Structure- and Ligand-Based Modeling of Beta-Secretase 1 (BACE1) Inhibitors

BY

BOOBALAN PACHAIYAPPAN

B.S., University of Madras, India, 1999

M.S. Indian Institute of Technology, Bombay, 2002

M.S., Virginia Tech, USA 2005

DISSERTATION

Submitted as partial fulfillment of the requirements

for the degree of Doctor of Philosophy in Medicinal Chemistry

in the Graduate College of the University of Illinois at Chicago, 2011

Chicago, Illinois

Defense Committee:

Dr. Pavel A. Petukhov, Chair and Advisor

Dr. Michael E. Johnson

Dr. Richard B. van Breemen

Dr. Scott G. Franzblau

Dr. Seungpyo Hong, Biopharmaceutical Sciences

DEDICATION

To my brother Shanmugam Pachaiyappan

ACKNOWLEDGEMENTS

It is a pleasure to thank the many people who made this PhD thesis possible.

It is difficult to overstate my gratitude to my advisor Dr. Pavel Petukhov, University of Illinois at Chicago (UIC). With his invaluable guidance, enthusiasm, inspiration, leadership efforts, mentorship and financial support during the course of my PhD, he helped me to envision my dream to be a Doctorate in Medicinal Chemistry. The molecular modeling and organic synthesis skills that I gained in his lab is instrumental in shaping my future research pursuits. Throughout my thesis-writing period, he provided constant encouragement, sound advice, and lots of good ideas. I would have been definitely lost without this advice.

I gracefully acknowledge the immense contribution of the rest of my PhD committee members namely Dr. Michael E. Johnson, Dr. Richard B. van Breemen, Dr. Scott G. Franzblau and Dr. Seungpyo Hong. I also would like to thank Dr. Qingbo Li for serving me in my PhD preliminary exam committee.

I am deeply indebted to the following people who have been instrumental for our collaboration needs: Dr. Yong Shen and his group members Hikmet F. Nural and Xin Cheng from Sun Health Institute, AZ (IC_{50} measurement of BACE1 inhibitors), Dr. Gregory R. J. Thatcher and Samer O. Abdul-Hay (measurement of Aβ levels), Dr. Gopal Thinakaran, University of Chicago for the generous gift of N2a cells, David C. Lankin (2D-NMR studies). I also would like to express my thanks to Openeye Scientific Software for proving academic license for the modeling software.

I am happy to acknowledge my debt to the financial support institutes – namely, National Institute of Health (R21 AG027454, PI - Pavel A. Petukhov) and University of Illinois at Chicago – without which this PhD might not have been possible.

The long, laborious, mind-boggling, journey of my PhD was made easy and comfortable, thanks due mainly to Dr. Petukhov's research group members. I feel very fortunate to be a part of the hard-working research group where knowledge transfer is smooth, and helping-each-other is everyday's philosophy. To name the group memebers: Dr. Srinivas Reddy Chirapu, Dr. Subash Velaparthi, Dr. Hongbin Yuan, Dr. Michael Brunsteiner, Dr. Gilles Piffet, Dr. Bhargava Reddy, Dr. Reaz Uddin, Dr. Raghupathi Neelarapu, Kohinoor Das, Kelvin Bai, Emma Mendonca, Hazem Abdelkarim and Aditya Vaidya.

I would like to thank College of Pharmacy at UIC for providing me a congenial atmosphere for research, Libraries at UIC for providing me a world of resources, UIC shuttle bus and red car service for providing transportation to my apartment.

Thanks are due in greater amount to my friends in USA, as well as in India, for their moral support throughout the course of my PhD, camaraderie, as well as for walking with me in life. To name a few: Priscilla Jiménez, S. Karthikeyan, V. Kumaresan, K. Manokaran, P. Anbarasu, S. Jayababu, R. Uma, V. Arun Prasad, G. Rajaram, R. Kalaivani and D. Partha Sarathy.

I also would like to thank my undergraduate chemistry instructor Mr. S. Srinivasa Chetty (Cheyyar) for his outstanding teaching style and sound knowledge that inspired me to pursue chemistry.

Last, but not least, I would like to express my deepest appreciation to my parents (S. Pachaiyappan, P. Vadivambal and P. Rajambal) for providing me love, education, care, support and genes. I would like to express my earnest gratitude to my siblings, P. Mahalakshmi, P. Shanmugam, P. Indirani and P. Ramanathan, and their extended families, for their boundless love and constant caring. In particular, I would like to thank my brother Shanmugam for acting as an in-house mentor during my formative years. The enthusiasm that he has envies me even today. His immense love, support, mentorship, and caring during the tough times made so many

things possible in my life, including my higher education in USA. I am honored to dedicate my PhD thesis to him.

TABLE OF CONTENTS

LIST OF TABLES

TABLE **PAGE**

LIST OF FIGURES

FIGURE **PAGE**

LIST OF SCHEMES

SCHEME **PAGE**

LIST OF ABBREVIATIONS

Ac	Acetyl
AchE	Acetyl Cholinesterase
AD	Alzheimer's Disease
ApoE	Apolipoporotein E
APP	Amyloid Precursor Protein
Aβ	Amyloid beta
BACE1	β-site Amyloid Precursor Protein Cleaving Enzyme 1
BDNF	Brain Derived Neurotrophic Factor
C logP	logarithmic Partition Coefficient
Ca^{2+}	Calcium Ion
CADD	Computer-Assisted Drug Design
CoMFA	Comparative Molecular Field Analysis
EDCI	1-Ethyl-3-(3-dimethylaminopropyl) Carbodiimide)
gCOSY	Gradient-Selected Correlation Spectroscopy
gHMBC	Gradient-Selected Heteronuclear Multiple Bond Correlation
gHSQC	Gradient-Selected Hetero Nuclear Single Quantum Coherence
GOLD	Genetic Optimization for Ligand Docking

H-bonding	Hydrogen-bonding
HIV	Human Immunodeficiency Virus
HOBt	Hydroxybenzotriazole
LBDD	Ligand-Based Drug Design
LOO	Leave One Out
MMFF	Merck Molecular Force Field
MOLCAD	Molecular Computer-Aided Design
NMR	Nuclear Magnetic Resonance Spectroscopy
PCA	Principal Component Analysis
PDB	Protein Data Bank
PHF	Paired Helical Fragments
PLS	Partial Least Square Analysis
PPAR	Peroxisome Proliferator-Activated Receptor
PS-1	Presenillin 1
PS-2	Presenillin 2
QSAR	Quantitative Structure-Activity Relationship
RACHEL	Real-Time Automated Combinatorial Heuristic Enhancement of Lead Compounds
ROCS	Rapid Overlay of Chemical Structures

RMSD	Root Mean Square Deviation
SAR	Structure-activity relationship
SBDD	Structure-Based Drug Design
TFA	Trifluoroacetic Acid
TMS	Trimethylsilyl
WT	Wild-type

SUMMARY

Understanding the binding of ligands to macromolecular targets of pharmacological relevance is an essential step for effective drug-design. The scenario gets complicated when small molecules bind to a target that is both solvent-accessible and flexible. Structural snapshots provided by experimental techniques namely X-ray and NMR are very useful to deal with targets that are not very labile by providing only the long-lived and more populated (low energy states) snapshots of binding partners. However, these techniques have limited scope when plasticity of the ligand-binding site has to be accounted for. In those cases, fast computational procedures have found to be a powerful supplementary tool to fill the paucity of information on receptor flexibility.

The two most commonly used computational methods in drug design are structure-based (SBDD) and ligand-based drug design (LBDD). The former method capitalizes on the structure of the biological target, knowledge of binding site and structure of co-crystal ligand obtained using either NMR or X-ray in order to predict the biological activity of the newly designed ligand. Whereas in the cases where it is difficult to obtain knowledge about the 3D structure of the protein, its binding site or the co-crystal ligand, LBDD is employed starting from the computer-generated 3D model of the active ligand which is generally used as a template to design new ligand whose binding affinity is forecasted before synthesis and development. There are cases where the 3D structure of the active ligand is available through X-ray or NMR, yet computational chemists performed LBDD to gain deeper insights that are otherwise difficult to obtain using SBDD method.

In this PhD dissertation, well known SBDD methods such as protein modeling, *de novo* ligand-design and molecular docking, and a well-known LBDD method, namely 3D-QSAR CoMFA has been used to gain deeper insights about the binding site of the enzyme β-secretase

(BACE1), the structure-activity relationship of its inhibitors, the predicted in silico blood-brain barrier profile (log BB), side-chain flexibility of the BACE1 binding site, and the steric and electrostatic characteristics of BACE1 inhibitors.

In Chapter 1, a brief introduction to Alzheimer's disease (AD) is presented. Very little is known about the etiology of AD, even though this disease was first diagnosed about 100 years ago. However, extensive research work carried out during the past few decades delineated the mechanism and understanding of production of amyloid plaques (Aβ) through the processing of membrane-bound amyloid precursor protein (amyloid pathway) by a set of secretases (β- and γ-secretase). Because the enzyme BACE1 is the rate-limiting step in this vicious cascade, it is considered to be one of the key targets for the therapeutic development of AD. The target BACE1 has been independently validated using several animal model studies.

The three-dimensional structure of BACE1 was immediately obtained since its isolation in 1999 by six different research groups. The first co-crystal structure of BACE1 bound to an active ligand (OM99-2), solved in the year 2000 by Hong et al., has provided the knowledge of BACE1 binding site, location and importance of catalytically active aspartic acid and other key residues required for inhibitor-design. The key challenges in BACE1 inhibitor-design has also been pointed out – very large active site, solvent-exposed regions, flexible binding site and importance of crossing blood-barrier in order to render the therapeutic effect. Despite these challenges, a BACE1 inhibitor (CTS-21166) has successfully completed Phase IIa clinical trials.

The beginning of Chapter 2 briefly summarizes the key SBDD methods that have been widely employed today, and examples of certain key drugs that resulted out of this method. Because previous knowledge of structural flexibility was available through perusal of literature, we attempted to superimpose various X-ray crystal structures of BACE1 publicly available in

protein databank. We particularly inspected the BACE1 S2 binding site because our preliminary modeling calculations and literature analysis suggests that the flexibility of this region is unexplored from drug discovery perspective.

Our hypothesis to design novel BACE1 ligand-scaffolds was: “selectively targeting charged amino acid residues present in the solvent-exposed flexible binding site would result in novel ligand scaffolds”. We proposed a three-step procedure to achieve our goals: (i) identify potentially flexible amino acid residues in the BACE1 S2 binding site, and develop and validate a computational procedure to design novel inhibitors that would target this residue; (ii) design novel inhibitors that target the flexible residue identified above and prioritize them for synthesis and biological tests; (iii) perform synthesis, biological testing and derive SAR of the novel scaffolds and test its consistency with the proposed hypothesis. Because the enzyme BACE1 is expressed in brain, we calculated log BB for all our novel inhibitors in order to compare the relative efficacy of these compounds in crossing the blood-brain barrier. As outlined in the first step above, superimposition of BACE1 PDB structures and further inspection on the S2 region revealed that Arg 235 undergoes greater flexibility compared to other residues such as Asn 233 and Ser 325 thereby making this residue ideal candidate for ligand targeting. The fact that the side-chain of Arg 235 has a positively charged guanidine group opens up the possibility of targeting ligands that can form directional, electrostatic interactions.

In order to design novel ligand scaffolds, the active ligand, compound **1**, reported by Stachel et al., was used as a template to probe the BACE1 S2 site. Using the software RACHEL (Real-time Automated Combinatorial Heuristic Enhancement of Lead compounds), novel ligand scaffolds were generated, docked (first using the tool FlexX and later using GOLD), and prioritized for synthesis.

Using standard organic chemistry reactions, we synthesized simple substituents (such as hydroxyl, methoxy, etc.), aryls, biaryls (such as phenyl, heteroaromatic), sultams and fused-ring compounds and were tested for its BACE1 inhibitory activity in a cell-lysate assay. Simple substituents fail to make good van der Waals (vdW) contacts with the residues in the S2 site and hence displayed poor activity. Aryl substituents too displayed poor inhibitory activity and we reasoned that these bulky substituents, under equilibrium conditions, have the tendency to collide with the nearby residues thereby rendering the enzyme-inhibitor complex unstable.

The analysis of fused-ring compounds provides interesting observations as well as strong evidence for our hypothesis. Prior to this, we characterized these compounds using 2D-NMR techniques (COSY, HSQC and HMBC). Between the isomers, **5a-5c** or **6a-6c**, we found that the degree of lipophilicity and flexibility incorporated at the P2 position correlates well with the BACE1 inhibitory activity, with **5c** and **6c** being more active. Based on the Gold docking results, we reasoned that this is due to the enhanced flexibility of the 7-membered ring and the vdW contacts that it forms with the residues in the S2 site.

To account for the biological activity difference between the regioisomers **5a-5c** compared to **6a-6c**, we employed side-chain flexibility of Arg235 during GOLD docking. Visualization of poses suggests that the regioisomers **5a-5c** has an oxygen atom locked in a position amenable for hydrogen bonding with side-chain NH_2 of Arg235 and hence offers a better BACE1 activity compared to the regioisomers **6a**-**c**. The synergistic combination of flexibility, lipophilicity, ability to form electrostatic interactions with Arg235 and improved vdW contacts with residues at the S2 site makes the fused-ring compounds **5a-5c** a much better candidate series compared to **6a-6c**. In addition, these results provide strong evidence to our hypothesis that it is advantageous to target the positively charged Arg235 at the BACE1 S2 site to design novel ligand scaffolds.

When mouse neuroblastoma cells (N2a) are treated with active BACE1 inhibitors **5b** and **5c**, a reduction in $A\beta_{40}$ production, ca. 65% and 35% respectively, compared to the control, was observed. Thus, our SBDD method pursuits to probe the S2 site of BACE1 resulted in the novel ligand scaffolds with improved inhibitory activity and a better log BB value.

The application of 3D-QSAR CoMFA method to understand the binding characteristics of BACE1 inhibitors is the focus of Chapter 3. The principle of CoMFA, the prerequisites, and the key statistical parameters used in the study are also discussed. Our strategy to perform CoMFA analysis consists of the following steps: (i) filter ligands whose activity is tested under identical conditions; (ii) assign charges; (iii) align molecules using ROCS; (iv) calculate CoMFA field values; (v) PLS analysis and deriving statistical parameters; and (vii) generate contour maps and explain the trend in activity.

Our CoMFA analysis dataset consist of 51 compounds – 44 training set and 7 test set compounds. Statistical analysis using PLS method yielded R^2_{NV} = 0.98, R^2_{CV} = 0.64, R^2_{LOO} = 0.67, SEE = 0.154, F = 287.219 and NOC = 7. The predictive power of the CoMFA model for the external dataset (R^2_{pred}) was found to be 0.74 and was in excellent agreement with the BACE1 binding characteristics. The ratio of steric to electrostatic field contributions was found to be 66:34 indicating the dominance of steric contribution. This high statistical significance illustrated that these diverse inhibitors share structural commonalities important for binding to BACE1. We strongly believe this model should be useful for the identification, design and development of novel BACE1 inhibitors.

1. INTRODUCTION

1.1. Background

The holy-grail of medicinal chemistry is the design and development of small molecules that can act as suitable drugs with minimum side-effects and maximum efficacy. Since enzymes are one of the important classes of drug targets, understanding the binding of small drug molecules to enzymes of pharmacological relevance is considered an essential step for effective drug-design. Even though the concept of drug-design has been used since 1910 during the time of Paul Ehrlich and Sacachiro Hata, who produced the antiprotozoal arsphemamime,[1] the three-dimensional (3D) structure of macromolecular target was not considered due to non-availability of sophisticated techniques. In fact, the first few X-ray crystal structures were solved in late 1950s and early 1960s for the enzymes myoglobin and hemoglobin.[2-5] Not only their findings helped understand the "tensed" and "relaxed" forms of hemoglobin (depending on its oxygenation status), but it also opened up a new avenue of research in medicinal chemistry – target flexibility. Multiple studies conducted later on proved the fact that enzymes indeed are flexible structures and can adopt different conformations on a wider range of scales, both in time and space, depending upon various conditions such as the presence of substrate, inhibitor, water molecules, cofactors, ions, etc. (for review, see reference).[6] However, the concept of target flexibility is still underutilized in drug design methods primarily due to inadequate computational tools and difficulty in modeling intrinsic protein flexibility.[7]

1.2. Role of molecular modeling methods in drug design

Classical drug discovery widely relied on atomic information extracted from X-ray and multi-dimensional NMR techniques. The 'time-averaged' static snapshots provided by these powerful techniques played an instrumental role to design and develop various drugs that do not require

details about target ‘plasticity’. However, emerging evidence posits the consideration of target flexibility during in drug design. In situations where the structural information obtained from X-ray and NMR are not sufficient enough, molecular modeling techniques have been found to be a powerful supplementary tool.[8-14]

Various modeling techniques have been successfully adopted to address the target flexibility, including: (i) molecular docking to an ensemble of protein structures – to predict the binding mode of putative ligand in a flexible binding site,[6] (ii) molecular docking that incorporates induced-fit; (iii) molecular dynamics – a procedure to obtain more complete set of protein conformers, especially those that are relatively high energy and undetectable,[15] (iv) dynamic pharmacophore modeling,[16] and (v) fragment-based computational solvent mapping.[17]

One of the critical points to consider while targeting flexible regions is the enthalpy-entropy compensation phenomenon. The well-known Gibbs’ free energy equation suggests that improving the enthalpic and/or entropic contributions results in favorable binding of small molecule binding to the protein, as long as these contributions are not counteracted by the opposite entropic or enthalpy changes. In brief, enthalpic optimization is achieved by positioning the h-bonding donor/acceptor groups for optimal interactions with the binding residues. However, if the structuring of these h-bonding groups inside the binding pocket result in significant loss of conformational entropy, then the binding enthalpy is compensated resulting in unaltered/reduced binding affinity. Thus, judicious incorporation of h-bonding groups in a conformationally constrained system is a crucial step to improving the binding affinity. On the other hand, entropic optimization is achieved by increasing the hydrophobicity of the ligand molecule and the enthalpic penalty in this case is profound. It is therefore believed that enthalpic optimization is a relatively difficult task compared to entropic optimization during design and development of high-affinity ligand molecule.[18]

Currently available state-of-the art computational programs have the ability to perform one or more of the modeling calculations described above. A body of literature is available where all these five approaches have been successfully applied.

For example, Huang et al. successfully validated a fast ensemble docking algorithm on 10 protein ensembles of 105 crystal structures and 87 ligands.[19] A success rate of about 93% was obtained with the criterion of root mean square deviation <2.5 Å if the top 5 poses of each ligand were considered. Sherman et al. applied the induced-fit docking protocol for 21 diverse flexible receptors and found that the RMSD is ≤ 1.8 Å for 18 cases.[20] Novak et al. applied molecular dynamics simulations and free-energy calculations to explore the role of ligand-induced conformational changes in modulating the activity of three generations of Bcl-X_L inhibitors.[21] In addition to the excellent agreement between predicted and measured binding affinities, they also pointed out the role of Asn100 in the binding site that forms stabilizing interactions with the ligand. Deng et al. developed a dynamic receptor-based pharmacophore model using a series of representative conformations of HIV-1 integrase (HIV-IN).[22] Conformations of the target were sampled through a molecular dynamics study of the catalytic domain of HIV-IN monomer and an ensemble of representative structures were collected via a probability-based representative conformer sampling method that considers both the potential energy and structural similarity of the protein conformations. This method was validated using a set of 128 known inhibitors with a 72% success rate. Application of this model resulted in inhibitors with significant improvement of binding affinity from high mM to low micromolar range. Sheu et al. applied the concept of computational solvent mapping to explore the “hot spots” of PPARγ ligand-binding domain by moving six solvent probe molecules (acetonitrile, methanol, t-butanol, urea, acetone, phenol and 2-butanol) around 12 receptor structures including two structures without ligand, two structures bound to a partial agonist and the eight structures bound to PPARγ agonist.[23] Their analysis revealed ten binding “hot spots”, four in the ligand-binding pocket, two in the coactivator-binding

region, one in the dimerization domain, two around the ligand entrance site, and one minor site without a known function. Along the lines of the docking method discussed above, we developed a strategy to target solvent-exposed flexible binding sites of interest to design novel scaffolds.

1.3. Solvent-exposed, flexible biological targets

Targeting flexible binding site becomes even more complicated when the target in question is membrane-bound and solvent exposed. To the best of our knowledge, targeting membrane-bound, solvent-exposed, flexible binding sites is not explored and understood as well as targeting the buried ones, however, the modeling approaches described earlier can be applied to this case as well.[24] Some of the key therapeutically relevant targets containing solvent-exposed flexible binding-site and the corresponding diseases that they are linked to are as follows: (1) β-Secretase 1 (Alzheimer's disease),[25] histone deacetylases or HDACs (cancer, inflammation and neurodegeneration),[26] Bcl-X_L system (cancer),[27] glutamate racemase (anti-microbial target),[28] tyrosine kinase receptor (cancer),[29] G-protein coupled receptors (wide-variety of physiological processes including visual sense, smell, mood behavior, regulation of immune-system, inflammation, and other diseases),[30] estrogen receptors (aging, cancer, obesity, and other diseases).[31]

In this PhD dissertation, emphasis will be given to structure- and ligand-based modeling of BACE1, a target for Alzheimer's disease therapy. A brief summary of Alzheimer's disease, BACE1 and its biological significance is due before discussing the computational modeling of BACE1 and its inhibitors.

1.4. Alzheimer's disease

Dementia is the disease characterized by the decline of reasoning, memory and other cognitive abilities. As the disease progresses, it results in impaired ability to perform everyday

function such as driving, household chores, and even personal care such as bathing, dressing, and feeding. There are over ten different types of dementia reported thus far: (i) vascular dementia; (ii) mixed dementia; (iii) dementia with Lewy bodies; (iv) Parkinson's disease; (v) fronto-temporal dementia; (vi) Creutzfeldt-Jacob dementia; (vii) normal pressure hydrocephalus; (viii) Huntington's disease; (ix) Wernicke-Korsakoff syndrome; (x) mild cognitive impairment. Among these types of dementia, AD, which normally follows mild cognition impairment, is considered to be the major type.[32]

AD was named after the German physician Dr. Alois Alzheimer. The disease was first diagnosed in a patient named Ms. Auguste D on April 8, 1906 after she died. Upon inspection of her brain, Dr. Alzheimer noted three major pathological features: (i) extracellular neuritic plaques; (ii) intracellular neurofibrillary tangles (fibrils) and (iii) loss of synapse. The findings were first reported by Dr. Alzheimer in a lecture he gave for South German neuropsychiatrists, entitled "Ueber eine eigenartige Erkrankung der Hirnrinde" or (About a remarkable illness of the cerebral cortex) on November 3, 1906.[33]

AD is a progressive, irreversible, brain disorder characterized by neurological and behavioral disabilities associated with the production of plaques and tangles, synaptic loss and neuroinflammation.[34] Over 5.3 million Americans are currently affected by this devastating brain disease and the worldwide AD patients exceed over 16 million.[35] AD puts a vast economic burden on the society costing $172 billion annually in USA. The two greatest risk factors for AD are age and genetics. The occurrence of this disease increases with age – most individuals affected are above 65 or older. The likelihood of getting AD doubles every five years after 65. About one-half of the population above 85 years old has AD.[35]

It has been over hundred years since AD was first diagnosed, but the scientific community still did not understand the etiology. However, an improved understanding of production of

plaques and tangles has been achieved.[25, 36-37] Amyloid plaques are extracellular protein fragments that are aggregated between nerve cells in the disease conditions, whereas neurofibrillary tangles (NFTs) are hyperphosphorylated forms of the microtubule-assisted tau protein. The well-known functions of tau protein are thought to be stabilization of microtubules and regulation of motor-driven axonal transport.[38] In Alzheimer's disease condition, the tau protein is excessively phosphorylated resulting in the formation of paired helical fragments (PHF). This abnormal phosphorylation weakens the binding affinity of tau-microtubule binding, thereby affecting the stabilization of microtubules and axonal transport. The production and development of PHFs is a whole area of research by itself and hence will not be discussed in this dissertation. The following discussion will be limited to our current understanding of production of amyloid plaques and ways to mitigate it.

1.5. APP processing

In human brain, APP is processed via two different pathways – namely, non-amyloid pathway (major) and amyloid pathway (minor) – as illustrated in figure 1. In the major non-amyloid pathway, APP is cleaved by α-secretase within the amyloid beta (Aβ) sequence to generate a large extracellular soluble fragment (sAPPα) and a smaller intracellular fragment (C83). These fragments appear to have no pathological significance to AD, although sAPPα may have neuroprotective characteristics.[39] Concomitant cleavage of C83 by γ-secretase results in non-amyloidogenic p3, a 30kD peptide, as well as Aβ intracellular domain (AICD). As mentioned above, since α-secretase cleavage of APP happens with the Aβ sequence, this pathway prevents the formation of Aβ.

In the amyloid pathway, where endoproteolytic processing of the APP by consecutive cleavages of BACE1 and γ-secretase takes place, Aβ is produced.[25] The enzyme BACE1 cleaves APP at the N-terminal side of the Aβ sequence to secrete sAPPβ, producing the cell-

bound, carboxyl-terminal fragment C99. The C99 fragment is then cleaved by the γ-secretase producing AICD and amyloid peptides 39-43. It was reported that the majority of Aβ produced via amyloid pathway is $A\beta_{40}$ which is soluble and non-neurotoxic. The relatively minor product $A\beta_{42}$ (10%), that is less soluble and more hydrophobic, has a tendency to undergo fast aggregation and was reported to be toxic.[40] The correlation between $A\beta_{42}$ and cognitive decline is reported.[41] Under normal physiological conditions, there is equilibrium between production and clearance of toxic $A\beta_{42}$.[42] However, under AD conditions, a pronounced decrease in the clearance of $A\beta_{42}$ was observed resulting in the excessive production and aggregation of $A\beta_{42}$ which in turn triggers a cascade of events such as oligomerization, fibril formation, neuroinflammation, reactive oxygen species production, oxidative damage, calcium dyshomeostasis, tau hyperphosphorylation, synaptic dysfunction, neuronal loss and eventually resulting in dementia.[40] This phenomenon has been referred as amyloid cascade hypothesis and it was first proposed by Hardy et al.[25] Because BACE1 is the rate-limiting enzyme in the amyloid pathway, it appears promising as a molecular target for the therapeutic intervention of AD.[43-45] It is worth mentioning here that the amyloid cascade hypothesis is recently challenged with a new, complex web of interactions model based on age proposed by Herrup.[46] The fact that brain structure and function decline over the age led the author to propose the three necessary steps that could lead to AD: (i) an initiating injury; (ii) a chronic inflammation; and (iii) a cellular change-of-state. Even though this model is in primitive stage requiring supporting research from elsewhere, it is important to state that lowering of amyloid plaques may not necessarily be considered as the gold standard for AD therapy.

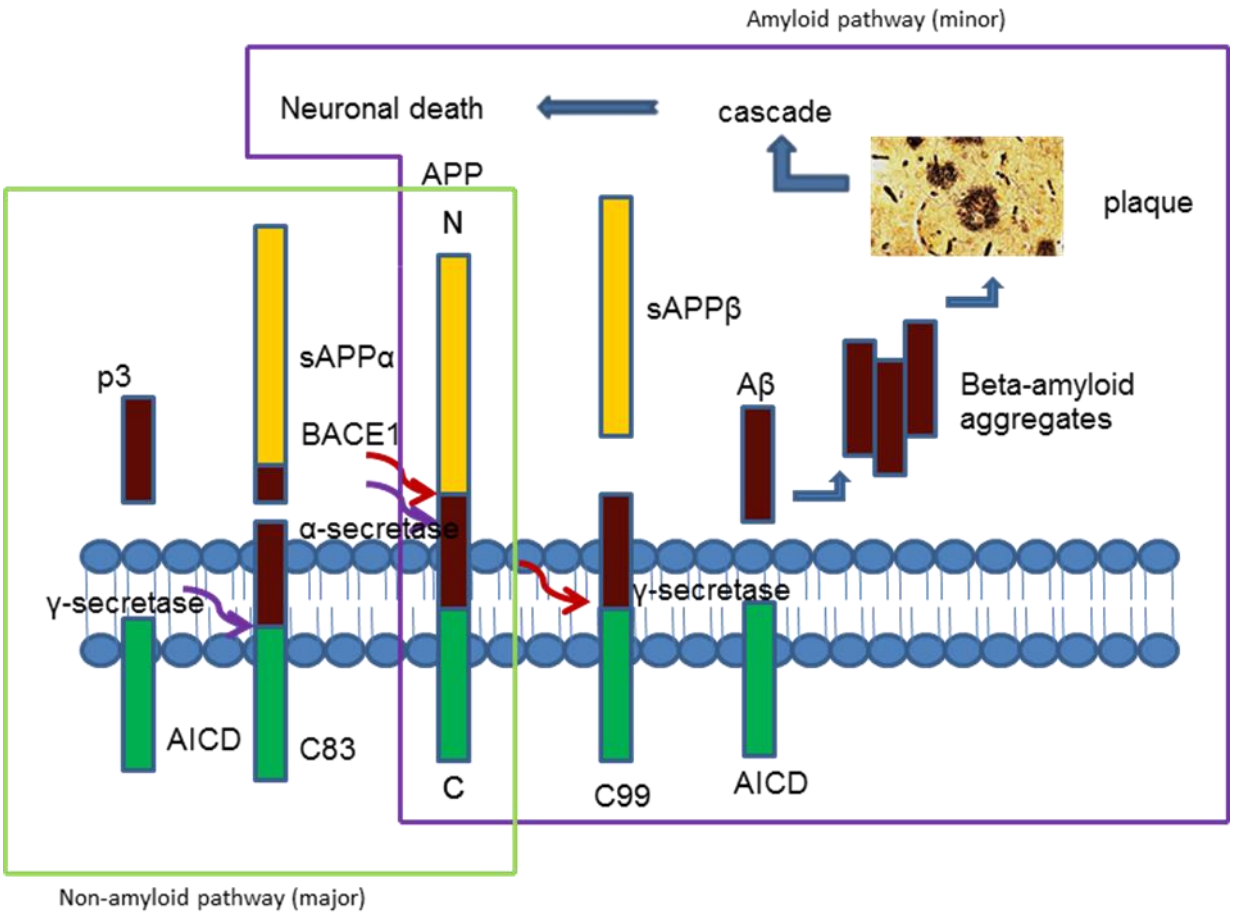


Figure 1. Processing of APP by secretases

1.6. Discovery of BACE1

BACE1 was discovered in 1999 by several research groups independently – Vassar et al., Hussain et al., Sinha et al., Yan et al., and Lin et al.[47-51] Several nomenclatures for BACE1 exist in the literature – such as, BACE1, β-Secretase 1, Asp2 and memapsin 2 - however, in this dissertation, only the term BACE1 will be consistently used. BACE1 belongs to the type-I membrane-bound aspartic protease of the pepsin family. Similar to other aspartic proteases, it contains the unique dual active-site motif (D-T/S-G-T/S) in its ectodomain. The structural organization of BACE1 is depicted as a cartoon in figure 2. A full length BACE1 has an N-

terminal signal peptide (residues 1-23), a Pro-peptide domain (residues 24-48), catalytic domain (residues 49-454), a membrane-spanning region (residues 455-478) and a C-terminal cytoplasmic tail (residues 479-501). Hussain et al. and Bennett et al. reported that mutation in one of the two active aspartic acid residues result in complete loss of enzyme activity.[47, 52]

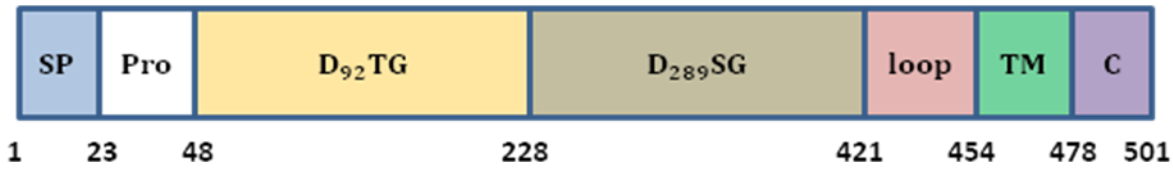


Figure 2. Sequence organization of BACE1

1.7. BACE1 expression levels in brain areas

The enzyme BACE1 is ubiquitously expressed, however its transcript levels are found highest in pancreas and neurons in the brain.[50] BACE1 activity in pancreas is rather low despite its abundance due to the generation of alternatively spliced transcripts that produce BACE1 variant with a reduced proteolytic activity.[53] In contrast, BACE1 activity is higher in hippocampus, cerebellum and frontal lobe regions of the brain.[53] Even though neurons are the major sources of BACE1, astrocytes are known to be the alternative source. The latter is known to be important for the clearance and degradation of Aβ and for acting as a barrier between neurons and toxic deposits of Aβ.[54] It was reported that resting astrocytes in brain do not express detectable levels of BACE1, whereas cultured astrocytes display BACE1 promoter activity, express BACE1 mRNA and active BACE1.[54]

1.8. Enzymatic activity of BACE1

Purified BACE1 has an optimal enzymatic activity at pH 4.5, which reflects its primary site of action inside the cell.[50] In fact, cell-surface expressed BACE1 is transported to early

endosomes and *trans* golgi network by endocytosis. It is believed that BACE1 catalyzes APP processing to release sAPPβ and C99 after endocytosis.[55-56]

1.9. Function of BACE1

The widely known function of BACE1 is APP processing and this has been described in detail. In addition to APP, other BACE1 substrates are also reported which suggests a variety of physiological functions of BACE1. It is beyond the scope of this dissertation to address the significance of other substrates. However, briefly, these substrates include: α-2,6-sialyltranferase (ST6Gal-1),[57] Neuregulin-1 (NRG1) implicated in myelination,[58] platelet selectin glycoprotein ligand 1 (PSGL-1),[59] and interlukin receptor type II.[60] The role of BACE1 in regulating the voltage dependent sodium channel is also reported.[61]

1.10. BACE1 as a target for AD therapy

It has been hypothesized that therapeutics targeting APP processing may mitigate the formation of toxic fragments and hence prevent plaque formation.[25, 36-37] Since BACE1 is the rate-limiting enzyme in this cascade, it is considered as one of the propitious targets for AD. Further studies have shown that -BACE1knock-out mice fail to produce Aβ thereby providing a strong evidence that BACE1 is indeed the sole enzyme with a bonafide β-secretase activity. It was also reported that these mice are healthy, fertile, and appear normal in gross anatomy, tissue histology, hematology and clinical chemistry.[62-65] Based on these results, it was suggested that inhibition of BACE1 in humans may not have mechanism-associated toxicity. Interestingly, overexpression of BACE1 levels in mice resulted in abnormal production of Aβ. However, results obtained from behavioral studies revealed memory impairment, subtle behavioral changes, being timid, anxious and less inclined to explore compared to mice over expressing BACE1 suggesting that only partial inhibition of BACE1 may be desired.[62-63, 65-67] In humans, it has been shown that BACE1 levels are significantly elevated *in vivo* in rapidly

autopsied brains of sporadic AD patients (< 3 hours) compared with age-matched non-AD patients,[68-70] and a correlation (R^2 = 0.54) between BACE1 activity and amyloid plaques count was reported.[71] It was, therefore, suggested that lowering of amyloid plaques *in vivo* may be achieved by a decrease in production of $A\beta_{40}$ and $A\beta_{42}$ through inhibition of BACE1, thus opening up a new avenue for AD drug discovery.[37, 44]

1.11. BACE1 inhibitors in clinical trials

Despite the fact that the enzyme BACE1 has been discovered a decade ago, clinical development of an ideal BACE1 inhibitor is still far from reality. BACE1 inhibitor design is impeded by various challenges including: (1) ability of inhibitors to penetrate the blood-brain barrier; (2) presence of an efflux transporter, phosphoglycoprotein (P-gp), at the blood-brain barrier; (3) large and conformationally flexible BACE1 active site; (4) homology to other aspartic proteases, specifically towards BACE2 even though its expression levels are higher in the periphery compared to that in brain.[45, 72] Even though both BACE1 and APP are membrane-tethered, it has been suggested that the physiologically relevant cleavage of APP by BACE1 occurs in endosomes where the pH is optimal for BACE1 enzymatic machinery (pH = 4.5)[73-77] or in cholesterol rich lipid rafts.[78] Therefore, further improvement of the cell membrane permeation properties of BACE1 inhibitors would be one of the most important tasks during lead optimization. A proof of concept for an alternative to the passive permeation approach, a BACE1 inhibitor conjugated with a lipid fragment targeting endosomes, was recently reported by Simons and co-authors.[55]

Despite these challenges, Ghosh et al. recently developed an orally bioavailable small molecule BACE1 inhibitor (CTS21166) that displayed excellent efficacy, selectivity, brain penetration and pharmacologic activity in preclinical studies.[45] The phase IIa studies of this inhibitor were promising and are awaiting further advancement.

1.12. Scope of the current study

A growing body of literature in the form of publications, patents and proceedings supports the significance of BACE1 in AD pathogenesis. The abundant structural information available for BACE1-inhibitor complexes, as well as the availability of powerful molecular modeling techniques, fueled the design and development of BACE1 inhibitors. Despite these advances, there is still an unmet clinical need for the AD therapy.

In the present study, we have employed both the structure- and ligand-based drug design methods to understand the binding characteristics of BACE1 inhibitors. Our ultimate goal is to further advance development of more active blood-brain permeable inhibitors of BACE1 as potential therapeutics for AD.

2. STRUCTURE-BASED MODELING OF BETA-SECRETASE1 INHIBITORS

2.1. Background

Contrary to traditional drug discovery, which relied heavily on serendipity, rational drug discovery utilizes the knowledge of three dimensional structures of proteins and ligands. One of the types of rational drug discovery is the structure-based drug design method (SBDD). In this method, three dimensional structures of biological targets obtained using X-ray or multi-dimensional NMR and knowledge of active site were are used as starting points for the effective design of active and selective ligands that would eventually be tested in clinical trials. In the case of difficulty in obtaining 3D structures of targets, computational homology modeling has been used. A typical drug discovery pipeline begins with validation of target. Once the biological target is validated and its atomic structure is solved (X-ray or NMR), several SBDD methods can be used to identify hit compounds. Some of these methods are: (1) target-based high-throughput screening that involves screening of compounds (several thousands to even millions); (ii) target-based virtual screening that involves molecular docking and scoring of a huge library of ligands on the target structure; and (iii) de novo design that involves growing fragment of ligands inside the active site of the target using a database containing small molecules and chemical fragments. In conjunction with various computational tools methods including molecular docking, hit to lead transformation and lead optimization can be envisioned before testing the compounds in clinical trials. Modern SBDD methods involve the synergistic combination of techniques including X-ray crystallography, NMR, computational methods and advanced chemical synthesis.

Explosion in structural information, proteomics and genomics resulted in hundreds of new targets that were otherwise unknown and have provided ample opportunities for SBDD drug discovery. In fact, SBDD is not limited to proteins, but can be applied to nucleic acids and

membranes as well. The success of drugs such as saquinavir, ritonavir, indinavir and nelfinavir can be attributed to the success of SBDD.

One of the key challenges in SBDD is to design drugs for flexible targets. Lead generation obtained by considering only from a single, rigid target structure is prone to give misleading results for a conformationally flexible target. Computational techniques have been judiciously employed in these cases, despite the difficulty in sampling entire protein conformer system. Modeling the side-chain flexibility of key residues in the binding pocket and accommodating multiple protein structures are some of the techniques employed by computational chemists to handle protein flexibility in SBDD.

In the present work, structure-based drug design using GOLD docking and de novo ligand design using RACHEL has been employed to identify novel ligand scaffolds targeting BACE1.

2.2. Crystal structures of BACE1

The validation of BACE1 as a target for AD has been discussed in the previous chapter. The first X-ray crystal structure of BACE1 in complex with an eight-residue inhibitor, OM99-2, was solved at 1.9 Å resolution by Hong et al in 1999 (figure 3).[79] Based on the atomic information obtained from this and multiple other crystal structures, it was found that the BACE1 is bilobal and pseudo 2-fold symmetric. The catalytic core of BACE1 is conserved compared to the other aspartic protease members, however, flexibility is pronounced in the binding regions. For example, the active site region contains a flexible β hairpin loop (residues 67-77) also called the flap region.[79] It is believed that this flap region controls the access of substrate to the BACE1 binding pocket as well as to orient the substrate in optimal geometry for catalysis.[79] In particular, the flap is reported to exist in open form (to allow the entry of substrate), close form (to lock the substrate inside the binding pocket of BACE1) and other intermediate positions depending upon the size of the substituents.[79-81] In addition, there is another flexible loop near

the N terminus comprised of residues 9-14 (herein referred as 10s loop). Compared to other aspartic proteases, the active site of BACE1 is large, more open, less hydrophobic and conformationally flexible.[79]

The majorities of BACE1 inhibitors, but not all, have tight hydrogen bonding network with the catalytic aspartic residues (Asp32 and Asp228). For example, the substrate-based inhibitor OM99-1, that occupies from S4 to S4' binding sites, was found to involve in significant hydrogen bonding and hydrophobic interactions with the residues present at the pertinent pockets (figure 3). In particular, a direct hydrogen bonding between the hydroxyl group of the hydroxyethylene moiety of the inhibitor with the catalytic aspartic residues Asp32 and Asp228 was observed. As of today, there are over a hundred publicly available BACE1 PDB structures.

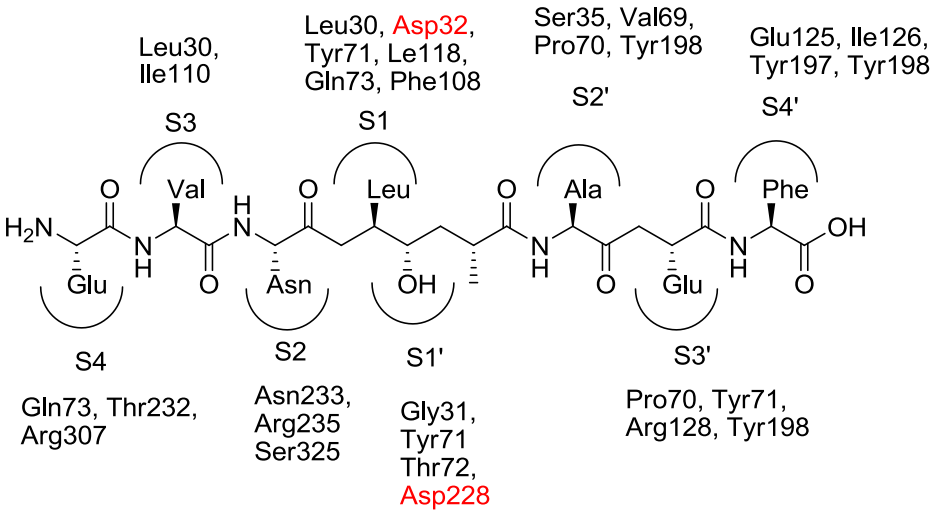


Figure 3. OM99-2 in BACE1 active site[79]

2.3. Conformational flexibility of BACE1

Superposition of multiple X-ray crystal structures revealed that BACE1 is a highly flexible enzyme with experimentally determined motion in the flap region, the catalytic residues, 10s loop, and the S2 site. The residues 67-77 present at the flap region adopt a β hairpin

conformation and reportedly exist in two conformations.[82] The apo enzyme that lacks the stabilizing interactions of these flap residues adopts an 'open' conformation, whereas the ligand-bound enzyme adopts a 'closed' conformation by h-bonding with Tyr71. Recently, a ligand-bound enzyme in a flap-open conformation was also reported.[83] McGaughey et al. showed that the loop comprising the residues 9-14 can adopt two or more different conformations depending on the size of the ligands and h-bonding with Thr232.[84] The authors also observed that ligands that possess a bulky P3 moiety modulate the enzyme to exist in an "up" conformation whereas the ligands that have a small P3 moiety induce the ligands to exist in a 'closed' conformation. Using this rationale, highly potent inhibitors has been developed.[85] The flexibility of BACE1 S2 site had not been rigorously addressed from computational perspective at the time we embarked on our research. In particular, the X-ray structure analysis suggests that the Arg235 present at the S2 site is highly flexible. Thus, the highly flexible, positively charged Arg235 could be explored for future ligand-design efforts.

2.4. Rational for targeting S2 site of BACE1

Our interests in BACE1 S2 site flexibility stem from the preliminary modeling calculations that we performed by superimposing various crystal structures of BACE1 publicly available in the protein data bank.[86] The major amino acid residues present in BACE1 S2 site are Asn233, Arg235 and Ser325 out of which the fluctuations observed in Asn233 and Ser325 are minimal. We believed Arg235 as a potential candidate for targeting due to the following reasons: (1) the side-chain guanidine group in Arg235 is positively charged in neutral pH compared to the neutral side-chains present in Asn233 and Ser325. Hence, targeting the guanidine side-chain with appropriate ligand atoms will result in strong, electrostatic interactions that are directional in nature; (2) the side-chain of Arg235 is highly flexible and hence adopts several conformations as evidenced by visualization of several BACE1 PDB structures. We reasoned that the size and electronic parameters of P2 substituents are the key determinants in controlling the Arg235

movement. The extreme displacements of Arg235 observed in the X-ray structures of 2B8L and 1W51 is shown in figure 4. Docking of crystallographic 1W51 ligand in 2B8L binding site did not provide any meaningful pose suggesting that the flexibility of the BACE1 must be accounted during *in silico* inhibitor design. This increased our curiosity to explore the side-chain movements of Arg235 during docking calculations. It is imperative that the ligand component that binds to the S2 site must be flexible enough in order to interact with Arg235 as well as with the rest of the residues that form the S2 site. Our modeling analysis resulted in identifying the attributes of an ideal P2 component of the ligand: (1) ability to interact with charged Arg235; (2) ability to adapt its conformation in a flexible binding site; (3) containing fewer polar atoms. Any compound that possesses more/all attributes listed above is expected to efficiently inhibit BACE1 and hence will provide strong evidence to our hypothesis. In fact, our modeling analysis resulted in scaffolds containing an oxygen atom locked in a conformation that is amenable for h-bonding with the Arg235 (enthalpic optimization). In addition, the fact that the fused-ring compounds already contain a constrained ring (5 or 6 or 7-membered), the entropic penalty during binding would be minimal (entropic optimization).

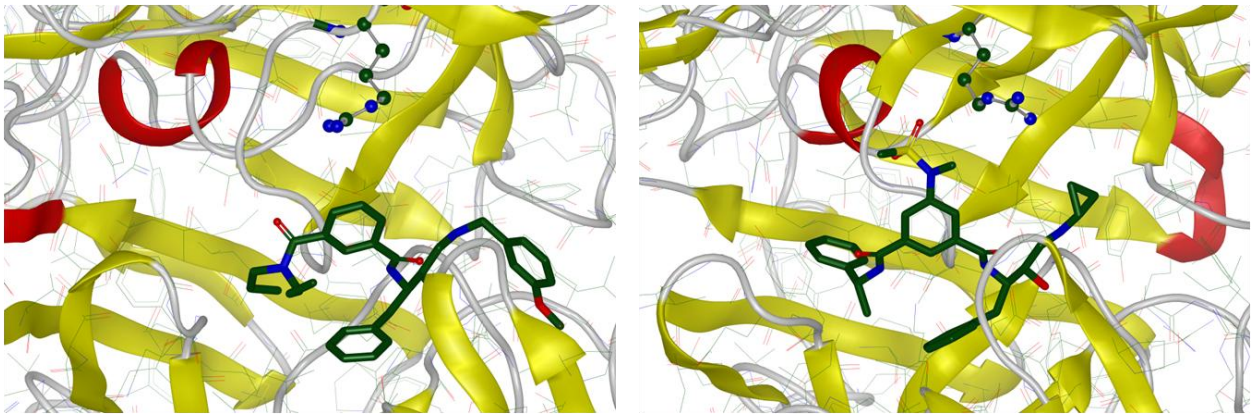


Figure 4. Two different conformations of Arg235 in BACE1 PDB structures

PDB code: 2B8L (left) and 1W51 (right)

2.5. Experimental methods

2.5.1. Overall design strategy

The overall strategy adopted during our pursuit of BACE1 inhibitor-design is summarized below (figure 5). The software RACHEL (Real-time Automated Combinatorial Heuristic Enhancement of Lead compounds) was used to generate the new scaffolds (fused-rings and biaryls). The RACHEL scoring function was amended in order to obtain cyclic structures as hits. FlexX software was used to dock the hits obtained using RACHEL. FlexX score and visualization of poses in the binding pocket of BACE1 was utilized to analyze the results. With inputs from medicinal chemists, two diverse series of fused-ring and biaryl compounds were identified as hits and proceeded forward for chemical synthesis and biological testing.

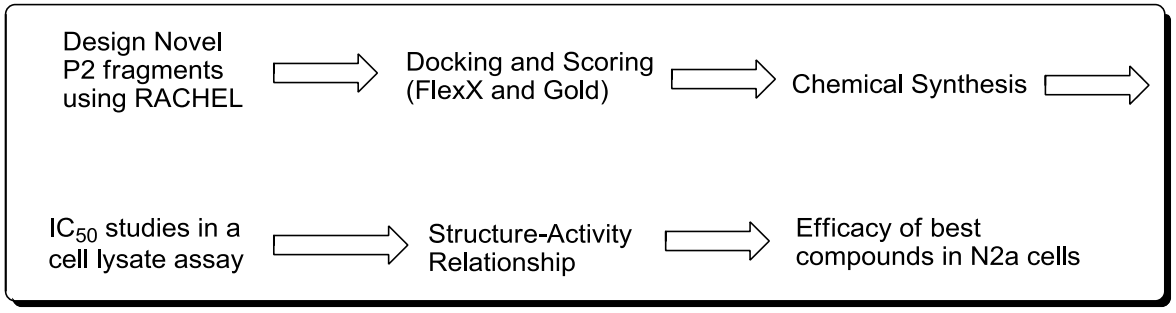


Figure 5. Strategy for BACE1 inhibitor design

2.5.2. De novo ligand-design using RACHEL

De *novo* ligand-design using RACHEL, a software package for automated de novo design, was selected to generate the new scaffolds. We elected to maintain the X and Y substituents in ligand **1**, while allowing changes in portions A, B, and C (figure 6). The bridging group C was designed to maintain orientation of the polar groups (if any) in positions A and B, to decrease the number of rotatable bonds, and to form an additional interaction with the binding

site. The A, B, C fragments were allowed to be either “nothing” (in case of the fragment C only), H, or a group selected from an in-house database of small molecule fragments. This database was generated by RACHEL from a set of drugs and druglike small molecules found on the DrugBank web-site.[87] The RACHEL scoring function was amended to promote selection of cyclic structures. Asn233 was selected as a target site for growing substituents. The molecular weight of the substituents was limited to 200. The newly generated by RACHEL ligands were re-docked to the binding site using FlexX docking software. Among all potential ligands generated by RACHEL bi-cycles and biaryls stood up as groups of ligands having the binding poses simi**la**r to those found in BACE1 X-rays**,** targeting S2 site, and having fewer or no hydrogen bond donors or acceptors. Upon visual inspection and evaluation of synthetic accessibility two scaffolds were selected for further synthetic elaboration: saturated bicycles with fused 6 or 7 ring systems containing a heteroatom and biaryls with 5- and 6-membered terminal aryl (Figure. 6). It was further decided not to use the actual FlexX docking score because (1) docking scores are typically less reliable compared to the accuracy of the binding poses prediction, and (2) we expected that the binding site residues in S2 may change their conformation to accommodate the new substituents. To explore if these new chemotypes are indeed compatible with the binding site of BACE1 two diverse series of biaryl and 5,6,7-membered bicyclic ligands shown in figure 6 were advanced for further medicinal chemistry efforts and biological testing.

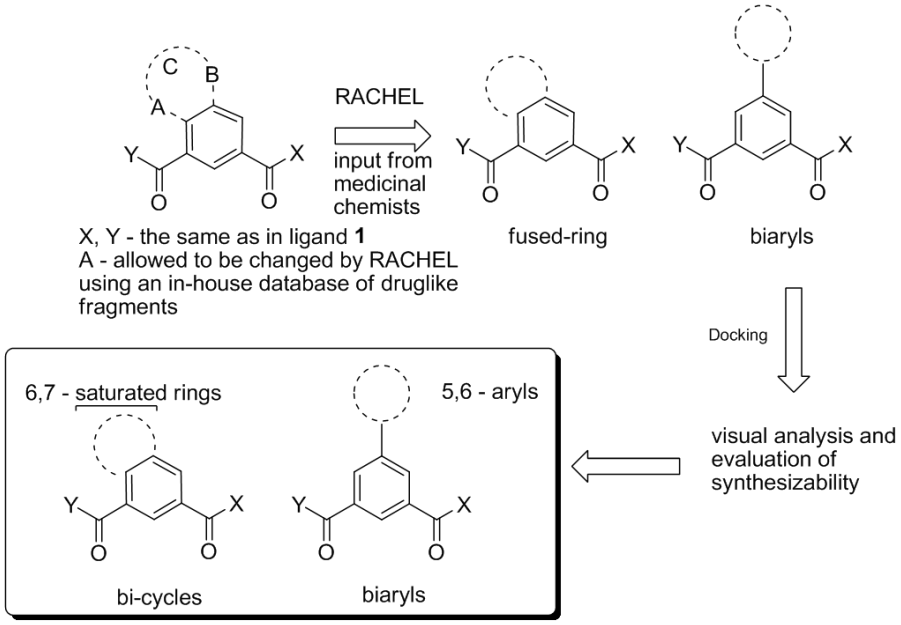


Figure 6. De novo design strategy utilized to find P2 fragments

2.5.3. Molecular docking using GOLD

The subsequent molecular modeling analysis was performed in the following way. The X-ray coordinates of BACE1 (PDB code: 2B8L) were downloaded from the Protein Databank. The protonation state of two catalytically active aspartic acids (Asp32 and Asp228) is the subject of debate. However, based on full-linear quantum chemical calculations carried out by Rajamani et al., and virtual screening studies of a library of BACE1 inhibitors conducted by Polgar et al., we protonated Asp32. [88-89] The active site was defined as a sphere enclosing residues within 10Å of the bound ligand. The 3D structures of ligands were built using Sybyl 7.3 and energy minimized using Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, 5000 iterations, and gradient of 0.005 kcal/mol/Å as the termination criterion. The resulting minimized molecules were docked to the binding site of BACE1 using the GOLD software.[90] All poses outputted by the docking program were visualized; however only the pose with the best fitness score was used for the following SAR analysis. Wherever appropriate, docking under induced-fit mode was

applied to accommodate the side-chain flexibility of Arg235. Prior to the SAR analysis, the ability of the docking program to successfully reproduce the binding mode of co-crystallized ligand **1** was evaluated. It was found that GOLD was able to reproduce the X-ray binding mode of **1** with RMSD of 0.36 Å.

2.5.4. In silico prediction of blood-brain barrier penetration (log BB) of BACE1 inhibitors

We also calculated the logBB values to confirm that the new ligands exhibit improved blood-brain permeation properties. The log BB values of all inhibitors **1, 2a-2c, 3a-3d, 4a-4b, 5a-5e, 6a-6c** were calculated using Clark's equation.[91] The original coefficients in the Clark equation were refitted using the original data in the Clark paper to account for the difference between the polar surface area and ClogP used by Clark and those generated by the Tripos software. The contact van der Waals surface area (cvdWSA) was calculated using MOLCAD separated surfaces between R_1 and the residues Thr231, Thr232, Asn233, Ser325, Arg235, Gln73 and Thr72 proximal to R_1. The figures were generated either with Vida 3.0 or with Sybyl8.0.

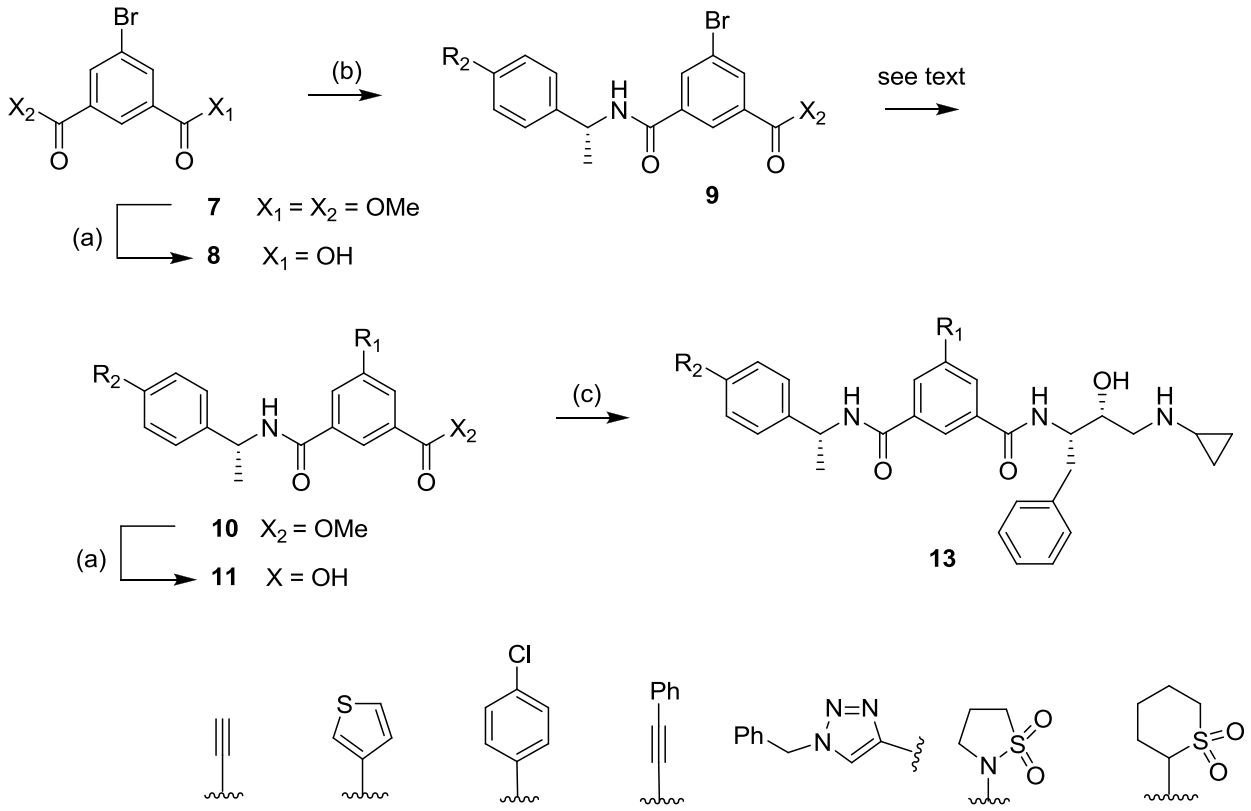
2.5.5. Synthesis of BACE1 inhibitors

The synthesis of BACE1 inhibitors were carried out by our colleague Dr. Srinivas Reddy Chirapu. The biaryl and fused-ring ligands were prepared using a common intermediate **12** according to a published procedure.[92] The synthesis of biaryl ligands is shown in Scheme 1. Commercially available substituted bromoisophthalic acid **7** was converted into **9** by 1) hydrolysis of **7** with 1N NaOH which afforded the monoacid **8** and 2) subsequent coupling of **8** with (R)-α-methylbenzylamine in the presence of EDCI and HOBt resulting in the formation of **9**. Intermediates **10** required for final products **2c** and **3c** were prepared by a Sonogashira coupling of bromide **9** with the mono TMS derivative of acetylene and phenylacetylene,

respectively. Suzuki coupling of 3-thiopheneboronic acid or 4-chlorophenylboronic acid with **9** resulted in intermediates **10** required for the final products **3a** and **3b**. Sultam intermediates **10** required for the final products **4a** and **4b** were prepared by coupling **10** with the corresponding ring sultams. The final compounds **2c, 3a-3d, 4a-4b** were obtained by hydrolysis of **10** to form the acid intermediate **11** followed by subsequent coupling with the TFA salt of (2*R*, 3S)-N-1-2-hydroxy-4-phenylbutane-1, 3-diamine.[93-95] Compounds **2a** and **2b** were prepared using hydroxy and methoxy derivatives of diethyl esters of isophthalic acid in four steps: hydrolysis of one of the ester group with 1N NaOH, coupling with the left side fragment (R)-α-methylbenzylamine, hydrolysis of the remaining ester group using 1N NaOH, and finally coupling with the TFA salt of (2*R*, 3S)-N-1-2-hydroxy-4-phenylbutane-1, 3-diamine.

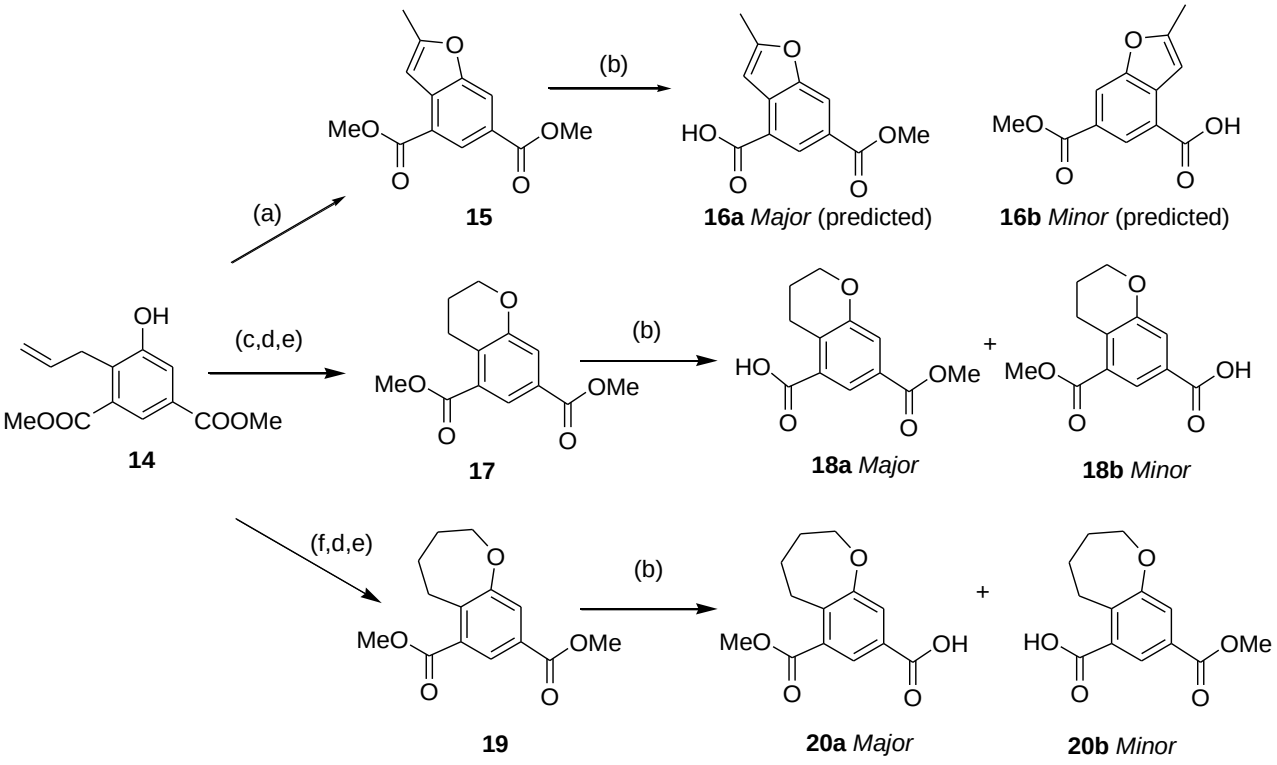
The bis-esters **15, 17** and **19** were prepared according to the reported procedures starting from **14** (Scheme 2).[96-97] Subsequent selective hydrolysis of **17** yielded monoester **18a** as a major product whereas **19** resulted in monoester **20a** as a major product. Further treatment of **16a**, **18a** and **20b** with (R)-α-methylbenzylamine led to the corresponding intermediates **21-23** that upon hydrolysis and subsequent reaction with **12** resulted in the final products **5a**, **5b** and **6c** (Scheme 3)**.** Reaction of **16b**, **18b** and **20a** with **12** resulted in intermediates **24-26** which upon hydrolysis and subsequent coupling with (R)-α-methylbenzylamine led to the final ligands **5c**, **6a**, and **6b** (Scheme 4).

Scheme 1. General Synthesis of BACE1 inhibitors



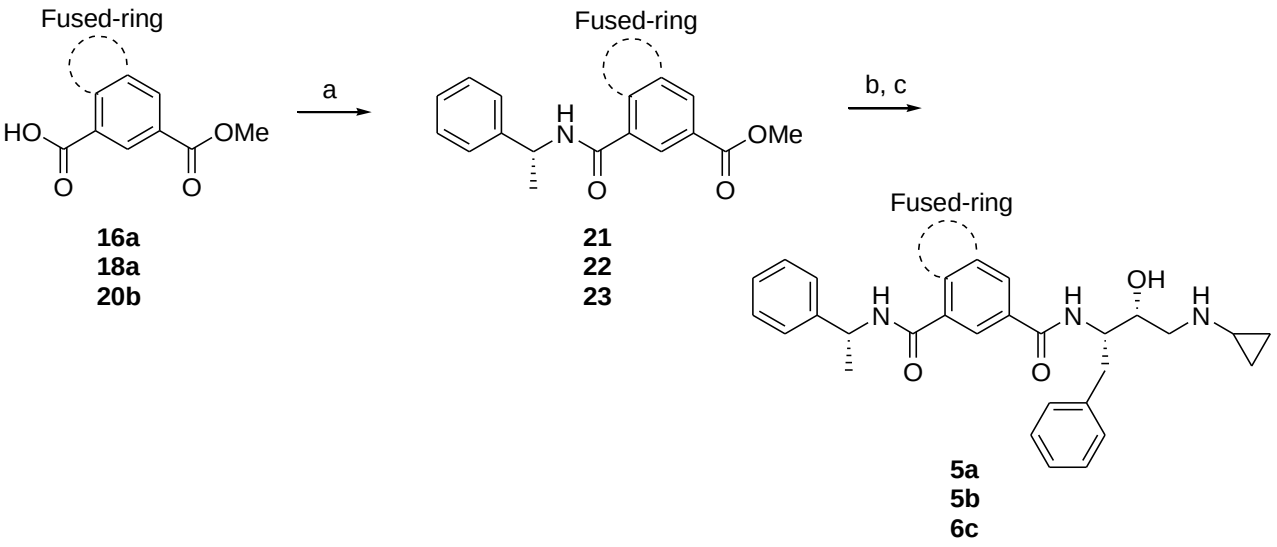
Reagents and conditions: (a) 1 N NaOH; THF: MeOH (50:50); (b) (R)-α-methylbenzylamine, EDCI/HOBt, 0°C to r.t.; (c) **8**, EDCI/HOBt, 0 °C to r.t.; (c) **12**, [92] BOP, DIPEA.

Scheme 2. Hydrolysis pattern of fused-ring esters



Reagents and conditions: (a) $PdCl_2$, $Cu(OAc)_2$, LiCl, MeOH:H_2O (2mL : 0.1mL); (b) 1 N NaOH; THF: MeOH (50:50), 0 °C to r.t.; (c) Tetravinyltin, $Cu(OAc)_2$, O_2, CH_3CN, 2 days; (d) 2nd generation Grubbs catalyst,[98] DCM, r.t.; (e) H_2, Pd/C, MeOH ; (f) allyl bromide, K_2CO_3, acetone, reflux.

Scheme 3. Synthesis of final compounds **5a-5c**



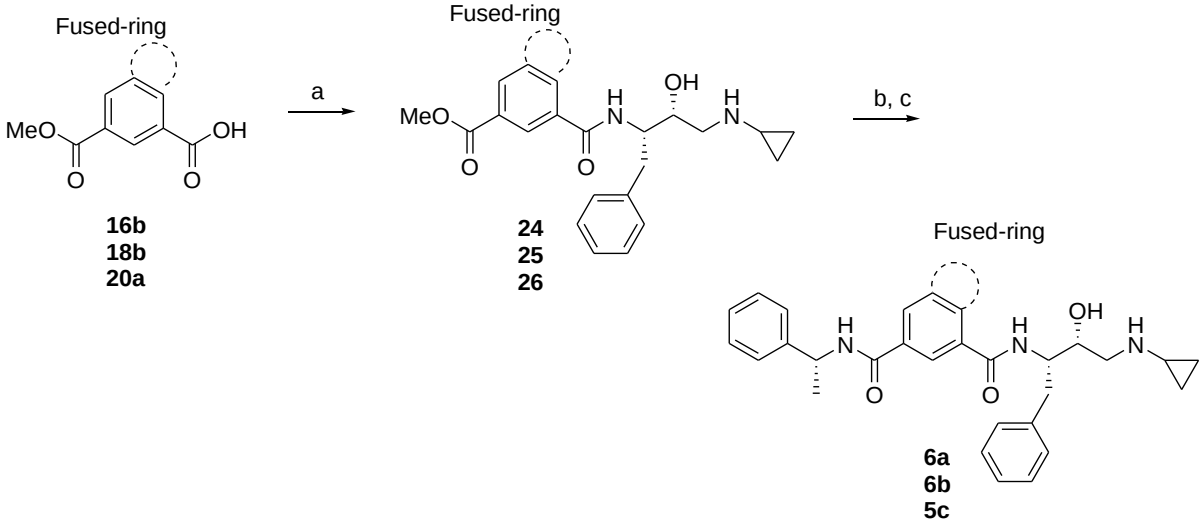
16a
18a
20b

21
22
23

5a
5b
6c

Reagents and conditions: (a) (R)-α-methylbenzylamine, EDCI/HOBt, 0 °C-r.t; (b) 1 N NaOH; THF: MeOH (50:50), 0 °C to r.t.; (c) **12**,[92] EDCI/HOBt, 0 °C to r.t.

Scheme 4. Synthesis of final compounds **6a-6c**



16b
18b
20a

24
25
26

6a
6b
5c

Reagents and conditions: (a) **12**,[92] EDCI/HOBt, 0 °C to r.t.; (b) 1 N NaOH; THF: MeOH (50:50), 0 °C to r.t.; (c) (R)-α-methylbenzylamine, EDCI/HOBt, 0 °C to r.t.

2.5.6. IC_{50} measurements

The IC_{50} studies were carried out by Hikmet F. Nural and Xin Cheng supervised by Dr Yong Shen our collaborators at the Sun Health Institute, Sun City, AZ. For IC_{50} measurements, 293T cells stably transfected with pcDNA-BACE1 were maintained in 200ug zeocin/ml DMEM, 10% FBS. Cell were lysed by lysis buffer (PBS with 1%TritonX100 and 0.1%SDS), lysate were adjusted to 4ug/ul. BACE1 inhibitors were diluted to desired concentration with reaction buffer (100mM Tris-HCl, 100mM NaCl, pH4.5). BACE1 substrate (EDANS-SEVNLDAEFR-DABCYL) was dissolved in DMSO as stock then diluted to 10μM working solution. 20ug of 293T/pcDNA-BACE1 cell lysate were mixed with BACE1 inhibitor and the substrate. The final substrate concentration was 5μM. Fluorescence was detected in microplate reader (Bio-tek) every 5min at emission length 500 nm as well as excitation at about 430 nm. Maximal velocities were calculated by the time point within 20min.

Similarly, 293 cells stably transfected with pcDNA-BACE2 were maintained in 200ug G418/ml DMEM, 10% FBS. Cell were lysed by lysis buffer (PBS with 1%TritonX100 and 0.1%SDS), 5ug protein were used in the assay. BACE2 inhibitors were diluted to desired concentration with reaction buffer (100mM Tris-HCl, 100mM NaCl, pH4.5). BACE2 substrate (MCA-ERHADGLALALEPA(K-Dnp) was dissolved in DMSO as stock. Inhibitors were further diluted to desired concentration with reaction buffer (100mM Tris-HCl, 100mM NaCl, pH4.5). Cell lysate were mixed with BACE2 inhibitors and the substrate (final concentration 5μM). Fluorescence was detected every 5min at emission length 430 nm as well as excitation at about 340 nm. Maximal velocities were calculated by the time point within 20 min.

The activity data of both BACE1 and BACE2 are presented in Table 1.

2.5.7. Assay to detect and quantify $A\beta_{40}$ reduction

The biological assay to quantify the $A\beta_{40}$ reduction was carried out by Samer O. Abdul-Hay supervised by Dr. Gregory R.J. Thatcher at UIC. The following procedure was implemented to determine the efficacy of our best BACE1 inhibitors in reducing the Aβ40 production. Mouse neuroblastoma cells (N2a) stably transfected with Myc-epitope tagged Swedish mutant APP 695 cDNA (a kind gift from Dr. Gopal Thinakaran) were cultured in 1:1 Opti-MEM/Dulbecco's modified Eagle's medium (high glucose) containing 5% fetal bovine serum, 1% penicillin/streptomycin and 0.2 mg/ml Geneticin (G418) in a humified air incubator at 37°C (5% CO_2 – 95% O_2). Cells were plated in 24 well plates for 24h at a concentration of $15x10^4$ cell/well. After 24h, the growth medium was discarded and replaced with 0.5 ml of serum reduced medium consisting of Dulbecco's modified Eagle's medium (high glucose) and 0.2% fetal bovine serum. Fifteen minutes after medium replacement, the cells were treated with drugs dissolved in DMSO and incubated for 24h. At the end of the incubation period, 100 µl of conditioned media were collected and the protease inhibitor AEBSF (4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride) was immediately added at a final concentration of 1mM. Aβ40 concentration was determined using human beta amyloid [1-40] colorimetric ELISA kit from Invitrogen. The procedure was followed as instructed in the supplied protocol.

2.6. Results and Discussions

2.6.1. NMR studies of fused-ring compounds

The 2D-NMR studies were carried out Dr. David C. Lankin from UIC. Theoretically, two regioisomeric monoesters could have been produced from each diester **15**, **17**, and **19** by hydrolysis of the right-side ester or the left-side ester (Scheme 2). The structure of the hydrolysis products produced from **17** and **19** were deduced in a self-consistent manner from the interpreted results of extensive NMR studies using1-D ^1H and ^{13}C, 2-D gradient-selected

correlation spectroscopy (gCOSY), gradient-selected heteronuclear single quantum coherence (gHSQC), and gradient-selected heteronuclear multiple-bond correlation spectroscopy(gHMBC). The assignment strategy employed for the determination of the regioisomeric structure of the monohydrolysis products represents a straight forward approach involving: 1) unambiguous assignment the ^{1}H and ^{13}C NMR spectra for each of the hydrolysis products and 2) the careful examination of the patterns of the $^{3}J_{C,H}$ correlations present in the gHMBC 2-D spectrum of each of the hydrolysis products, specifically the correlations pertaining to the respective carbonyl carbons. In the generalized structure (n = 1 or 2; figure 7a), the aromatic ring proton located between both carboxylate carbons will exhibit $^{3}J_{C,H}$ correlations in the gHMBC to both carbonyl carbon signals (red and blue). The remaining aromatic ring proton (right side) will show the indicated $^{3}J_{C,H}$ correlation to only the carbonyl carbon on the right side (in red). If a carbonyl exists as an ester moiety, i.e., with a methoxy group bonded to the carbonyl carbon, there will be an additional correlation cross peak in the gHMBC 2-D spectrum arising from a $^{3}J_{C,H}$ coupling to the protons of the methoxy group to that carbonyl carbon. In this way, unambiguous assignment of the structure of the hydrolysis products can be made with confidence. This concept is illustrated for the starting dimethyl ester **17** (n = 1, R = R' = Me), for which all of the ^{1}H and ^{13}C resonances were assigned (figures 7b and 7c).

In the case of the hydrolysis product **18a** derived from **17**, the 1-D proton NMR spectrum (400 MHz, DMSO-d_6) confirmed the presence of three chemically distinct methylene groups with resonances centered at δ1.921 (m), 3.027 (triplet, J = 5.2), and 4.175 (triplet, J = 6.6), each integrating for two (2) protons and all associated with the 6-membered ring present in the monohydrolyzed product. The gCOSY spectrum confirmed that the methylene groups constituted an isolated mutually coupled proton spin system. A singlet (3H) appearing at δ 3.837 was assigned to the methyl group of the monoester moiety. Two doublets were observed centered at δ 7.979 and δ 7.428 (J = 1.5 Hz), which were shown to be spin coupled (gCOSY)

and were assigned to the two aromatic protons present on the aromatic ring. The magnitude of the J-coupling between the two aromatic protons confirmed the indicated *meta*–relationship. The protonated carbons could be assigned from the gHSQC spectrum ($^{1}J_{C,H}$) and gHMBC spectrum ($^{2}J_{C,H}$ and $^{3}J_{C,H}$). A broad resonance (1H) appeared at ~ δ 13 ppm and was assigned to the proton of the free acid.

The gHMBC of hydrolysis product **18a** from **17** showed correlations from the proton at δ 7.979 to both of the carbonyl carbons at δ 166.2 and 166.3 ppm. There was a correlation observed from both the proton signal at δ 7.428 to the carbonyl carbon appering at δ 166.3 as well as a correlation cross peak to the ester methyl group at δ 3.837 indicating that the hydrolysis of **17** proceeded at the left-side carbonyl in the generalized structure. In contrast, the 1-D proton NMR spectrum (400 MHz, DMSO-d_6) of the hydrolysis product **20a** from **19** confirmed the presence of four chemically distinct methylene groups with slightly broadened resonances centered at (δ 1.673, 1.899, 2.992-3.018 m, and 4.012 t (J = 6.6 Hz) each integrating for two (2) protons and all associated with the 7-membered ring present in the mono-hydrolyzed product from **19**. As in the case of **20a**, the gCOSY spectrum confirmed that these resonances also constituted an isolated mutually coupled proton spin system. A singlet (3H) at δ 3.854 was assigned to the methyl group of the monoester moiety. Two signals were observed at δ 7.937 and δ 7.582, which were shown to be spin coupled (gCOSY) (J = 1.7 Hz) and which were assigned to the two aromatic protons present on the aromatic ring. The magnitude of the J-coupling between the two aromatic protons is also consistent with the indicated *meta* – relationship. Similar to the NMR of **18a**, the hydrolysis product of **17**, a broad resonance (1H) at ~ δ 13 ppm was also observed and was assigned to the proton of the expected free acid functionality. The protonated carbons could be easily assigned from the gHSQC spectrum ($^{1}J_{C,H}$). The gHMBC of hydrolysis product **20a** from **19** showed correlations from the proton at δ 7.937 to the both of the carbonyl carbons appearing at δ 166.0 and 167.2 ppm. There was a

correlation observed from the proton signal at δ 7.582 to the carbonyl carbon at δ 166.0. The carbonyl at δ 167.2 also showed a correlation in the gHMBC spectrum to a methyl singlet (δ 3.854) indicating that the hydrolysis of **19** proceeded at the right-side carbonyl (red, in the generalized structure). Analysis of the 2-D gCOSY, gHSQC, and gHMBC data together with the 1-D ^{1}H and ^{13}CNMR data obtained for both hydrolysis products permitted a self-consistent assignment of the ^{1}H chemical shifts and of the ^{13}C chemical shifts for both regioisomers. The ^{1}H and ^{13}C NMR assignments of intermediates **17** and **19** and their hydrolysis products **18a** and **20a** are shown in figures 7d and 7e.

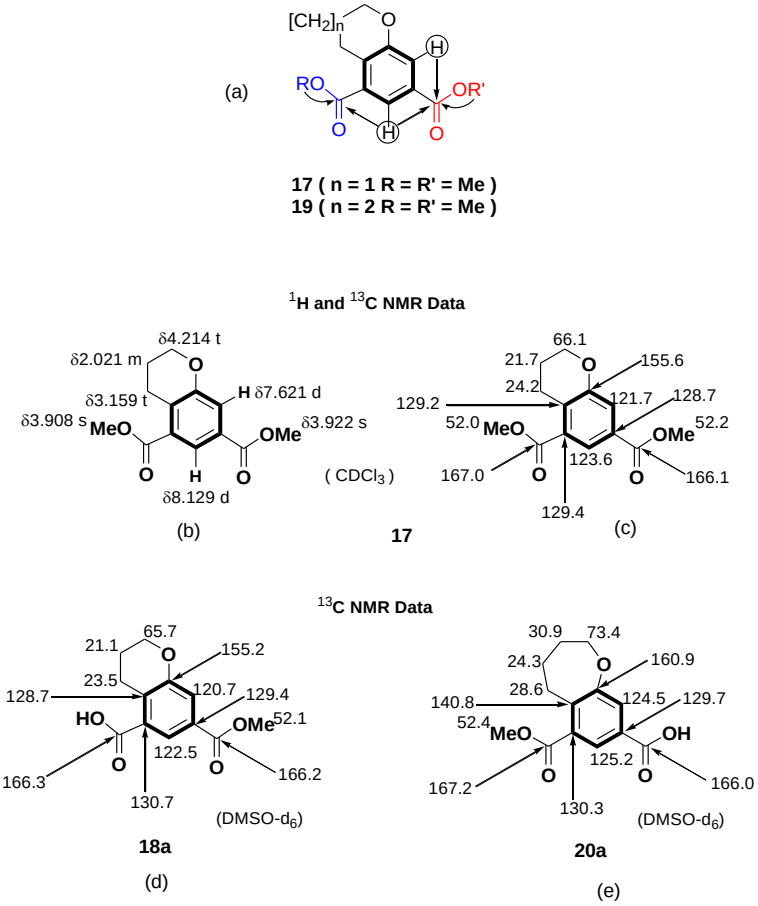
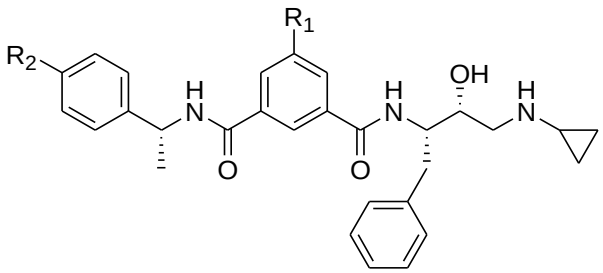
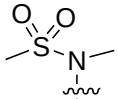
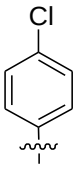
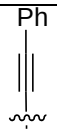
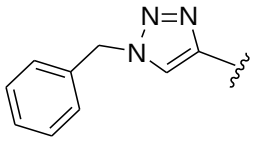
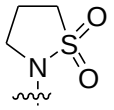
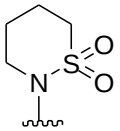


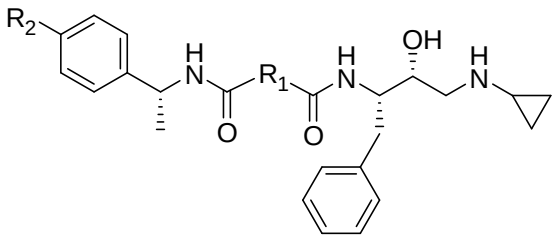
Figure 7. General Structure of ligands and NMR structures

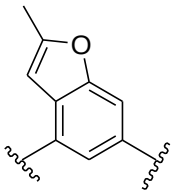
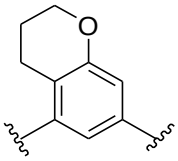
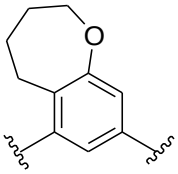
(a) General structure of ligands **17** and **19**; (b) ^{1}H NMR assignments for **17**; (c) ^{13}C NMR assignments for **17**; (d) ^{13}C NMR assignments for **18a**; (e) ^{13}C NMR assignments for **20a**.

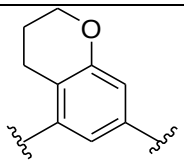
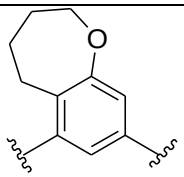
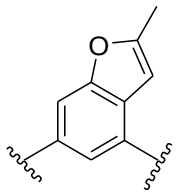
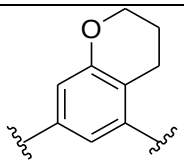
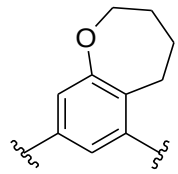


Ligands	R_1	BACE1 IC_{50} (nM)	BACE2 IC_{50} (nM)	log BB[a]
1		281 ± 42	132 ± 17	-0.9
2a	OH	1581 ± 269	> 100000	-0.8
2b	OMe	2506 ± 576	> 100000	-0.5
2c		889 ± 108	> 100000	-0.3
3a		2233 ± 384	> 100000	-0.6
3b		811 ± 102	> 100000	-0.1

3c		1119 ± 173	> 100000	-0.1
3d		1256 ± 140	> 100000	-0.5
4a		1774 ± 218	141 ± 16	-0.9
4b		158 ± 19	251 ± 33	-0.8



Ligands	-R_1-	R_2	BACE1 IC_{50} (nM)	BACE2 IC_{50} (nM)	log BB[a]
5a		H	524 ± 150	> 100000	-0.3
5b		H	63 ± 8	> 100000	-0.3
5c		H	56 ± 7	> 100000	-0.3

5d		F	740 ± 95	2089 ± 311	-0.3
5e		F	262 ± 36	1949 ± 207	-0.2
6a		H	998 ± 78	> 100000	-0.
6b		H	112 ± 12	> 100000	-0.4
6c		H	89 ± 13	> 100000	-0.2

[a]calculated log BB

Table 1. BACE1 and BAC2 inhibition profile of P2 substituents

2.6.2. Activity profile of simple substituents

The best docking poses were visualized in the binding site (figure 8a), and the SAR of the synthesized compounds is discussed below. Substitution of R_1 by a simple hydroxyl or a

methoxy group in **2a** and **2b** (Table 1) resulted in approximately 6- and 9-fold loss in BACE1 activity compared to ligand **1**, respectively, a relatively small trade-off in exchange of improvement of the druglike profile and BACE1/BACE2 selectivity. The docking shows that the solvent-exposed hydroxyl or methyl group in **2a** and **2b** do not form favorable contacts with the binding site. In the induced-fit mode docking, the side-chain guanidine group of Arg235 is able to interact with the hydroxyl or methoxy groups in **2a** and **2b**, respectively. Substitution of R_1 in **1** with an acetylene moiety in **2c** led to an only 3–fold decrease in BACE1 activity. As well as in the case of ligands **2a** and **2b** this is a small trade-off for replacing methyl sulfonamide group of **1** with a small acetylene group in **2c**. This observation supports the notion that the loss of hydrogen- bonding contribution may be compensated by non-directional vdW interactions of the P2 substituents as a result of possible induced fit changes. Indeed, the docking shows that the P2 acetylene moiety forms extensive vdW contact with the alkyl side-chains of Arg235 and Thr231 located in S2. None of **2a**-**c** ligands showed BACE2 inhibition below 100 μM concentration. As expected, the relatively less polar **2a**-**c** are predicted to have better log BB of -0.3 compared to -0.9 for **1**.

2.6.3. Activity profile of biaryl substituents

Encouraged by these initial findings, we synthesized and tested ligands **3a-3d** to explore whether the S2 pocket can accommodate larger biaryl substituents. The docking poses of these compounds are shown in Figure 8b. The 3-thiophene moiety in **3a** did not influence significantly the activity compared to **2b**. A replacement of R_1 with the bulkier *p*-chlorophenyl moiety in ligand **3b** led to a 3-fold increase in potency compared to **3a** and an almost equal activity to **2c**, a ligand with a smaller R_1 acetylene substituent. An analysis of the docking pose of **3b** shows that binding of its *p*-chlorophenyl substituent requires a displacement of the conserved water molecule WAT43 (in 2B8L) typically mediating hydrogen bonding between the sidechains of Ser325, Arg235 and Gln326 (discussed below). It is likely that an induced fit disorder caused by

the removal of the conserved water molecule may contribute to improved binding energy of **3b**.[99] When Gold program was allowed to determine whether WAT43 should be bound or displaced upon binding of **3b**, the resulting ligand-protein complex did not contain WAT43. Probing of the binding site with larger, extended substituents as in the phenylacetylene- and N-benzyltriazole-containing ligands **3c** and **3d** resulted in a 1.4- and 1.6-fold loss of activity compared to **3b**, respectively. The loss was more pronounced in **3d** where the benzyl appendage appears to be completely solvent-exposed. Unlike **3c** and **3d**, ligands **3a** and **3b** fit well within the cleft of the S2 pocket. The predicted log BB values for the biaryl series range from modest -0.6 and -0.5 for compounds **3a** and **3d**, respectively, to -0.1 for both **3b** and **3c**.

2.6.4. Activity profile of sultams

To explore the impact of rigidification/cyclization of the sulfonamide group of ligand **1** on BACE1 and BACE2, compare the resulting ligands with biaryls without sulfonamide group, and to find a future pharmacophore for the hot spot Asn233 (Leu246 in BACE2), we have tested two ring sultams **4a** and **4b**. The six-membered sultam **4b** is found to be about 11-fold more active than five-membered sultam **4** and about 1.8-fold more active than ligand **1** at BACE1. The five-membered sultam **4a**, on the other hand, was 12.6-fold more active at BACE2 than at BACE1. An analysis of the binding modes of ligands shows that the SO_2 group in sultams point towards the S2 site whereas the alicyclic ring is solvent-exposed (figure 8c). This predicted mode of binding is very similar to the one that is observed in X-ray crystal structure of the 6-membered sultam published during the preparation of this manuscript (PDB code: 2VNM). In **4b**, one of the SO_2 oxygen atoms forms a hydrogen-bond with backbone NH of Asn233 (N-O distance = 3.3 Å, N…H…O angle = 170°), whereas the other oxygen atom forms ion-dipole interactions with epsilon nitrogen (NE) of Arg235 (distance = 2.8 Å). Ligand **4a**, owing to its constrained 5-membered ring, forms weak hydrogen bond with Asn233 (N-O distance = 3.7 Å, N…H…O angle = 142°) and weak ion-dipole interactions with NE of Arg235 (distance= 3.7 Å) and hence less

active compared to **4b**. Further, the cvdWSA in $Å^2$ of **4a** and **4b** are 278 and 292 respectively. The calculated logBB values of compounds **4a** and **4b** are similar to that of ligand **1**.

2.6.5. Activity profile of fused-ring compounds

Next, we synthesized a series of fused-ring compounds **5a-5c, 6a-6c** which were subsequently tested for inhibition of BACE1 (figures 8d and 8e). We synthesized and tested three sets of regioisomers that differ by the size and position of the ring fused to the central phenyl moiety. We found that the degree of lipophilicity and flexibility incorporated at P2 position in **5a-5c** and their regioisomers **6a-6c** correlates with the BACE1 inhibitory activity of the ligands: (less lipophilic, less active) **5a** < **5b** < **5c** (more lipophilic, more active) and similarly **6a** < **6b** < **6c**. Upon docking using Gold under induced-fit mode, it was found that ligands **5c** and **6c,** containing the more flexible 7-membered ring, form better vdW contact with the enzyme as compared with the ligands **5b** and **5b**, which contain only a moderately flexible 6-membered ring. In contrast, ligands **5b** and **6b** possess more efficient vdW interactions with the binding site than their 5-membered ring analogs **4a** and **4a** (figures 8d and 8e). The cvdWSA in $Å^2$ calculated under the induced-fit mode for **5a-5c** are 183, 258, 266, whereas for compounds **6a-6c** it is 209, 197, and 227, respectively. The cvdWSA correlates with activity of compounds **5a-5c** but not **6a-6c** suggesting that this parameter alone is not sufficient to explain the trend in activity.

To find a plausible explanation for the difference in activity profiles of **5a-5c** and their regioisomers **6a-6c**, we analyzed the X-ray co-crystal structures of BACE1 protein available in Protein Data Bank. Two interesting observations result from this analysis. First, the group at the P2 position appears to control the position of Arg235 through an induced-fit effect. Depending on the steric bulk and electronic properties of the P2 substituent, Arg235 is forced to adopt different conformations. The maximum change in the location of Arg235 was observed between

1W51 and 2B8L with a deviation of 4.4Å between the two guanidine carbon atoms in the Arg235 sidechain. Second, in those cases where the P2 has N-methylsulfonamide moiety (PDB codes: 1TQF, 2IRZ, 2B8L, 2IS0, 2NTR, 2OAH, 2P4J, 2P8H, and 2PH6), a conserved water molecule mediates hydrogen bond between the sidechains of Arg235, Ser325 and Gln326. On the other hand, if Arg235 changes its position as in 1W51 the water molecule is not observed in the crystal structures. It appears that both observations are interconnected – those ligands responsible for the movement of Arg235 also contribute to the displacement of the water molecule and loss of the hydrogen bonds between this water molecule and residues Ser325, Arg235 and Gln326. To account for the induced fit effect Arg235 was allowed to change conformation during docking with Gold. Fused-ring compounds **5a-5c** has an oxygen atom locked in a position amenable for hydrogen bonding with side-chain NH_2 of Arg235. Our docking experiments suggest that in order to facilitate this hydrogen bond, Arg235 must lose its contact with water (WAT43 in 2B8L). Similar hydrogen-bonding interaction is not possible in regioisomers **6a-6c** because the ring precludes their contact with Arg235 and hence a near 2-fold drop in potency compared to **5a-5c** is observed. The calculated logBB is -0.3 for ligands **5a**-**5c** and -0.3, -0.4 and -0.2 for **6a**-**6c**, respectively. It is improved compared to the ligand with the sulfonamide moiety bound to S2 pocket. As with the other ligands in the biaryl and fused ring series that according to the docking experiments require an induced fit, ligands **5a**-**c** and **6a**-**c** did not show inhibition of BACE2.

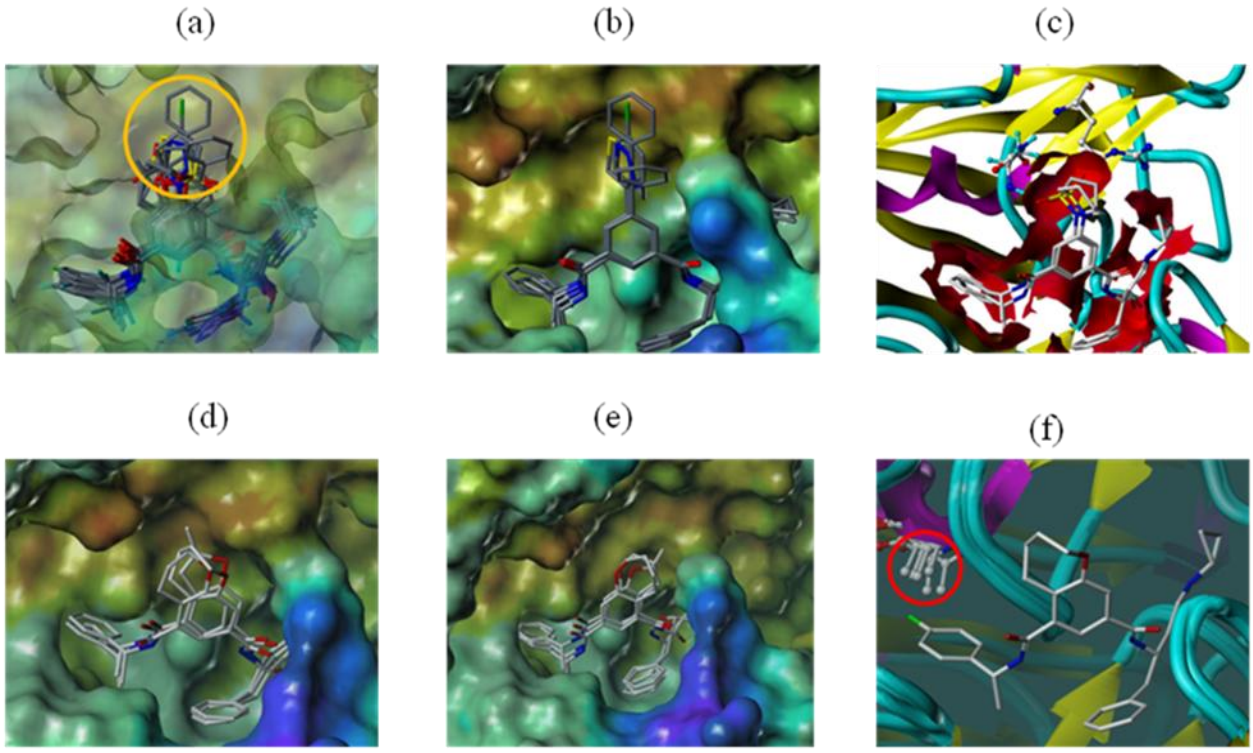


Figure 8. Gold docked poses of all fused-ring ligands in BACE1 binding site

(a) Gold docked conformations of all ligands in BACE1 binding site. The P2 moieties are circled yellow; (b) Biaryl compounds 3a-3d in BACE1 binding site; (c) Cyclic ring sultam compounds 4a-4b in BACE1 binding site rendered as ribbon and tube representation. Residues Asn233 and Arg235 are rendered as ball-and-stick representations; (d) Overlay of docked poses of ligands 5a-5c when side-chain flexibility of Arg235 is considered; (e) Overlay of docked poses of ligands 6a-6c when side-chain flexibility of Arg235 is considered; (f) Ligand 5e in BACE1 binding site. Different orientations of methyl group in Ala335 observed in X-ray crystal structure is rendered as a ball-and-stick representation and circled in red for clarity.

2.6.6. Activity profile of fluorinated compounds

Several publications have reported that fluorinated BACE1 inhibitors exhibited better activity than the non-fluorinated ones due to increased lipophilicity.[92, 100] Unlike the compounds described in these publications, both fluorinated compounds **5d** and **5e** are found to be less active than non-fluorinated analogs **5b** and **5c** by c.a. 5- and 12-fold, respectively. Ligand **5e** exhibits 2.8-fold better activity than **5d** perhaps for the same reasons **5c** is more active than **5a**

and **5b** – larger area of contact with the binding site due to flexibility of the seven-membered ring. To understand what could be the reason for overall lower activity of **5d** and **5e** compared to their non-fluorinated analogs we visually inspected the crystal structures of BACE1 with different P2 substituents and found that the induced-fit effect of P2 affects the orientation of the methyl group in Ala335 (Figure 8f). Of particular interest, the RMSD between CB atoms for crystal structures 2P8H (P3 = p-fluoro-α-methylbenzylamino, P2 = N-methylsulfonamide) and 2IQG (P3 = N, N-dipropanoyl and P2 = Me) is 1.2 Å. Because of the proximity of the methyl group in Ala335 and the fluorine atom in the BACE1 inhibitors, any conformational changes in S2 may result in mutual steric clashing. If an induced fit caused by the fused rings in series **5** and **6** takes place it may, indeed, affect the position of Ala 335 and explain the overall lower activity of the fluorinated compounds **5d** and **5e** compared to their non-fluorinated analogs. The predicted log BB values of **5d** and **5e** are almost the same as those of the non-fluorinated compounds.

2.6.7. Comparison of BACE1 and BACE2 inhibition profile

To find a possible explanation for BACE1 selectivity of the ligands we tried to compare qualitatively the flexibility of the BACE1 and BACE2 proteins using their b-factors. A quantitative comparison of the b-factors is not possible since there is large number of the same residues in different X-ray structures of BACE1 proteins that differ in their b-factors depending on the co-crystallized ligand. It is clear, however, that the flexibility patterns for the residues in the S2 pocket and adjacent to it areas of the BACE1 and BACE2 proteins are different, suggesting that induced fit may play a role in an improved selectivity toward BACE1.

2.6.8. Effect of BACE1 inhibitors on APP processing

To investigate the effect of BACE1 inhibitors on APP processing the most active ligands **5b** and **5c** were tested for reduction in $A\beta_{40}$ production in mouse neuroblastoma cells (N2a) transfected with Swedish mutant APP695 (figure. 9). Compounds **5b** and **5c** reduce ca. 65%

and 35% of Aβ production, respectively, compared to the control. At the moment it is unclear why almost equally active **5b** and **5c** exhibit 1.9-fold difference in reduction of Aβ40. It may be related to their metabolic stability or delivery to the site of action. This observation is currently actively investigated.

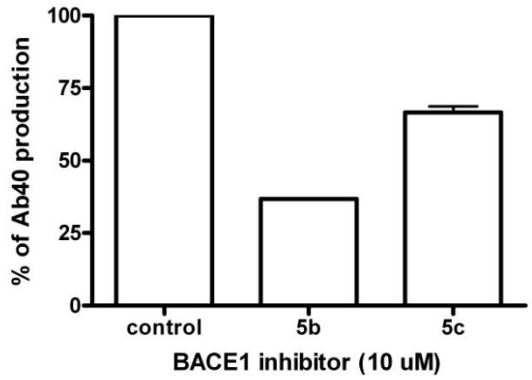


Figure 9. Reduction of Aβ40 levels by BACE1 inhibitors detected in N2a cells stably transfected with Swedish human APP

2.7. Conclusions

In summary, by using computer-aided drug design methods, we have designed, synthesized, tested the activity of the fused-ring and biaryl compounds against BACE1, and analyzed the resulting SAR using docking protocols. The fused-ring compounds are in general more active than the biaryl-based ligands with an activity range from 56 nM to 998 nM. This is comparable to the activity of the ligands with polar substituents occupying the S2 binding pocket, which lends support for our initial hypothesis that the S2 site may be targeted with less polar substituents. Most of the ligands displayed more favorable calculated logBB. The side-chain flexibility of Arg235 and, perhaps, adjacent residues, and possibly the presence of a water molecule mediating hydrogen bond interactions between Arg235, Ser325 and Gln326 in S2 appears to play an important role in accommodating the fused-ring and biaryl-based ligands in

the binding site. The fused-ring ligands **5b** and **5c** combine the best attributes of acceptable logBB, BACE1 inhibitory activity, BACE2 selectivity, and ability to reduce the Aβ production. In addition, the enthalpic optimization achieved by positioning oxygen atom amenable for h-bonding with Arg235, and entropic optimization achieved by incorporating a constrained ring that would undergo minimum conformation entropy loss during binding, makes compounds **5b** and **5c** a successful example. The current work that describes computer-assisted design, synthesis and screening of BACE1 inhibitors has been published.[101] Finally, a similar structure-guided molecular modeling approach may be used to explore other binding pockets of BACE1 to create new chemotypes for further development of therapeutics to treat AD and to serve as general probes for drug discovery.

3. LIGAND-BASED MODELING OF BETA-SECRETASE1 INHIBITORS

3.1. Introduction

3.1.1. Importance of 3D-QSAR in Drug Design

Despite the fact that the X-ray based SBDD has contributed to the discovery of a number of drugs and late-stage clinical candidates, there exists some serious limitations: ambiguities in identifying nitrogen and oxygenatoms, flexibility of ligand/proteins, induced-fit effects, scoring accuracy, position of water molecules, effect of crystallization conditions on protein conformation, etc.[102-103] In such cases, 3D-QSAR methods have proven to be a powerful supplement to predict the binding affinity of an unknown compound. It is based on the simple premise that medicinal chemists have known for several decades: compounds with similar physical and chemical properties also have similar biological properties. The easy-to-use 3D-QSAR method has resulted in several thousands of publications during the last decade itself (for the most recent review, see references).[104-106] During the early stages, 3D-QSAR was employed to predict the binding properties of focused structural analogs, however later advancements such as pharmacophore modeling expanded the horizons to predict the binding properties of diverse molecules.[107-108]

One of the hot topics of debate in the field of CADD is whether 3D-QSAR is necessary if one has the 3D-structure of the protein.[109] Ambiguities in X-ray structures, time-intensive free-energy perturbation calculations, approximations in force fields and electrostatic calculations, inefficient parameterization of scoring functions have made 3D-QSAR an excellent, surrogate method for the prediction of binding affinity of molecules. For a series of molecules and their corresponding binding affinities (IC_{50} or K_i), a 3D-QSAR equation may be derived in order to forecast the binding affinity of unknown molecule(s) within the structural class. Like any

other methods, the 3D-QSAR methods do have limitations:[106] (i) sensitivity of external dataset prediction depends upon the nature of the bioactive conformation and alignment; (ii) influence of basis set selection on predictive power; (iii) nature of molecular descriptors used in the study; (iv) inability to extrapolate the model. Nevertheless, the role of 3D-QSAR methods in combinatorial chemistry and high-throughput screening is continuously growing.[105] In this chapter, one of the powerful 3D-QSAR techniques, CoMFA, will be discussed in detail followed by its applications in understanding the binding requirements of a series of BACE1 inhibitors.

3.1.2. CoMFA – an introduction and overview

The inception of Comparative Molecular Field Analysis (CoMFA) is traced back to the days when Cramer and Milne made a first attempt to compare the molecules by aligning them in space and mapping their molecular fields in a 3D grid.[110] The technique has taken a full form when the first CoMFA application was published in predicting the binding affinity of steroids to human corticosteroid and testosterone binding globulins. [111] Since then over 1000 publications have been reported worldwide that highlighted the improvement in methodology, scope and applications of this technique (for review, see the references).[106, 112-113]

The underlying idea of CoMFA is that differences in target property are often related to differences in the shape of the non-covalent fields surrounding the tested molecules. The following steps are required to derive a valid CoMFA model:[114] (i) selection of molecules for training and test set. The major prerequisite in this step is that all the molecules included in the dataset must interact with the same receptor in the same manner. Also, the biological activities of the molecules must have been tested in identical conditions and the activity may span about three or more log units of K_i or IC_{50}; (ii) generation of multiple low-energy conformations of the molecules in the dataset. The X-ray structure of the bioactive conformation of a ligand is a good starting point. Some of the computational programs for the efficient, fast-generation of

conformations are Omega, Tripos and MOE; (iii) Calculation of partial atomic charges of molecules. The widely used charges are Gasteiger-Marsili, Gasteiger-Huckel, MMFF94, AM1BCC; (iv) Alignment of molecules using a template ligand or a set of alignment rules. A good alignment is the single most important requirement for doing a traditional CoMFA analysis. The common substructure usually has the same conformation in all molecules, and other remaining parts of the molecule are superimposed by adjusting the internal torsional angles. Some of the widely employed alignment methods include FlexS, docking-based alignment, database alignment, rapid overlay of chemical structures (ROCS),[115] etc.; (v) Placement of the training set molecules in a three-dimensional grid to calculate the interaction energy using probe molecules such as sp^3 carbon with a +1 charge. In this step, the interaction between the training set molecules and the probe atoms at each grid point is calculated by calculating the magnitude of the steric (Lennard-Jones) and electrostatic (Coloumbic) fields throughout the defined region. (vi) Partial least-square (PLS) analysis of the calculated energy values and the descriptor values to calculate the conventional r^2, q^2 (cross-validated r^2), standard error of prediction (SEP) in order to identify the optimum number of components. (vi) Validation of CoMFA model by predicting the activity of the unknowns. A good CoMFA model should show satisfactory statistical significance, explanatory capability of the variance in the activity of the training set and the predictive power of the potency of the new compounds.[114] Once an acceptable CoMFA model is generated, then it can be manipulated using various graphic techniques such as coefficient contour maps to understand the binding characteristics of the protein and the ligand.

3.1.3. Points to consider before CoMFA analysis

The application of any statistical method depends on a proper experimental design – in this case, an appropriate selection of training set from which a QSAR model is derived and a test set whose binding properties are to be predicted. When this is neglected in scenarios such as biased dataset or including molecules containing diverse functionalities, severe problems can

arise.[116] Enough care must be taken such that the training set compounds should span at least 3 to 4 log units to obtain a meaningful model. The structures of training and test set compounds should not be drastically different – usually congeneric series of ligands are preferred in order to derive meaningful models.

3.1.4. Important concepts and their definitions used in QSAR analysis

The following are the some of the important concepts and definitions used in 3D-QSAR analysis.

Correlation Coefficient (R^2): A parameter which measures how closely the observed data fits with the regression line. The value of R^2 falls between 0 and 1. An R^2 of 0 means there is no correlation between observed biological activity and the parameters selected for the study.

Cross-validated R^2 (q^2): A measure of ”goodness of internal prediction”. If $q^2 > 0.6$, the model is considered fairly good. If $q^2 = 0.4 – 0.6$, the model is questionable. If $q^2 < 0.4$, the model is poor.

Predictive correlation coefficient (R^2_{pred}): For test set,

$$R^2pred = \frac{SSD - PRESS}{SSD}$$

where SSD = sum of squared deviations = Σ [(target data value) – (target data mean)]^2 PRESS = predictive sum of squares = Σ [(target data value) – (corresponding predicted value)]^2.

Leave One Out (LOO) method: A method to determine how accurately a model will be able to predict data that it was not trained on. This is carried by training the model multiple times using all but one of the training set data points. This method is commonly used for cross-

validation analysis. Sometimes, leave-N-out method is also used where more than one data points.

Bootstrapping: A validation method used in QSAR analysis to calculate the confidence intervals (mean and standard deviation) for the parameters to be estimated.

Root mean square error (s): A measure of the target property uncertainty that is still unexplained after deriving the QSAR equation. The magnitude of ***s*** is by the measurement scale of a target property.

$$s = \sqrt{\frac{PRESS}{n - c - 1}}$$

where n = number or rows and c = number of components

QSAR equation: An equation that describes prediction of a target property based on an explanatory property, a set of coefficients and an intercept. For example,

Prediction = intercept + (explanatory1 * coeff1) + (explanatory2 * coeff2) + …

Fisher Statistics (F-ratio or F-value): A weighted ratio of explain/unexplained target property (R^2/1-R^2). A higher F-ratio in QSAR analysis is desired as this is an indication of significantly greater ability of the QSAR model to predicting the targeting property.

Residual: The difference between actual and predicted value of a target property. A large residual value means that that particular compound is not appropriately modeled.

Partial Least Square Analysis (PLS): A powerful tool for deriving multi-linear relationships among columns of data. For a given data, PLS will look for linear expression relating column variance in target properties (Y-block) to variations in explanatory properties (X-block) so as to minimize the sum of squares of deviations from the model thus produced.

Principal Component Analysis (PCA): A redundancy reduction mathematical procedure that transforms a number of possible correlated variables into a smaller number of uncorrelated variables called principal components.

3.1.5. Scope of our study

The aim of the present study is to understand the SAR of BACE1 inhibitors for which the IC_{50} values were measured in the same conditions. This analysis will help us not only to elucidate the binding characteristics of the enzyme, but also to predict binding affinity of future ligands.

3.2. Experimental Methods

3.2.1. Overall scheme for CoMFA analysis

The overall scheme for our CoMFA modeling analysis is shown in figure 10. Our strategy to perform CoMFA analysis consists of the following steps: (i) filter ligands whose activity is tested under identical conditions and sort them out in the training or test dataset; (ii) minimize the ligands and align them using ROCS; (iii) assign Gasteiger-Huckel charges; (iv) calculate CoMFA field values; (v) PLS analysis and deriving statistical parameters; and (vii) generate contour maps and explain the trend in activity.

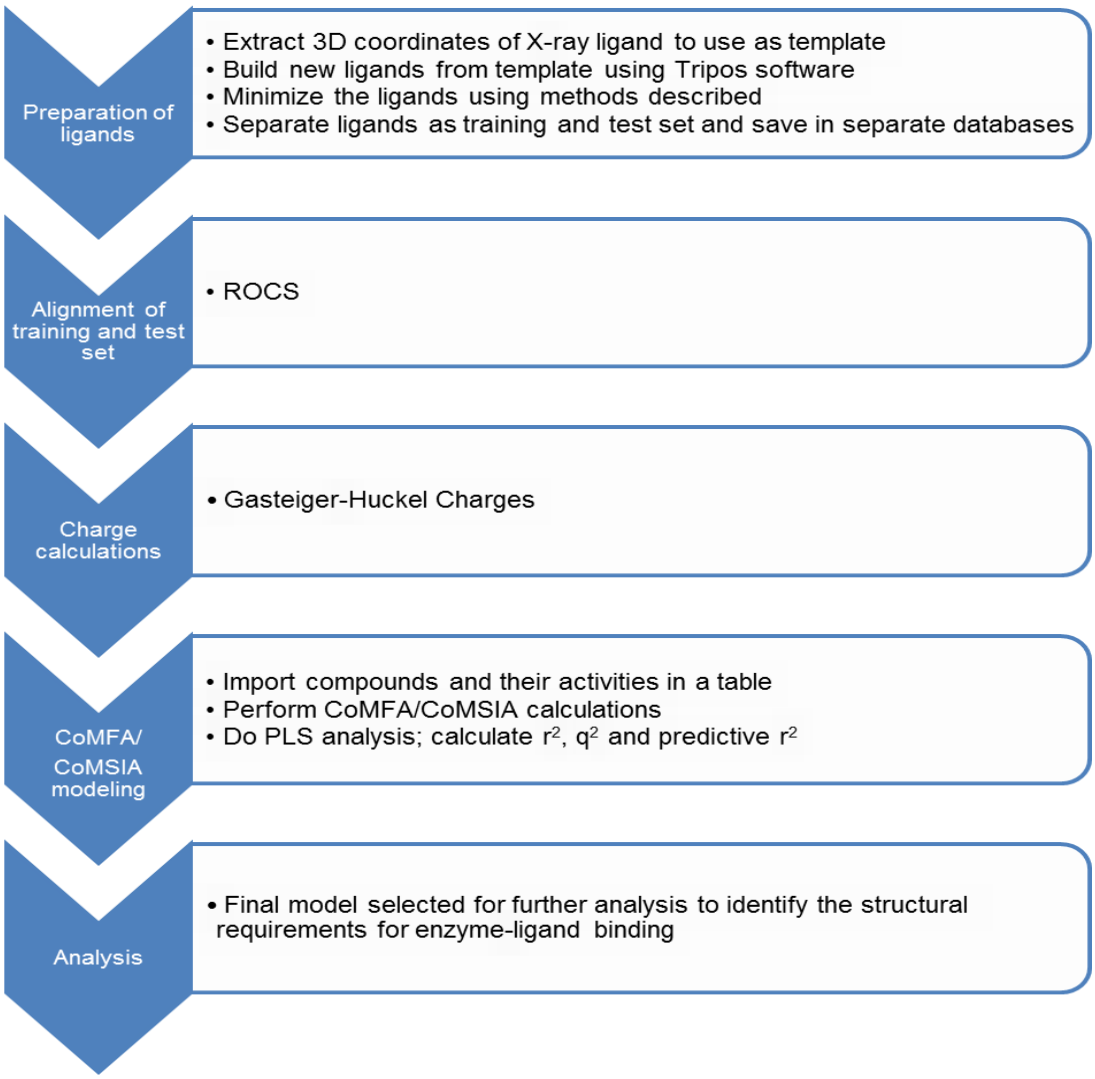


Figure 10. Overall scheme of CoMFA modeling of BACE1 inhibitors

3.2.2. Dataset for CoMFA analysis

In the present work, a total of 51 BACE1 inhibitors were selected for CoMFA analysis.[84-85, 117-122] The general template for these compounds is illustrated in Figure 11. The reported compounds albeit belong to a congeneric series showed variations in the nature of the substituents and a broad inhibitory activity profile ranging from 2nM to 17 μM (4 logarithmic units) thereby making this dataset ideal for performing 3D-QSAR analysis. Because the

parameter q^2 appears to be a necessary but not a sufficient condition for a model to have high predictive power, an emphasis has been given in the current study for validation of the developed models using an external test. Of the 51 compounds in the dataset, 44 of them were randomly chosen as training set and the remaining 7 as test set (bold entries in Table 2).

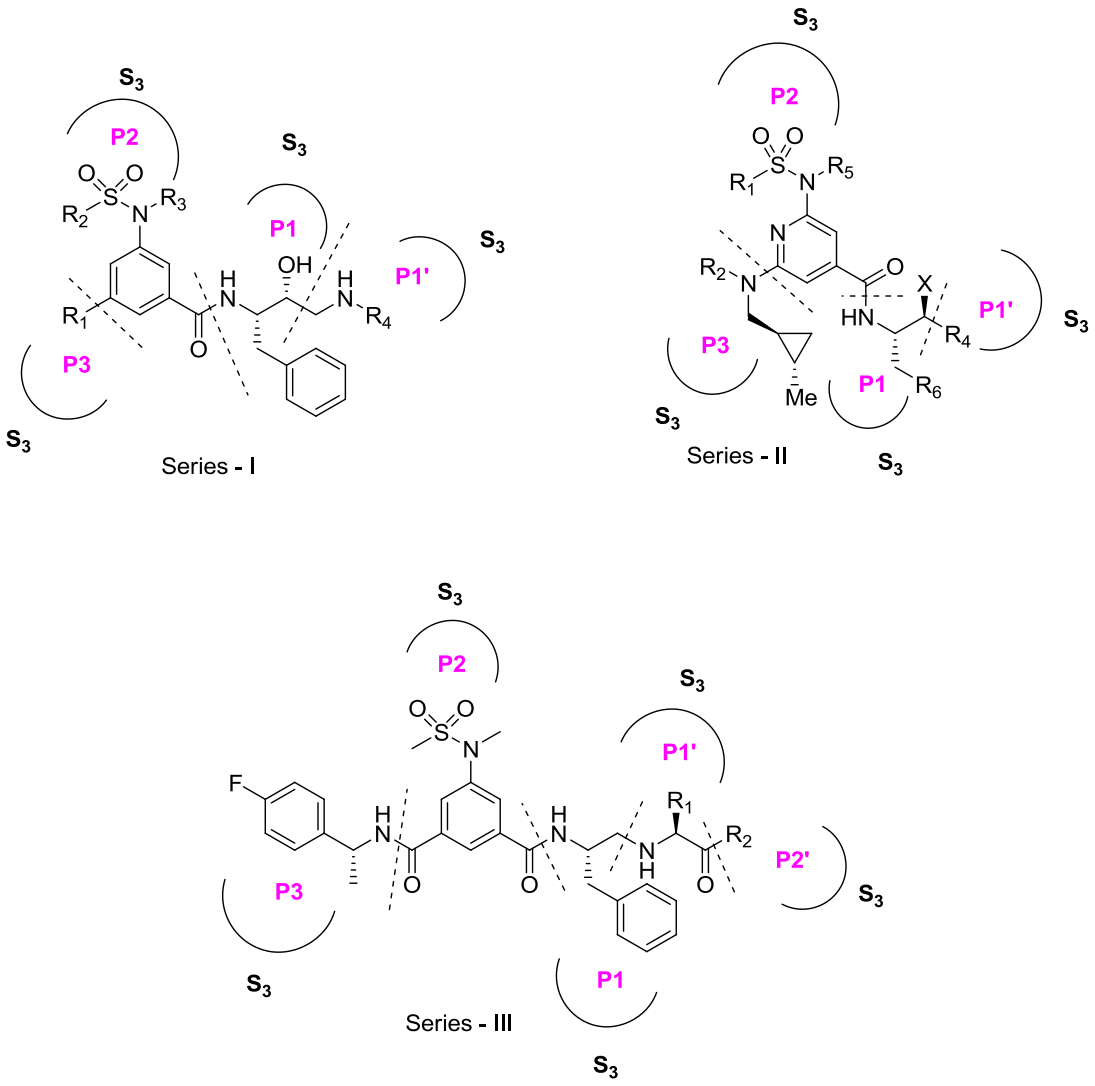


Figure 11. Template of compounds and their corresponding BACE1 interaction sites

3.2.3. Preparation of ligands

All our calculations were performed on a Linux workstation running Sybyl 9.0. The three-dimensional structures of inhibitors were built starting from the X-ray structure of 2B8L, one of the most active BACE1 inhibitors. The X-ray structure of 2B8L ligand is used as a reference ligand for future alignment purposes. Initial optimization of the structures was carried out using the following settings: Tripos force field, BFGS method, gradient 0.0001 kcal/mol/Å, Gasteiger-Huckel charges and 5000 iterations. Conformational energies were computed with electrostatic terms; the lowest energy structures finally minimized were used in alignment.

3.2.4. Alignment methods for training and test dataset

The most crucial point in 3D-QSAR analysis is the alignment of molecules in the test dataset because the CoMFA field values differ depending on the nature of the ligand pose. In our study, we used a relatively new alignment methods called Rapid Overlay of Chemical Structures (ROCS).[115] In brief, ROCS performs fast, shape-based comparison based on the premise that molecules have similar shapes if their volumes overlay well and any mismatch in volume is a measure of dissimilarity. Prior to ROCS analysis, Gasteiger-Huckel charges were assigned. In this method, a smooth Gaussian function is used to represent the molecular volume in order to routinely minimize to the best global match. ROCS uses both shape and chemistry (donor, acceptor, cation, hydrophobicity, etc.) to identify the conformations that matches with the reference ligand. The best 3D overlay of each molecule in the data set, obtained using ROCS, was used as an input for CoMFA calculations.

3.2.5. Biological Activity (IC50) and pIC50 calculations

The biological activity of BACE1 inhibitors used in our calculations was tested in identical conditions.[84-85, 117-122] The IC_{50} values were converted into pIC_{50} using: $pIC_{50} = -\log IC_{50}$.

3.2.6. Parameters for CoMFA analysis

The CoMFA calculations were carried out by applying the default settings in the Sybyl program. To derive the CoMFA descriptor fields, a 3D cubic lattice with grid spacing of 2Å in x, y and z were created to encompass the aligned molecules. CoMFA descriptors were calculated using an sp^3 carbon as a probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to steric (Lennard-Jones 6-12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric energy values at each lattice point were truncated at a default value of 30 kcal/mol.

3.2.7. Partial least-square (PLS) calculations and validations

To derive the 3D-QSAR, a partial least-square analysis (PLS) approach was used. The calculated CoMFA molecular fields were used as independent variables and pIC_{50} as dependent variable. The optimal number of components was determined using cross-validation and leave-one-out method. To speed up the calculations, columns with values below 2.0 kcal mol^{-1} were filtered off, i.e., those columns (lattice points) whose energy variance is less than 2.0 kcal/mol is omitted from the analysis. The cross-validated R^2 (q^2) that resulted in optimum number of components and lowest standard error of prediction were taken for final analysis to calculateR^2_{pred}, F-value and standard error of estimate (SEE). To further assess the robustness and the statistical confidence of the derived model, boot strapping analysis was performed.

3.2.8. CoMFA contour maps

The CoMFA contour maps were generated as a scalar product of coefficients and standard deviation associated with each column. The favored and disfavored levels are fixed at 80% and 20%. Various colored fields are produced during visualization of CoMFA analysis. In CoMFA contour maps, the steric fields are shown in green (more bulk favored) and blue (less

bulk favored), whereas the electrostatic field contours are displayed in red (electronegative substituents favored) and blue (electropositive substituents favored).

3.3. Results and Discussions

The purpose of 3D-QSAR studies using CoMFA is to build statistical and graphical models in order to relate the properties of molecules (for example, IC_{50}) to their chemical structures.[123] The statistical tools in 3D-QSAR include principal component analysis (PCA or factor analysis) for uncovering relationships between descriptors, partial least squares (PLS) regression in order to analyze continuous regression data (IC_{50}, K_i, etc.). Bootstrapping and cross-validation techniques are used at the later stages of the analysis in order to test the predictive power of the 3D-QSAR model, to diagnose chance correlation and to ensure the robustness of a model. In 3D-QSAR analysis using CoMFA, the data and the results of the statistical analyses can be displayed as scatter plots, distributions or histograms for improved understanding. Several publications have been reported that highlights the success of CoMFA. A 3D-QSAR model built using CoMFA can be used not only to predict the binding affinity of the unknowns but also as an aid during lead optimization/design process.[104, 112-113, 123-126] A typical CoMFA analysis requires about 20 to 60 chemical compounds that meet the following requirements: (i) all the ligands bind to the same receptor in the same binding pocket in the same binding mode; (ii) the biological activity of these compounds must have been tested under identical conditions; (iii) ligands should be a part of a congeneric series and thus must be efficiently aligned; (iv) the compounds used in the dataset should have diverse functionalities and must display a broad activity range (at least 3 or more logarithmic units). In a typical 3D-QSAR analysis, a potent ligand is used as a template for aligning the remaining ligands in the database.

In the current study, 51 hydroxyethylamine-, amine-, hydroxyl-derived BACE1 inhibitors has been used. All these inhibitors not only bind to the catalytic site of the enzyme BACE1, but

also display similar binding modes (based on X-ray data of the representative candidate ligands). The IC_{50} values of these inhibitors are measured using identical conditions. For simplification purposes, these 51 BACE1 inhibitors were divided into three series: Series – I comprises hydroxyethylamine derivatives, Series – II comprises of primary amine or hydroxyl derivatives and Series III comprises of 2° amine derivatives. Overall, all these inhibitors display diverse functionalities including polar, hydrophilic, hydrophobic, bulk groups. Further, the pIC_{50} of these inhibitors range from 4.76 to 8.7 (about four logarithmic units) thereby making these inhibitors an ideal dataset for 3D-QSAR analysis. Table 2 lists the compounds chosen for our study. Once the 3D structures of these inhibitors have been built, they have been randomly divided into training and test set (entries in bold in Table 2) and proceeded for alignment. In the present study, we used a relatively new alignment tool called ROCS that takes into account both shape and chemistry during superposition of molecules. For alignment purposes, the X-ray crystal structure of **compound 38** has been used as a reference. The output conformations of ROCS were ranked according to their similarity with the template molecule using Tanimoto Coefficients. Every conformation of the inhibitors is then visually analyzed to identify the best conformer for CoMFA analysis. Figure 12 shows the alignment of minimized conformations of compounds in the CoMFA dataset generated using ROCS.

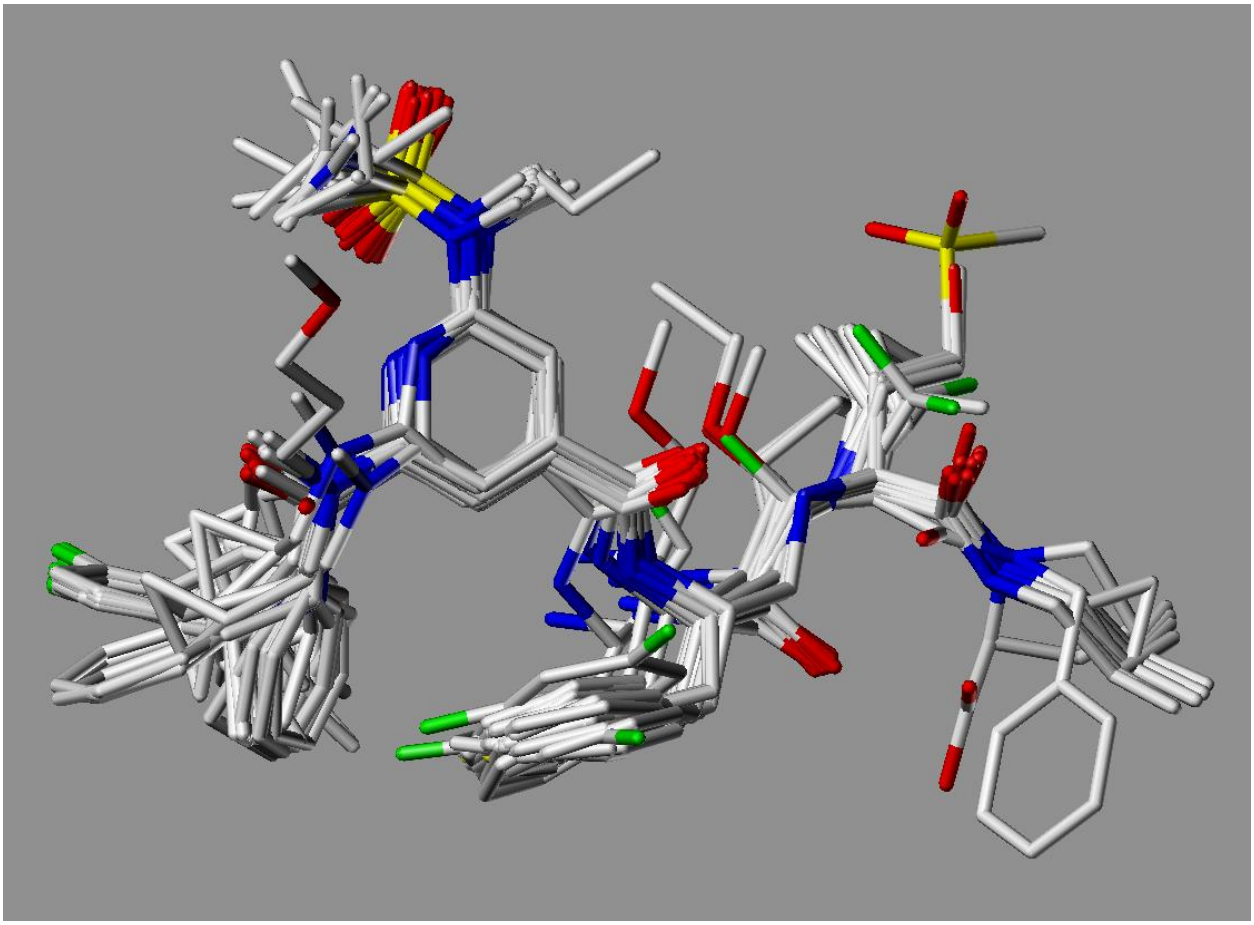


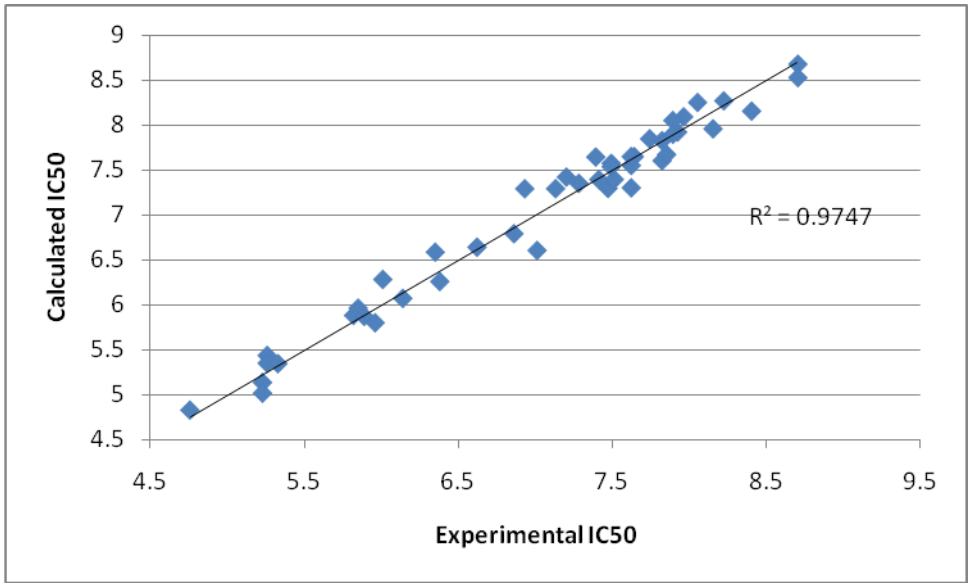
Figure 12. ROCS alignment of compounds used in training and test dataset

3.3.1. CoMFA statistical analysis

The correlation between experimental and predicted IC_{50} values of both training and test dataset used for CoMFA analysis is shown as a scattered plot (Figure 13) and tabulated in Table 2. The best CoMFA model yielded R^2_{NV} = 0.97, R^2_{CV} = 0.64, R^2_{LOO} = 0.67, SEE = 0.154, F = 287.219 and NOC = 7. The fact that the residual values of the training set compounds is within ± 0.4 pIC_{50} represents the goodness fit of the QSAR model. The ultimate test for the predictability of a CoMFA model in the drug design process is to predict the biological activity of compounds in the test dataset. In our study, the predictive power of the CoMFA model for the

external dataset, calculated using Microsoft Excel based on the statistical parameters generated using CoMFA analyis (R^2_{pred}) was found to be 0.74. Thus, our CoMFA model displays higher predictivity both in regular cross-validation as well as in the prediction of test set compounds indicating the robustness of the model. Furthermore, the statistical validity and the stability of the CoMFA model was assessed by performing bootstrap analysis for 10 runs. An R^2_{bs} value of 0.99 is a strong indication of the predictive power of our 3D-QSAR model. Finally, the ratio of steric to electrostatic field contributions was found to be 65.8: 34.2 indicating the dominance of steric contribution.

Table 2 lists the experimental, predicted and the residual IC_{50} values of both the training and test dataset used for our CoMFA analysis. The histogram of residual values for the test molecules for CoMFA is shown in Figure 14. Compound **34** is predicted to have a maximum residual value. Inspection of aligned poses revealed that the **34** has a cyclopentanoyl group, a small size group compared to the bulky P3 group in the template compound **38**. As mentioned in Chapter 2.3, the S3 site of BACE1 can adopt different conformations depending upon the nature of the P3 substituents. Both small and bulky P3 groups can be successfully accommodated and was shown to possess nM activity. For example, compound **36** has a smaller P3 group, yet better activity (IC_{50} = 35 nM).



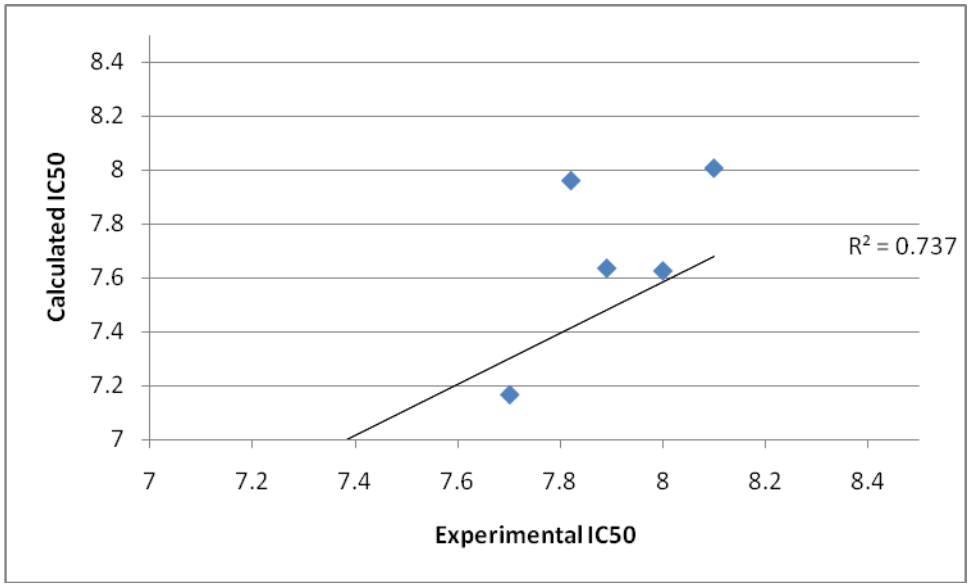


Figure 13. Graph of actual vs. predicted activities of our best CoMFA model obtained for: (top) training dataset; (bottom) test dataset

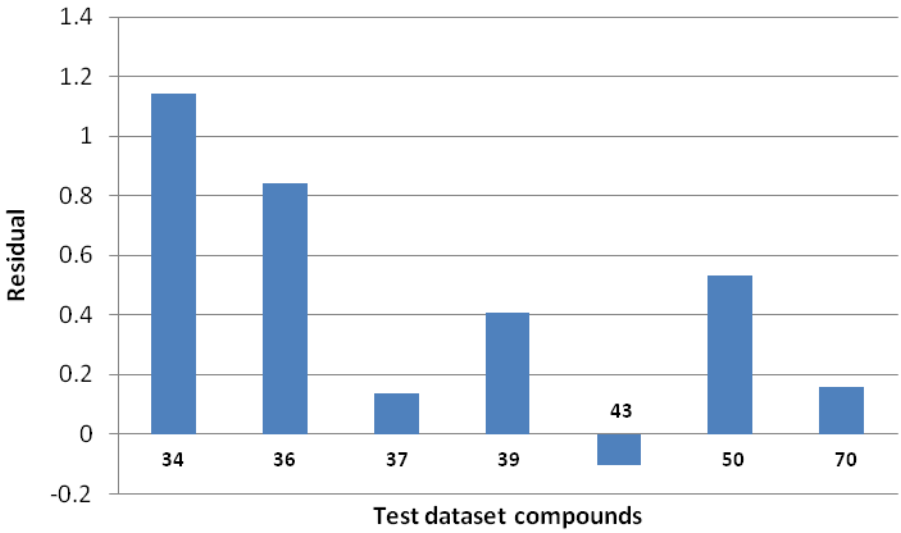


Figure 14. Histogram of the residual values of test set compounds obtained for our CoMFA model

3.3.2. CoMFA steric contour maps

The 3D-QSAR contour maps from CoMFA analysis of BACE1 inhibitors in the training dataset illustrate clearly the steric and electrostatic requirements for ligand binding. It also tells how the variation in the physiochemical interactions (steric and electrostatic) of the training set compounds with the enzyme improves or decreases the activity.

In the CoMFA steric contour maps, the favorable areas that can accommodate more bulk are indicated by green contours and the unfavorable steric areas in yellow. In figure 15, prominent green contours present at the vicinity of the P3 and P1' and P2' positions and a smaller green contour present at the P1 position indicate that steric bulk is favored at these positions. This is consistent with the experimental results: for example, in series-I, the smaller size of P3 in compound **24** (IC_{50} = 980 nM) is responsible for 10-fold less potency compared to the bulky P3 position in compound **39** (IC_{50} = 10 nM). Interestingly, the most active members, compounds **51** and **64** (IC_{50} = 2 nM) have less bulky steric substituents and are about five-fold

more active than **39**. This is because the residues 9-14 forming the S3 site (10s loop) of BACE1 is believed to exist in two or more conformations depending upon the size of the P3 substitutents and h-bonding with Thr232.[84] Ligands with a bulky P3 moiety modulate the enzyme and binds to it in an “up” conformation, whereas the ligand that has a smaller P3 moiety binds and binds to a “down” conformation. Compound **38** makes tight vdW contact with the ‘up’ conformation of the BACE1 whereas compounds **51** and **64** binds to the “down” conformation of the enzyme but still maintains tight vdW contacts. The two green contours present at the P3 positions is a reflection of two different classes of P3 substituents.

The prominent green contours present at the P1’ position parallels the trend observed in experimental activities as a result of altered steric bulk. For example, in series I, compound **41** (R_4 = H, IC_{50} = 23 nM) is about 1.5-fold less active than compound **39** (R_4 = cyclopropyl, IC_{50} = 10 nM). Similarly, for series III, as the size decreases from compounds **65** to **68**, the activity increases: **65** (H, 117 nM), **66** (Me, 8 nM), **67** (Et, 7 nM) and **68** (n-Pr, 4 nM). Improved steric bulk at the P1’ position favors inhibitory activity.

In series III, compound **66** with the bulky isobutyl substituent at the P2’ position has a better activity (IC_{50} = 8 nM) compared to the ethyl (**71**, 52 nM) or cyclopropyl (**72**, 24 nM). This is consistent with the green contour present at the S2’ site.

There is also a large yellow polyhedron at the P2’ site which means steric bulk is unfavorable at this region. Inspection of compounds **73** and **74** revealed that the P2’ groups of these compounds, N-benzyl and carboxy group of the prolyl moiety, respectively, are embedded in this unfavorable yellow contour region. Thus, the prediction from our CoMFA model that an increase in steric bulk in this yellow region leads to decreased activity is consistent with the low activity of **73** (139 nM) and **74** (17200 nM) compared to **72** (24 nM).

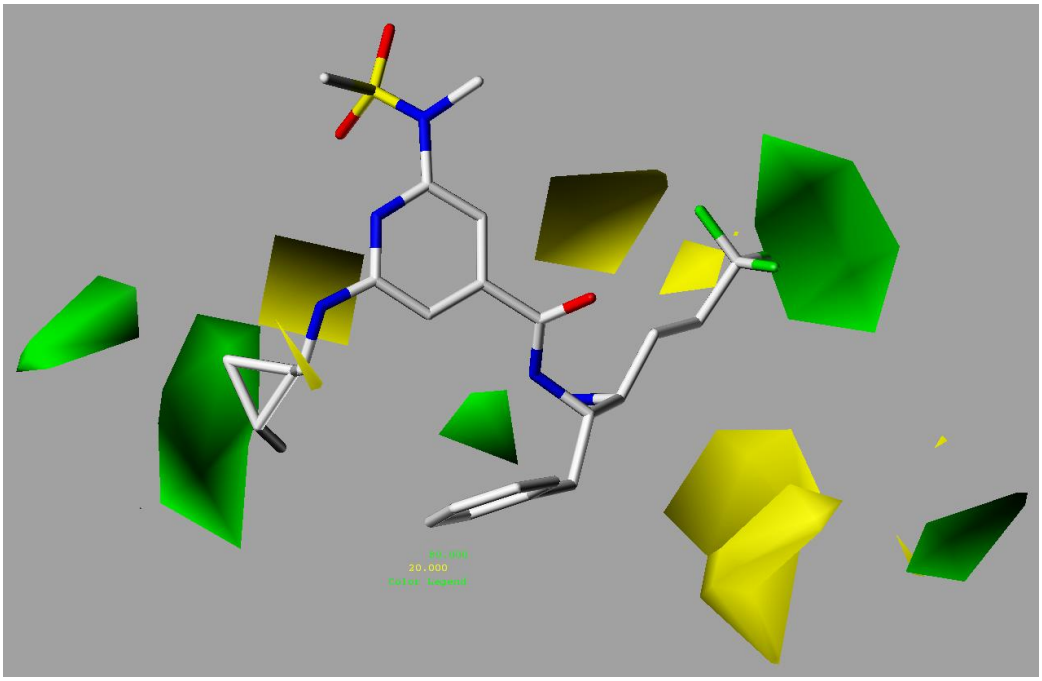


Figure 15. CoMFA steric contour map with the most active compound 51. Green and yellow polyhedra indicate regions where more steric bulk or less steric bulk, respectively, will enhance the activity

3.3.3. CoMFA electrostatic contour maps

In the CoMFA electrostatic maps, blue contours indicate the regions where electropositive groups increase activity, whereas the red areas indicate the regions where electronegative regions increase activity. The electrostatic contour map of our CoMFA model with the most active compound **51** is shown in figure 16.

Majority of the compounds in our dataset, including the most active compounds **51** and **64**, has a 1° or 2° nitrogen at the P1' that perfectly entrenched into the positive charge favored blue contour region. It has been shown in X-ray analysis that these positively charged 1° or 2° nitrogen bind to the catalytic aspartic acid present in the binding site.

Interestingly, the negatively carboxylic acid moiety (P2' position) of compound **74** lies on the blue contour where positively charged moiety is preferred. This mismatch in electrostatic complementarity is one of the contributing factors for the poor activity of compound **50** in addition to the steric hindrance.

The red contour in the vicinity of the P2 and P3 positions signifies the preference for electronegative groups at this region. Consistent with this, compounds **38** to **45** that contains electronegative oxygen (from keto group at R_1) display better activity. From the X-ray crystal structure analysis, it was found that these keto oxygen atoms indeed hydrogen bond with Thr-232 in the BACE1 binding site.

Similarly, the red contour present the vicinity of P2 and P1 position is consistent with the presence of oxygen atom, which indeed was reported to bind to the backbone NH of Gln73.[85] All the compounds in the dataset have oxygen at this position which is believed to be responsible for activity.

The presence of oxygen at the vicinity of the P2' position in compounds **65** to **74** is reflected in the red contour map. Indeed, this oxygen is believed to engage in hydrogen bonding with the NH backbone group of Thr72.[85]

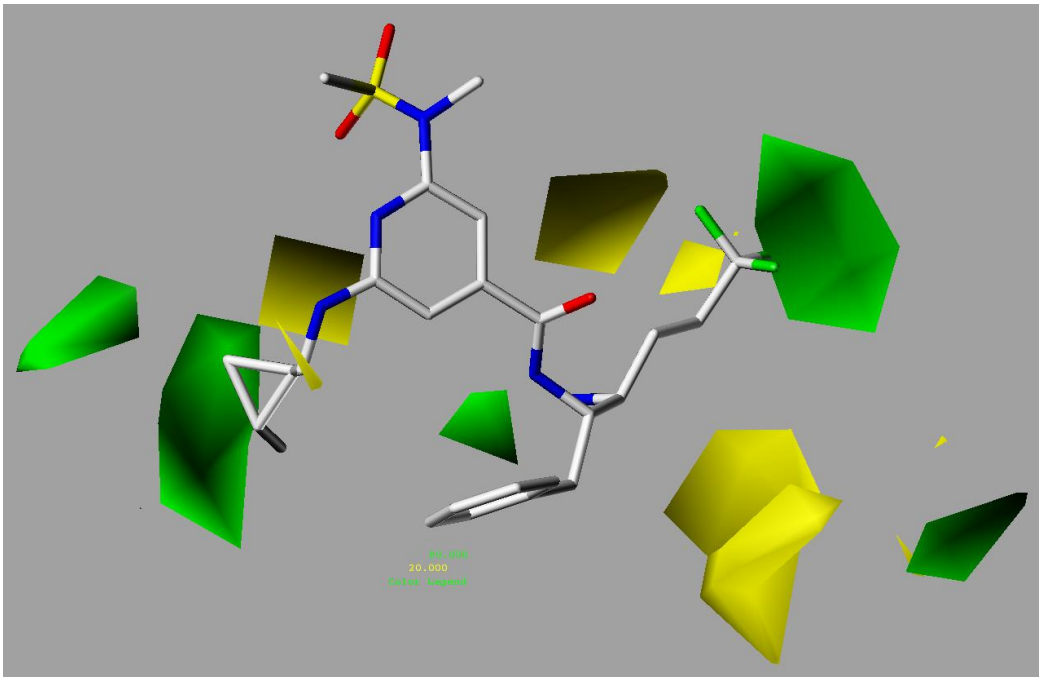


Figure 16. CoMFA steric contour map with the most active compound 51. Green and yellow polyhedra indicate regions where more steric bulk or less steric bulk, respectively, will enhance the activity

3.3.4. CoMFA electrostatic contour maps

In the CoMFA electrostatic maps, blue contours indicate the regions where electropositive groups increase activity, whereas the red areas indicate the regions where electronegative regions increase activity. The electrostatic contour map of our CoMFA model with the most active compound **51** is shown in figure 16.

Majority of the compounds in our dataset, including the most active compounds **51** and **64**, has a 1° or 2° nitrogen at the P1' that perfectly entrenched into the positive charge favored blue contour region. It has been shown in X-ray analysis that these positively charged 1° or 2° nitrogen bind to the catalytic aspartic acid present in the binding site.

Interestingly, the negatively carboxylic acid moiety (P2' position) of compound **74** lies on the blue contour where positively charged moiety is preferred. This mismatch in electrostatic complementarity is one of the contributing factors for the poor activity of compound **50** in addition to the steric hindrance.

The red contour at the vicinity of the P2 and P3 positions signifies the preference for electronegative groups at this region. Consistent with this, compounds **38** to **45** that contains electronegative oxygen display better activity. From the X-ray crystal structure analysis, it was found that these oxygen atoms indeed hydrogen bond with Thr-232 present at the BACE1 binding site.

Similarly, the red contour present the vicinity of P2 and P1 position is consistent with the presence of electronegative oxygen atom, which indeed was reported to bind to the backbone NH of Gln73.[85] All the compounds in the dataset have electronegative oxygen at this position which is believed to be responsible for activity.

The presence of electronegative oxygen at the vicinity of the P2' position in compounds **65** to **74** is reflected in the red contour map. Indeed, this oxygen is believed to hydrogen bond with the NH backbone of Thr72.[85]

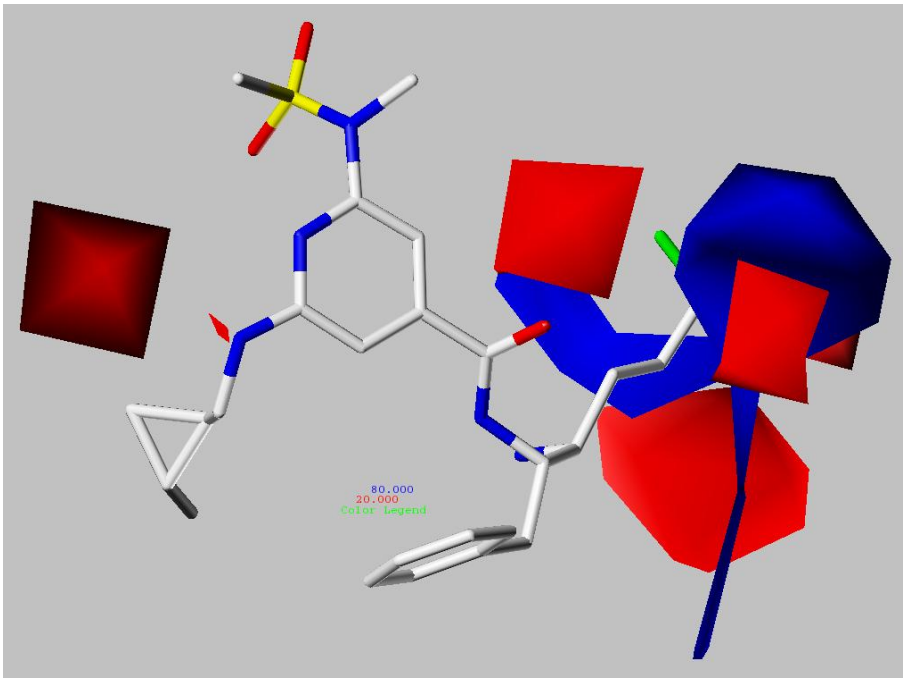
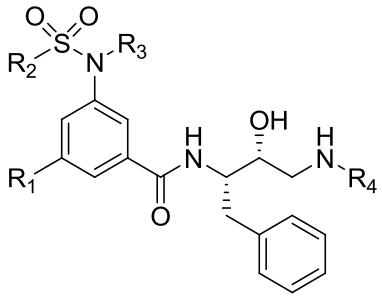
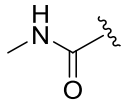
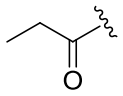
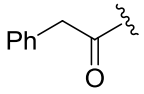
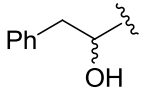
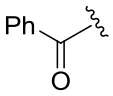
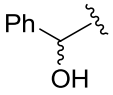
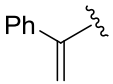
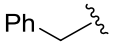
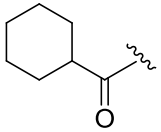
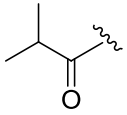
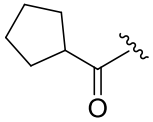


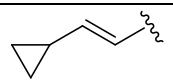
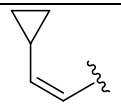
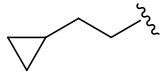
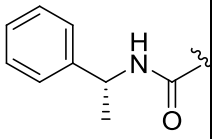
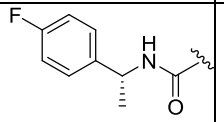
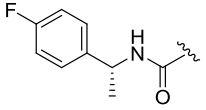
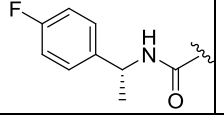
Figure 17. CoMFA electrostatic contour map with the most active compound 51. Blue contours indicate regions where electropositive groups increase activity, whereas red contours indicate regions where electronegative groups increase activity

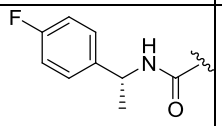
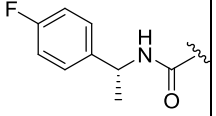
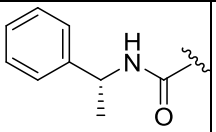
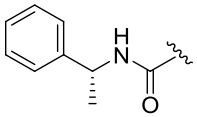
Series-I



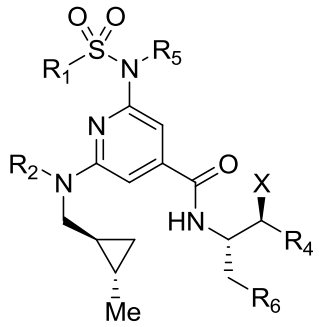
S.No	R_1	R_2	R_3	R_4	IC_{50} (nM)	IC_{50}		Residuals
						Actual	CoMFA	
24		Me	Me		980	6.01	6.28	-0.27
25		Me	Me		1500	5.82	5.88	-0.06
26		Me	Me		1400	5.85	5.96	-0.11
27		Me	Me		5500	5.26(R)	5.44 (R)	-0.18 (R)

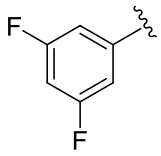
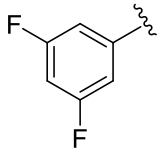
						5.26 (S)	5.35 (S)	-0.09 (S)
28		Me	Me		98	7.01	6.60	0.41
29		Me	Me		5900	5.23 (R) 5.23 (S)	5.14 (R) 5.02 (S)	0.09 (R) 0.21 (S)
30		Me	Me		1300	5.89	5.87	0.02
31		Me	Me		4630	5.33	5.35	-0.02
32		Me	Me		450	6.35	6.59	-0.24
33		Me	Me		240	6.62	6.64	-0.02
34		**Me**	**Me**		**180**	**6.74**	**5.65**	**1.09**

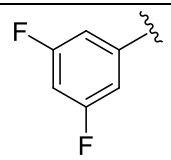
35		Me	Me		730	6.14	6.07	0.07
36		**Me**	**Me**		**35**	**7.46**	**6.74**	**0.72**
37		Me	Me		420	6.38	6.26	0.22
38		Me	Me		15	7.82	7.60	0.22
39		**Me**	**Me**		**10**	**8.00**	**7.62**	**0.38**
40		Me	H		63	7.2	7.42	-0.22
41		Me	Me	H	23	7.64	7.65	-0.01

42		Me	Me	Et	15	7.82	7.82	0
43		**Me**	**Me**	**t-Bu**	**15**	**7.82**	**7.96**	**-0.14**
44		i-Pr	Me		41	7.39	7.64	-0.25
45		NMe_2	Me		14	7.85	7.67	0.18

Series-II

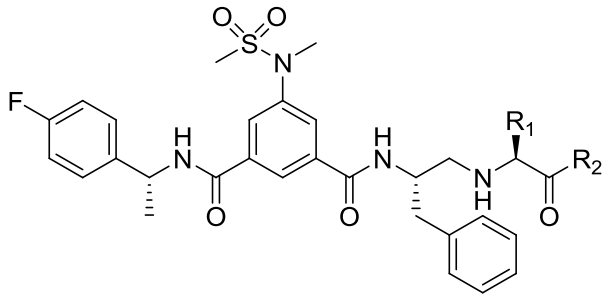


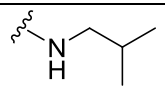
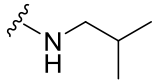
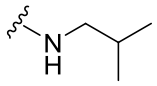
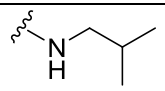
S.No	R_1	R_2	R_4	R_5	R_6	X	IC_{50} (nM)	IC_{50} (nM)		Residual
								Actual	CoMFA	
46	Me	H	H	Me		OH	74	7.13	7.29	-0.16
47	**Me**	**H**	**n-Bu**	**Me**		**OH**	**13**	**7.89**	**7.63**	**0.26**

48	Me	H	n-Bu	Me		NH_2	9	8.05	8.25	-0.20
49	Me	H	H	Me	Ph	NH_2	34	7.47	7.29	0.18
50	**Me**	**H**	**n-Bu**	**Me**	**Ph**	**NH_2**	**20**	**7.7**	**7.17**	**0.54**
51	Me	H	-$(CH_2)_2$CF3	Me	Ph	NH_2	2	8.7	8.67	0.03
52	Me	H	n-Bu	Me	3-Thiophenyl-	NH_2	11	7.96	8.09	-0.13
53	Me	H	Me	Me	3-Thiophenyl-	NH_2	32	7.49	7.57	-0.08
54	Me	Me	n-Bu	Me	Ph	NH_2	31	7.51	7.39	0.13
55	Me	Me	n-Bu	n-Pr	Ph	NH_2	39	7.41	7.39	0.02
56	Me	Me	(R)-CH_2OMe	Me	Ph	NH_2	24	7.62	7.55	0.07
57	Me	Me	(R)-CH_2OEt	Me	Ph	NH_2	24	7.62	7.65	-0.03
58	Me	$(CH_2)_3$ OMe	(R)-CH_2OMe	Me	Ph	NH_2	6	8.22	8.27	-0.05

59	Me	H	H	Me	Ph	OH	1100	5.96	5.80	0.16
60	Me	H	H	i-Pr	Ph	OH	14	7.85	7.669	0.18
61	Me	H	H	Me	Ph	NH_2	34	7.47	7.53	-0.06
62	Me	Me	$MeOCH_2$	Me	Ph	NH_2	12	7.92	7.91	0.01
63	Me	Me	CH_2F	i-Pr	Ph	NH_2	18	7.74	7.84	-0.10
64	Me	H	CH_2F	i-Pr	Ph	NH_2	2	8.7	8.52	0.18

Series III



S.No	R_1	R_2	IC_{50} (nM)	IC_{50}		Residual
				Actual	CoMFA	
65	H		117	6.93	7.29	0.36
66	**Me**		**8**	**8.1**	**8.01**	**0.09**
67	Et		7	8.15	7.96	0.19
68	*n*-Pr		4	8.4	8.15	0.25

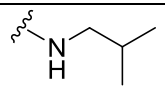
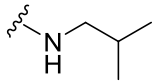
69	-CH_2CH_2OH		13	7.89	8.05	-0.16
70	-$CH_2CH_2SO_2Me$		13	7.89	7.90	-0.01
71	Me	NHEt	52	7.28	7.35	-0.07
72	Me	NHc-Pr	24	7.62	7.3	0.32
73	Me	NHBn	139	6.86	6.79	0.07
74	Me	Proline	17200	4.76	4.83	-0.07

Table 2. Structures and activities of BACE1 inhibitors used in CoMFA modeling. Test set compounds are highlighted in bold.

3.4. Conclusions

The application of 3D-QSAR models for the prediction of binding affinity, hit-to-lead optimization, and evaluation of toxicity is continuously emerging. In the current study, a 3D-QSAR CoMFA model was developed using a diverse set of BACE1 inhibitors that displayed a broad range of potency. This study further enhances our current understanding of BACE1 inhibitors implicated in the treatment of Alzheimer's disease. The main conclusions emerged from this study are:

The reported 3D-QSAR models are statistically significant thereby demonstrating a sound SAR for inhibitors that bind to the catalytic site of BACE1. For CoMFA analysis, the statistical parameters are: R^2_{NV} = 0.98, R^2_{CV} = 0.64, R^2_{LOO} = 0.67, SEE = 0.154 and F = 287.219. This high statistical significance illustrates that these diverse inhibitors share structural commonalities important for binding to BACE1. This model should be useful for the identification, design and development of potential BACE1 inhibitors.

The BACE1 binding site regions derived from the CoMFA analysis are consistent with the physical structure of the BACE1 ligand binding site as explained in the discussion section. The CoMFA contour map correctly identifies the impact of more/less bulk group or electropositive/electronegative groups on the IC_{50} values.

We demonstrated that the CoMFA model generated in this study has good predictive capability with low residuals to estimate the BACE1 binding affinity of an external test set. We would like to point out that the best CoMFA model generated for the current training and test dataset can be further refined by scrambling the compounds to increase the activity range of the test dataset.

CITED LITERATURE

1. Lloyd, N. C.; Morgan, H. W.; Nicholson, B. K.; Ronimus, R. S. The composition of Ehrlich's salvarsan: resolution of a century-old debate. *Angew. Chem. Int. Ed. Engl.* **2005,** 44, 941-4.
2. Bolton, W.; Perutz, M. F. Three dimensional fourier synthesis of horse deoxyhaemoglobin at 2.8 Angstrom units resolution. *Nature* **1970,** 228, 551-2.
3. Fermi, G.; Perutz, M. F.; Shaanan, B.; Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 A resolution. *J. Mol. Biol.* **1984,** 175, 159-74.
4. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **1958,** 181, 662-6.
5. Safo, M. K.; Abraham, D. J. The enigma of the liganded hemoglobin end state: a novel quaternary structure of human carbonmonoxy hemoglobin. *Biochemistry* **2005,** 44, 8347-59.
6. Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. Target flexibility: an emerging consideration in drug discovery and design. *J. Med. Chem.* **2008,** 51, 6237-55.
7. Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Neurol.* **2003,** 2, 527-41.
8. Barcellos, G. B.; Pauli, I.; Caceres, R. A.; Timmers, L. F.; Dias, R.; de Azevedo, W. F., Jr. Molecular modeling as a tool for drug discovery. *Curr. Drug Targets* **2008,** 9, 1084-91.
9. De Benedetti, P. G.; Fanelli, F. Ligand-receptor communication and drug design. *Curr. Prot. Peptide Sci.* **2009,** 10, 186-93.
10. Kumar, N.; Hendriks, B. S.; Janes, K. A.; de Graaf, D.; Lauffenburger, D. A. Applying computational modeling to drug discovery and development. *Drug Discov. Today* **2006,** 11, 806-11.
11. Muster, W.; Breidenbach, A.; Fischer, H.; Kirchner, S.; Muller, L.; Pahler, A. Computational toxicology in drug development. *Drug Discov. Today* **2008,** 13, 303-10.
12. Pitera, J. W. Current developments in and importance of high-performance computing in drug discovery. *Curr Opin Drug Discov Devel.* **2009,** 12, 388-96.
13. Schnur, D. M. Recent trends in library design: 'rational design' revisited. *Curr Opin Drug Discov Devel.* **2008,** 11, 375-80.
14. Sun, H. Pharmacophore-based virtual screening. *Curr. Med. Chem.* **2008,** 15, 1018-24.
15. Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000,** 57, 213-8.
16. Amaro, R. E.; Li, W. W. Emerging Methods for Ensemble-Based Virtual Screening. *Curr. Top. Med. Chem.* **2009**.
17. Vajda, S.; Guarnieri, F. Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr Opin Drug Discov Devel.* **2006,** 9, 354-62.
18. Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Mario Amzel, L.; Freire, E. Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem. Biol. Drug Des.* **2007,** 69, 413-22.
19. Huang, S. Y.; Zou, X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007,** 66, 399-421.
20. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006,** 49, 534-53.
21. Novak, W.; Wang, H.; Krilov, G. Role of protein flexibility in the design of Bcl-X(L) targeting agents: insight from molecular dynamics. *J. Comput. Aided Mol. Des.* **2009,** 23, 49-61.

22. Deng, J.; Sanchez, T.; Neamati, N.; Briggs, J. M. Dynamic pharmacophore model optimization: identification of novel HIV-1 integrase inhibitors. *J. Med. Chem.* **2006,** 49, 1684-92.
23. Sheu, S. H.; Kaya, T.; Waxman, D. J.; Vajda, S. Exploring the binding site structure of the PPAR gamma ligand-binding domain by computational solvent mapping. *Biochemistry* **2005,** 44, 1193-209.
24. Brenk, R.; Vetter, S. W.; Boyce, S. E.; Goodin, D. B.; Shoichet, B. K. Probing molecular docking in a charged model binding site. *J. Mol. Biol.* **2006,** 357, 1449-70.
25. Hardy, J. A.; Higgins, G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **1992,** 256, 184-185.
26. Minucci, S.; Pelicci, P. G. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat. Rev. Drug Discov.* **2006,** 6, 38-51.
27. Kang, M. H.; Reynolds, C. P. Bcl-2 inhibitors: targeting mitochondrial apoptotic pathways in cancer therapy. *Clin. Cancer Res.* **2009,** 15, 1126-32.
28. Lundqvist, T.; Fisher, S. L.; Kern, G.; Folmer, R. H.; Xue, Y.; Newton, D. T.; Keating, T. A.; Alm, R. A.; de Jonge, B. L. Exploitation of structural and regulatory diversity in glutamate racemases. *Nature* **2007,** 447, 817-22.
29. Sharma, P. S.; Sharma, R.; Tyagi, T. Receptor tyrosine kinase inhibitors as potent weapons in war against cancers. *Curr. Pharm. Des.* **2009,** 15, 758-76.
30. Congreve, M.; Marshall, F. The impact of GPCR structures on pharmacology and structure-based drug design. *British Journal of Pharmacology* **2009**.
31. Dechering, K.; Boersma, C.; Mosselman, S. Estrogen receptors alpha and beta: two receptors of a kind? *Curr. Med. Chem.* **2000,** 7, 561-76.
32. Gilbert, H. 10 Types of Dementia That Aren't Alzheimer's and How They're Diagnosed *http://www.caring.com/articles/types-of-dementia*.
33. Alzheimer, A.; Stelzmann, R. A.; Schnitzlein, H. N.; Murtagh, F. R. An English translation of Alzheimer's 1907 paper, "Uber eine eigenartige Erkankung der Hirnrinde". *Clin. Anat.* **1995,** 8, 429-31.
34. Reitz, C.; Brayne, C.; Mayeux, R. Epidemiology of Alzheimer disease. *Nat Rev Neurosci.* **2011**.
35. Changing the Trajectory of Alzheimer's Disease: A National Imperative. *Changing the Trajectory of Alzheimer's Disease: A National Imperative*.
36. Selkoe, D. J. Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.* **2001,** 81, 741-66.
37. Hardy, J.; Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **2002,** 297, 353.
38. Ittner, L. M.; Gotz, J. Amyloid-beta and tau--a toxic pas de deux in Alzheimer's disease. *Nat Rev Neurosci.* **2011,** 12, 65-72.
39. Mattson, M. P.; Cheng, B.; Culwell, A. R.; Esch, F. S.; Lieberburg, I.; Rydel, R. E. Evidence for excitoprotective and intraneuronal calcium-regulating roles for secreted forms of the beta-amyloid precursor protein. *Neuron* **1993,** 10, 243-54.
40. LaFerla, F. M.; Green, K. N.; Oddo, S. Intracellular amyloid-beta in Alzheimer's disease. *Nat Rev Neurosci.* **2007,** 8, 499-509.
41. Parvathy, S.; Davies, P.; Haroutunian, V.; Purohit, D. P.; Davis, K. L.; Mohs, R. C.; Park, H.; Moran, T. M.; Chan, J. Y.; Buxbaum, J. D. Correlation between Abetax-40-, Abetax-42-, and Abetax-43-containing amyloid plaques and cognitive decline. *Arch. Neurol.* **2001,** 58, 2025-32.
42. Kurz, A.; Perneczky, R. Amyloid clearance as a treatment target against Alzheimer's disease. *J. Alzheimers Dis.* **2011,** 24, 61-73.
43. Willem, M.; Lammich, S.; Haass, C. Function, regulation and therapeutic properties of beta-secretase (BACE1). *Sem. Cell Dev. Biol.* **2009,** 20, 175-82.
44. Vassar, R. ß-Secretase (BACE) as a drug target for alzheimer's disease. *Adv. Drug Deliv. Rev.* **2002,** 54, 1589-1602.

45. Ghosh, A. K.; Kumaragurubaran, N.; Hong, L.; Koelsh, G.; Tang, J. Memapsin 2 (beta-secretase) inhibitors: drug development. *Curr. Alzheimer Res.* **2008,** 5, 121-31.
46. Herrup, K. Reimagining Alzheimer's disease--an age-based hypothesis. *J. Neurochem.* **2010,** 30, 16755-62.
47. Hussain, I.; Powell, D.; Howlett, D. R.; Tew, D. G.; Meek, T. D.; Chapman, C.; Gloger, I. S.; Murphy, K. E.; Southan, C. D.; Ryan, D. M.; Smith, T. S.; Simmons, D. L.; Walsh, F. S.; Dingwall, C.; Christie, G. Identification of a novel aspartic protease (Asp 2) as beta-secretase. *Molecular and Cellular Neuroscience* **1999,** 14, 419-27.
48. Lin, X.; Koelsch, G.; Wu, S.; Downs, D.; Dashti, A.; Tang, J. Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein. *Proc. Natl. Acad. Sci. USA* **2000,** 97, 1456-60.
49. Sinha, S.; Anderson, J. P.; Barbour, R.; Basi, G. S.; Caccavello, R.; Davis, D.; Doan, M.; Dovey, H. F.; Frigon, N.; Hong, J.; Jacobson-Croak, K.; Jewett, N.; Keim, P.; Knops, J.; Lieberburg, I.; Power, M.; Tan, H.; Tatsuno, G.; Tung, J.; Schenk, D.; Seubert, P.; Suomensaari, S. M.; Wang, S.; Walker, D.; Zhao, J.; McConlogue, L.; John, V. Purification and cloning of amyloid precursor protein beta-secretase from human brain. *Nature* **1999,** 402, 537-40.
50. Vassar, R.; Bennett, B. D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E. A.; Denis, P.; Teplow, D. B.; Ross, S.; Amarante, P.; Loeloff, R.; Luo, Y.; Fisher, S.; Fuller, J.; Edenson, S.; Lile, J.; Jarosinski, M. A.; Biere, A. L.; Curran, E.; Burgess, T.; Louis, J. C.; Collins, F.; Treanor, J.; Rogers, G.; Citron, M. Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **1999,** 286, 735-41.
51. Yan, R.; Bienkowski, M. J.; Shuck, M. E.; Miao, H.; Tory, M. C.; Pauley, A. M.; Brashier, J. R.; Stratman, N. C.; Mathews, W. R.; Buhl, A. E.; Carter, D. B.; Tomasselli, A. G.; Parodi, L. A.; Heinrikson, R. L.; Gurney, M. E. Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity. *Nature* **1999,** 402, 533-7.
52. Bennett, B. D.; Denis, P.; Haniu, M.; Teplow, D. B.; Kahn, S.; Louis, J. C.; Citron, M.; Vassar, R. A furin-like convertase mediates propeptide cleavage of BACE, the Alzheimer's beta -secretase. *J. Biol. Chem.* **2000,** 275, 37712-7.
53. Mowrer, K. R.; Wolfe, M. S. Promotion of BACE1 mRNA alternative splicing reduces amyloid beta-peptide production. *J. Biol. Chem.* **2008,** 283, 18694-701.
54. Rossner, S.; Lange-Dohna, C.; Zeitschel, U.; Perez-Polo, J. R. Alzheimer's disease beta-secretase BACE1 is not a neuron-specific enzyme. *J. Neurochem.* **2005,** 92, 226-34.
55. Rajendran, L.; Schneider, A.; Schlechtingen, G.; Weidlich, S.; Ries, J.; Braxmeier, T.; Schwille, P.; Schulz, J. B.; Schroeder, C.; Simons, M.; Jennings, G.; Knolker, H. J.; Simons, K. Efficient inhibition of the Alzheimer's disease beta-secretase by membrane targeting. *Science* **2008,** 320, 520-3.
56. Rajendran, L.; Knolker, H. J.; Simons, K. Subcellular targeting strategies for drug design and delivery. *Nat. Rev. Drug Discov.* **2010,** 9, 29-42.
57. Kitazume, S.; Tachida, Y.; Oka, R.; Shirotani, K.; Saido, T. C.; Hashimoto, Y. Alzheimer's beta-secretase, beta-site amyloid precursor protein-cleaving enzyme, is responsible for cleavage secretion of a Golgi-resident sialyltransferase. *Proc. Natl. Acad. Sci. USA* **2001,** 98, 13554-9.
58. Willem, M.; Garratt, A. N.; Novak, B.; Citron, M.; Kaufmann, S.; Rittger, A.; DeStrooper, B.; Saftig, P.; Birchmeier, C.; Haass, C. Control of peripheral nerve myelination by the beta-secretase BACE1. *Science* **2006,** 314, 664-6.
59. Lichtenthaler, S. F.; Dominguez, D. I.; Westmeyer, G. G.; Reiss, K.; Haass, C.; Saftig, P.; De Strooper, B.; Seed, B. The cell adhesion protein P-selectin glycoprotein ligand-1 is a substrate for the aspartyl protease BACE1. *J. Biol. Chem.* **2003,** 278, 48713-9.
60. Kuhn, P. H.; Marjaux, E.; Imhof, A.; De Strooper, B.; Haass, C.; Lichtenthaler, S. F. Regulated intramembrane proteolysis of the interleukin-1 receptor II by alpha-, beta-, and gamma-secretase. *J. Biol. Chem.* **2007,** 282, 11982-95.

61. Kim, D. Y.; Carey, B. W.; Wang, H.; Ingano, L. A.; Binshtok, A. M.; Wertz, M. H.; Pettingell, W. H.; He, P.; Lee, V. M.; Woolf, C. J.; Kovacs, D. M. BACE1 regulates voltage-gated sodium channels and neuronal activity. *Nat. Cell Biol.* **2007,** 9, 755-64.
62. Cai, H.; Wang, Y.; McCarthy, D.; Wen, H.; Borchelt, D. R.; Price, D. L.; Wong, P. C. BACE1 is the major beta-secretase for generation of Abeta peptides by neurons. *Nat. Neurosci.* **2001,** 4, 233-4.
63. Luo, Y.; Bolon, B.; Kahn, S.; Bennett, B. D.; Babu-Khan, S.; Denis, P.; Fan, W.; Kha, H.; Zhang, J.; Gong, Y.; Martin, L.; Louis, J. C.; Yan, Q.; Richards, W. G.; Citron, M.; Vassar, R. Mice deficient in BACE1, the Alzheimer's beta-secretase, have normal phenotype and abolished beta-amyloid generation. *Nature Reviews Cancer* **2001,** 4, 231-2.
64. Ohno, M.; Sametsky, E. A.; Younkin, L. H.; Oakley, H.; Younkin, S. G.; Citron, M.; Vassar, R.; Disterhoft, J. F. BACE1 deficiency rescues memory deficits and cholinergic dysfunction in a mouse model of Alzheimer's disease. *Neuron* **2004,** 41, 27-33.
65. Roberds, S. L.; Anderson, J.; Basi, G.; Bienkowski, M. J.; Branstetter, D. G.; Chen, K. S.; Freedman, S. B.; Frigon, N. L.; Games, D.; Hu, K.; Johnson-Wood, K.; Kappenman, K. E.; Kawabe, T. T.; Kola, I.; Kuehn, R.; Lee, M.; Liu, W.; Motter, R.; Nichols, N. F.; Power, M.; Robertson, D. W.; Schenk, D.; Schoor, M.; Shopp, G. M.; Shuck, M. E.; Sinha, S.; Svensson, K. A.; Tatsuno, G.; Tintrup, H.; Wijsman, J.; Wright, S.; McConlogue, L. BACE knockout mice are healthy despite lacking the primary beta-secretase activity in brain: implications for Alzheimer's disease therapeutics. *Hum. Mol. Genet.* **2001,** 10, 1317-24.
66. Laird, F. M.; Cai, H.; Savonenko, A. V.; Farah, M. H.; He, K.; Melnikova, T.; Wen, H.; Chiang, H. C.; Xu, G.; Koliatsos, V. E.; Borchelt, D. R.; Price, D. L.; Lee, H. K.; Wong, P. C. BACE1, a major determinant of selective vulnerability of the brain to amyloid-beta amyloidogenesis, is essential for cognitive, emotional, and synaptic functions. *J. Neurochem.* **2005,** 25, 11693-709.
67. McConlogue, L.; Buttini, M.; Anderson, J. P.; Brigham, E. F.; Chen, K. S.; Freedman, S. B.; Games, D.; Johnson-Wood, K.; Lee, M.; Zeller, M.; Liu, W.; Motter, R.; Sinha, S. Partial reduction of BACE1 has dramatic effects on Alzheimer plaque and synaptic pathology in APP Transgenic Mice. *J. Biol. Chem.* **2007,** 282, 26326-34.
68. Fukumoto, H.; Cheung, B. S.; Hyman, B. T.; Irizarry, M. C. Beta-secretase protein and activity are increased in the neocortex in Alzheimer disease. *Arch. Neurol.* **2002,** 59, 1381-9.
69. Holsinger, R. M.; McLean, C. A.; Beyreuther, K.; Masters, C. L.; Evin, G. Increased expression of the amyloid precursor beta-secretase in Alzheimer's disease. *Ann. Neurol.* **2002,** 51, 783-6.
70. Li, R.; Lindholm, K.; Yang, L. B.; Yue, X.; Citron, M.; Yan, R.; Beach, T.; Sue, L.; Sabbagh, M.; Cai, H.; Wong, P.; Price, D.; Shen, Y. Amyloid beta peptide load is correlated with increased beta-secretase activity in sporadic Alzheimer's disease patients. *Proc. Natl. Acad. Sci. USA* **2004,** 101, 3632-7.
71. Yang, L. B.; Lindholm, K.; Yan, R.; Citron, M.; Xia, W.; Yang, X. L.; Beach, T.; Sue, L.; Wong, P.; Price, D.; Li, R.; Shen, Y. Elevated beta-secretase expression and enzymatic activity detected in sporadic Alzheimer disease. *Nat. Med.* **2003,** 9, 3-4.
72. Stachel, S. Progress Toward the Development of a Viable BACE-1 Inhibitor. *Drug Dev. Res.* **2009,** 70, 101-110.
73. He, X.; Cooley, K.; Chung, C. H.; Dashti, N.; Tang, J. Apolipoprotein receptor 2 and X11 alpha/beta mediate apolipoprotein E-induced endocytosis of amyloid-beta precursor protein and beta-secretase, leading to amyloid-beta production. *J. Neurochem.* **2007,** 27, 4052-60.
74. Kinoshita, A.; Fukumoto, H.; Shah, T.; Whelan, C. M.; Irizarry, M. C.; Hyman, B. T. Demonstration by FRET of BACE interaction with the amyloid precursor protein at the cell surface and in early endosomes. *J. Cell Sci.* **2003,** 116, 3339-46.
75. Koo, E. H. The beta-amyloid precursor protein (APP) and Alzheimer's disease: does the tail wag the dog? *Traffic* **2002,** 3, 763-70.

76. Rajendran, L.; Honsho, M.; Zahn, T. R.; Keller, P.; Geiger, K. D.; Verkade, P.; Simons, K. Alzheimer's disease beta-amyloid peptides are released in association with exosomes. *Proc. Natl. Acad. Sci. USA* **2006,** 103, 11172-7.
77. Small, S. A.; Gandy, S. Sorting through the cell biology of Alzheimer's disease: intracellular pathways to pathogenesis. *Neuron* **2006,** 52, 15-31.
78. Cheng, H.; Vetrivel, K. S.; Gong, P.; Meckler, X.; Parent, A.; Thinakaran, G. Mechanisms of disease: new therapeutic strategies for Alzheimer's disease--targeting APP processing in lipid rafts. *Nature Clinical Practice Neurology* **2007,** 3, 374-82.
79. Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; Ghosh, A. K.; Zhang, X. C.; Tang, J. Structure of the protease domain of memapsin 2 (beta-secretase) complexed with inhibitor. *Science* **2000,** 290, 150-3.
80. Gorfe, A. A.; Caflisch, A. Functional plasticity in the substrate binding site of beta-secretase. *Structure* **2005,** 13, 1487-98.
81. Hong, L.; Tang, J. Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry* **2004,** 43, 4689-95.
82. Patel, S.; Vuillard, L.; Cleasby, A.; Murray, C. W.; Yon, J. Apo and inhibitor complex structures of BACE (beta-secretase). *J. Mol. Biol.* **2004,** 343, 407-16.
83. Cole, D. C.; Manas, E. S.; Stock, J. R.; Condon, J. S.; Jennings, L. D.; Aulabaugh, A.; Chopra, R.; Cowling, R.; Ellingboe, J. W.; Fan, K. Y.; Harrison, B. L.; Hu, Y.; Jacobsen, S.; Jin, G.; Lin, L.; Lovering, F. E.; Malamas, M. S.; Stahl, M. L.; Strand, J.; Sukhdeo, M. N.; Svenson, K.; Turner, M. J.; Wagner, E.; Wu, J.; Zhou, P.; Bard, J. Acylguanidines as small-molecule beta-secretase inhibitors. *J. Med. Chem.* **2006,** 49, 6158-61.
84. McGaughey, G. B.; Colussi, D.; Graham, S. L.; Lai, M. T.; Munshi, S. K.; Nantermet, P. G.; Pietrak, B.; Rajapakse, H. A.; Selnick, H. G.; Stauffer, S. R.; Holloway, M. K. Beta-secretase (BACE-1) inhibitors: accounting for 10s loop flexibility using rigid active sites. *Bioorg. Med. Chem. Lett.* **2007,** 17, 1117-21.
85. Stachel, S. J.; Coburn, C. A.; Steele, T. G.; Crouthamel, M. C.; Pietrak, B. L.; Lai, M. T.; Holloway, M. K.; Munshi, S. K.; Graham, S. L.; Vacca, J. P. Conformationally biased P3 amide replacements of ß-secretase inhibitors. *Bioorg. Med. Chem. Lett.* **2006,** 16, 641-644.
86. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000,** 28, 235-42.
87. Wishart, D. DrugBank. http://www.drugbank.ca/
88. Rajamani, R.; Reynolds, C. H. Modeling the Protonation States of the Catalytic Aspartates in b-Secretase. *J. Med. Chem.* **2004,** 47, 5159-5166.
89. Polgár, T.; Keserü, G. M. Virtual Screening for-Secretase (BACE1) Inhibitors Reveals the Importance of Protonation States at Asp32 and Asp228. *J. Med. Chem.* **2005,** 48, 3749-3755.
90. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997,** 267, 727-748.
91. Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci.* **1999,** 88, 807-814.
92. Stachel, S. J.; Coburn, C. A.; Steele, T. G.; Jones, K. G.; Loutzenhiser, E. F.; Gregro, A. R.; Rajapakse, H. A.; Lai, M. T.; Crouthamel, M. C.; Xu, M.; Tugusheva, K.; Lineberger, J. E.; Pietrak, B. L.; Espeseth, A. S.; Shi, X. P.; Chen-Dodson, E.; Holloway, M. K.; Munshi, S.; Simon, A. J.; Kuo, L.; Vacca, J. P. Structure-based design of potent and selective cell-permeable inhibitors of human beta-secretase (BACE-1). *J. Med. Chem.* **2004,** 47, 6447-50.
93. Thorand, S.; Krause, N. Improved Procedures for the Palladium-Catalyzed Coupling of Terminal Alkynes with Aryl Bromides (Sonogashira Coupling). . *J. Org. Chem.* **1998,** 63, 8551-8553.

94. Steinhuebel, D.; Palucki, M.; Askin, D.; Dolling, U. Synthesis of N-arylated sultams: palladium-and copper-catalyzed cross coupling of aryl halides with 1, 4-butane and 1, 3-propanesultams. *Tetrahedron Lett.* **2004,** 45, 3305-3307.
95. Miroshnikova, O. V.; Hudson, T. H.; Gerena, L.; Kyle, D. E.; Lin, A. J. Synthesis and Antimalarial Activity of New Isotebuquine Analogues. *J. Med. Chem.* **2007,** 50, 889-896.
96. van Otterlo, W. A. L.; Ngidi, E. L.; Kuzvidza, S.; Morgans, G. L.; Moleele, S. S.; de Koning, C. B. Ring-closing metathesis for the synthesis of 2H-and 4H-chromenes. *Tetrahedron* **2005,** 61, 9996-10006.
97. Blouin, M.; Frenette, R. A new method for the preparation of aryl vinyl ethers. *J. Org. Chem.* **2001,** 66, 9043-5.
98. Groaz, E.; Banti, D.; North, M. Synthesis of Oxygen-Containing Medium and Large Rings using One-Pot Combinations of Sequential Alkene, Enyne and Alkyne Metathesis Reactions. *Adv. Synth. Catal.* **2007,** 349, 142-146.
99. Crespo, A.; Fernández, A. Induced Disorder in Protein–Ligand Complexes as a Drug-Design Strategy. *Mol. Pharmacol.* **2008**.
100. Freskos, J. N.; Fobian, Y. M.; Benson, T. E.; Bienkowski, M. J.; Brown, D. L.; Emmons, T. L.; Heintz, R.; Laborde, A.; McDonald, J. J.; Mischke, B. V.; Molyneaux, J. M.; Moon, J. B.; Mullins, P. B.; Bryan Prince, D.; Paddock, D. J.; Tomasselli, A. G.; Winterrowd, G. Design of potent inhibitors of human beta-secretase. Part 1. *Bioorg. Med. Chem. Lett.* **2007,** 17, 73-7.
101. Chirapu, S. R.; Pachaiyappan, B.; Nural, H. F.; Cheng, X.; Yuan, H.; Lankin, D. C.; Abdul-Hay, S. O.; Thatcher, G. R.; Shen, Y.; Kozikowski, A. P.; Petukhov, P. A. Molecular modeling, synthesis, and activity studies of novel biaryl and fused-ring BACE1 inhibitors. *Bioorg. Med. Chem. Lett.* **2009,** 19, 264-74.
102. Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem. Int. Ed. Engl.* **2003,** 42, 2718-36.
103. Verlinde, C. L.; Hol, W. G. Structure-based drug design: progress, results and challenges. *Structure* **1994,** 2, 577-87.
104. Debnath, A. K. Quantitative structure-activity relationship (QSAR) paradigm--Hansch era to new millennium. *Mini-Rev. Med. Chem.* **2001,** 1, 187-95.
105. Martin, Y. C. 3D QSAR: Current State, Scope, and Limitations. *Perspect. Drug Discovery Des.* **1998**, 3-23.
106. Verma, J.; Khedkar, V. M.; Coutinho, E. C. 3D-QSAR in drug design--a review. *Curr. Top. Med. Chem.* **2010,** 10, 95-115.
107. John, S.; Thangapandian, S.; Sakkiah, S.; Lee, K. W. Identification of potent virtual leads to design novel indoleamine 2,3-dioxygenase inhibitors: pharmacophore modeling and molecular docking studies. *European Journal of Medicinal Chemistry* **2010,** 45, 4004-12.
108. Nagarajan, S.; Ahmed, A.; Choo, H.; Cho, Y. S.; Oh, K. S.; Lee, B. H.; Shin, K. J.; Pae, A. N. 3D QSAR pharmacophore model based on diverse IKKbeta inhibitors. *J. Mol. Model.* **2010**.
109. Ajay; Murcko, M. A. Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.* **1995,** 38, 4953-67.
110. Cramer, R. D. a. M., M. The Lattice Model: A General Paradigm for Shape-Related Structure/Activity Correlation. *Abstracts of Papers of the Am. Chem. Soc. Meeting, Honolulu* **1979**.
111. Cramer, R. D., 3rd; Patterson, D. E.; Bunce, J. D. Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Biol. Res* **1989,** 291, 161-5.
112. Cross, S.; Cruciani, G. Molecular fields in drug discovery: getting old or reaching maturity? *Drug Discov. Today* **2010,** 15, 23-32.

113. Du, Q. S.; Huang, R. B.; Chou, K. C. Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Curr. Prot. Peptide Sci.* **2008,** 9, 248-60.
114. Cramer, R. D. Rethinking 3D-QSAR. *J. Comput. Aided Mol. Des.* **2011,** 25, 197-201.
115. Rush, T. S., 3rd; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005,** 48, 1489-95.
116. Kubinyi, H. Comparative Molecular Field Analysis (CoMFA). *The Encyclopedia of Computational Chemistry* **1998,** 1, 448-460.
117. Barrow, J. C.; Rittle, K. E.; Ngo, P. L.; Selnick, H. G.; Graham, S. L.; Pitzenberger, S. M.; McGaughey, G. B.; Colussi, D.; Lai, M. T.; Huang, Q.; Tugusheva, K.; Espeseth, A. S.; Simon, A. J.; Munshi, S. K.; Vacca, J. P. Design and synthesis of 2,3,5-substituted imidazolidin-4-one inhibitors of BACE-1. *ChemMedChem* **2007,** 2, 995-9.
118. Coburn, C. A.; Stachel, S. J.; Jones, K. G.; Steele, T. G.; Rush, D. M.; DiMuzio, J.; Pietrak, B. L.; Lai, M. T.; Huang, Q.; Lineberger, J.; Jin, L.; Munshi, S.; Katharine Holloway, M.; Espeseth, A.; Simon, A.; Hazuda, D.; Graham, S. L.; Vacca, J. P. BACE-1 inhibition by a series of psi[CH2NH] reduced amide isosteres. *Bioorg. Med. Chem. Lett.* **2006,** 16, 3635-8.
119. Nantermet, P. G.; Rajapakse, H. A.; Stanton, M. G.; Stauffer, S. R.; Barrow, J. C.; Gregro, A. R.; Moore, K. P.; Steinbeiser, M. A.; Swestock, J.; Selnick, H. G.; Graham, S. L.; McGaughey, G. B.; Colussi, D.; Lai, M. T.; Sankaranarayanan, S.; Simon, A. J.; Munshi, S.; Cook, J. J.; Holahan, M. A.; Michener, M. S.; Vacca, J. P. Evolution of tertiary carbinamine BACE-1 inhibitors: Abeta reduction in rhesus CSF upon oral dosing. *ChemMedChem* **2009,** 4, 37-40.
120. Rajapakse, H. A.; Nantermet, P. G.; Selnick, H. G.; Munshi, S.; McGaughey, G. B.; Lindsley, S. R.; Young, M. B.; Lai, M. T.; Espeseth, A. S.; Shi, X. P.; Colussi, D.; Pietrak, B.; Crouthamel, M. C.; Tugusheva, K.; Huang, Q.; Xu, M.; Simon, A. J.; Kuo, L.; Hazuda, D. J.; Graham, S.; Vacca, J. P. Discovery of oxadiazoyl tertiary carbinamine inhibitors of beta-secretase (BACE-1). *J. Med. Chem.* **2006,** 49, 7270-3.
121. Stauffer, S. R.; Stanton, M. G.; Gregro, A. R.; Steinbeiser, M. A.; Shaffer, J. R.; Nantermet, P. G.; Barrow, J. C.; Rittle, K. E.; Collusi, D.; Espeseth, A. S.; Lai, M. T.; Pietrak, B. L.; Holloway, M. K.; McGaughey, G. B.; Munshi, S. K.; Hochman, J. H.; Simon, A. J.; Selnick, H. G.; Graham, S. L.; Vacca, J. P. Discovery and SAR of isonicotinamide BACE-1 inhibitors that bind beta-secretase in a N-terminal 10s-loop down conformation. *Bioorg. Med. Chem. Lett.* **2007,** 17, 1788-92.
122. Zhu, H.; Young, M. B.; Nantermet, P. G.; Graham, S. L.; Colussi, D.; Lai, M. T.; Pietrak, B.; Price, E. A.; Sankaranarayanan, S.; Shi, X. P.; Tugusheva, K.; Holahan, M. A.; Michener, M. S.; Cook, J. J.; Simon, A.; Hazuda, D. J.; Vacca, J. P.; Rajapakse, H. A. Rapid P1 SAR of brain penetrant tertiary carbinamine derived BACE inhibitors. *Bioorg. Med. Chem. Lett.* **2010,** 20, 1779-82.
123. Podlogar, B. L.; Ferguson, D. M. QSAR and CoMFA: a perspective on the practical application to drug discovery. *Drug Des. Disc.* **2000,** 17, 4-12.
124. Akamatsu, M. Current state and perspectives of 3D-QSAR. *Curr. Top. Med. Chem.* **2002,** 2, 1381-94.
125. Debnath, A. K. Application of 3D-QSAR techniques in anti-HIV-1 drug design--an overview. *Curr. Pharm. Des.* **2005,** 11, 3091-110.
126. Garcia, I.; Fall, Y.; Gomez, G. QSAR, docking, and CoMFA studies of GSK3 inhibitors. *Curr. Pharm. Des.* **2010,** 16, 2666-75.

VITA

Boobalan Pachaiyappan (Email: boobalanp@gmail.com)

Professional Summary

- **Computational medicinal chemistry professional having sound knowledge in life sciences, pharmaceuticals, and chemical areas** including organic chemistry, homology modeling, *de novo* chemical library design, molecular docking, virtual screening, drug likeness, ligand-, fragment- and structure-based drug design, hit-to-lead optimization, ADMET modeling, 3D-QSAR
- **Hands-on experience on molecular modeling tools** such as Tripos®, GOLD®, MOE®, Schrodinger®, Openeye®, ChemAxon® (including Marvin and JChem), VMD® and Schrodinger®, and scientific search tools such as SciFinder®, Delphion®, MedTRACK® and EndNote®
- **Recognized both within and outside the organization** for very good resource, project & deadline management, in-depth research and team-coordination skills

Selected Professional Experience

- **Graduate Research/Teaching Assistant, University of Illinois at Chicago, IL (Fall 2006 – Fall 2010)**
  Performed medicinal chemistry research involving computer-assisted inhibitor/probe design and synthesis targeting Alzheimer's β-secretase (BACE1) and Histone Deacetylase 8 (HDAC8) involved in cancer.

- **Graduate Research/Teaching Assistant, Virginia Tech, Blacksburg, VA (Fall 2003 – Fall 2005)**
  Performed computational medicinal chemistry research involving ligand design and synthesis of Peptidyl Prolyl Isomerase1 (Pin1) inhibitors implicated in cell-cycle regulation. Instructed undergraduate chemistry labs

- **Summer Intern, Ciba Specialty Chemicals India Ltd., Mumbai, India (June – August, 2001)**
  Responsible for operation and maintenance of ISIS Base reaction database. Provided technical support to various R&D projects and assisted in PC- and server-based chemoinformatics capabilities related to the optimization of small molecules.

Publications / Posters

- Molecular modeling, synthesis and activity studies of novel biaryl and fused-ring BACE1 inhibitors. **Boobalan Pachaiyappan** et al. *Bioorg Med Chem Lett*., 63, 264-274 (2008)
- Molecular modeling, synthesis and activity studies of inhibitors targeting Alzheimer's disease β-secretase 1 (BACE1). **Boobalan Pachaiyappan** et al. *The 46th Annual MIKI meeting, Iowa City, IA. (Apr 25-27, 2007)*
- A novel method for the synthesis of highly functionalized 3,4-dihydropyrimidin-2(1H)-ones through the 1,4-addition on pyrimidin-2(1H)-ones. Co-Author: **P**. **Boobalan**. *Tetrahedron*, 63, 12215-12219 (2007)
- On the pursuit of new β-Secretase 2 (BACE2) inhibitors using structure-based *de novo* design methods. **Boobalan Pachaiyappan** et al. *The 45th Annual MIKI meeting, Kansas City, KS. (Apr 13-15, 2007)*
- Docking study of two Pin1 inhibitors into the catalytic site of Pin1. **Boobalan Pachaiyappan** et al. *The 230th American Chemical Society National meeting, Washington D.C. (Aug28-Sep1, 2005)*
- Synthesis of single strapped 21-thia tetraphenylporphyrin systems. **P. Boobalan** et al. *Indian J. Chem*., 41A, 1860-1964 (2002)

Speaking Engagements

- Structure- and Ligand-based Modeling Methods: Principles and Applications. **Boobalan Pachaiyappan**. Department of Chemical Biology and Therapeutics, St Jude Children's Research Hospital. (Jan 21, 2011)
- Role of molecular techniques in the design and development of Alzheimer's disease β-secretase inhibitors. **Boobalan Pachaiyappan.** *The 48th Annual MIKI meeting, Chicago, IL. (Apr 9-11, 2010)*
- Molecular modeling, synthesis and activity studies of inhibitors targeting Alzheimer's β-secretase. **Boobalan Pachaiyappan.** *The 236th American Chemical Society National Meeting*, *Philadelphia, PA.* (*Aug 17-21, 2008)*
  Inhibition of β-secretase (BACE1) as a therapeutic intervention strategy for Alzheimer's disease – Revisited. **Boobalan Pachaiyappan**. University of Illinois at Chicago, Chicago, IL (Oct 22, 2009

Education

- 2011 (Summer); PhD Medicinal Chemistry, University of Illinois at Chicago (USA)
- 2005; M.S. Organic Chemistry, Virginia Tech (USA)
- 2002; M.Sc Indian Institute of Technology, Bombay (India)
- 1999; B.Sc. Chemistry (*summa cum laude*), University of Madras, Chennai (India)