been used as the dependent variable for the QSAR studies. Applicability domain has been checked, which is a theoretical approach for ensuring the predictability of the model property for the entire set of chemicals[12]. It is a theoretical region in a chemical space defined by the model descriptor and model response. Predictions for only those chemicals that fall into that space are considered reliable. Various statistical parameters[13-15], such as R^2_{pred}, r^2_m, Δr^2_m and s_p (standard error of prediction) of test set compounds have been used for validating the generated model. Further, all the ligand-based models are adjudged by docking in protein crystal structure (PDB code: 3OHH)[16] in order to analyze the ligand-receptor interactions in 3D space.

Application of 3D QSAR: Generation of CoMFA and CoMSIA models

The 3D QSAR, CoMFA model depicts the steric (s) and electrostatic (e) features[17] of the scaffold for selective inhibition of the target, whereas CoMSIA model helps to understand the hydrophobic (p), HB acceptor (a) and donor (d) features in addition to the 's' and 'e' requirement for inhibitory activity. The contour maps of both models are used to get general insights into the 3D topological features of the molecule for imparting bioactivity.

Proper alignment of ligands is a crusial factor for development of reliable 3D QSAR model[18]. For the development and subsequent validation of CoMFA/CoMSIA models, structurally diverse molecules of the training and test sets were aligned by shape alignment algorithm[19]. Initially, the individual molecules of the dataset were docked with the crystal structure of BACE (PDB code: 3OHH)[16], obtained from Protein Data Bank (PDB)[20], using LigandFit protocol of Discovery Studio 2.5 (DS)[21]. Then the best docked conformational posture at the active site of the receptor of individual ligands was aligned to each other by shape alignment algorithm[19]. These aligned molecules were further considered for development of 3D QSAR models using the Sybyl 7.2[22].

On a regular space grid of 3 Å, 's' and 'e' fields interactions were calculated for field analyses. For similarity analyses, a common probe atom on regularly placed grid points of pre-aligned molecules was employed[7]. Value of the generated fields has been truncated at 30.0 kcal/mol and Gasteiger-Huckle method has been used to calculate partial atomic charges[23]. Partial least square (PLS)[24] method has been used to develop QSAR models with field and similarity factors being the independent variables and inhibitory activity (pIC_{50}) as the dependent variable, respectively. The statistical parameters, such as R^2 (correlation coefficient of PLS analysis without validation), Q^2 (crossvalidated correlation coefficient by LOO method), *se* (standard error of estimate), r^2_m matrices and R^2_{bs} (bootstrapped correlation coefficient) have been used to judge the model quality. The models were further validated with four test sets for robustness.

Application of HQSAR: Generation of molecular finger prints

In HQSAR technique, fragment fingerprints or molecular holograms are employed to correlate predictive variables of structurally related data with biological activity[25]. Linear, branched, cyclic and the overlapping features of the molecular fragments are included in the extended form of fingerprints, known as molecular holograms. In this study, the HQSAR model has been derived on the basis of various combinations of fragment distinction and fragment generation parameters for each hologram length using Sybyl 7.2[22]. The selection of the HQSAR model has been carried out on the basis of best Q^2 and the cross-validated standard error (se_{cv}) value. Bin length, fragment size and specific fragment distinction parameters determine the optimum component number to obtain the best PLS HQSAR model. The optimal HQSAR model has been derived by screening through the 12 default hologram length values, which are a set of 12 prime numbers ranging from 53 to 401 fragment length and different fragment types. The developed HQSAR model is validated through prediction of activity of the four folds test compounds.

Pharmacophore mapping

The ligand-based pharmacophore model has been generated by hypogen algorithm[26] in DS[21]. The hydrogen bond (HB) acceptor (a) and donor (d), hydrophobic (p), positive ionization (i) and aromatic ring (r) features were used to optimise the pharmacophore mapping. The training set comprising 30 compounds, selected through k-MCA[11], has been used to generate the model. Maximum number of conformers have been kept as 250, generation type has been checked as best quality, spacing limitation between features has been fixed into 300 and threshold energy has been kept as 20 kcal/mol above the calculated global minimum. The different control parameters used for hypothesis generation (Hypogen process)[26] include uncertainty, weight variation and spacing (minimum inter feature distance for hypothesis). For the purpose of hypothesis optimization, the difference between total and null costs has been considered to be 60 bits[27]. The hyporefine process[26], where steric (ev) feature has been considered for bioactivity, was further used to nullify over-

prediction of inactive ligands in the generated hypothesis. The selected hypothesis has been validated through a cross-validation technique using CatScramble, based on Fischer's randomization test[28] by randomly reassigning activity value among the training set compounds. Four folds test sets were used for external validation of the model.

De novo design: Generation of new fragments

The *de novo* drug design involves searching an immense space of feasible drug-like molecules[10]. The molecule was designed and generated in LigBuilder 2.0[29]. Based on a "seed" structure database, a desired ideal molecule can be built within the binding pocket. A genetic algorithm[30] is used to grow ligands and to link building blocks for ligand generation. The *de novo* design approach has inherent potential to generate novel molecule that blocks a protein functional site[31]. Design of ligand involved the use of CAVITY, GROW and PROCESS modules step by step. The CAVITY module has been used for analyzing binding site and generating grid with pharmacophoric features essential to build ligand. Information of the grid file was used for the grow step[32]. The co-crystal ligand, which bound in the crystal structure of BACE enzyme (PDB: 3OHH)[16], has been taken as the reference structure for building the ligand. The seed ligand has been used to generate the pocket[29]. The genetic algorithm was used to control the process of building up the ligand through growing, linking and exploring[29]. The growing strategy started from a seed structure that has been preplaced into the binding pocket. This newly formed structure served as the seed for the next growing cycle. Compounds obtained from these seeds were subsequently submitted for evaluation of drug-likeness (ADMET) properties[33], such as MW, logPo/w, loghERG (potential hERG channel blockage leads to QT syndrome), logBB (the ability to cross the BBB), #metab (predicting the number of metabolic reactions likely to occur), and human oral absorption. Human Ether-a`-go–go-Related Gene (hERG) toxicity of the selected compounds was predicted using QikProp[34].

Docking interaction

Structure-based flexible molecular docking study has been performed to depict the important interactions between the ligands and the residues at the receptor's active site. Crystal structure of BACE enzyme is defined as the receptor. There are numerous (>250) x-ray crystal structures available in PDB[20]. Among them PDB id: 3OHH[16] was selected for docking analysis, based on resolution of the structure (< 2.5 Å) with bound potential co-crystal ligand (IC_{50} = 18 nM), and structural validation through self docking analysis (rmsd < 2). The docking study has been performed in LigandFit of 'Receptor-Ligand interactions' protocol of DS 2.5[21]. Pre-treatment processes before molecular docking include ligand preparation and defining active site of the receptor, i.e. binding site for the prepared ligands. Ionization change, tautomer and isomer generation have been used as constraint parameters in ligand preparation. The duplicate structures have been removed by Lipinski filter and 3D generator. The active site has been selected based on the ligand binding domain of the bound co-crystal ligand. In order to validate the docking protocols, self docking was also performed[35]. The docked receptor-ligand complex has been analyzed to investigate the type of interactions and to compare the dock score. Moreover, related crystal structures (PDB ids: 3K5C[36], 3K5D[37], 3K5F[37], 3K5G[37], 3DV5[38], 3MSJ[39], 3MSK[39], 3OHF[16], 3PI5[40], 3QBH[40]) with the bound ligand are compared for the interaction study.

Results and Discussion

The *in vivo* assay of piperazine sulfonamides as BACE inhibitor analogous[41] justify a relationship between binding affinity and IC_{50} values. The chemometric models for binding affinity to BACE have been already reported[42]. In the present work, important bio-phoric features required for BACE inhibitory activity have been explored using various ligand-based chemometric tools, followed by validation through structure-based docking analysis. The *de novo* design molecules were screened to expedite potential BACE inhibitors.

3D QSAR studies

The results of 3D QSAR (CoMFA/ CoMSIA) studies are listed in Table 1.

The CoMFA model

Combination of 's' and 'e' developed the best model (Model 1: n_{Tr} = 557, R^2= 0.880, *se*= 0.402, Q^2= 0.596, R^2_{bs} = 0.936, r^2_m = 0.880, Δr^2_m = 0.108) in CoMFA study, whereas individually 's' and 'e' features failed to develop validated contour maps. The contour map of Model 1 is depicted in Figure 1 (I). Validation has been performed by n-fold external validation technique. The predictive ability of the model with four-folds test compounds (R^2_{pred} = 0.708, 0.716, 0.703 and 0.724 with n_{Ts} = 100, 100, 100 and 123, respectively) justify the robustness of the model. The contributions of 's' and 'e' in Model 1 are found to be 46.70% and 53.30%, respectively, which indicate that both 's' and 'e' fields play essential roles. The predictiveness of the Model 1 is depicted in Figure 2 and also listed in Supp. file (Table S1).

In the field analysis, electrostatic fields expressed as blue (favorable) and red (unfavorable) contours represent 80% and 20% level of contributions for BACE inhibition, respectively. The highly active compounds (comp **560**) fitted in the contour map appropriately (Figure 1(I) A), whereas the least active compound (comp **118**) (Figure 1(I) B) does not fit the model at all. The region around the blue contour indicates that increased positive charges favor the activity, while negative charges near the red contours may also be favorable for activity. Presence of electron rich chains are favourable; while positive inductive effect of alkyl chain increases the positive charge, favourable for bioactivity as revealed by CoMFA electrostatic contour map. The regions of green contour (steric favorable) in Model 1 suggest that bulky substituent like macrocyclic ring in that position may improve the biological activity due to conformational rigidity, while the yellow region (steric unfavourable) indicates that an increased steric bulk is unfavorable for the activity.

Table 1

The CoMSIA model

In similarity studies, the best model is generated through 's', 'e', 'd', 'a' and 'p' factors (Model 2: n_{Tr} = 557, R^2= 0.903, *se*= 0.362,

Q^2= 0.578, R^2_{bs} = 0.947, r^2_m = 0.903, Δr^2_m = 0.088). The test compounds were divided into 4 sets and obtained the predictive statistics R^2_{pred} = 0.643, 0.752, 0.763 and 0.703 with n_{Ts} = 116, 107, 124, 76 respectively, justifies the robustness of the model. The contributions of 's', 'e', 'd', 'a' and 'p' in Model 2 are found to be 18.10%, 19.40%, 20.90%, 19.40% and 22.20% respectively, which indicate that 's', 'e', 'd', 'a', 'p' together play essential roles for the inhibition. The prediction of the compounds as per Model 2 is portrayed in Figure 2 and also listed in the Supp. file (Table S1).

In CoMSIA model, the contour maps show different favourable and unfavourable regions in the molecular area. Highly active compound (comp **560**, Figure 1(II) A) is fitted in the contour map very well, however the least active compound (comp **118**, Figure 1(II) B) could not able to map accurately with the contour. The green contour demonstrates the steric favourable region, whereas the yellow contour indicates the unfavorable steric zone. Blue contour around aromatic hydrocarbon suggests positive charged atmosphere is favorable for bio-activity due to its electron donating property. The red contour suggests that electron rich motifs, like hydroxyl with its highly electronegative oxygen, primary amine with its electron donating property and sulfonyl group are favourable for electrostatic activity. Heteroatoms along with alkyl chain possess positive inductive effect, and cyclic alkane for its non-polarity plays important role for hydrophobic interactions. Magenta contour describes promising HB donor favourable region at the nitrogen atom of the primary amines, however orange contour confirms the negative impact of the HB donating effect of sulfonyl group. The cyan contour suggests that heteroatoms are favorable for HB acceptors. The purple contour recommends the electron deficient atmosphere and alkyl chains are unfavorable for HB acceptors.

Figures 1
Figure 2

Hologram QSAR model

The analysis has been performed on the basis of different fragment length and the fragment distinctions, A/B/C/H/Ch/D-A[A (atom type), B (bond type), C (connectivity) H (hydrogens), Ch (chirality) and D-A (donor and acceptor)] of the training set compounds (n_{Tr} = 557) (Table 2). The most suitable fragment size was selected based on Q^2, low cross-validated standard error (se_{cv}) and the best combination of the fragment features. The fragment size and combination of fragment distinctions have been optimized for selection of significant hologram length, which was based on the PLS analysis that yields the lowest se_{cv} and the highest Q^2.

The best model is obtained by fragment size = 5 – 10, hologram length = 401, and optimum component number = 10 (Model 3: n_{Tr} = 557, R^2 = 0.823, se = 0.488, Q^2 = 0.823, r^2_m = 0.822, Δr^2_m = 0.155). The four folds test compounds obtained the R^2_{pred} = 0.850, 0.789, 0.745, and 0.687 with n_{Ts} = 100, 103, 103 and 117 respectively, indicated the good predictive ability of the model. The prediction as per Model 3 is depicted in Figure 2 and also listed in Supp. file (Table S1).

Contour maps of the HQSAR analysis (Figure 3) show the different colors of the atoms or fragments, which determine the overall contribution to the activity profiles of the molecules. The contributions of the different colors are – (i) red-orange color indicates a bad contribution ranging from -0.107 to -0.064, (ii) the white color indicates an average contribution ranging from −0.043 to 0.102, (iii) a yellow color indicates a good contribution of 0.102 to 0.153, and (iv) the green color signifies the maximum contribution of 0.255 or above. The most important green contribution is observed in the macrocyclic and phenyl rings (Figure 3A). Backbone alkyl chains also depict the average contribution as per the most active compound. The benzoimidazole scaffold of the least active compound is mapped by red-orange and yellow color (Figure 3B), indicating the most unfavorable fragment contributing to the activity and has negative impact towards the bio-activuty. The HB donor, acceptor and chiral atoms signify their contributions for potential BACE inhibition. The favorable and unfavorable fragments for BACE inhibitory activity are provided in the Supp. file (Figure S1).

Figure 3
Table 2

Pharmacophore mapping

In ligand-based pharmacophore mapping, the best hypogen hypothesis (hypo 5, Table 3) demonstrates 81% correlation for inhibition of BACE, but the hyporefine model demonstrates better correlation of 83% (hypo 10, Table 3) and prediction on the same parameters. The hyporefine model (Model 4) shows good null cost of 319.198, Δcost of 205.715 and *rmsd* of 1.578. The HB a and d, p, and i features along with steric influence (ev) obtained in Model 4 might function as prime bio-phores for the activity. The inter-feature distances of the pharmacophore features in 3D space are critical for selective inhibitory activity of BACE, represented in Figure 4A. The quality of the generated hypothesis, adjudged through Fischer's randomization test[28] at 99% confidence, indicates the superiority of the hypothesis considered. For external predictability, the test set compounds (nTs = 950) were divided into 4 sets of which R^2_{pred} was obtained in the range of 0.689–0.849, indicated the robustness of the model. The observed and predicted inhibitory activity (pIC_{50}) of the training and test sets compounds have been represented in Figure 2 and also tabulated in supp. file (Table S1). Oxygen atom of sulfonyl group behaves as HB acceptor for its lone pair of electron and has significant contribution to inhibitory action. Electronegative atoms, oxygen and nitrogen of the hydroxyl group and secondary amines, respectively behave as HB donors at the receptor site cavity. Acyclic ring and the alkyl chains play important role as hydrophobic fragments for their non-polarity. Electron donating hetero groups confirm the importance of positively ionised feature and favour the bio-activity. The highly active compound (comp **560**) is mapped thoroughly with all the features in the model (Figure 4B), but the least active compound (comp **118**) is fitted with only three features of the model (Figure 4C). Presence of excluded volumes in between the pharmaphoric features suggest that steric contributions in the cavity sites are essential for inhibitory activity. The model 4 accounts 10.82 % compounds as active in the dataset (n = 980), 77.86 % compounds accurately predicted, 4.80 % compounds underestimated and 17.35 % compounds over-estimated, justify the reliability of the model for virtual screening.

Figure 4
Table 3

***De novo* design**

The *de novo* design involves development of those fragments[10] which contribute their importance in ligand-based drug design and show their interaction at the enzyme's catalytic site cavity. Three seed structures (Figure 5) have been obtained from GROW module, depict the importance of 3 specific fragments. These fragments were further used to figure new hit molecules. These hit molecules when used for docking, one of the seeds, benzene (S1) and its associated chain, show the interaction with Arg128 and Asp228 of the receptor. Second seed, piperidine-1-carbaldehyde (S2) having interaction with the Thr72 at the catalytic site. Third seed, 3-(1, 2-Dihydro-pyridin-3-yl)-4,5-dihydro-1H-pyrrolo[3,2-c]pyridine (S3) has an interaction with Gly11 at the catalytic site cavity. These molecules (Supp. file, Figure S2) also passed the ADMET[34] analysis. Screening includes properties, such as logPo/w (-2.0-6.5), loghERG (<-5), logBBB (-3.0-1.2), #metab (1-8), human oral absorption, and Jorgensen's rule of three that includes aqueous solubility >-5, Caco-2 cell permeability >22 nm/s, and primary metabolites less than 7, justify acceptance of the molecules. The fragments or seeds can be successfully used to design new hit molecules with oral absorptivity and less toxicity.

Figure 5

Binding interaction

Docking study has been adjudged through the docking score and the binding interactions at the receptor site. Most of the compounds have shown similar binding orientations and interactions. The highly active molecule (comp **560**) has good affinity toward BACE in terms of docking score (Supp. file, Table S1). But the least active compound (comp **118**) neither participated in docking interactions nor provided any recordable dock score. In Figure 6A, the comp **560** has been portrayed for the analysis at the catalytic cleft of crystal structure of the receptor (PDB code: 3OHH)[16]. The amino acids Ser10, Gln12, Asp32, Ser35, Tyr71, Thr72, Lys107, Trp115, Tyr198, Asp228, Ser229, Thr231, Thr232, Arg235 (polar amino acids), and Gly11, Gly13, Leu30, Gly34, Val69, Pro70, Gly73, Gly74, Phe108, Ile110, Ile118 and Gly230 (non-polar amino acids) are responsible for important interactions at the active site cavity (within 4 Å). The interactions are observed in amino group of Arg235 and sulfonyl group of propane -1- sulfonic acid methyl-amide chain at a distance of 1.723 Å, carboxyl group of Asp228 and two different hydrogen of amino group (N_{27}) at a distances of 2.404 Å and 1.032 Å respectively, keto group of Gly230 and a hydroxyl group associated with C_{25} at a distance of 2.182 Å of the compound, confirm the prominent HB interactions. But the comp **118** does not show any interactions with the amino acids (Figure 6B). The Asp228 and Gly230 formed HB interactions with the originally bound ligand in the catalytic site. Related X-ray crystal structures of BACE were also compared for the interactions of docked pose of the highly active compounds (comp **560**) of the dataset. The crystal structures 3K5C[36], 3K5D[37], 3K5F[37], 3K5G[37], 3DV5[38], 3MSJ[39], 3MSK[39], 3OHF[16], 3PI5[40], 3QBH[40] showed similar kind of interactions with the ligands which are already bound with them, compared to the highly active compound with 3OHH[16]. The residues Asp228 and Gly230 are the most common catalytic amino acids, which interact with the bound ligands. The bound ligand of the PDB structure 3K5C[36] is structurally analog to comp **560**, which possesses similar kinds of interactions that include hydrogen bond interactions of Asp228 and Gly32 with amino group and hydroxyl groups of the ligands. These interactions justify that the docked poses are reliable for understanding crucial interactions in the binding site.

The features of QSARs, pharmacophore space mapping and *de novo* models are adjudged through all these interaction sites. Importance of the acceptor functionality of heteroatom in Model 4 is substantiated by HB interaction with Arg235. The nucleophilic substituent that imparts electron donating property in Model 2 is important through corroborating two HB interactions of Thr72 and hydroxyl group. It is also observed that aromatic ring is important for steric property both in Model 1 and 2. Importance of phenyl ring along with alkyl chain in Models 2, 3 and 4 confirm the significance of hydrophobicity through possessing the inductive effects. The HB interactions of amino functional group with Asp228 and Thr231 adjudge the significance of HB donor feature in Models 2 and 4. All these models establish that these chemical features and groups are prime structural features, and can be effectively utilized to formulate the BACE inhibitory hit compound that may potentially inhibit the β-secreatase and limit the Aβ precipitation.

Figure 6

The *de novo* molecules also showed good interactions with the catalytic amino acids of the receptor cavity, like Arg128, Asp228, Thr72 and Gly11(Supp. Figure S2). The molecules obtained from the fragments also succeeded to pass the ADMET as well as the blood brain barrier analyses. It implicates that the explored fragments (Figure 5) can be successfully used for designing new orally absorb and less toxic hit molecules in the future.

Conclusions

From both ligand– and structure–based studies, it can be concluded that aromatic ring and associated alkyl chain, sulfonyl alongwith alkyl and methyl-amide chains, amide backbone and macrocyclic ring are prime molecular scaffolds for imparting BACE inhibitory activity. The significance of the *de novo* designed seeds, benzene, piperidine-1-carbaldehyde and 3-(1, 2-Dihydro-pyridin-3-yl)-4,5-dihydro-1H-pyrrolo [3,2-c] pyridine fragments has been adjudged through the pharmacophore features obtained from the ligand–based models, and the designed hit molecules are showing good interactions pattern at the catalytic site of the receptor. The ADMET screening and BBB penetration property also proved the drug likeness of the hit molecules. The study depicts that HB donor and acceptor, hydrophobic, electrostatic and steric properties of ligands are essential pharmacophores (Figure 7) for BACE inhibitory activity.

Figure 7

Acknowledgments

Financial assistanceship for Major Research Project (MRP) from University Grants Commission (UGC) is thankfully acknowledged. One of the authors, Tabassum Hossain wishes to thank UGC-MANF for awarding Junior Research Fellowship.

Notes and references

[a] *Department of Chemical Technology, University of Calcutta, 92, A.P.C. Road, Kolkata-700009, India Fax: +91-33-2351-9755; Tel: +91-33-2350-8386; E-mail: achintya_saha@yahoo.com*

† Electronic Supplementary Information (ESI) available: Figure S1 and S2, Table S1. See DOI: 10.1039/b000000x/

1. R. Vassar, B. D. Bennett, S. Babu-Khan, S. Kahn, E. A. Mendiaz, P. Denis, D. B. Teplow, S. Ross, P. Amarante, R. Loeloff, Y. Luo, S. Fisher, J. Fuller, S. Edenson, J. Lile, M. A. Jarosinski, A. L. Biere, E. Curran, T. Burgess, J. C. Louis, F. Collins, J. Treanor, G. Rogers and M. Citron, *Science, New York, N.Y*, 1999, **286**, 735-741.
2. S. J. Choi, J. H. Cho, I. Im, S. D. Lee, J. Y. Jang, Y. M. Oh, Y. K. Jung, E. S. Jeon and Y. C. Kim, *Eur. J. Med. Chem.*, 2010, **45**, 2578-2590.
3. R. Cacabelos, *Methods Mol. Biol., Clifton, N.J*, 2008, **448**, 213-357.
4. G. Tresadern, F. Delgado, O. Delgado, H. Gijsen, G. J. Macdonald, D. Moechars, F. Rombouts, R. Alexander, J. Spurlino, M. Van Gool, J. A. Vega and A. A. Trabanco, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 7255-7260.
5. U. Norinder, *Recent Progress in CoMFA Methodology and Related Techniques*, Springer Netherlands, Sweden, 2002.
6. S. Yu, P. Wang, Y. Li, Y. Liu and G. Zhao, *SAR QSAR Environ. Res.*, 2013, **24**, 819-839.
7. M. Bohm, J. St rzebecher and G. Klebe, *J. Med. Chem.*, 1999, **42**, 458-477.
8. C. R. Rodrigues, T. M. Flaherty, C. Springer, J. H. McKerrow and F. E. Cohen, *Bioorg. Med. Chem. Lett.*, 2002, **12**, 1537-1541.
9. O. Guner, *Pharmacophore Perception, development and Use in Drug Design*, International University Line, La Jolla, CA, 2000.
10. Y. Yuan, J. Pei and L. Lai, *J. Chem. Inf.Model.*, 2011, **51**, 1083-1091.
11. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, *IEEE T. Pattern. Anal.*, 2002, 24, 881-892.
12. P. Gramatica, *QSAR Comb. Scie.*, 2007, **26**, 694–701.
13. A. Tropsha and S. X. Wang, *Ernst Schering Foundation symposium proceedings*, 2006, 49-73.
14. P. P. Roy, and K. Roy, *QSAR Comb. Scie.*, 2008, **27**, 302-313.
15. K. Roy, P. Chakraborty, I. Mitra, P. K. Ojha, S. Kar and R. N. Das, *J. Comp. Chem.*, 2013, **34**, 1071-1082.
16. L. R. Marcin, M. A. Higgins, F. C. Zusi, Y. Zhang, M. F. Dee, M. F. Parker, J. K. Muckelbauer, D. M. Camac, P. E. Morin, V. Ramamurthy, A. J. Tebben, K. A. Lentz, J. E. Grace, J. A. Marcinkeviciene, L. M. Kopcho, C. R. Burton, D. M. Barten, J. H. Toyn, J. E. Meredith, C. F. Albright, J. J. Bronson, J. E. Macor and L. A. Thompson, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 537-541.
17. A. K. Debnath, *Mini-rev. Med. Chem.*, 2001, **1**, 187-195.
18. G. F. Yang and X. Huang, *Curr. Pharm. Des.*, 2006, **12**, 4601-4611.
19. C. M. Venkatachalam, X. Jiang, T. Oldfield and M. Waldman, *J. Mol. Graphics Modell.*, 2003, **21**, 289-307.
20. *RCSB Protein Data Bank* (accessed October 2013).
21. a) LigandFit, b) Pharmacophore; *DISCOVERY STUDIO 2.5*, (2009) Accelrys Software Inc., San Diego.
22. *SYBYL 7.2*, (2006) *Tripos Inc.*.
23. R. H. Kroemer, P. Liedl, KR., *J. Comp. Chem.*, 1996, **17**, 1296–1308.
24. V. Vinzi, L. Trinchera and S. Amato, *PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement*, Springer, Napoli, 2010.
25. M. R. Doddareddy, Y. J. Lee, Y. S. Cho, K. I. Choi, H. Y. Koh and A. N. Pae, *Bioorg. Med. Chem.*, 2004, **12**, 3815-3824.
26. H. Li, J. Sutter and R. Hoffmann, in *Pharmacophore perception, development, and use in drug design*, ed. O. Guener, International University Line, La Jolla, CA, 2000.
27. D. Schuster, C. Laggner, T. M. Steindl, A. Palusczak, R. W. Hartmann and T. Langer, *J. Chem. Inf. Model.*, 2006, **46**, 1301-1311.
28. J. A. Jacquez and G. M. Jacquez, *Math. Biosci.*, 2002, **180**, 23-28.
29. P. Kare, J. Bhat and M. E. Sobhia, *Mol. Div.*, 2013, **17**, 111-122.
30. S. C. Pegg, J. J. Haresco and I. D. Kuntz, *J. Comput. Aid. Mol. Des.*, 2001, **15**, 911-933.
31. A. Ishchenko, Z. Liu, P. Lindblom, G. Wu, K. C. Jim, R. D. Gregg, D. A. Claremon and S. B. Singh, *J. Chem. Inf. Model.*, 2012, **52**, 2089-2097.
32. S. B. Sagar, K. Roy, A. Saha, *Med. Chem. Res.*, 2014, in press.
33. M. Brustle, B. Beck, T. Schindler, W. King, T. Mitchell and T. Clark, *J. Med. Chem.*, 2002, **45**, 3345-3355.
34. Schrodinger, *QikProp 3.8*, (2013) Schrodinger LLC, New York.
35. M. O. Taha, M. Habash, Z. Al-Hadidi, A. Al-Bakri, K. Younis and S. Sisan, *J. Chem. Inf. Model.*, 2011, **51**, 647-669.
36. A. Lerchner, R. Machauer, C. Betschart, S. Veenstra, H. Rueeger, C. McCarthy, M. Tintelnot-Blomley, A. L. Jaton, S. Rabe, S. Desrayaud, A. Enz, M. Staufenbiel, P. Paganetti, J. M. Rondeau and U. Neumann, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 603-607.
37. S. Hanessian, Z. Shao, C. Betschart, J. M. Rondeau, U. Neumann and M. Tintelnot-Blomley, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 1924-1927.
38. R. Machauer, K. Laumen, S. Veenstra, J. M. Rondeau, M. Tintelnot-Blomley, C. Betschart, A. L. Jaton, S. Desrayaud, M. Staufenbiel, S. Rabe, P. Paganetti and U. Neumann, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 1366-1370.
39. J. Madden, J. R. Dod, R. Godemann, J. Kraemer, M. Smith, M. Biniszkiewicz, D. J. Hallett, J. Barker, J. D. Dyekjaer and T. Hesterkamp, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 5329-5333.
40. H. Rueeger, J. M. Rondeau, C. McCarthy, H. Mobitz, M. Tintelnot-Blomley, U. Neumann and S. Desrayaud, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 1942-1947.
41. J. Cumming, S. Babu, Y. Huang, C. Carrol, X. Chen, L. Favreau, W. Greenlee, T. Guo, M. Kennedy, R. Kuvelkar, T. Le, G. Li, N. McHugh, P. Orth, L. Ozgur, E. Parker, K. Saionz, A. Stamford, C. Strickland, D. Tadesse, J. Voigt, L. Zhang and Q. Zhang, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 2837-2842.
42. T. Hossain, M. A. Islam, R. Pal, A. Saha, *Med. Chem. Res*, 2013, **22**, 4766–4774.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

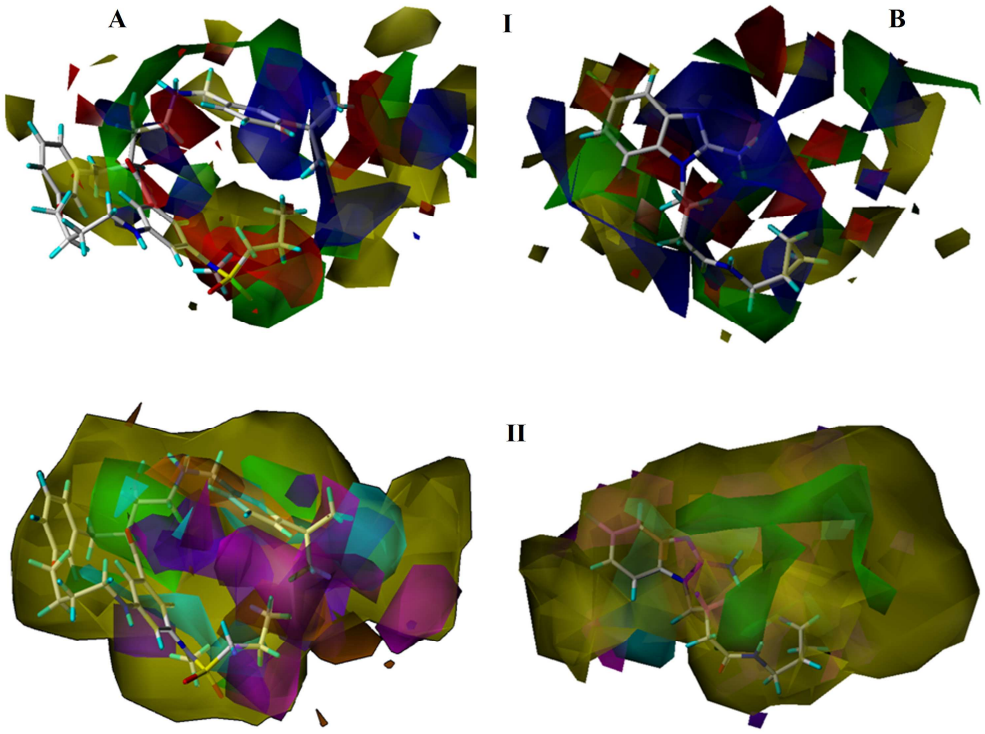


Figure 1 (I) CoMFA and (II) CoMSIA models mapped with (A) most active (comp **560**) and (B) least active (comp **118**) compounds.

Steric: *Green* favorable, *yellow* unfavorable; Electrostatic: *Blue* favorable, *red* unfavorable. (C) & (D) Acceptor: *Cyan* favorable, *purple* unfavorable; Donor: *Magenta* favorable, *orange* unfavorable; Hydrophobic: *pink* favorable, *white* unfavorable.

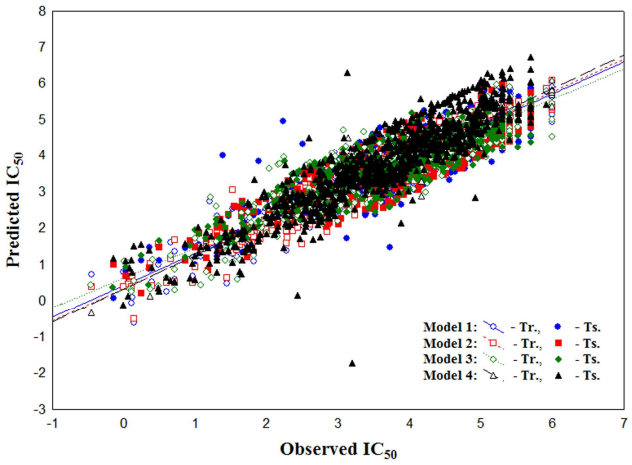
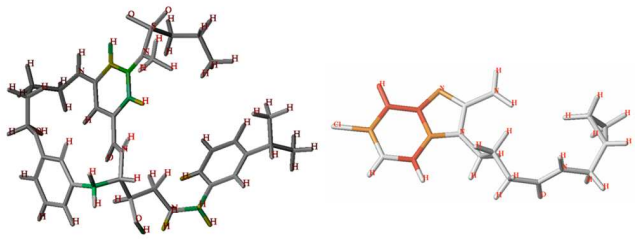


Figure 2 Observed versus predicted binding affinity as per QSAR and pharmacophore models of BACE inhibitors



A B

Figure 3 HQSAR contribution map of (A) comp **560** (most active) and (B) comp **118** (least active)

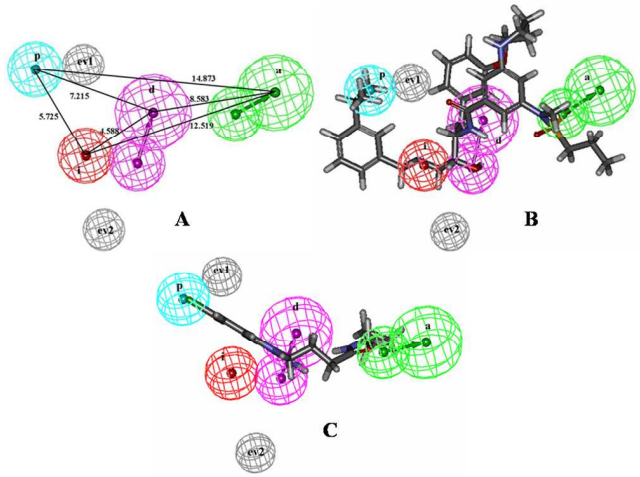
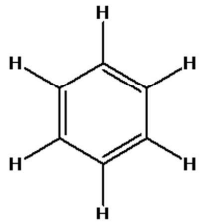
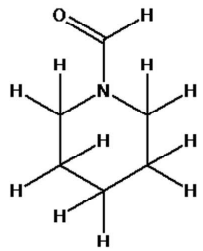


Figure 4 (A) Pharmacophoric features with inter - feature distance (Å) of Model 4, mapped with (B) most active and (C) least active compounds.

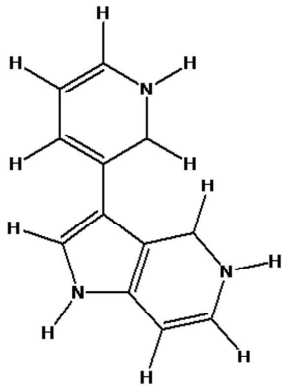
Features are HB acceptor (a), HB donor (d), hydrophobic (p), positive ionization (i) and excluded volume (ev).



Benzene (S1)

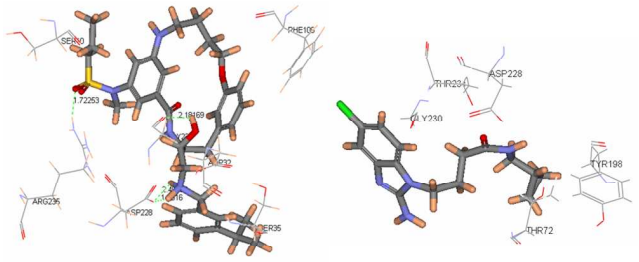


Piperidine-1-carbaldehyde (S2)



3-(1,2-Dihydro-pyridin-3-yl)-4,5-dihydro-1*H*-pyrrolo[3,2-*c*]pyridine (S3)

Figure 5 Important fragments obtained through *de novo* design



A B

Figure 6 Molecular docking interactions at the binding site with the (A) comp **560** and (B) comp **118**

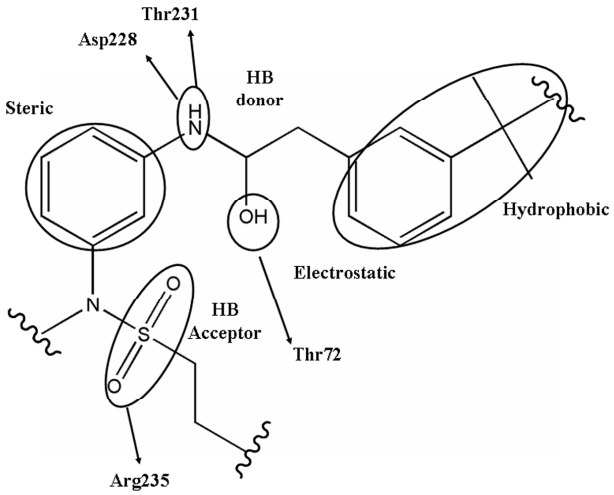


Figure 7 Schematic representation of pharmacophore features of potent BACE inhibitors

Table 1 Statistical parameters of QSAR and pharmacophore models

Study	Parameters	Models		
		1	2	3
QSAR	Components	10	10	10
	n_{tr}	557	557	557
	R^2	0.880	0.903	0.823
	se	0.402	0.362	0.488
	F (df)	401.914	510.381	-
	Q^2	0.596	0.578	0.823
	R^2_{bs}	0.936	0.947	-
	s_{bs}	0.298	0.270	-
	r^2_{m}	0.880	0.903	0.823
	$r^2_{\mathrm{m}(rev)}$	0.772	0.815	0.668
	$r^2_{\mathrm{m}(avg)}$	0.826	0.859	0.745
	Δr^2_{m}	0.108	0.088	0.155
	n_{ts}	423	423	423
	R^2_{pred}	0.713	0.715	0.768
	s_p	0.597	0.589	0.580
	r^2_{m}	0.694	0.723	0.688
	$r^2_{\mathrm{m}(rev)}$	0.509	0.499	0.449
	$r^2_{\mathrm{m}(avg)}$	0.602	0.611	0.569
	Δr^2_{m}	0.184	0.225	0.239
	Contribution (%)			
	s	46.70	18.10	-
	e	53.30	19.40	-
	d	-	20.90	-
	a	-	19.40	-
	p	-	22.20	-

Model 4		
Pharmacophore	n_{tr}	30
	R^2	0.833
	rmsd	1.578
	Q^2	0.845
	Cost analysis	
	Config.	15.949
	Null	319.198
	Δ	205.715
	Output features	a, d, p, i, 2xev
	r^2_{m}	0.790
	$r^2_{\mathrm{m}(rev)}$	0.785
	$r^2_{\mathrm{m}(avg)}$	0.788
	Δr^2_{m}	0.005
	n_{ts}	950
	R^2_{pred}	0.764
	s_p	0.509
	r^2_{m}	0.707
	$r^2_{\mathrm{m}(rev)}$	0.754
	$r^2_{\mathrm{m}(avg)}$	0.729
	Δr^2_{m}	0.048

Table 2 Results of HQSAR analysis

Fragment size	Fragment distinction	Q^2	se_{cv}	R^2	se	Component	Hologram length
4-7	A/B/C/D&A	0.820	0.492	0.820	0.492	10	401
1-12	A/B/C/D&A	0.827	0.483	0.827	0.483	10	401
5-10	A/B/C/D&A	0.823	0.488	0.823	0.488	10	401
1-8	A/B/C/D&A	0.816	0.497	0.816	0.497	10	353
3-9	A/B/C/D&A	0.819	0.494	0.819	0.494	10	401
4-8	A/B/C/D&A	0.814	0.500	0.814	0.500	10	353

Table 3Hypothesis parameters observed in successive runs

Hypothesis No.	Uncertainty	Weight variance	Spacing (pm)	Pharmacophore features in generated hypothesis	Cost		R^2	*rmsd*
					Null	Δcost		
1	2.5	0.302	200	a, d, p, i	319.198	147.439	0.814	1.706
2	1.5	0.302	200	a, d, p, i	1213.600	879.882	0.811	3.803
3	2.5	1.500	200	a, d, p, i	319.198	158.914	0.794	1.752
4	2.5	2.000	200	a, d, p, i	319.198	163.465	0.814	1.667
5	1.5	1.000	250	a, d, p, i	1213.600	887.676	0.795	3.953
6*	2.5	2.000	200	a, d, p, i, 2ev	319.198	205.715	0.833	1.578