β-secretase inhibitors for Alzheimer's disease: identification using pharmacoinformatics

*Md Ataul Islam[1,2], Tahir S. Pillay[*1,3]*

[1]*Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria and National Health Laboratory Service Tshwane Academic Division, Pretoria, South Africa.*
[2]*School of Health Sciences, University of Kwazulu-Natal, Westville Campus, Durban, South Africa.*
[3]*Division of Chemical Pathology, University of Cape Town, South Africa.*

Email IDs: MA Islam: ataul.islam80@gmail.com
TS Pillay: tspillay@gmail.com

This work was supported by the National Research Foundation (NRF) Innovative post-doctoral fellowship of University of Pretoria, South Africa

Correspondence should be addressed to T.S. Pillay, Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia, Pretoria, 0007
Email: tspillay@gmail.com
Phone: +27-123192114
Fax: +27-123283600

Abstract

In this study we searched for potential β-site amyloid precursor protein cleaving enzyme1 (BACE1) inhibitors using pharmacoinformatics. A large dataset containing 7155 known BACE1 inhibitors were evaluated for pharmacophore model generation. The final model (R = 0.950, $RMSD$ = 1.094, Q^2 = 0.901, se = 0.332, r^2_m = 0.901, R^2_{pred} = 0.756, sp = 0.468, $r_m{}^2_{\text{-}test}$ = 0.667) was revealed with the importance of spatial arrangement of hydrogen bond acceptor and donor, hydrophobicity and aromatic ring features. The validated model was then used to search NCI and InterBioscreen databases for promising BACE1 inhibitors. The initial hits from both databases were sorted using a number of criteria and finally three molecules from each database were considered for further validation using molecular docking and molecular dynamics studies. Different protonation states of Asp32 and Asp228 dyad were analysed and best protonated form used for molecular docking study. Observation of the number of binding interactions in the molecular docking study supported the potential of these molecules being promising inhibitors. Values of RMSD, RMSF, Rg in molecular dynamics study and binding energies unquestionably explained that final screened molecules formed stable complexes inside the receptor cavity of BACE1. Hence, it can be concluded that the final screened six compounds may be potential therapeutic agents for Alzheimer's disease.

Keywords: BACE1, BACE1 inhibitors, Pharmacophore, Virtual screening, Molecular Docking, Molecular dynamics

Introduction

Alzheimer's disease (AD) is an incurable age-related neurodegenerative syndrome of the central nervous system which alters the mental capacity(Ghosh & Osswald, 2014) resulting in senile dementia, loss of memory, disorientation, difficulty in speaking or writing, loss of reasoning skills etc.(Selkoe, 2001). The disease was first identified more than hundred years ago by Alois Alzheimer, but about seventy years passed before it was recognized as the most common cause of dementia, as well as a major cause of death(Albert et al., 2011). According to World Health Organization (WHO), in United States an estimated 5.4 million people had AD in 2016. Globally, about 44 million people are living with AD or a related dementia in 2016, while only 1 out of 4 people with AD have been diagnosed. It is observed that AD is very common in Western Europe followed by North America. Statistics indicate that AD prevalence is increasing exponentially which will result in the number of affected people doubling by 2030 and tripling by 2050(*World Health Organization and Alzheimer's Disease International. Dementia: a public health priority*, 2012). Moreover, with progression of AD, other age-related risk factors also develop including hypertension, dyslipidaemia, metabolic syndrome and diabetes(Castello & Soriano, 2013, 2014; Drachman, 2014). To date there are no effective curative agents for the treatment of such a devastating disease except for management of symptoms(Ghosh & Osswald, 2014).

Accumulation and deposition of amyloid β (Aβ) is the most popular and accepted hypothesis for the development of AD(Vassar et al., 1999). Aβ is a neurotoxic species produced by the consecutive cleavage of β-amyloid precursor protein (APP) by two aspartyl protease, beta-site APP cleaving enzyme1 (BACE1) and finally by γ secretase(Cui et al., 2011; Sinha et al., 1999; Vassar et al., 1999; Yan et al., 1999; Zhang, Thompson, Zhang, & Xu, 2011). This generates oxidative stress in the brain, changingthe cortico-cortical connectivity that leads to termination of the cerebral cortex and ultimate death of brain cells(Jellinger & Bancher, 1998). BACE1 is the trans-membrane aspartic protease and predominantly found in neurons. This enzyme is behaved like membrane-bound or related to membrane protein as it proficiently cut membrane-bound substrate. Several studies proved that BACE1 is an important drug target and active site of the enzyme covered by a flexible antiparallel β-hairpin, called a flap(Hussain et al., 2000). It is elucidated that the flap can control substrate access to the receptor site and set the substrate into the perfect orientation for the catalytic process(Lin et al., 2000). Hence, the inhibition of proteases such as BACE1 may represent modifying treatment for AD by controlling the production of Aβ(Tresadern et al., 2011). BACE1 has already been recognized as an important drug target for the treatment of AD(Butini et al., 2013; Eketjall et al., 2013; Zou et al., 2013). Therefore BACE1 inhibitors may be used to treat AD.

In this study, pharmacophore-based virtual screening of BACE1 inhibitors was performed to identify potential and less toxic chemical entities from the National Cancer Institute (NCI)(NCI, 2013) and InterBioScreen (IBS)(IBS, 2015) databases for the treatment of AD. Initial hit molecules from the databases were screened by imposing several criteria and the final proposed molecules were subjected to molecular docking study to observe binding interactions between molecules and catalytic amino residues at the active site cavity of BACE1. Finally, the docked complexes of final screened compounds and most active compound of the dataset were used for molecular dynamics simulation and binding energy calculation studies.

Materials and method

The 3D QSAR pharmacophore study is one of the most extensively used and versatile techniques to identify novel chemical entities for various targets. Pharmacophore modelling can mainly be categorised into two ways viz. ligand-based and structure-based. In the present study ligand-based pharmacophore modelling approach was considered for a set of BACE1 inhibitors with inhibitory activity (K_i). Among several commercially available tools the Discovery Studio 2016 (DS)(BIOVIA, 2016) is one of the important pharmacoinformatics tools for the development of the 3D QSAR model. The DS was used for the pharmacophore, virtual screening, molecular docking and binding energy calculation studies whereas Gromacs 5.1.2(Abrahama et al., 2015) used for molecular dynamics simulation. The DS contains several module packages widely used in pharmacoinformatics drug discovery(Al-Balas et al., 2013; Amin, Adhikari, Jha, & Gayen, 2016; Amin, Bhargava, Adhikari, Gayen, & Jha, 2017; Chhabria, Brahmkshatriya, Mahajan, Darji, & Shah, 2012; Huang et al., 2012; Middha et al., 2013; Pavadai et al., 2016). The *3D QSAR Pharmacophore Generation* module enables the use of structure and activity data for a set of potential BACE1 ligands to create hypotheses. Gromacs is a freely available versatile package to perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles.

Dataset

A total 7155 compounds belonging to a collection of BACE1 inhibitors were downloaded from Binding DB (http://www.bindingdb.org/) with inhibition constant (K_i) activity in nM range. 2206 molecules were found to be duplicates and were deleted. Molecules with no activity and without definite activity were removed and it was found that 379 molecules possessed biological activity. Further the Lipinski's rule of five(Lipinski, Lombardo, Dominy, & Feeney, 2001) and Viber's(Veber et al., 2002) rules were checked and it was observed that 155 molecules failed to satisfy both rules. Therefore the remaining 224

molecules were considered for the study. The molecules of the dataset have a wide range of K_i, from 0.494 to 82600.000 nM. The experimental inhibitory activity (K_i) of entire dataset were converted into logarithm value [$pK_i = \log((1/K_i)x10^7)$].

The whole dataset was divided into three categories on the basis of inhibitory activities values; highly active (pK_i >5.0 nM, +++), moderately active (3.0 < $pK_i \leq$ 5.0 nM, ++) and least active/inactive (3.0 nM $\leq$ pK_i, +). Basic strategies of Li et al.(H. Li, Sutter, & Hoffman, 1999) were followed to select the training set molecules. The guidelines described as (a) molecules should be selected to provide clear and brief information with structure features and range of activity, (b) at least 16 diverse molecules for training set should be considered to ensure the statistical significance and avoid chance correlation, (c) the training set must include the most and the least active molecules and (d) the biological activity data of the molecules should have spanned at least 4 orders of magnitude. Following the above guidelines, the whole dataset was randomly divided into eight training sets (Tr1, Tr2, Tr3, ……, Tr8) containing 30 compounds each, except Tr8 which contains 28 molecules. It was also kept in mind that no compounds were common in any two training sets except for the most active and least active molecules. The remaining 294 compounds were taken as test set molecules for each training set (Ts1, Ts2, Ts3, ……., Ts8) and used for assessing the performance of pharmacophore model. The observed and predicted activity along with SMILES representation of Tr1, Tr2, Tr3, Tr4, Tr5, Tr6 and Tr8 are given in Table S2 (Supplementary file) whilst 2D chemical representations of Tr7 with inhibition constant values are depicted in Figure 1. The 2D/3D visualizer of DS was used to generate 3D coordinates of the compounds. For each compound, the coordinates were corrected, atoms were typed and energy was minimized using the modified CHARMm force field.(Brooks et al., 1983; Momany & Rone, 1992) The several packages of DS were used for pharmacophore, virtual screening, molecular docking and binding energy calculation studies.

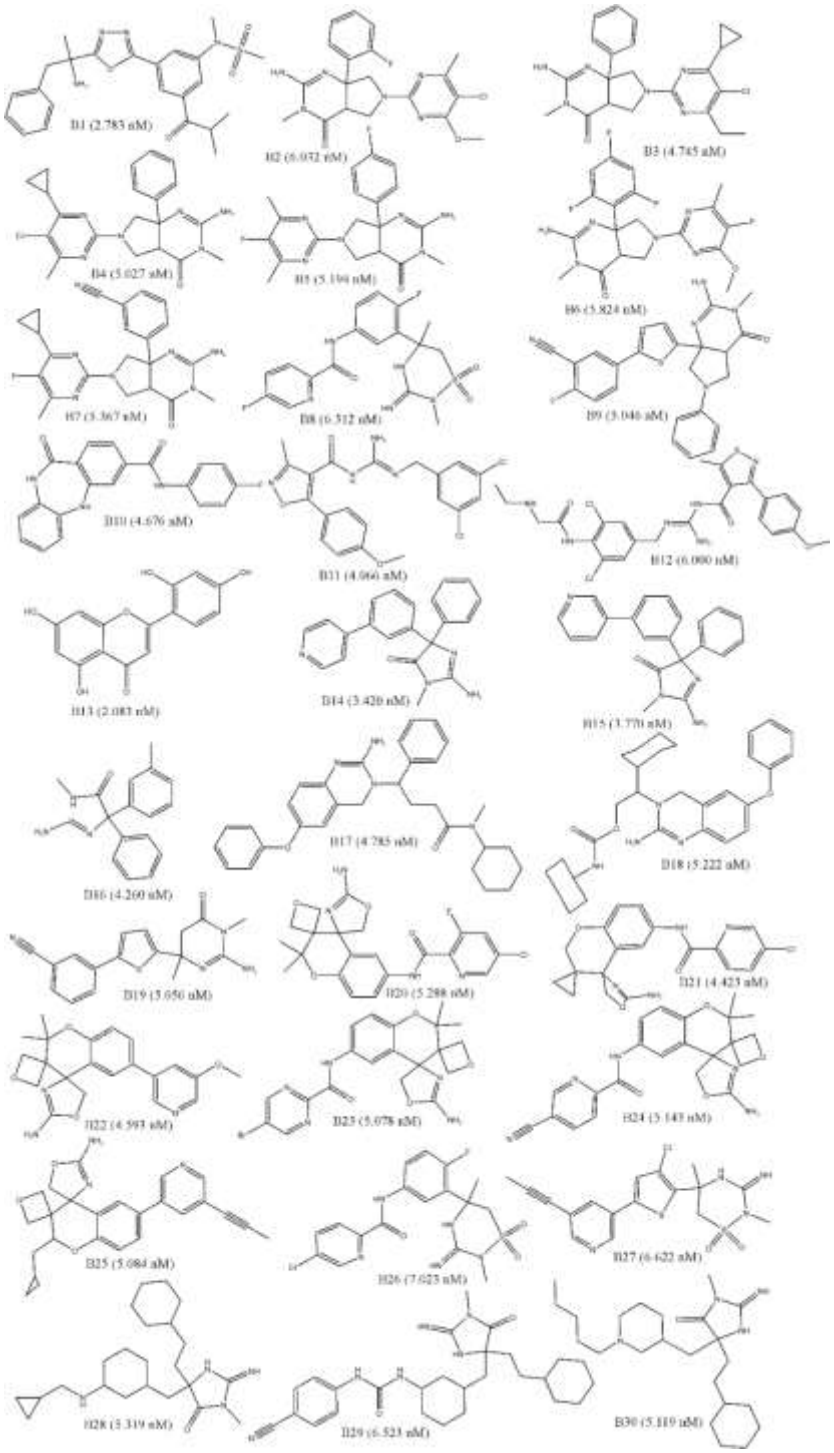

Figure 1. *2D chemical structures of the training set compounds and the inhibition constant values (pK_i) are given in the parentheses. ($pK_i = log[(1/K_i)x10^7]$).*

Pharmacophore model generation

In order to select the best training set molecules, the *3D QSAR Pharmacophore Model Generation* module of DS was used and the pharmacophore models were developed from the each of eight training sets (Tr1, Tr2, Tr3, ……, Tr8). Conformations of each training set compounds were generated by *Cat-*

Conf program of the DS software package. Among BEST/FAST, the BEST method was adopted to achieve multiple acceptable conformations that provide complete and enhanced coverage of conformational space with help of rigorous energy minimization and optimizing the conformations by the poling algorithm(Smellie, Teig, & Towbin, 1995). In the BEST algorithm, the chemical features are arranged in space instead of simply by the arrangement of atoms (Kristam, Gillet, Lewis, & Thorner, 2005). For prediction of the favourable features for the highly active compounds of the dataset the *Feature mapping* protocol was considered. It was observed that hydrogen bond (HB) acceptor ('a') and donor ('d'), hydrophobic ('p') and aromatic ring ('r') were crucial pharmacophoric features present in the BACE1 inhibitors. Mapped features were given as input features for pharmacophore model generation. With the help of conformers along with chemical features the modules operates in two modes such as *HipHop* and *HypoGen*. The *HipHop* approach generates the pharmacophore models by using active compounds only, while the *HypoGen* approach considers both active and inactive compounds in order to find hypotheses which are common in the active molecules and absent in the inactive compounds(Kristam et al., 2005). The top ten hypotheses were generated by *HypoGen* with consideration of the training set, conformational models and chemical features through three steps: constructive, subtractive and optimization(Sadler, Cho, Ishaq, Chae, & Korach, 1998). In the first step, hypotheses are generated that are common in the most active compounds; in subtractive phase, inactive compounds are removed from those that fit the hypotheses. In final step, the remaining hypotheses improve the score with help of small perturbations (Kristam et al., 2005; H. Li, Sutter, & Hoffmann, 2000). The best hypothesis was selected based on the best correlation coefficient (*R*), low root mean square deviation (*RMSD*), cost function analysis and good predictive ability.

All eight training sets models were utilized separately to obtain statistically robust pharmacophore models. Furthermore, the pharmacophore models developed from each training set were used to predict the biological activity of corresponding test compounds. All training set molecules are depicted in the supplementary file.

Validation of pharmacophore model

The validation of any pharmacoinformatics model is an essential and crucial step to check the predictivity and applicability as well as the robustness of the model. The pharmacophore hypotheses developed from all training sets were validated by four different methods, (1) internal validation, (2) cost function analysis, (3) test set prediction and (4) decoy set. Moreover, the model developed from best training set molecules was also validated by Fischer's randomization test.

Internal validation

The process of cross-validation using the training molecules used to develop the model is one of the important internal validation protocols. This is known as the Leave-one out (LOO) cross-validation technique, in which one compound is randomly deleted from the training set in each cycle and the model redeveloped using the rest of the compounds with the same parameters used in original model. The model developed with the new set is used to predict the biological activity of deleted molecule. All molecules of the training set were predicted as per above procedure and predicted activity recorded. The predicted activity of training set compounds based on the LOO method were used to calculate , the LOO cross-validated correlation coefficient (Q^2) and error of estimation (*se*). As described in previous reports high Q^2 (>0.5) and low *se* explain better predictive ability (Kubinyi, Hamprecht, & Mietzner, 1998) of the model. Further, the modified r^2 ($r^2_{m(LOO)}$) and Δr^2_m reported by Roy *et al.*(Ojha, Mitra, Das, & Roy, 2011; K. Roy et al., 2012) were also calculated to confirm the good predictive ability of the training set molecules. Both parameters measure the degree of deviation of the predicted activity from the observed ones. It was stated that model may be considered with $r^2_{m(LOO)}>0.5$ and $\Delta r^2_m<0.2$.

Test set prediction

In order to check the external prediction capability of the model developed from all eight training sets, the biological activity of corresponding test set (194 compounds) were predicted using the *Ligand Pharmacophore Mapping* protocol in DS. The quality of prediction of each pharmacophore model was adjudged based on statistical parameters, R^2_{pred} (correlation coefficient) and *sp* (error of prediction)(Golbraikh & Tropsha, 2002; Mitra, Saha, & Roy, 2010). The parameter, R^2_{pred} relies on the mean actual activity of the training set molecules. As both R^2_{pred} and *sp* depend on mean value, it might be achieved for compounds with a wider range of activity value, but this may not be assured that the predicted biological activity values are very close to those actual activity. Therefore, instead of a good overall correlation being maintained, there is chance of a significant numerical difference between the two values. In order to better indicate the prediction capability of the pharmacophore model, modified r^2 [$r^2_{m(test)}$](P. P. Roy, Paul, Mitra, & Roy, 2009; P. P. Roy & Roy, 2008) values were calculated (threshold value=0.5).

Decoy set

Decoy set validation is the approach to verify efficiency of screening capability of the selected pharmacophore model. Decoys are small compounds that are supposed to be inactive against the receptor or are not likely to interact with the target i.e. this method verifies how the pharmacophore model can retrieve active molecules over inactive molecules on virtual screening from a combined set of active and

inactive molecules. Hence, a set of decoys was generated by online DecoyFinder2.0(Cereto-Massague et al., 2012) tool. DecoyFinder tool selects decoys based on five parameters including molecular weight, number of rotational bonds, hydrogen-bond donor count, hydrogen-bond acceptor count, and the octanol–water partition coefficient of the active inhibitors. In order to distinguish decoys and active molecules chemically the MACCS fingerprints were calculated based on the maximum Tanimoto coefficient values. The amalgamated decoys and active inhibitors were screened by best pharmacophore selected from each training set. Screened molecules based on pharmacophore model were ranked on basis of fit value. Based on screening results different statistical parameters were calculated to validate the model. The enrichment factor (EF) denotes total known active inhibitors retrieved from the part of screened database. In present study, EF (1%) was calculated from top 1% hits. The Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC) was calculated which gives the significance of the dataset screening. The BEDROC is a comprehensive form of receiver operating characteristic (ROC), which recognises problems in the screening method. Calculation of the enrichment factor and BEDROC are as described by Bhayye et. al(Bhayye, Roy, & Saha, 2016).

Fischer's randomization test

The best model from final selected training set was used to check the strong relationship between the chemical compound and the biological activity of the training set molecules with the help of the Fischer's randomization test. In this method, the activity value was scrambled and assigned new values to the compound. Using compounds with new activity value, pharmacophore hypotheses were generated using the original pharmacophoric features and constraints used to generate the original pharmacophore hypotheses. If the randomization run generates improved correlation coefficient and/or better statistical parameters than the original hypothesis may be considered to be developed by chance. Different number of spreadsheets are generated based on the statistical significance randomization run. The statistical significance is given by following equation.

$$Significance = [1-(1+a)/b] \quad (1)$$

Where, *a* represents the number of hypotheses with a total cost less than the best hypothesis, whereas *b* denotes a collection of *HypoGen* runs and random runs. For example, at 95% confidence level, the total number of random spreadsheets are generated as 19 (b = 20) and each generated spreadsheet is submitted to *HypoGen* using the same parameters as the initial run. In the present study, the developed pharmacophore model was tested at 99% confidence level which produced 99 spreadsheets.

Virtual screening

The identification of potential small ligands from molecular databases is known as virtual screening and is one of the crucial techniques in the pharmacoinformatics approach. The best model from the best training set molecules was used to screen the NCI and IBS databases to discover novel chemical molecules that could act as BACE1 inhibitors. The NCI database contains 265,242 compounds while IBS database consist of 523,421 molecules. Best pharmacophore model was submitted to the NCI and IBS databases with set 'Limit Hits' as 'Best N' and 'Maximum Hits' as 'All'. The initial hit molecules from both databases were filtered separately with a number of criteria to find final promising BACE1 inhibitors. Moreover, molecular docking study was carried out to analyse binding interactions between the potential BACE1 inhibitors and catalytic residues of the active site. Finally, molecular dynamics analysis of docked complexes of final screened compounds along with most active compound of the dataset was performed.

Molecular docking

Molecular docking is one of the crucial filtering approaches and an important method in the drug discovery pipeline. The *LigandFit* protocol of DS was adopted in order to determine how the screened drug-like virtual hits bind to the receptor through a molecular docking approach. First of all the *LigandFit* detects the cavity to find out and select the region of the protein as the active site followed by fitting the ligands to the selected active site. In order to find out the active site, the 3D regular grids of points were employed. The crystal structure of BACE1 enzyme was downloaded from RCSB Protein Data Bank (RCSB-PDB) for the molecular docking study. It has been elucidated that for molecular docking study, protein structures may have low resolution (<2.5Å) and R-factor (<0.28)(Anderson, 2003). In this study, the structure PDB ID: 4X7I(May et al., 2015) was selected from several BACE1 structures keeping in mind the above criteria along with receptor size and date of deposit. The resolution and R-factor of the selected enzyme are found to be 1.770Å and 0.208 respectively. Prior to molecular docking protein and ligands were prepared using the *Prepare Protein* and *Prepare Ligand* tools of DS respectively. The CHARMm force field(Vanommeslaeghe et al., 2010) was used to minimize both protein and ligands. In case of protein preparation by *Prepare Protein* module of DS the 'Build Loop' and 'Protonate' parameters were fixed to 'True' while, dielectric constant, pH, ionic strengths and energy cut-off were taken as default value. While, for *Prepare Ligand* module, preparation 'Change ionization', 'Generate Tautomers' and 'Generate isomers' were considered as 'False', and 'Generate Coordinates' was set to '3D'. The active site cavity was identified on the prepared protein on the basis of volume occupied by the bound ligand at the receptor site. The validation of docking protocol prior to docking is an important step

to avoid false positive results of molecular docking. For this purpose the co-crystal small molecule in the PDB complex file was initially redrawn and the same docked into the active site of BACE1 (PDB ID: 4X7I(May et al., 2015)). The binding interactions between the best docked pose of co-crystalized ligand was explored followed by superimposing the docked pose and the co-crystal. To verify docking parameters that were capable of regenerating a comparable conformation to that of the co-crystal at the active site of BACE1, the RMSD value was recorded. After validation, the same parameters of co-crystalized docking were adopted for molecular docking studies of potential molecules. For the analysis of binding interactions and dock score values, top ten poses for each ligand were considered.

BACE1 active site contains Asp dyad of two aspartate amino residues such as Asp32 and Asp228(Hong et al., 2000). Site directed mutagenesis studies revealed that BACE1 loses potency on mutation of either one of the aspartate dyad(Hussain et al., 1999). Asp32 and Asp228 behave as acid-base by frequently altering their protonation states(Barman, Schurer, & Prabhakar, 2011). Therefore, the determination of the exact protonation states of aspartate dyad is crucial to explore the binding interactions of ligands towards BACE1. Several studies have been done to explore the role of dyad Asp32 and Asp228 on catalytic activity of the BACE1. Park and Lee explained that a protonated state of OD1 (inner un-protonated hydrogen atom) in Asp32 was energetically more favourable compare to protonated state of OD1 in Asp228(Park & Lee, 2003). In another study Rajamani and Reynolds reported that protonated state of OD1 in Asp228 at high pH; and, protonated state of OD2 (outer un-protonated hydrogen atom) in Asp32 and OD1 in Asp228 at low pH were crucial for BACE1 for the same inhibitor(Rajamani & Reynolds, 2004). Merz et al. performed QM/MM refinement technique on the eight different protonation states and reported that protonation state of OD1 in Asp32 was the most important state in the presence of considered inhibitor (Yu et al., 2006). Polgár and Keserü reported that protonation state of OD1 in Asp32 is the most suitable protonation state for virtual screening (Polgar & Keseru, 2005). In a recent study by Ellis et. al. reported that the BACE1 flap is flexible and occupies pH-dependent conformational state and Asp32 visibly acts as the acid in peptide hydrolysis, while Asp228 acts as the base in active pH(Ellis, Tsai, Lin, & Shen, 2017). From the above observations it was unambiguously revealed that the chemical nature of the small molecules played a significant role to determine the protonation state of the Asp dyad in BACE1.

In order to check the important protonated state of the Asp dyad, different combinations of the protonation states of oxygen atoms in Asp32 and Asp228 were generated using the Yet Another Scientific Artificial Reality Application (YASARA)(Krieger & Vriend, 2002) tool. Detailed different protonation states are given in Figure S1 (supplementary file). According to the Figure S1, Prot1 represents un-protonated state of Asp32-Asp228 dyad. OD1 of Asp32; and, OD1, OD2 and OD1-OD2 atoms of Asp228 were protonated for Prot2, Prot3 and Prot4 respectively. Similarly, OD2 of Asp32; and, OD1, OD2 and

OD1-OD2 of Asp228 were protonated to generate Prot5, Prot6 and Prot7 respectively. In case of Prot8 and Prot9 both oxygen atoms of Asp32 were protonated along with OD1 and OD2 of Asp228 protonated for Prot8 and Prot9 respectively. Finally, Prot10 was generated by protonation of both oxygen atoms of each Asp32 and Asp228 dyad. All protein molecules (Prot1, Prot2,…Prot10) were used for docking using final proposed molecules and most active compound of the dataset to find critical protonation state of Asp32-Asp228 dyad. The best docked complex of all screened molecules including most active compound were considered for MD study for 20ns of time span. The protocol MD simulation is explained somewhere else(Islam & Pillay, 2017).

Free energy calculation using MM-PBSA

In order to quantitatively measure the binding strength between enzyme and ligands, the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA)(Kumari, Kumar, Open Source Drug Discovery, & Lynn, 2014) method was adopted through Gromacs version 5.0.6. The binding free energy ($\Delta G_{binding}$) can be expressed as below.

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand}) \quad (2)$$

where, $G_{complex}$ is the combined free energy of the protein-ligand complex, $G_{protein}$ and G_{ligand} are the total energy of individual enzyme and ligand in solvent, respectively. Individual free energy of each $G_{complex}$, $G_{protein}$ and G_{ligand} can be calculated as,

$$G_{complex} = E_{MM} + G_{solvation} \quad (3)$$

$$G_{protein} = E_{MM} + G_{solvation} \quad (4)$$

$$G_{ligand} = E_{MM} + G_{solvation} \quad (5)$$

where, E_{MM} is the average molecular mechanics potential energy in vacuum and $G_{solvation}$ is the free energy of solvation. The E_{MM} can be calculated using the following equation:

$$E_{MM} = E_{bonded} + E_{non-bonded} = E_{bonded} + (E_{vdw} + E_{elec}) \quad (6)$$

where, E_{bonded} is bonded interactions including bond length, angle, dihedral angle and improper interactions. $E_{non\text{-}bonded}$ represents two terms, van der Waals (E_{vdw}) and electrostatic (E_{elec}) interactions. The ΔE_{bonded} is always considered as zero.

Results and discussion

Pharmacophore

All training set molecules were considered for pharmacophore model development using the *HypoGen* module of DS. In order to get statistically robust models the different input parameters ‘Spacing’, ‘Uncertainty’ and ‘Weight variation’ were varied. The best model of each training set was selected based

on Debnath's analysis(Debnath, 2002, 2003) which describes that the best model should have the low RMSD, high correlation coefficient, low cost value and high cost difference. It is also reported that a well validated hypothesis should have the overall cost of the hypothesis distant from the null cost and close to the fixed cost, and differences between null cost and total known as *Δcost* in the range of 40–60 bits explains the probability of the predictive correlation of 75–90%, while the cost difference more than 60 bits defines the hypothesis and has a correlation probability of more than 90%(Sakkiah, Thangapandian, John, Kwon, & Lee, 2010). The best model of all training sets is reported in the Table S1 in supplementary file. The best model of each set was used to predict the inhibitory activities of corresponding test set compounds and the statistical parameters were calculated (Table S1 in supplementary file).

In detail, it was observed that the correlation values (*R*) of all training sets were found to be from 0.865 to 0.950. The highest correlation value was observed in case of Tr7. Table 1 explained that the highest LOO correlation coefficient values was found for Tr7 (Q^2 = 0.901) and lowest for Tr8 (Q^2 = 0.767). Further r^2_m and Δr^2_m values were calculated for all training sets. It was observed that r^2_m and Δr^2_m were found to be 0.837 and 0.439; 0.757 and 0.189; 0.792 and 0.176; 0.889 and 0.093; 0.713 and 0.228; 0.881 and 0.109; 0.901 and 0.092; and, 0.765 and 0.198 for Tr1, Tr2, Tr3, Tr4, Tr5, Tr6, Tr7 and Tr8 respectively. The *Δcost* (Null cost – Total cost) were recorded and it was found that all training sets had values greater than 60 except Tr6 which explained that except Tr6 *Hypo1* of all training sets were not generated by chance. Observed and predicted activity (pK_i) of all training sets except Tr7 are depicted in Figure S2 (Supplementary file).

Table 1. Statistical parameters of pharmacophore models derived from Training set 7

Run No.	Spacing	[1]Unc.	[2]Wt. Var.	[3]Hypo. No.	[4]*R*	*Rmsd*	Costs				Out Features
							Null	Fixed	Δ	[5]Config.	
1	3.0	3.0	0.3	*Hypo1*	0.925	0.854	174.566	120.626	41.36	16.596	a, d, p, r
2	2.5	3.0	0.3	*Hypo1*	0.894	0.994	174.566	120.996	48.314	16.966	d, 2xp
3	2.0	3.0	0.3	*Hypo1*	0.930	0.846	174.566	120.699	39.541	16.669	a, 2xd, p
4	1.5	3.0	0.3	*Hypo1*	0.911	0.913	174.566	120.740	41.317	16.710	a, 2xd, p
5	1.0	3.0	0.3	*Hypo1*	0.904	0.950	174.566	120.749	40.098	16.719	2xd, 2xp
6	0.5	3.0	0.3	*Hypo1*	0.918	0.897	174.566	120.964	39.148	16.934	a, 2xd, p
7	3.0	2.5	0.3	*Hypo1*	0.914	1.080	201.353	115.182	67.605	16.600	a, 2xd, p
8	3.0	2.0	0.3	*Hypo1*	0.936	1.248	271.132	106.822	137.485	16.609	2xa, d, p
9	3.0	1.5	0.3	*Hypo1*	0.934	2.113	611.772	90.010	442.067	16.883	a, d, p, r
10	3.0	2.0	0.4	*Hypo1*	0.899	1.541	272.132	107.077	128.557	16.609	a, 2xd, p
11	3.0	2.0	1.0	*Hypo1*	0.950	1.094	272.132	107.424	145.882	16.609	a, d, p, r
12	3.0	2.0	1.5	*Hypo1*	0.928	1.306	272.132	107.626	138.513	16.609	a, d, p, r
13	3.0	2.0	2.0	*Hypo1*	0.920	1.373	272.132	107.770	135.881	16.609	a, 2xd, p
14	3.0	2.0	2.5	*Hypo1*	0.907	1.481	272.132	107.882	131.282	16.609	a, 2xd, p

[1]Uncertaunty; [2]Weight variation; [3]Hypothesis number; [4]Correlation coefficient; [5]Configuration cost

Furthermore, the corresponding test compounds of each training set were predicted. The statistical data are given in the Table S1. It was observed that few molecules in each training set were either extremely over estimated or extremely underestimated. Therefore these molecules were considered as outliers and deleted to calculate the R^2_{pred}, *sp*, $r^2_{m\text{-}test}$ and $\Delta r^2_{m\text{-}test}$. In case of Ts1, Ts2, Ts3, Ts4, Ts5, Ts6, Ts7 and Ts8 the number of outliers were found to be 14, 9, 7, 13, 12, 8, 8 and 7 respectively. The highest value of both R^2_{pred} and $r^2_{m\text{-}test}$ were found to be 0.756 and 0.667 for Ts7 (Test of Tr7). The $\Delta r^2_{m\text{-}test}$ value of all test sets satisfies the criteria (<0.20). The observed and predicted activity (pK_i) of all test sets except Ts7 are plotted and given in Figure S3 (Supplementary file).

Moreover, the best model of each training set was used to screen the amalgamation of 1448 decoy and 30 active compounds. Based on the screening results different parameters were recorded and depicted in Table S3. The AUC curve was also plotted and it found that value under curve was maximum in case of Tr7 (0.680). As shown in Table 2, all test sets have AUC value more than 0.500 except Ts1, Ts2 and Ts4. The accuracy of all test sets were found more than 0.700 while maximum value was found for Ts2 (0.970) followed by Ts7 (0.960). The highest specificity was found for Ts2 as 0.979 followed by of Ts7 (0.967), but the specificity of Ts2 was 0.400 whereas for Ts7 it was 0.700.

Therefore by analysing statistical parameters of all training, test and decoy sets it was observed that model developed from Tr7 was found to be more competent in case of predictivity and screening capability compared to models developed from other training sets. Hence, the model developed from Tr7 was deliberated as the best model and considered for further validation and virtual screening. Statistical and cost parameters of several runs on Tr7 are given in Table 1. The observed and predicted activities of all training and test sets (except Tr7 and Ts7) are given in Figures S1 and S2 respectively. The structure of Tr7 is depicted in Figure 1 and inhibition constant given within parentheses.

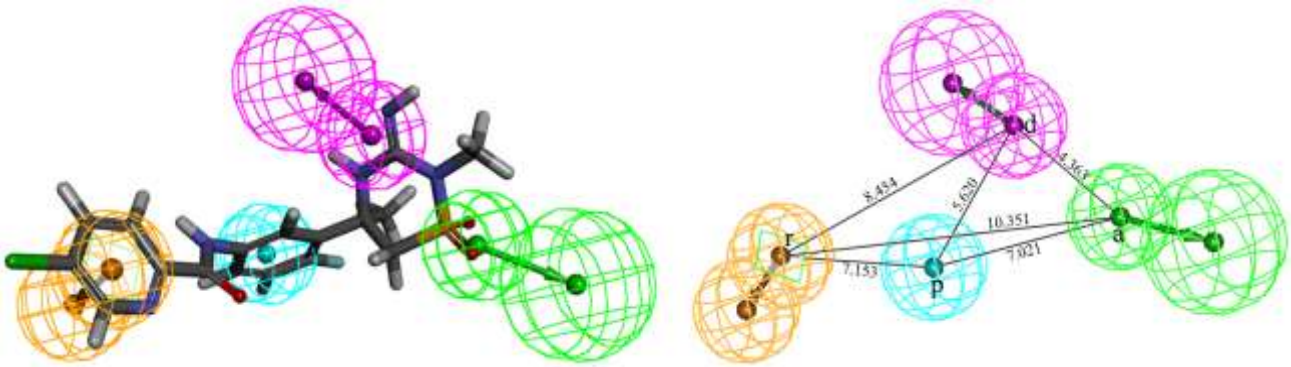

Figure 2. *Mapped pharmacophore features with most active compound and the inter-feature distances of Hypo 1 of run no. 11 (Table 1) of the best model of Tr7.*

The detailed stepwise development of pharmacophore model from Tr7 with successive variation of input parameters is given in Table 2. From the Table 2 it can be seen that *Hypo1* of run number 11 was found to have the highest correlation value and high cost difference. Hence, this model was considered as the best model and considered for further validation. Mapping between thebest model and most active compound of the dataset along with inter-features distances are given in Figure 2. The best model suggested (Figure 2a) that one of each hydrogen bond (HB) acceptor and donor, hydrophobicity and aromatic ring features were crucial for inhibitory activity. The most active compound of the dataset was mapped with the pharmacophore model. The mapped features explained that sulphonamide group present in the molecule behaved as HB acceptor whereas imino group present in the same ring was revealed as HB donor. The benzene ring present in between two non-aromatic rings imparts the hydrophobicity of the molecule. The pyridine ring in the molecular system was found to be crucial as an aromatic ring feature. Therefore, from the best model it was revealed that pharmacophoric features namely, HB acceptor and donor, hydrophobicity and aromatic ring are critical for the design and discovery of new potential BACE1 inhibitors.

The observed and predicted activity of individual compounds of Tr7 set molecules are depicted in Table 2 and Figure 3. Table 2 explained that one highly active compound (**B25**) (pK_i > 5 nM, +++) was underestimated as moderately active molecule, while two moderately active (**B3** and **B7**) molecules (3 < $pK_i \leq 5$ nM, ++) overestimated as highly active. The remaining molecules in the training set were found to be estimated correctly within their range. Therefore, from the above findings it can be explained that the *Hypo 1* of run number 11 (Table 1) projected inhibitory activity of the training set molecules correctly which is echoed by the high correlation between experimental and predicted inhibitory activities.

Validation of pharmacophore model

In order to check the robustness of best pharmacophore model following validation procedures were adopted.

Internal validation

The *Hypo 1* of the best selected model was used to calculate the predicted inhibitory activities. The observed and predicted activities of the training set are given Table 2 and Figure 3. The ratio between the observed and predicted activities were calculated which is represents as error value of all training molecules and reflects the consistency between both activities. The error values of each molecules were found in reasonable range (Table 2). Furthermore, the cross-validated correlation coefficient was calculated and it was observed that the best hypothesis gives Q^2 of 0.901 with *se* of 0.332. Moreover, r^2_m and Δr^2_m parameters were also calculated and found to be 0.901 and 0.092 respectively. It is reported that

for acceptance of the model r^2_m and Δr^2_m should be more than 0.500 and less than 0.200 respectively. The high Q^2 and r^2_m, and low *se* and Δr^2_m of the *Hypo 1* (run no. 11 in Table 1) suggested that model is statistically robust in nature.

Cost value analysis

During the generation of pharmacophore hypotheses the *HypoGen* algorithm in DS computes and gives a number of parameters for preliminary assessment of the model. These parameters includes the Δ_{cost} which is the difference between null and total costs, the configuration cost, and RMSD between the estimated and the experimental inhibitory activities of the training set molecules. It was observed that the cost difference of the best model was found to be 145.882 which clearly explains that the selected hypothesis was not generated by chance. A consistent and robust pharmacophore model should also have a configuration cost value less than 17. *Hypo 1* (Table 1) generated configuration cost of 16.609. It was also previously reported that lower differences between total and fixed costs is an indication of a robust model. For *Hypo 1* of best model the difference between total and fixed cost was found to be 12.580 and this is significant for the model.

Test set prediction

It is crucial to check the predictive ability of compounds not involved in model generation. In this purpose *Pharmacophore Mapping* module of DS was used to predict inhibitory activities of the 194 test set compounds (Ts7). After prediction it was found that eight compounds were either extremely overestimated or underestimated and these molecules were deleted as outliers for the statistical parameter calculations.

The remaining 186 test compounds in SMILES representation along with pK_i values is given in the supplementary file (Table S4). From the observed and predicted (Table S4 and Figure 3) activity of test compounds it was revealed that five highly active and two moderately active molecules were overestimated as moderately active and least active respectively. Four moderately active compounds were underestimated as highly active molecules. Another two moderately active overestimated as least active compounds. Hence out of 186 compounds eleven compounds were either underestimated or overestimated from their original activity, whilst the remaining 175 were estimated correctly within their range. Furthermore the correlation coefficients (R^2_{pred}) between observed and predicted inhibitory activities and error of prediction (*sp*) were calculated and values found to be 0.756 and 0.468 respectively. The above observations clearly indicate that *Hypo 1* of run number 11 (Table 1) is competent enough to estimate the inhibitory biological activity of the compounds beyond the training set.

Table 2. Observed, predicted activities values of the training set molecules (Tr7), obtained using the pharmacophore model *Hypo 1* of run number 11

Mol. No.	Activity (pK_i nM)		Error	Activity scale	
	[a]Obs.	[b]Pred.		[a]Obs.	[b]Pred.
B1	2.783	2.704	+1.029	+	+
B2	6.032	5.302	+1.138	+++	+++
B3	4.745	5.300	-0.895	++	+++
B4	5.027	5.270	-0.954	+++	+++
B5	5.194	5.401	-0.962	+++	+++
B6	5.824	5.405	+1.078	+++	+++
B7	5.367	5.467	-0.982	+++	+++
B8	6.312	6.380	-0.989	+++	+++
B9	5.046	5.016	+1.006	+++	+++
B10	4.676	4.870	-0.960	++	++
B11	4.066	4.386	-0.927	++	++
B12	6.000	5.187	+1.157	+++	+++
B13	2.083	2.329	-0.894	+	+
B14	3.420	3.896	-0.878	++	++
B15	3.770	3.651	+1.032	++	++
B16	4.260	4.271	-0.997	++	++
B17	4.785	5.074	-0.943	++	+++
B18	5.222	5.291	-0.987	+++	+++
B19	5.056	5.139	-0.984	+++	+++
B20	5.288	5.047	+1.048	+++	+++
B21	4.423	4.764	-0.928	++	++
B22	4.593	4.325	+1.062	++	++
B23	5.078	5.185	-0.979	+++	+++
B24	5.143	5.062	+1.016	+++	+++
B25	5.084	4.661	+1.091	+++	++
B26	7.023	7.343	-0.956	+++	+++
B27	6.622	6.147	+1.077	+++	+++
B28	5.319	5.087	+1.045	+++	+++
B29	6.523	6.877	-0.948	+++	+++
B30	5.119	5.236	-0.978	+++	+++

[a]Observed. [b]Predicted. $pK_i = \log[(1/K_i)x10^7]$

Further, to confirm better determination of predictive ability of the selected model another important statistical parameter $r^2_{m(test)}$ was calculated which explains how the predicted inhibitory activities are contiguous to the equivalent experimental values as a high correlation coefficient value (R^2_{pred}) cannot always put forward a low residual between the experimental and predicted activity data. In this regard, two parameters $r^2_{m(test)}$ and $\Delta r^2_{m(test)}$ were identified and values found to be 0.667 and 0.079 correspondingly which explains that selected hypothesis (*Hypo 1*) has adequate predictive potential.

Consequently, the above findings explained that the selected hypothesis can reasonably predict the biological activities of new molecules.

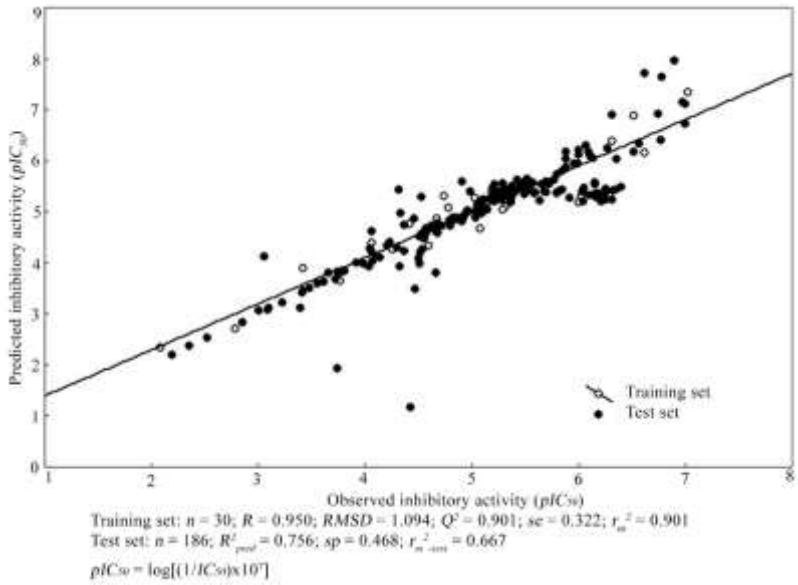

Figure 3. *Observed and predicted activity of training and test set compounds as per pharmacophore model.*

Fischer randomization test

Fischer randomization test was performed on the selected best model to evaluate the quality of the hypothesis by assigning a particular confidence level. In the present work, the model was considered for 99% confidence level. In this process a total of 99 random spreadsheets (random hypotheses) were generated according to the equation (1) to accomplish a confidence level of 99%. The total costs and the correlation values of all 99 spreadsheets were analysed. Total cost of top 25 random spreadsheets along with best model is given in Figure 4. In addition, the lowest total cost and highest correlation value of all 99 randomized runs and selected model were plotted and given in Figures S3 and S4 (Supplementary file) respectively. From Figures 4, S3 and S4 it can be noted that no one hypothesis perceived higher correlation and lower total cost value compare to *Hypo 1* of run number 11 in Table 1. Moreover it can be claimed that the statistics of the best model was far more superior to the top 25 random hypotheses as well as the other 74 random hypotheses. Therefore, the above observations of Fischer's randomization approach unquestionably explained superiority of the hypothesis and *Hypo 1* of run number 11 was not generated by chance.

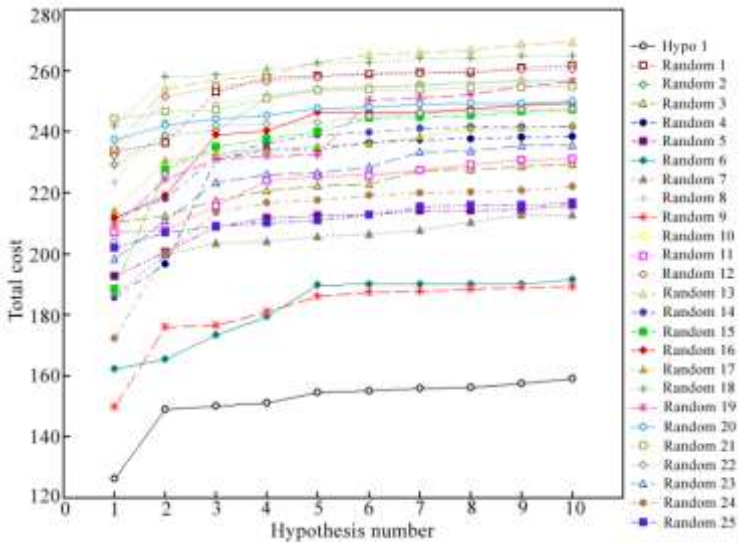

Figure 4. *Total cost of top 25 randomized runs in Fischer's randomization test.*

Decoy set

The decoy set validation was performed to verify the screening competence of the best selected pharmacophore model. A set of 30 active BACE1 inhibitors was used as input for DecoyFinder2.0(Cereto-Massague et al., 2012) to retrieve decoys from ZINC database(Irwin, Sterling, Mysinger, Bolstad, & Coleman, 2012). A total 1448 decoys were found and 30 input active compounds amalgamated with decoys to screen through the best pharmacophore model to discriminate between actives and decoys. The accuracy value was found to be 0.960. The true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values were found to be 21, 48, 1400 and 9 respectively. From the screening output the ROC plot of the model was plotted by the true positive rate of actives vs. false positive rate of inactive compounds and given in Figure 5.

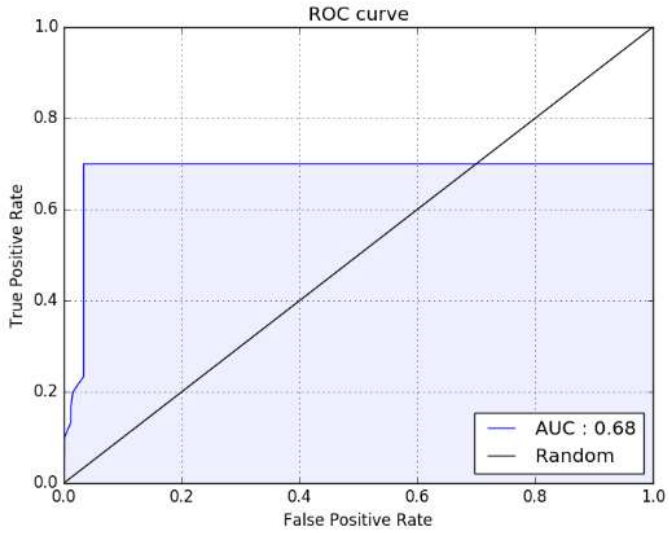

Figure 5. *ROC curve for pharmacophore model, derived from true positive rate of actives vs. false positive rate of inactive compounds.*

The Figure 5 undoubtedly clarified that actives and decoys are well-classified. Furthermore, the area under curve (AUC) was calculated and the value found to be 0.680 that indisputably explained that more true positives had been verified. Further, the enrichment factor (EF 1%) and Boltzmann-enhanced discrimination of ROC were also calculated and the value of average EF (1%) value for pharmacophore model was found to be 9.85 which showed that hypothesis has acknowledged active compounds very well and the top 1% hit is enriched with active molecules. The above-mentioned findings of decoy validation strongly explain that the developed pharmacophore features in the selected model are impeccably acceptable for the mapping of BACE1 inhibitors.

Virtual screening

Pharmacophore-based virtual screening of small molecule databases is one of the powerful techniques to search for novel and potential inhibitors. With regard to obtaining promising BACE1 inhibitors, *Hypo 1* of run number 11 was used to explore the NCI and IBS databases. In this purpose the '*Search Database*' protocol under '*Pharmacophore*' module of DS package was used. In the parameter list the protocol 'Search Method' and 'Limit Hits' were set to 'Best' and 'All' respectively. After searching using the best model we retrieved 32102 hits from the NCI database while 198 692 were retrieved from the IBS database. Molecules obtained from both databases were further screened separately using the number of criteria. First of all the "*Ligand Pharmacophore Mapping*" protocol of DS with "Maximum Omitted Feature" set to '0' to calculate the predicted inhibitory activity. After successful prediction, the estimated activities were compared with the experimental inhibitory activity (pK_i = 7.023 nM) of the most active compound (**B26** in Table 2) of the dataset. Compounds with better prediction than **B26** i.e. estimated inhibitory activity less than 0.949 were considered for further analysis. It was observed that 86 from NCI and 299 from IBS fulfilled the above criteria. These compounds were further considered for the Lipinski's Rule of Five and Viber's rule validation. 32 and 213 compounds from NCI and IBS respectively passed both rules. These compounds were considered for the molecular docking studies. After completion of the molecular docking studies it was found that 3 compounds from NCI failed to fit into the active site of BACE1 but all compounds from IBS successfully fitted. It was observed that all docked compound from both databases were given higher dock score values compare to the most active compound of the dataset. Moreover, the ADMET descriptors were calculated using the *ADMET Descriptor* protocol of DS. The human intestinal absorption (HIA), aqueous solubility and blood brain barrier (BBB) were analysed and it was found that three compounds from each NCI (**NSC11408**, **NSC367261** and **NSC694875**) and IBS (**STOCK1N-03331**, **STOCK3S-74396** and **STOCK6S-74113**) databases show good absorption, aqueous solubility and penetration values (Figure 6).

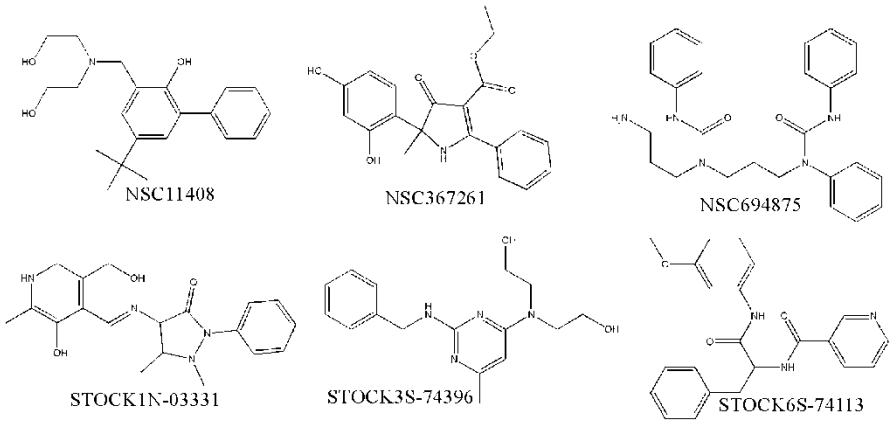

Figure 6. *Screened promising BACE1 inhibitors from NCI (**NSC11408**, **NSC367261** and **NSC694875**) and IBS (**STOCK1N-03331**, **STOCK3S-74396** and **STOCK6S-74113**) databases.*

Finally, all the above six compounds were checked for synthetic accessibility using Sylvia. The synthetic accessibility score were found to be less than 5 for all of them which clearly indicated that these molecules will not be difficult to synthesize. Hence the above six compounds were considered to be promising BACE1 inhibitors and were analysed further to assess the critical interactions with the catalytic amino residues of BACE1.

Molecular docking

The best docked poses of final six compounds along with most active compound of the dataset were analysed to observe the preferred orientation and binding interactions at the receptor cavity of BACE1. The crystal structure of the BACE1 (PDB ID: 4X7I(May et al., 2015)) was obtained from RCSB-Protein Data Bank. In order to validate the docking protocol the self-docking(Taha et al., 2011) approach was adopted. In this method, the already bound small molecule was re-docked at the catalytic site of macro molecule and the conformer of the original bound small molecule was overlaid to the re-docked pose to calculate RMSD value. It is reported that the RMSD < 2 Å value of original bound ligand validates the docking procedure(Taha et al., 2011). After self-docking it was observed that the RMSD value between co-crystal and docked conformer found as 0.356Å, which clearly indicated that the protocol selected in the docking method was validated. The superimposition between bound crystal ligand and docked pose of the same is given in Figure S6 in supplementary file.

Total ten different protein molecules (Figure S1, supplementary file) were generated by adding hydrogen atoms to the OD1 and OD2 of the aspartate dyad to find the critical protonated states of Asp32-Asp228 dyad. All six proposed and most active molecule of the dataset were docked to the ten protein molecules separately. The dock score, binding energy and number of bonding interactions with Asp32-Asp228 were

recorded and given in the Table S5 (supplementary file). Detailed analysis of Table S5 explained that almost all screened molecules give higher dock score and binding energy when docked to Prot1. The most active compound also showed more affinity towards Prot1 compare to others. The number of hydrogen bond and non-hydrogen bond analysis also revealed that **NSC11408**, **NSC367261**, **STOCK1N-03331**, **STOCK3S-74396** and **STOCK6S-74113** formed 2, 5, 3, 3 and 2 bonding interactions respectively with Prot1 which is higher than any other protein molecule. Hence protonated state of Prot1 that is un-protonated form of both Asp32 and Asp228 dyad was found to be crucial for the proposed molecule. Subsequently, Prot1 was considered for following docking analysis and MD simulations study. Best docked poses of **B26**, **NSC11408**, **NSC367261**, **NSC694875**, **STOCK1N-03331**, **STOCK3S-74396** and **STOCK6S-74113** are given in Figure 7. From the Figure 7 it can be explained that most active compound (**B26**) of the dataset formed two binding interactions through hydrogen bond with Lys222 and bump interaction with Ile110. It was observed that all six final screened molecules were successful in forming a number of interactions with the catalytic amino acid residues of BACE1. In detail, each of Gly34 and Arg235 were found to form one hydrogen bond with **NSC11408**. Asp228 and Thr231 were observed to be crucial for interaction with **NSC367261** and **STOCK1N-0333** via hydrogen bond interactions.

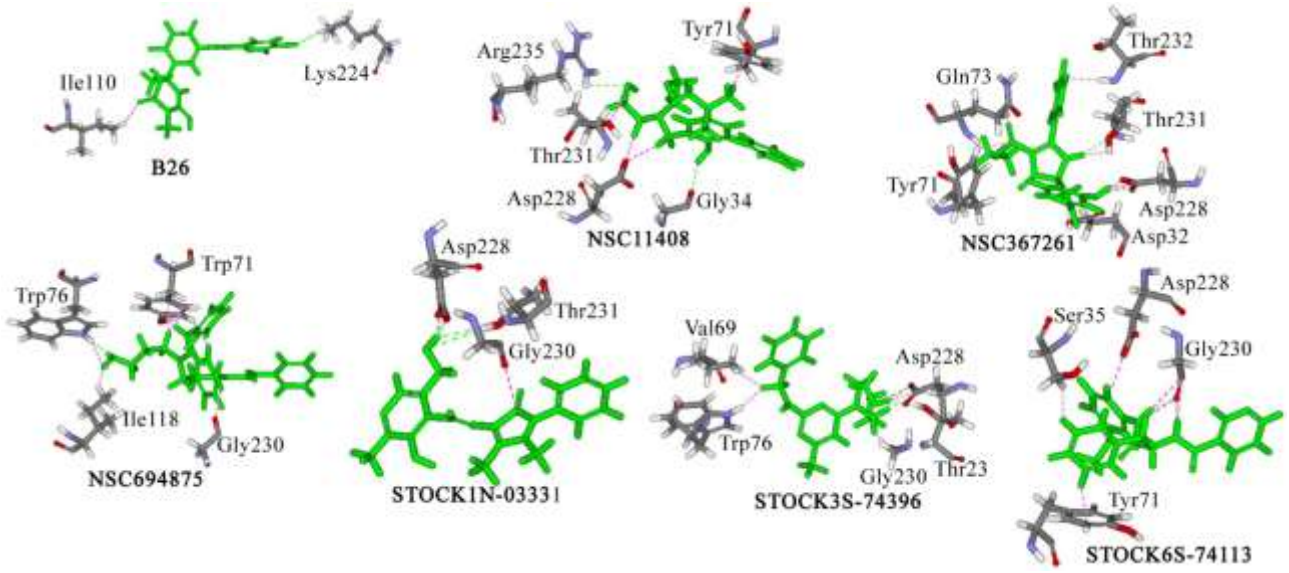

Figure 7. *Binding modes of the most active molecule of the dataset and final screened compounds from databases.*

Tyr71 catalytic amino residue at the active site clashed with **NSC11408**, **NSC367261**, **NSC694875** and **STOCK6S-74113** with one, one, three and one bump interactions respectively. **NSC11408** and **NSC367261** interacted with Thr231 through bump interactions. Asp228 was sufficient enough to form a number of bump interactions with all screened molecules except **NSC694875**. Catalytic amino residue, Trp76 was formed one of each hydrogen bond and bump interaction with **NSC694875** and also formed one bump interaction with **STOCK3S-74396**. Gly230 was connected with **NSC694875**, **STOCK1N-**

03331 and **STOCK3S-74396** via bump interaction, whereas same amino residue formed one hydrogen bonding and four bump interactions with **STOCK6S-74113**. Moreover, Asp32, Glu73 and Thr232 were connected via bump interactions with **NSC367261**. The side chains of **STOCK3S-74396** interacted with Thr23 and Asp228 via two and one hydrogen bond interactions respectively. **NSC367261**, **NSC694875**, **STOCK3S-74396** and **STOCK6S-74113** were interacted with Glu73, Ile118, Val69 and Ser35 respectively via one bump interaction each. Therefore the above observation undoubtedly explained that final screened molecules interacted with a larger number of binding interactions compared to the most active compound of the dataset.

Molecular dynamics

In order to analyse stability of complex between screened molecules and most active compound of the dataset with BACE1 the best docked poses of each were considered for molecular dynamics simulation study. For this purpose, Gromacs5.1.2 was used for a time span of 20ns. In order to understand the simulation output backbone RMSD, RMSF and radius of gyration (Rg) and RMSD of ligands were explored to analyse the complex constancy during simulation time.

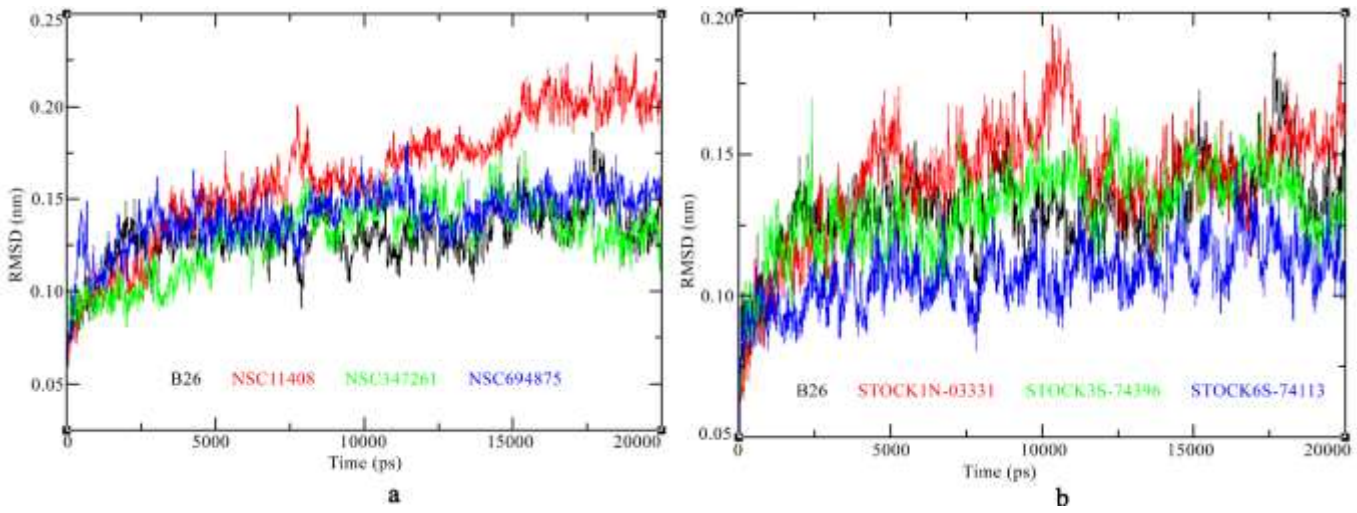

Figure 8. *RMSD vs simulation time (ns): a) complex of most active compound of the dataset and screened molecules from NCI database; b) complex of most active compound of the dataset and screened molecules from IBS database.*

The mean RMSD value of protein backbone were calculated and found to be 0.131, 0.164, 0.130, 0.141, 0.141, 0.130 and 0.108 nm for the BACE1 complex with **B26**, **NSC11408**, **NSC367261**, **NSC694875**, **STOCK1N-03331, STOCK3S-74396** and **STOCK6S-74113** respectively. The Figure 8 explained that protein backbone with complex **B26** initially fluctuated but after 3 ns of time it became stable. On analysis of RMSD trajectories of screened compounds from NCI database (Figure 8a) it was found that RMSD value of protein backbone of complex with **NSC11408** increased sharply up to about 16ns of time

span of simulation and then formed a stable complex at RMSD value approximately 2 nm. During the simulation of complex between BACE1 and **NSC367261**, it was observed that RMSD of the backbone of the BACE1 enzyme initially increased till about 8ns thereafter it formed stable complex at around 0.125 nm of RMSD. The complex between BACE1 and **NSC694875** was showed almost constant stability during the simulation study. The RMSD vs time plot of the complex with screened compounds from IBS database (Figure 8b) explained that **STOCK1N-03331** initially showed disturbance in stability but achieved constancy in about 0.15 nm of RMSD value. The backbone of the protein molecule bound to **STOCK3S-74396** was found to be stable throughout the simulation with minimum fluctuations. In the case of **STOCK6S-74113** and BACE1 complex the trajectory was shown to fluctuate much less compared to the others and maintained consistency throughout the simulation.

Furthermore, the variation of RMSD of screened molecules from both databases and **B26** itself were analysed to verify how comfortably these molecules can fit inside the receptor cavity of BACE1. The plot of RMSD vs time of the ligands is portrayed in the Figure 9.

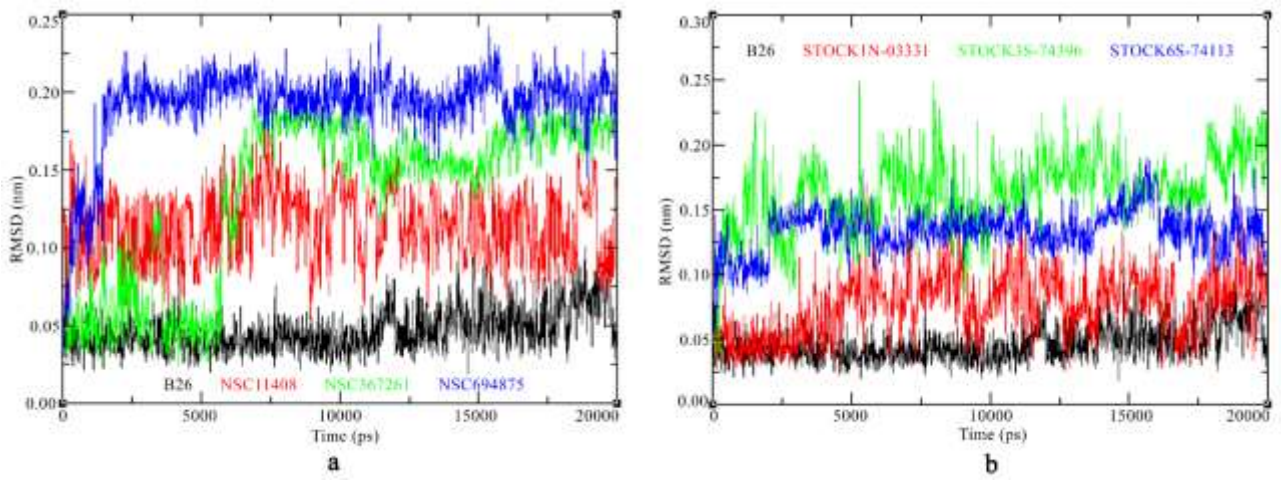

Figure 9. *RMSD vs time: a) most active compound of the dataset and screened molecules from NCI database; b) most active compound of the dataset and screened molecules from IBS database.*

The trajectories explained that maximum and minimum RMSD values were found to be 0.108 and 0.0005 nm, 0.186 and 0.0005 nm, 0.198 and 0.0005 nm, 0.244 and 0.0005 nm, 0.172 and 0.0005 nm, 0.250 and 0.000048, and, 0.191 and 0.0005 nm for **B26**, **NSC11408**, **NSC367261**, **NSC694875**, **STOCK1N-03331, STOCK3S-74396** and **STOCK6S-74113** respectively. The average RMSD value of the molecules were found to be 0.047, 0.115, 0.134, 0.192, 0.078, 0.163 and 0.134 nm for **B26**, **NSC11408**, **NSC367261**, **NSC694875**, **STOCK1N-03331, STOCK3S-74396** and **STOCK6S-74113** respectively. Figure 9 demonstrates that all molecules attained consistency during the simulation. Therefore, from the RMSD

trajectories of ligands itself inside the receptor cavity, it was clearly indicated that molecules formed extensive binding interactions with catalytic amino residues to form stable complex.

Furthermore, the RMSF were analysed to verify the role of amino acid in forming the stable complex between screened ligands and BACE1. On analysis of RMSF outputs it was observed that average and difference between maximum and minimum RMSF values were found to be 2.131 and 0.060 nm, 2.137 and 0.074 nm, 2.118 and 0.048 nm, 2.130 and 0.061 nm, 2.131 and 0.063 nm, 2.130 and 0.062 nm, and, 2.124 and 0.054 nm for **B26**, **NSC11408**, **NSC367261**, **NSC694875**, **STOCK1N-03331, STOCK3S-74396** and **STOCK6S-74113** respectively. The plot between RMSF and residue number is given in Figure 10. In Figure 10a it can be seen that the backbone of protein complexes with all screened molecules from NCI and IBS databases fluctuated more around the amino residues Ala313, Thr314, Ser315 and Gln316. Backbone of BACE1 enzyme around amino residues Phe159, Pro160 and Leu161 fluctuated about 0.13 nm when bound to **NSC11408** and **STOCK1N-03331**. Further, it was observed that backbone of complex with **STOCK1N-03331** and **STOCK3S-74396** fluctuated about 0.2 to 0.3 nm scale around the amino acids Pro70, Tyr71, Thr72 and Gln73. The possible reason of fluctuation of these region may be a lack of interactions between amino residues and ligands in these regions.

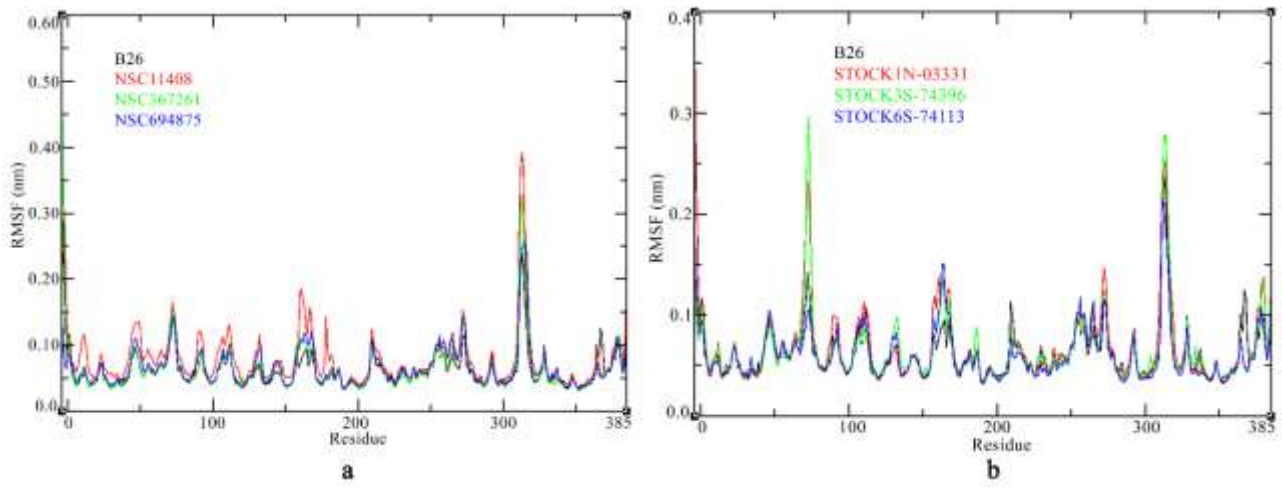

Figure 10. *RMSF vs number of residues:): a) complex of most active compound of the dataset and screened molecules from NCI database; b) complex of most active compound of the dataset and screened molecules from IBS database.*

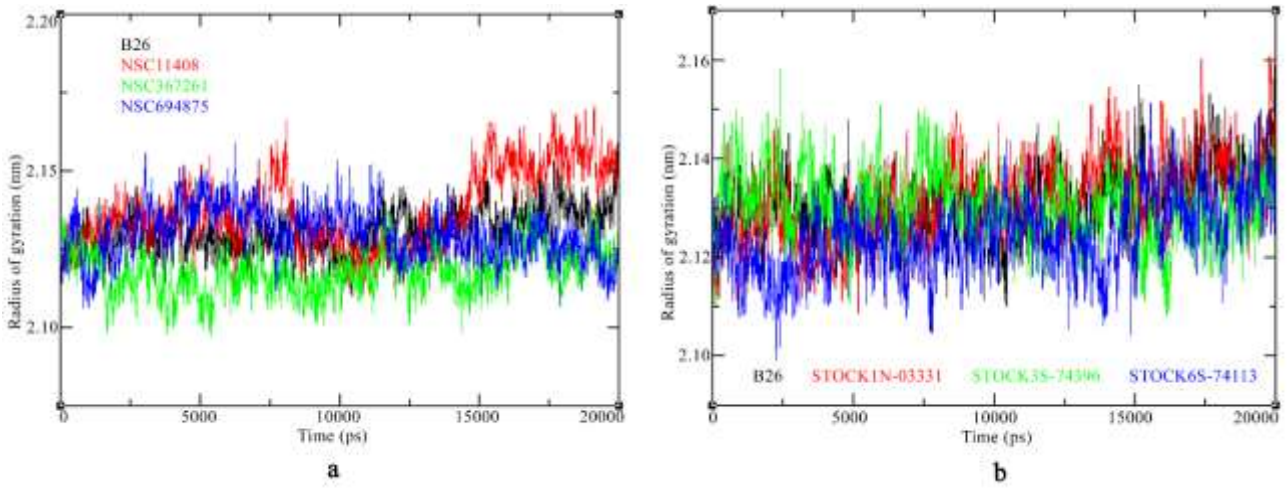

Figure 11. *Radius of gyration of Cα atoms of BACE1 over the simulation time.*

Finally the trajectories of the Rg was calculated and the plot is depicted in Figure 11. The trajectories explained that complexes with **B26** show no noticeable fluctuation. Initially both complexes with **NSC11408** and **NSC694875** were found to achieve almost similar Rg about 2.12 nm but after about 14ns of time of simulation complex NSC11408 increases the Rg value up to 2.15 nm. In case of complexes with compounds screened from IBS database (**STOCK1N-03331**, **STOCK3S74396** and **STOCK6S-74113**) a similar type fluctuation around 2.13 nm of Rg was found. The above findings of MD simulation clearly explained that the complex with all screened molecules achieved stable conformation at low RMSD except **NSC11408** and **STOCK1N-03331**. However, these compounds also attained stability with BACE1 at higher RMSD values. It was observed that both RMSD and RMSF analysis successfully correlated with findings of the molecular docking study. Fluctuation and the high Rg value in case of all complexes indicates the accessibility of the ligand to receptor cavity by opening of the binding pocket of the protein molecule.

Comparison of drug-likeness with standard BACE1 inhibitors

Drug-likeness of the screened compounds were compared with a four standard BACE1 inhibitors currently used in clinical trials. These molecules included **AZD3289**(Geschwindner et al., 2007), **AZD3293(Geschwindner et al., 2007)**, **Ly2886721**(May et al., 2015) and **MK-8931**(Scott et al., 2016). Different parameters including dock score, estimated activity, fit value, hydrophobicity, molecular weight, violation of Lipinski's rule of five, molecular volume, molecular refractivity, number of H-bonds and number of bump interactions were recorded using the DS and online MolInspiration (http://www.molinspiration.com/cgi-bin/properties) and depicted in the Table 3.

From the Table 3 it was revealed that all four standard BACE1 inhibitors were successfully docked into the BACE1 and also mapped on the best pharmacophore model of our study. The dock score of **AZD3289**, **AZD3293**, **Ly2886721** and **MK-8931** was found to be 61.907, 67.953, 28.911 and 63.911 respectively. The dock scores of **B26**, **NSC11408**, **NSC367261**, **NSC694875**, **STOCK1N-03331, STOCK3S-74396** and **STOCK6S-74113** were recorded as 4.011, 68.541, 67.239, 74.056, 72.746, 69.897 and 73.444 respectively. The dock score values of screened compounds and standard BACE1 inhibitors clearly explain that screened molecules show higher affinity for binding inside the receptor cavity of BACE1. Higher fit score value of screened compounds compared to standard BACE1 compounds also indicate that screened compounds are more efficient at fitting the best model. The estimated inhibitory activity of all compounds were predicted. It was observed that predicted activity of standard BACE1 inhibitors were much higher compared to the screened compounds which undoubtedly indicated that screened compounds might be more active than standard BACE1 inhibitors. The binding interactions were analysed from the best dock poses of screened and standard BACE1 inhibitors along with **B26**. Total number of interactions (hydrogen bonds and bump interactions) were found to be higher in number in case of screened compounds compared to the standard BACE1 inhibitors. Other parameters including molecular weight, molecular volume and molecular refractivity were also recorded and listed in Table 3. The above findings in comparison study indisputably explained that **NSC11408**, **NSC367261**, **NSC694875**, **STOCK1N-03331, STOCK3S-74396** and **STOCK6S-74113** may be promising BACE1 inhibitors for therapeutic application in AD.

Table 3. Comparative analysis of standard BACE1 inhibitors and screened compounds

Molecules	[1]DS	[2]EA	[3]FV	logP	[4]MW	[5]vROF	[6]MV	[7]MR	[8]HBond	[9]Bump
B26	4.011	0.454	11.117	1.72	425.875	0	335.65	101.64	1	1
AZD3289	61.907	99.616	9.076	4.28	451.530	0	368.21	118.43	3	4
AZD3293	67.953	44.708	9.424	3.69	412.540	0	391.24	125.75	1	5
LY2886721	63.370	393.058	8.480	2.04	390.420	0	316.26	96.07	4	3
MK-8931	28.911	99.238	9.077	2.35	405.470	0	344.54	101.26	1	4
NSC11408	68.541	0.020	12.780	3.62	343.470	0	342.49	103.86	2	4
NSC367261	67.239	0.298	11.600	2.61	353.370	0	313.84	96.81	3	8
NSC694875	74.056	0.022	13.729	4.11	445.570	0	427.30	129.15	1	7
STOCK1N-03331	72.746	0.317	11.573	1.87	352.390	0	318.20	101.65	4	3
STOCK3S-74396	69.897	0.013	12.976	1.41	302.38	0	288.93	88.76	3	5
STOCK6S-74113	73.444	0.031	13.581	2.52	375.43	0	344.65	103.59	2	6

[1]Dockscore; [2]Estimated activity; [3]Fit value; [4]Molecular weight; [5]Violation of Lipinski's rule of five; [6]Molecular volume; [7]Molecular refractivity; [8]Number of H-bonds; [9]Number of bump interactions

Free energy calculation using MM-PBSA

The binding free energy calculation based on MM-PBSA method was performed with all seven complexes. The average values of van der Waals, electrostatic and binding energy are given in the Table 4. Table 4 clearly indicates that all molecules have higher binding energy than complex with **B26** except complex with **NSC367261** and **STOCK3S-74396**. The highest binding energy was found in the complex

when STOCK6S-74113 bound to BACE1 and the value is -302.991 kcal/mol. Further it was noticed that the van der Waals and electrostatic energies negatively contributed to total energy of binding. As per Table 4, van der Waals interactions were given much higher contribution compared to electrostatic interactions for all cases. The total binding energies of all complexes are given in Figure 12. Therefore binding energy analysis between BACE1 and screened compounds undoubtedly explained that screened compounds have enough potential to form a number of bonded and non-bonded interactions with the catalytic amino residue at the active site cavity of BACE1.

Table 4. Average van der Waals, electrostatic and total binding energies of complexes of screened compounds and most active compound of the dataset with BACE1

Complex with compound	van dar Waals	Electrostatic	Binding energy
B26	-142.126	-13.571	-155.697
NSC11408	-130.844	-32.233	-163.077
NSC367261	-112.192	-36.210	-148.489
NSC694875	-160.391	-37.124	-197.515
STOCK1N-03331	-85.346	-22.148	-179.080
STOCK3S-74396	-82.389	-23.038	-106.223
STOCK6S-74113	-156.805	-146.187	-302.991

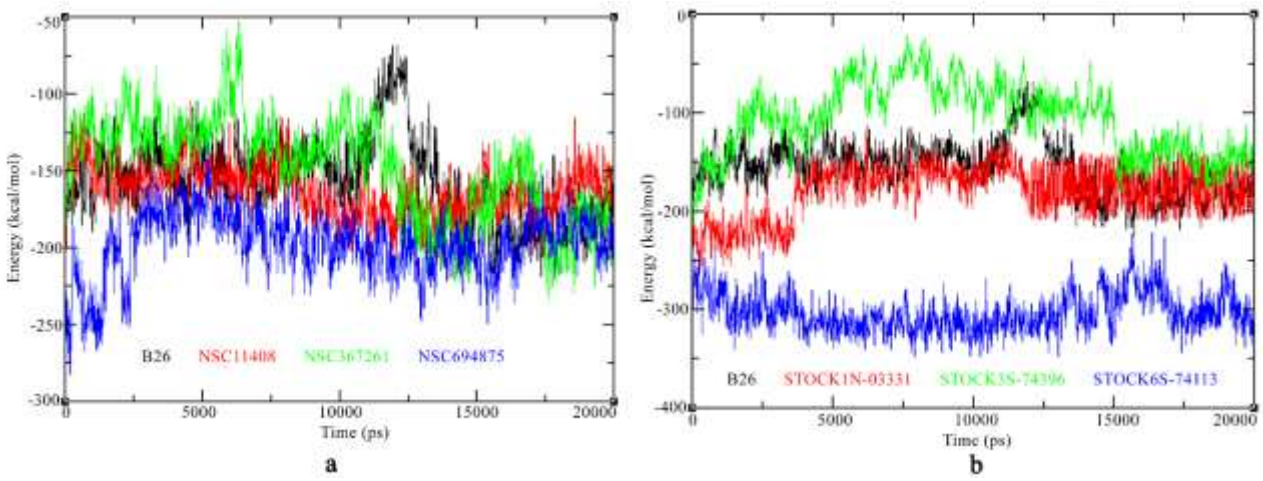

Figure 12. *Total binding energy of complexes with a) most active and screened molecules from NCI and b) most active and screened molecules from IBS.*

Conclusion

The current study developed ligand-based pharmacophore models followed by virtual screening, molecular docking and molecular dynamics study to identify promising BACE1 inhibitors for the treatment of AD. A total of eight training sets were generated and finally Tr7 was found to be more appropriate to develop models compared to others. The best model developed from Tr7 explained the

importance of HB acceptor and donor, hydrophobicity and ring aromatic features. The best model was validated internally and externally using R, Q^2, se, r_m^2, R_{pred}^2, sp, $r_{m(test)}^2$, $\Delta r_{m(test)}^2$, Fischer's randomization and decoy set. The NCI and IBS databases were used for virtual screening using best pharmacophore model. Initial hits from NCI and IBS respectively were passed through a number of criteria separately and finally three from each NCI (**NSC11408**, **NSC367261** and **NSC694875**) and IBS database (**STOCK1N-03331, STOCK3S-74396** and **STOCK6S-74113**) were found to be promising for inhibition of BACE1. The binding interactions between screened compounds and catalytic amino residues of BACE1 were analysed which revealed that screened compounds were able to form a number of binding interactions with the catalytic amino residues of BACE1. Furthermore, the molecular dynamics simulation study was performed with the complex between BACE1 and screened compounds along with most active compound of the dataset. Based on RMSD, RMSF and Rg values from MD simulation it can be proposed that all screened compounds might be promising BACE1 inhibitors. Comparison of the binding energy of complex between BACE1 and screened compounds along with most active molecule of the dataset and standard BACE1 inhibitors indicate that screened compounds have more potential to bind inside the BACE1 cavity. Therefore, it can be concluded that the final screened compounds might be potent and safer lead candidates for the treatment of AD but further confirmation will require experimental validation, in vitro.

Acknowledgment

MA Islam and TS Pillay were funded by the National Research Foundation (NRF), South Africa Innovation post-doctoral fellowship scheme. Authors are thankful to the CHPC (www.chpc.ac.za) for providing computational resources and tools.

Conflict of Interest

The authors declares that they have no conflicts of interest with the contents of this article.

References

Abrahama, M. J., Murtolad, T., Schulzb, R., Pálla, S., Smithb, J. C., Hessa, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX, 1-2*, 19-25

Al-Balas, Q. A., Amawi, H. A., Hassan, M. A., Qandil, A. M., Almaaytah, A. M., & Mhaidat, N. M. (2013). Virtual lead identification of farnesyltransferase inhibitors based on ligand and structure-based pharmacophore techniques. *Pharmaceuticals (Basel), 6*(6), 700-715. doi: 10.3390/ph6060700

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic

guidelines for Alzheimer's disease. *Alzheimers Dement, 7*(3), 270-279. doi: 10.1016/j.jalz.2011.03.008

Amin, S. A., Adhikari, N., Jha, T., & Gayen, S. (2016). First molecular modeling report on novel arylpyrimidine kynurenine monooxygenase inhibitors through multi-QSAR analysis against Huntington's disease: A proposal to chemists! *Bioorg Med Chem Lett, 26*(23), 5712-5718. doi: 10.1016/j.bmcl.2016.10.058

Amin, S. A., Bhargava, S., Adhikari, N., Gayen, S., & Jha, T. (2017). Exploring pyrazolo[3,4-d]pyrimidine phosphodiesterase 1 (PDE1) inhibitors: a predictive approach combining comparative validated multiple molecular modelling techniques. *J Biomol Struct Dyn*, 1-19. doi: 10.1080/07391102.2017.1288659

Anderson, A. C. (2003). The Process of Structure-Based Drug Design. *Chemistry & Biology, 10*(9), 787–797

Barman, A., Schurer, S., & Prabhakar, R. (2011). Computational modeling of substrate specificity and catalysis of the beta-secretase (BACE1) enzyme. *Biochemistry, 50*(20), 4337-4349. doi: 10.1021/bi200081h

Bhayye, S. S., Roy, K., & Saha, A. (2016). Development of Energy-based Pharmacophore Model and Stepwise Virtual Screening of LRRK2 Inhibitors Through Molecular Dynamics and Mechanics. *Letters in Drug Design & Discovery, 13*(24-32)

BIOVIA, D. S. (2016). Discovery Studio. San Diego: Biovia.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry, 4*(2), 187–217

Butini, S., Gabellieri, E., Brindisi, M., Giovani, S., Maramai, S., Kshirsagar, G., . . . Gemma, S. (2013). A stereoselective approach to peptidomimetic BACE1 inhibitors. *Eur J Med Chem, 70*, 233-247. doi: 10.1016/j.ejmech.2013.09.056

Castello, M. A., & Soriano, S. (2013). Rational heterodoxy: cholesterol reformation of the amyloid doctrine. *Ageing Res Rev, 12*(1), 282-288. doi: 10.1016/j.arr.2012.06.007

Castello, M. A., & Soriano, S. (2014). On the origin of Alzheimer's disease. Trials and tribulations of the amyloid hypothesis. *Ageing Res Rev, 13*, 10-12. doi: 10.1016/j.arr.2013.10.001

Cereto-Massague, A., Guasch, L., Valls, C., Mulero, M., Pujadas, G., & Garcia-Vallve, S. (2012). DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics, 28*(12), 1661-1662. doi: 10.1093/bioinformatics/bts249

Chhabria, M. T., Brahmkshatriya, P. S., Mahajan, B. M., Darji, U. B., & Shah, G. B. (2012). Discovery of novel acyl coenzyme a: cholesterol acyltransferase inhibitors: pharmacophore-based virtual screening, synthesis and pharmacology. *Chem Biol Drug Des, 80*(1), 106-113. doi: 10.1111/j.1747-0285.2012.01384.x

Cui, H., Hung, A. C., Klaver, D. W., Suzuki, T., Freeman, C., Narkowicz, C., . . . Small, D. H. (2011). Effects of heparin and enoxaparin on APP processing and Abeta production in primary cortical neurons from Tg2576 mice. *PLoS One, 6*(7), e23007. doi: 10.1371/journal.pone.0023007

Debnath, A. K. (2002). Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of Mycobacterium avium complex dihydrofolate reductase. *J Med Chem, 45*(1), 41-53

Debnath, A. K. (2003). Generation of predictive pharmacophore models for CCR5 antagonists: study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. *J Med Chem, 46*(21), 4501-4515. doi: 10.1021/jm030265z

Drachman, D. A. (2014). The amyloid hypothesis, time to move on: Amyloid is the downstream result, not cause, of Alzheimer's disease. *Alzheimers Dement, 10*(3), 372-380. doi: 10.1016/j.jalz.2013.11.003

Eketjall, S., Janson, J., Jeppsson, F., Svanhagen, A., Kolmodin, K., Gustavsson, S., . . . Falting, J. (2013). AZ-4217: a high potency BACE inhibitor displaying acute central efficacy in different in vivo

models and reduced amyloid deposition in Tg2576 mice. *J Neurosci, 33*(24), 10075-10084. doi: 10.1523/JNEUROSCI.1165-13.2013

Ellis, C. R., Tsai, C. C., Lin, F. Y., & Shen, J. (2017). Conformational dynamics of cathepsin D and binding to a small-molecule BACE1 inhibitor. *J Comput Chem, 38*(15), 1260-1269. doi: 10.1002/jcc.24719

Geschwindner, S., Olsson, L. L., Albert, J. S., Deinum, J., Edwards, P. D., de Beer, T., & Folmer, R. H. (2007). Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J Med Chem, 50*(24), 5903-5911. doi: 10.1021/jm070825k

Ghosh, A. K., & Osswald, H. L. (2014). BACE1 (beta-secretase) inhibitors for the treatment of Alzheimer's disease. *Chem Soc Rev, 43*(19), 6765-6813. doi: 10.1039/c3cs60460h

Golbraikh, A., & Tropsha, A. (2002). Beware of q2! *J Mol Graph Model, 20*(4), 269-276

Hong, L., Koelsch, G., Lin, X., Wu, S., Terzyan, S., Ghosh, A. K., . . . Tang, J. (2000). Structure of the protease domain of memapsin 2 (beta-secretase) complexed with inhibitor. *Science, 290*(5489), 150-153

Huang, D., Zhu, X., Tang, C., Mei, Y., Chen, W., Yang, B., . . . Huang, W. (2012). 3D QSAR pharmacophore modeling for c-Met kinase inhibitors. *Med Chem, 8*(6), 1117-1125

Hussain, I., Powell, D., Howlett, D. R., Tew, D. G., Meek, T. D., Chapman, C., . . . Christie, G. (1999). Identification of a novel aspartic protease (Asp 2) as beta-secretase. *Mol Cell Neurosci, 14*(6), 419-427. doi: 10.1006/mcne.1999.0811

Hussain, I., Powell, D. J., Howlett, D. R., Chapman, G. A., Gilmour, L., Murdock, P. R., . . . Christie, G. (2000). ASP1 (BACE2) cleaves the amyloid precursor protein at the beta-secretase site. *Mol Cell Neurosci, 16*(5), 609-619. doi: 10.1006/mcne.2000.0884

IBS. (2015). InterBioScreen Database. Chernogolovka, Russia: InterBioScreen Ltd.

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model, 52*(7), 1757-1768. doi: 10.1021/ci3001277

Islam, M. A., & Pillay, T. S. (2017). Identification of promising DNA GyrB inhibitors for Tuberculosis using pharmacophore-based virtual screening, molecular docking and molecular dynamics studies. *Chem Biol Drug Des, 90*(2), 282-296. doi: 10.1111/cbdd.12949

Jellinger, K. A., & Bancher, C. (1998). Neuropathology of Alzheimer's disease: a critical update. *J Neural Transm Suppl, 54*, 77-95

Krieger, E., & Vriend, G. (2002). Models@Home: distributed computing in bioinformatics using a screensaver based approach. *Bioinformatics, 18*(2), 315-318

Kristam, R., Gillet, V. J., Lewis, R. A., & Thorner, D. (2005). Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst. *J Chem Inf Model, 45*(2), 461-476. doi: 10.1021/ci049731z

Kubinyi, H., Hamprecht, F. A., & Mietzner, T. (1998). Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. *J Med Chem, 41*(14), 2553-2564. doi: 10.1021/jm970732a

Kumari, R., Kumar, R., Open Source Drug Discovery, C., & Lynn, A. (2014). g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model, 54*(7), 1951-1962. doi: 10.1021/ci500020m

Li, H., Sutter, J., & Hoffman, R. (1999). An automated system for generating 3D predictive pharmacophore models. In O. F. Guner (Ed.), *Pharmacophore*

Perception, Development, and Use in Drug Design (pp. 173-189). La Jolla, CA: International University Line.

Li, H., Sutter, J., & Hoffmann, R. (2000). *Pharmacophore Perception, Development, and Use in Drug Design*. La Jolla, CA: International University Line.

Lin, X., Koelsch, G., Wu, S., Downs, D., Dashti, A., & Tang, J. (2000). Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein. *Proc Natl Acad Sci U S A, 97*(4), 1456-1460

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev, 46*(1-3), 3-26

May, P. C., Willis, B. A., Lowe, S. L., Dean, R. A., Monk, S. A., Cocke, P. J., . . . Mergott, D. J. (2015). The potent BACE1 inhibitor LY2886721 elicits robust central Abeta pharmacodynamic responses in mice, dogs, and humans. *J Neurosci, 35*(3), 1199-1210. doi: 10.1523/JNEUROSCI.4129-14.2015

Middha, S. K., Goyal, A. K., Faizan, S. A., Sanghamitra, N., Basistha, B. C., & Usha, T. (2013). In silico-based combinatorial pharmacophore modelling and docking studies of GSK-3beta and GK inhibitors of Hippophae. *J Biosci, 38*(4), 805-814

Mitra, I., Saha, A., & Roy, K. (2010). Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers. *J Mol Model, 16*(10), 1585-1596. doi: 10.1007/s00894-010-0661-4

Momany, F. A., & Rone, R. (1992). Validation of the general purpose QUANTA ®3.2/CHARMm® force field. *Journal of Computational Chemistry, 13*(7), 888–900

NCI. (2013). National Cancer Institute Database: Center for Cancer Research, NCI, NIH.

Ojha, P. K., Mitra, I., Das, R. N., & Roy, K. (2011). Further exploring rm2 metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems, 107*(1), 194–205

Park, H., & Lee, S. (2003). Determination of the active site protonation state of beta-secretase from molecular dynamics simulation and docking experiment: implications for structure-based inhibitor design. *J Am Chem Soc, 125*(52), 16416-16422. doi: 10.1021/ja0304493

Pavadai, E., El Mazouni, F., Wittlin, S., de Kock, C., Phillips, M. A., & Chibale, K. (2016). Identification of New Human Malaria Parasite Plasmodium falciparum Dihydroorotate Dehydrogenase Inhibitors by Pharmacophore and Structure-Based Virtual Screening. *J Chem Inf Model, 56*(3), 548-562. doi: 10.1021/acs.jcim.5b00680

Polgar, T., & Keseru, G. M. (2005). Virtual screening for beta-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J Med Chem, 48*(11), 3749-3755. doi: 10.1021/jm049133b

Rajamani, R., & Reynolds, C. H. (2004). Modeling the protonation states of the catalytic aspartates in beta-secretase. *J Med Chem, 47*(21), 5159-5166. doi: 10.1021/jm049817j

Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. *J Chem Inf Model, 52*(2), 396-408. doi: 10.1021/ci200520g

Roy, P. P., Paul, S., Mitra, I., & Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules, 14*, 1660-1701

Roy, P. P., & Roy, K. (2008). On Some Aspects of Variable Selection for Partial Least Squares Regression Models. *QSAR & Combinatorial Science, 27*(3), 302–313

Sadler, B. R., Cho, S. J., Ishaq, K. S., Chae, K., & Korach, K. S. (1998). Three-dimensional quantitative structure-activity relationship study of nonsteroidal estrogen receptor ligands using the comparative molecular field analysis/cross-validated r2-guided region selection approach. *J Med Chem, 41*(13), 2261-2267. doi: 10.1021/jm9705521

Sakkiah, S., Thangapandian, S., John, S., Kwon, Y. J., & Lee, K. W. (2010). 3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors. *Eur J Med Chem, 45*(6), 2132-2140. doi: 10.1016/j.ejmech.2010.01.016

Scott, J. D., Li, S. W., Brunskill, A. P., Chen, X., Cox, K., Cumming, J. N., . . . Stamford, A. W. (2016). Discovery of the 3-Imino-1,2,4-thiadiazinane 1,1-Dioxide Derivative Verubecestat (MK-8931)-A

beta-Site Amyloid Precursor Protein Cleaving Enzyme 1 Inhibitor for the Treatment of Alzheimer's Disease. *J Med Chem, 59*(23), 10435-10450. doi: 10.1021/acs.jmedchem.6b00307

Selkoe, D. J. (2001). Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev, 81*(2), 741-766

Sinha, S., Anderson, J. P., Barbour, R., Basi, G. S., Caccavello, R., Davis, D., . . . John, V. (1999). Purification and cloning of amyloid precursor protein beta-secretase from human brain. *Nature, 402*(6761), 537-540. doi: 10.1038/990114

Smellie, A., Teig, S. L., & Towbin, P. (1995). Poling: Promoting conformational variation. *Journal of Computational Chemistry, 16*(2), 171-187

Taha, M. O., Habash, M., Al-Hadidi, Z., Al-Bakri, A., Younis, K., & Sisan, S. (2011). Docking-based comparative intermolecular contacts analysis as new 3-D QSAR concept for validating docking studies and in silico screening: NMT and GP inhibitors as case studies. *J Chem Inf Model, 51*(3), 647-669. doi: 10.1021/ci100368t

Tresadern, G., Delgado, F., Delgado, O., Gijsen, H., Macdonald, G. J., Moechars, D., . . . Trabanco, A. A. (2011). Rational design and synthesis of aminopiperazinones as beta-secretase (BACE) inhibitors. *Bioorg Med Chem Lett, 21*(24), 7255-7260. doi: 10.1016/j.bmcl.2011.10.050

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., . . . Mackerell, A. D., Jr. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem, 31*(4), 671-690. doi: 10.1002/jcc.21367

Vassar, R., Bennett, B. D., Babu-Khan, S., Kahn, S., Mendiaz, E. A., Denis, P., . . . Citron, M. (1999). Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science, 286*(5440), 735-741

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem, 45*(12), 2615-2623

World Health Organization and Alzheimer's Disease International. Dementia: a public health priority. (2012). Geneva.

Yan, R., Bienkowski, M. J., Shuck, M. E., Miao, H., Tory, M. C., Pauley, A. M., . . . Gurney, M. E. (1999). Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity. *Nature, 402*(6761), 533-537. doi: 10.1038/990107

Yu, N., Hayik, S. A., Wang, B., Liao, N., Reynolds, C. H., & Merz, K. M., Jr. (2006). Assigning the protonation states of the key aspartates in beta-Secretase using QM/MM X-ray structure refinement. *J Chem Theory Comput, 2*(4), 1057-1069. doi: 10.1021/ct0600060

Zhang, Y. W., Thompson, R., Zhang, H., & Xu, H. (2011). APP processing in Alzheimer's disease. *Mol Brain, 4*, 3. doi: 10.1186/1756-6606-4-3

Zou, Y., Xu, L., Chen, W., Zhu, Y., Chen, T., Fu, Y., . . . Shen, J. (2013). Discovery of pyrazole as C-terminus of selective BACE1 inhibitors. *Eur J Med Chem, 68*, 270-283. doi: 10.1016/j.ejmech.2013.06.027