Molecular Modelling of Potential anti-Alzheimer Agents Using Chemoinformatics Tools

Thesis submitted

by

VINAY KUMAR

Doctor of Philosophy (Pharmacy)

Department of Pharmaceutical Technology
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata, India

2022

DEDICATED TO SCIENCE

JADAVPUR UNIVERSITY
KOLKATA – 700 032
INDIA

INDEX NO. - 288/19/Ph

1. Title of the thesis:

"Molecular Modelling of Potential anti-Alzheimer Agents Using Chemoinformatics Tools"

2. Name, Designation & Institution of the Supervisor(s):

Dr. Kunal Roy
Professor
Drug Theoretics and Cheminformatics Laboratory
Department of Pharmaceutical Technology
Jadavpur University
Kolkata – 700 032
West Bengal, INDIA

Dr. Achintya Saha
Professor
Pharmaceutical & Fine Chemical Division
Department of Chemical Technology
University of Calcutta, Kolkata-700009
West Bengal, INDIA

3. List of publications:

Papers related to dissertation

Research Articles

6. Vinay Kumar**, Achintya Saha, and Kunal Roy. Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease. *Chemometrics and Intelligent Laboratory Systems* 233 (2023): 104734.

5. *Vinay Kumar, Achintya Saha, and Kunal Roy***. In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease. *Computational Biology and Chemistry* 88 (2020): 107355.

4. ***Vinay Kumar****, Achintya Saha**.* Chemometric Modeling of Structurally Diverse Carbamates for the Inhibition of Acetylcholinesterase (AChE) Enzyme in Alzheimer's disease. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)* 5(3) (2020): 6-60.

3. ***Vinay Kumar****, Probir K. Ojha, Achintya Saha, and Kunal Roy**.* Cheminformatic modelling of β-amyloid aggregation inhibitory activity against Alzheimer's disease. *Computers in Biology and Medicine* 118 (2020): 103658.

2. ***Vinay Kumar****, Probir K. Ojha, Achintya Saha, and Kunal Roy**.* Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease. *SAR and QSAR in Environmental Research* 31(2) (2020): 87-133.

1. ***Vinay Kumar****, Priyanka De, Probir K. Ojha, Achintya Saha, and Kunal Roy**.* A multi-layered variable selection strategy for QSAR modeling of butyrylcholinesterase inhibitors. *Current Topics in Medicinal Chemistry* 20(18) (2020): 1601-1627.

Papers not related to dissertation

Research and Review Articles

10. ***Vinay Kumar****, Supratik Kar, Priyanka De, Kunal Roy**,* Jerzy Leszczynski. Identification of potential antivirals against 3CLpro enzyme for the treatment of SARS-CoV-2: A multi-step virtual screening study. *SAR and QSAR in Environmental Research* 33, no. 5 (2022): 357-386.

9. *Priyanka De,* ***Vinay Kumar****, Supratik Kar, Kunal Roy**,* and Jerzy Leszczynski. Repurposing FDA approved drugs as possible anti-SARS-CoV-2 medications using ligand-based computational approaches: sum of ranking difference-based model selection. *Structural Chemistry* (2022): 1-13.

8. *Arkaprava Banerjee, Priyanka De,* ***Vinay Kumar****, Supratik Kar, Kunal Roy**.* Quick and efficient quantitative predictions of androgen receptor binding affinity for screening Endocrine Disruptor Chemicals using 2D-QSAR and Chemical Read-Across. *Chemosphere* 309, Part 1, (2022): 136579.

7. *Kabiruddin Khan,* ***Vinay Kumar****, Erika Colombo, Anna Lombardo, Emilio Benfenati, and Kunal Roy**.* Intelligent consensus predictions of bioconcentration factor of pharmaceuticals using 2D and fragment-based descriptors. *Environment International* (2022): 107625.

6. *Rajendra Kumar Mukherjee,* ***Vinay Kumar****, and Kunal Roy**.* Chemometric modeling of plant protection products (PPPs) for the prediction of acute contact toxicity against honey bees (A. mellifera): A 2D-QSAR approach. *Journal of Hazardous Materials* 423 (2022): 127230.

5. *Priyanka De, Sagar Bhayye,* ***Vinay Kumar****, and Kunal Roy**. In silico modeling for quick prediction of inhibitory activity against 3CLpro enzyme in SARS CoV diseases. *Journal of Biomolecular Structure and Dynamics* 40, no. 3 (2022): 1010-1036.

4. *Pathan Mohsin Khan,* ***Vinay Kumar****, and Kunal Roy**. In silico modeling of small molecule carboxamides as inhibitors of SARS-CoV 3CL protease: an approach towards combating COVID-19. *Combinatorial Chemistry & High Throughput Screening* 24.8 (2021): 1281-1299.

3. *Rajendra Kumar Mukherjee,* ***Vinay Kumar****, and Kunal Roy**. Ecotoxicological QSTR and QSTTR modeling for the prediction of acute oral toxicity of pesticides against multiple avian species. *Environmental Science & Technology* 56.1 (2021): 335-348.

2. *Probir Kumar Ojha,* ***Vinay Kumar****, Joyita Roy, and Kunal Roy**. Recent advances in quantitative structure–activity relationship models of antimalarial drugs. *Expert Opinion on Drug Discovery* 16.6 (2021): 659-695.

1. ***Vinay Kumar****, and Kunal Roy**. Development of a simple, interpretable and easily transferable QSAR model for quick screening antiviral databases in search of novel 3C-like protease (3CLpro) enzyme inhibitors against SARS-CoV diseases. *SAR and QSAR in Environmental Research* 31.7 (2020): 511-526.

Book chapters not related to dissertation

2. ***Vinay Kumar****, and Kunal Roy**. Computational Modeling of Chloroquine Analogues for Development of Drugs against Novel Coronavirus (nCoV). In Silico Modeling of Drugs against Coronaviruses. Humana, New York, NY, 2021. 579-614.

1. ***Vinay Kumar****, and Kunal Roy**. "Computational Modeling of RdRp Inhibitors for the Development of Drugs against Novel Coronavirus (nCoV)." In Silico Modeling of Drugs against Coronaviruses. Humana, New York, NY, 2021. 541-578.

4. List of patents: **Nil**

5**.** List of Presentation in National/ International:

International conferences

1. ***Vinay Kumar****, Achintya Saha, and Kunal Roy*. Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease*. **Oral** presentation in *23rd European Symposium on Quantitative Structure-Activity Relationship* held at Heidelberg, Germany on September 26-30, 2022.

2. ***Vinay Kumar****, Probir K. Ojha, Achintya Saha, and Kunal Roy*. *Exploring 2D-QSAR, Pharmacophore Mapping and Molecular Docking Studies of Beta-secretase 1 (BACE1) Enzyme Inhibitory Activity Against Alzheimer's Disease*. **Poster** presentation in *Current Trends in Pharmaceutical and Medical Sciences (CTPMS-2020)*, held at GIPER, Kashipur-244713, Uttarakhand, India on 26 - 29 Feb 2020.

3. ***Vinay Kumar****, Achintya Saha, and Kunal Roy*. *Cheminformatic modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease*. **Poster** presentation in *58th Annual Convention of Chemists (ACC), Indian Chemical Society (ICS) Organic & Bio-Chemistry Section*, Jointly Organized by Dept. of Chemistry, IIT Kharagpur & Dept. of Chemical Sciences, IISER Kolkata on 22-24 December, 2020.

National conference/ workshops

1. ***Vinay Kumar****, Achintya Saha, and Kunal Roy*. In silico analysis for the inhibition of Butyrylcholinesterase (BuChE) Enzyme in Alzheimer's disease*: A 2D-QSAR approach. **Poster** presentation in *Indian Conference on Bioinformatics (INBIX -2021)* held during November 11-13, 2021. (**"Won best poster Award"**)

2. International conference and workshop "Indo-Australian workshop on rational drug design" on February 26-27, 2019, conducted by the CSIR-Indian Institute of Chemical Biology, Kol-700 032, West Bengal.

3. International conference and workshop "*Chemometric data processing for analytical system in Agricultural and Environmental studies*" on December 2-7, 2019, conducted by the Department of Instrumentation and Electronics Engineering, Jadavpur University, Salt lake campus, Kolkata.

4. National Seminar on Center for sustainable development and resources efficiency management. During January 13-14, 2020. FET, Jadavpur University, Kolkata-700 032, West Bengal.

5. International conference "Health Informatics Summit" organized by the Department of Computational Biology, Indraprastha Institute of Information Technology Delhi (IIIT-D) and APbians (Bioinformatics Society) on 16th - 19th of October 2021.

“Statement of Originality”

I **Vinay Kumar** registered on **24th June 2019** do hereby declare that this thesis entitled “**Molecular Modelling of Potential anti-Alzheimer Agents Using Chemoinformatics Tools**” contains a literature survey and original research work done by the undersigned candidate as part of Doctoral studies.

All information in this thesis has been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work.

I also declare that I have checked this thesis as per the “Policy on Anti Plagiarism, Jadavpur University, 2019”, and the level of similarity as checked by iThenticate software is ***7%****.

Signature of Candidate:

Date:

Certified by Supervisor(s):

(Signature with date, seal)

1.

2.

*Ignoring smaller match of up to 14 words (as per the UGC) and ignoring own publications as the source.

CERTIFICATE FROM THE SUPERVISOR/S

This is to certify that the thesis entitled "***Molecular Modelling of Potential anti-Alzheimer Agents Using Chemoinformatics Tools***" submitted by Shri **Vinay Kumar,** who got his name registered on **24th June, 2019** for the award of Ph.D. (Pharmacy) degree of Jadavpur University is absolutely based upon his own work under the supervision of **Professor Kunal Roy** and co-supervision of **Professor Achintya Saha** that neither his thesis nor any part of the thesis has been submitted for any degree/ diploma or any other academic award anywhere before.

1._____

Signature of the Supervisor

(Signature with date, seal)

2._____

Signature of the Supervisor

(Signature with date, seal)

Acknowledgements

It has been my privilege to work under the guidance of Dr. Kunal Roy, Professor, Drug Theoretics & Cheminformatics Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata-32 and co-guidance of Dr. Achintya Saha, Professor, Pharmaceutical & Fine Chemical Division, Department of Chemical Technology, University of Calcutta Kolkata-09. I express my deep gratitude and regards to my guides for suggesting the subject of this thesis and rendering me his thoughtful and rational approach to this thesis work. I am greatly indebted to Dr. Kunal Roy for his valuable guidance throughout the work that enabled me to complete the work. His constant inspiration, intellectual support, helpful suggestions and innovative approaches have tended me to accomplish this work in time.

I am thankful to the authorities of Jadavpur University for providing all the facilities needed for my work. I express deep gratitude to Head of the Department, Vice-Chancellor, Registrar, Dean, Faculty of Engineering and Technology, Jadavpur University for facilities provided to carry out this work. I also thank ICMR, New Delhi for the financial support provided by the organizations to carry out this work. Additionally, I thank DBT and Heidelberg Institute for Theoretical Studies (HITS), Heidelberg Germany for providing travel grant to attend conference in abroad. Furthermore, I thank DST-SERB and CSIR for granting travel grant to attend conferences in abroad.

I am indeed glad to convey cordial thanks to my seniors and juniors in my lab especially to Dr. Probir Kumar Ojha, Dr. Pravin Ambure, Dr. Rahul Balasaheb Aher, Dr. Supratik Kar, Dr. Kabiruddin Khan, Priyanka De, Joyita Roy, Mainak Chatterjee, Dr. Pathan Mohsin Khan, Dr. Jillella Gopala Krishna, Dr. Sagar Bhayye, Sulekha Ghosh, Sapna Kumari Pandey, Rajendra Kumar Mukherjee, Arnab Seth, Arkaprava Banerjee, Rahul Paul, Ankur Kumar, Souvik pore, Shilpayan Ghosh and Aniket Nath who have extended their helping hands and friendly cooperation to do my work. Finally, I would like to thank my parents, Mr. Dinesh Kumar and Mrs. Suneeta Devi, my brothers Vijay, Ajay and Abhishek for all the love, patience and support throughout my entire course of work. ***Last but not the least I would also like to thank all the people I used to meet on a regular basis in this 'city of joy' for their kind hospitality which surely facilitated my pleasant stay in Kolkata.***

VINAY KUMAR

Date:

Place:

Preface

The presented work in this dissertation is spread over a span of four and half years. The current dissertation employed, numerous *in silico* techniques to study the potential leads against Alzheimer's disease. Alzheimer's disease (AD) is a progressive neuropathological disorder, found in the most common form of dementia, which causes severe brain deterioration and cognitive function loss. AD is a degenerative ailment that is thought to begin decades before symptoms appear. Clinicians are only able to identify even the first signs of AD after significant damage has already been done to crucial biological components. The root cause of AD is still uncertain, which is one of the main reasons for being incurable, while various hypotheses, such as, cholinergic hypothesis, amyloid hypothesis and tau hypothesis, oxidative stress hypothesis *etc*. have been proposed. At present only symptomatic treatment is available, which is based on several hypotheses that were found to be associated with AD. Based on these hypotheses twelve biological targets were studied in this work, where twelve main biological targets, namely, AChE, BuChE, BACE1, 5-HT6, CDK-5, Gamma (γ)-secretase, Glutaminyl Cyclase (QCs), GSK-3β, NMDA receptor, amyloid-beta plaques (Aβ), PDE 10A and MAO-B enzyme are targeted to find appropriate treatment in AD.

In the modern world, drugs that only target one enzyme or receptor are thought to be insufficient for the treatment of complex diseases, whose causes are frequently multifactorial, including neurodegenerative disorders. An emerging approach is to design a drug that is capable of hitting multiple biological targets. Such molecules are named multi-target-directed ligands (MTDLs). This strategy is challenging to implement, but it is far superior to earlier approaches such as multiple-medication therapy or multiple-compound medication because the MTDLs approach avoids any potential drug-drug interactions as well as the complications associated with the simultaneous use of numerous drugs, which may have varied solubility, absorption, bioavailability, distribution, and metabolism rate. An emerging area of pharmaceutical science called polypharmacology is involved in developing drugs that can target numerous disease pathways and multiple targets at once. The fundamental of polypharmacology is to study and recognize the interactions between potential drugs and different target proteins. Though, studying the drug-protein interactions at the experimental level is highly challenging, expensive, and time-consuming. *In silico* or computational methods play an important role in finding MTDLs, saving both cost and time. In the present thesis work, numerous *in silico* techniques were employed to study the potential leads against AD. The main objective was to use different in silico approaches to find and improve potential anti-Alzheimer's leads against several crucial targets involved in AD. Along with the single-target drug designing approach, we have also focused on identifying or designing dual-binding site AChE inhibitors, as well as multi-target inhibitors. Further, we have explored the selectivity issue of inhibitors against AChE over BuChE, which is a commonly observed issue while designing molecules against enzymes.

Furthermore, the development of inhibitors against AD is a challenging procedure due to the complication of the molecular pathways involved in the progression of the disease. Computer-Aided Drug Design (CADD) uses computer power, three-dimensional graphics, mathematics, and statistics to understand and predict the binding mode and energy of small molecule inhibitors with potential targets. The most common in-silico techniques employed by medicinal chemists to help them rationalize the selection of hit compounds and to perform hit-to-lead optimization include structure-based design like molecular docking and dynamics and ligand-based design like quantitative structure-activity

relationships (QSAR), chemical Read-Across and pharmacophore mapping. Among these methodologies, the quantitative structure-activity relationship (QSAR) and molecular docking have great applications in the area of in-silico search. The QSAR methods are essential for the exploration of important structural features and prediction of the biological activity of novel compounds based on mathematical and statistical relations. The idea of QSAR is based on the concept that endpoint values of compounds change systematically with modification of the structural attributes. In the present thesis work, numerous *in silico* techniques were employed to study the potential leads against AD. The main objective was to use different in silico approaches to find and improve potential anti-Alzheimer's leads against several crucial targets involved in AD. Along with the single-target drug designing approach, we have also focused on identifying or designing dual-binding site AChE inhibitors, as well as multi-target inhibitors. Further, we have explored the selectivity issue of inhibitors against AChE over BuChE, which is a commonly observed issue while designing molecules against enzymes. Although we used a variety of *in silico* methods, such as QSAR, molecular docking, pharmacophore modeling, virtual screening, and so on, the majority of our work is focused on developing predictive and statistically robust QSAR models. The QSAR approach is used extensively in the lead optimization step of any drug development effort to reduce time, money, and, most importantly, animal sacrifice. A QSAR model is used to identify the structural features responsible for the activity as well as to achieve selectivity. Additionally, we have also developed the quantitative structure activity-activity relationship (QSAAR) and selectivity-based models to explore the most important features contributing to the dual inhibition against the respective targets. Furthermore, the model provides significant information for designing new compounds with improved activity, and it is used to predict the activity of a query or newly designed compound.

The following studies have been performed in this dissertation:

A Multi-layered Variable Selection Strategy for QSAR Modeling of Butyrylcholinesterase Inhibitors.

Exploring 2D-QSAR for prediction of Beta-secretase 1 (BACE1) inhibitory activity against Alzheimer’s disease.

Cheminformatic modeling of β-amyloid aggregation inhibitory activity against Alzheimer's disease.

Chemometric modeling of structurally diverse carbamates for the inhibition of acetylcholinesterase enzyme (AChE) in Alzheimer’s disease

In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer’s disease

Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease

The work has been presented in this dissertation under the following sections:

Chapter 1 : Introduction
Chapter 2 : Present work
Chapter 3 : Materials and methods
Chapter 4 : Results and discussions
Chapter 5 : Conclusion
References
Appendix : Reprints

In the "***Introduction***" section, the basic information is provided on Alzheimer's disease, its present status, diagnosis, and the factors associated with the disease. This section also includes current approaches to design MTDLs, detail information on the computational methods implemented in this work and a short review on QSAR studies that were performed on anti-Alzheimer's agents. The sources of the employed datasets and the planned work have been discussed in the ***'Present Work'*** section. The detailed information associated to the descriptors and the methodologies has been provided in the section **'*Materials and Methods*'**, while the results have been thoroughly discussed in **'*Results and Discussions*'** section. Finally, ***'Conclusion'*** has been incorporated followed by **'*References*'**. The studies thus performed have been published in different ***reputed international journals*** and also presented in different national and international conferences which have been included under the section '***Reprints'***. Though, the work done and presented in this dissertation constitutes a small part of the broad spectrum of envisaged work. Considering the stipulated time limit, only some representative and relevant studies could be performed. Many other interesting aspects arising out of this work could have been investigated in a far more meaningful way, which might be planned in future.

Abbreviations

2D QSAR	Two dimensional QSAR
3D QSAR	Three dimensional QSAR
4D QSAR	Four dimensional QSAR
5D QSAR	Five dimensional QSAR
5-HT6	5-hydroxytryptamine 6
6D QSAR	Six dimensional QSAR
7D QSAR	Seven dimensional QSAR
7-MEOTA	7-mehtoxytacrine
AChE	Acetylcholinesterase
AChEIs	AChE inhibitors
AD	Alzheimer's disease
ADMET	Adsorption, distribution, metabolism, excretion and toxicity
AE	Absolute error
ANN	Artificial neural network
APP	Amyloid precursor protein
AUC	Area under the curve
Aβ	Amyloid-beta
BACE1	Beta (β)-site amyloid precursor protein cleaving enzyme 1
BuChE	Butyrylcholinesterase
CADD	Computer aided drug design
CCC	Concordance correlation Coefficient
CDK2	Cyclin-dependant kinase 2
CDK5	Cyclin-dependant kinase 5
CK1	Casein kinase 1
CNS	Central nervous system
CoMFA	Comparative molecular field analysis
CoMSIA	Comparative molecular similarity indices
DF	Discriminant function
DTDLs	Dual-target directed ligands
E-state	Electrotopological state
ETA	Extended topochemical atom
FDA	Food and drug administration
G/PLS	Genetic partial least squares
GA	Genetic algorithm
GFA	Genetic function approximation
GQSAR	Group based quantitative structure activity relationship
GSK-3β	Glycogen synthase kinase 3β
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
HYD	Hydrophobic
KNIME	Konstanz Information Miner
LBDD	Ligand-based drug design
LFER	Linear free energy relationships
LMO	Leave-many-out
LOF	Lack of fit
LOO	Leave-one-out
LR	Linear Regression
MAE	Mean absolute error
MAO-B	Monoamine oxidase-B
MLR	Multiple linear regression
MMPs	Matched molecular pairs
MOE	Molecular operating environment software
MAPK	Mitogen-activated protein kinases
MSA	Molecular spectrum analysis
MTDD	Multi-target drug designing
MTDLs	Multi-target-directed ligands
NMDA	N-methyl-D-aspartate
OECD	Organization for economic co-operation and development
PCA	Principal Component analysis
PDB	Protein data bank
PDE 10A	Phosphodiesterase 10A
PLS	Partial least squares
PRESS	Predicted residual sum of squares
PSEN1	Presenilin 1
QAAR	Quantitative activity-activity relationship
QCs	Glutaminyl Cyclase
QSAR	Quantitative structure-activity relationship
QSPR	Quantitative structure-property relationship
QSTR	Quantitative structure-toxicity relationship
RA	Ring aromatic
RDD	Rational drug design
REACH	Registration, evaluation, authorisation and restriction of chemicals
RMSD	Root mean square deviation
rmsep	Root mean square error in prediction
ROC	Receiver operating characteristics curve
ROS	Reactive oxygen species
RTO	Regression through origin
SBDD	Structured-based drug design
SD	Standard deviation
SDEP	Standard deviation of error of prediction
SDF	Structure data format/file
SEE	Standard error of estimate
S-MLR	Stepwise- multiple linear regression
SVM	Support vector machine
TSMs	target-specific medicines
VS	Virtual screening
WHIM	Weighted holistic invariant molecular descriptor
WHO	World health organization

Table of Contents

CHAPTER - 1

INTRODUCTION

Chapter 1: Introduction

1.1. Alzheimer's disease (AD)-a multi-factorial disease

Alzheimer's disease (AD) is a complex, progressive neurological disease, which contributes to about 60–80% of all dementia cases[1]. Most people are unsure about the distinction between Alzheimer's and dementia. Dementia is a broad term that refers to a set of symptoms[1-2]. Dementia is characterized by concerns with memory, language, problem-solving, and other intellectual abilities that limit a person's ability to carry out daily activities[3]. Dementia has various causes such as AD, cerebrovascular disease, Lewy body disease, Frontotemporal lobar degeneration (FTLD), Parkinson's disease (PD), Hippocampal sclerosis (HS), and mixed pathologies[2]. AD is the most prevalent type of dementia[1-3]. The initial signs of the disease may be a constant decline in loss of short-term memory and intellectual functions, repeatedly accompanied by abnormal behavior such as aggression and depression, and lastly, the ability to perform basic activities of daily living (ADLs), which ultimately lead to death[1-3].

1.1.1. History and the major hallmarks

Dr. Alois Alzheimer, a German psychiatrist, noticed the progressive decline in cognitive abilities such as memory loss, language problems, and unpredictable behavior in Auguste D., one of his female patients, in 1901[4-5]. Dr. Alois investigated and studied her brain tissues after she died in 1906 and discovered two types of aberrant deposits inside and between nerve cells, which are now known as amyloid plaques and neurofibrillary tangles, respectively[4-5]. Dr. A. Alzheimer named this disorder "presenile dementia", first in 1906. Afterward, Dr. Emil Kraepelin, a coworker of Dr. Alzheimer suggested the name "Alzheimer's disease" instead of "presenile dementia"[4-5]. More than a century after the disease was first described, the two fundamental pathology hallmarks continue to be the primary explanation for its pathophysiology. Other prominent markers include brain atrophy, dementia, and inflammation.

1.1.2. Current status (statistics)

It is challenging to predict how many deaths are caused by AD each year due to the way the causes of death are recorded. Statistics show that AD is the fifth leading cause of death for people aged 65 and above, and the sixth main cause of death worldwide[2, 5]. However, it might be too responsible for many more fatalities than what has been reported in official sources. Among elderly people, AD is a significant contributor to the morbidity-a state of poor health and disability. Before dying from Alzheimer's, a person has years of morbidity as the disease advances. According to preliminary data from the Centers for Disease Control and Prevention (CDC)[2, 5], excess mortality (the difference between the observed and projected number of deaths for a specific period) was extremely high in 2020 compared to prior years, particularly among older persons. Most of these increased deaths happened among vulnerable aged persons suffering from AD and other dementias[2, 5]. Preliminary CDC records show that in 2020, there were at least 42,000 more deaths from AD and other dementias than there were on average in the five years before 2020[2, 5]. This is about 16% higher than expected[2, 5]. Furthermore, the number of COVID-19 deaths for which the death certificate stated more than one condition as a cause of death was also counted by the CDC: in 4% of death certificates stating COVID-19 as the main reason for death, AD was also indicated as one of the various reasons of death, and in 11% of death certificates identifying COVID-19 as the prime cause of demise, there was also a record of an unspecified form of dementia[2, 5]. AD was identified as one of several reasons for death for 8% of

patients over the age of 85 who died from COVID-19, whereas vascular dementia was indicated for 20%[2, 5]. As a result, we anticipate that the substantial impact of this epidemic on Alzheimer's death patterns will be evident in the coming years. According to the World Alzheimer Report 2010[2, 5], 35.6 million people were living with dementia worldwide. Further, according to the World Alzheimer Report 2015[2, 5], it is estimated that around the world, there have been ≈9.9 million new cases of dementia in 2015, i.e., one case was reported every 3 seconds, and there were around 46.8 million people worldwide living with dementia. According to the World Alzheimer Report 2021[5], there were around 55 million people worldwide living with dementia, and the figure is expected to exceed 80 million in 2030 and 152 million in 2050 worldwide.

1.1.3. Signs and symptoms

AD is estimated to start 20 years or more before it manifests symptoms[2, 5-6]. It begins with undetectable alterations in the brain of the individual who is affected [2, 5-6]. Individuals only begin to notice symptoms such as memory loss and language difficulties after years of brain changes. Symptoms occur as a result of nerve cell damage or destruction in areas of the brain involved in intellectual functions, learning, and memory (cognitive abilities)[2, 5-6]. Other sections of the brain's neurons are also damaged as the disease progresses [2, 5-6]. Neurons in areas of the brain that allow a person to walk and swallow become impaired with time. Individuals become bedridden and demand 24-hour care. Alzheimer's disease is fatal in the end. The onset of symptoms in more than 90% of AD patients occurs after the age of 60, and the disease becomes more common as people get older[2, 5-6]. The consequences of such AD-related factors differ from person to person. Rarely, do patients with early-onset AD experience AD symptoms at 30, 40, or 50 years old[2, 5-6]. AD is a developing healthcare concern, with increased life expectancy as the primary risk cause. The disease prevalence is expected to more than double over the next several decades in the absence of adequate prevention and treatment strategies. Complex AD pathogenesis and progression are influenced by various pathophysiological events occurring sequentially in the brain depicted in **Figure 1.1**.

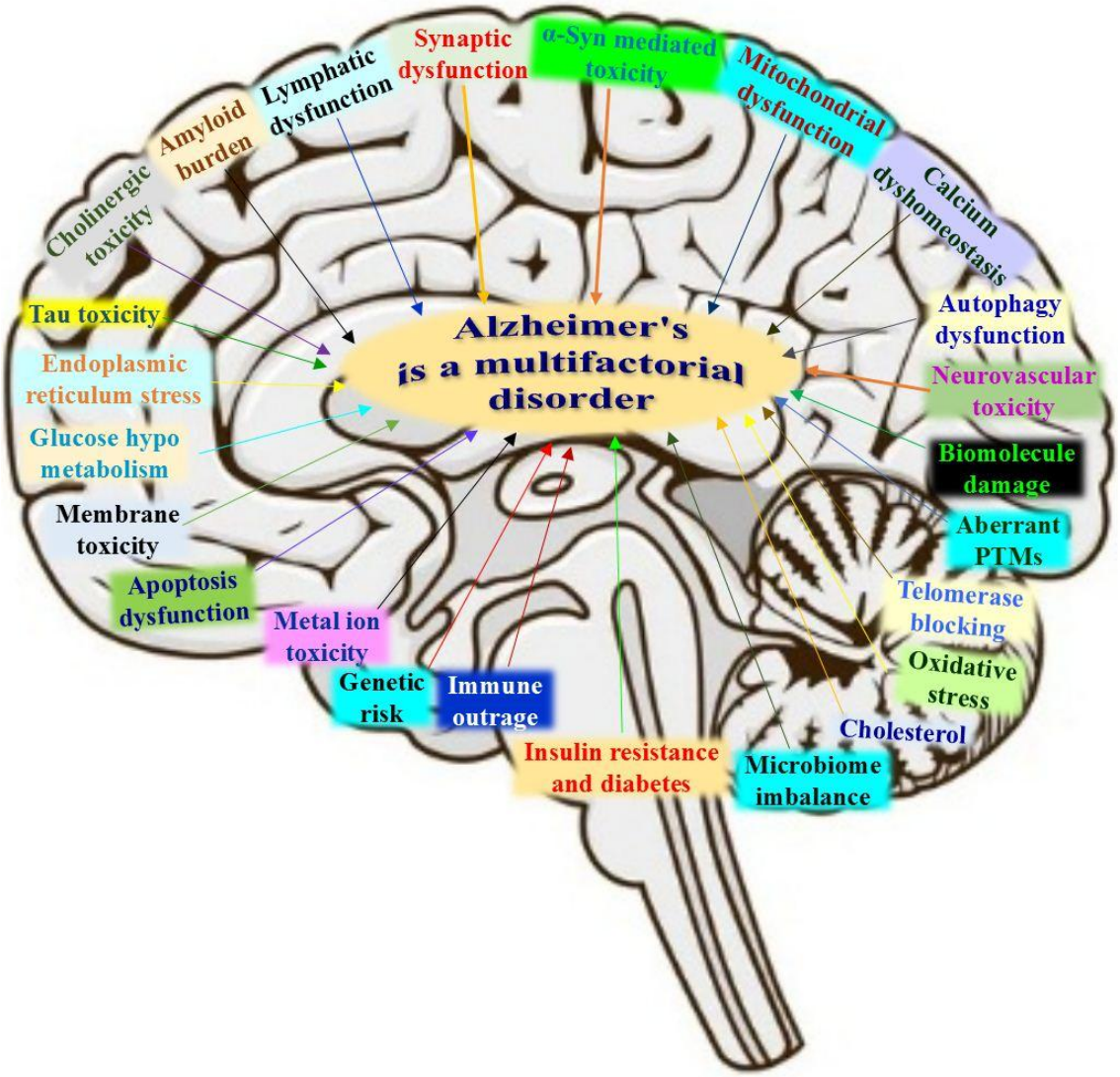


Figure 1.1. Complex AD pathogenesis and progression are influenced by various pathophysiological events occurring sequentially in the brain.

1.1.4. Multiple factors linked with AD

1.1.4.1. Genetic factors

AD can be genetic as an autosomal dominant disease with nearly perfect penetrance. The disease's autosomal dominant type is caused by mutations in three genes: the AAP gene on chromosome 21, the Presenilin 1 (PSEN1) gene on chromosome 14, and the Presenilin 2 (PSEN2) gene on chromosome 1[7-9]. The formation and aggregation of the beta-amyloid peptide may increase as a result of APP mutations. PSEN1 and PSEN2 mutations cause beta-amyloid aggregation by interfering with gamma-secretase processing[7-9]. Most cases of early-onset AD and 5 to 10% of all cases are caused by mutations in these three genes[7-9]. Apolipoprotein E is an additional genetic marker that raises the risk of AD[7-9]. It is a lipid metabolism regulator with an affinity for beta-amyloid protein[7-9]. The APOE isoform e4 gene, which is located on chromosome 19, has been linked to more sporadic and familial types of AD that manifest after age 65[7-9]. The existence of one APOEe4 allele does not always result in AD, but roughly 50% of people with one APOE-e4 allele have AD, and 90% of people with two alleles develop AD[7-9]. The age

of disease onset decreases with each APOEe4 allele. The APOEe4 allele is a significant risk factor for developing AD[7-9]. Both familial and sporadic types of AD have been linked to variations in the sortilin receptor SORT1 gene, which is crucial for moving APP from the cell surface to the Golgi-endoplasmic reticulum complex

1.1.4.2. Non-genetic factors

Although various hypotheses have been proposed to explain AD, none of them can pinpoint the precise aetiology of the disease[5, 7-9]. The most prevailing theory for treating AD is the cholinergic hypothesis, which refers to the enzymes acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE)[5, 7-9]. A healthy brain needs acetylcholine to accomplish cognitive processes, making it a crucial neurotransmitter[5, 7-9]. BuChE is a sister enzyme of AChE [5, 7-9]. The usual ratio of BuChE to AChE in the brain can change from 0.6 to 11 in AD when the activity level of the AChE enzyme decreases and the level of BuChE activity increases[5, 7-9]. These facts have led to the dual inhibition approach for these enzymes being suggested to improve the efficacy of the therapeutic strategy and improve the indications[5, 7-9]. Therefore, AChE and BuChE inhibitors have become significant tools in the treatment of AD. Though the effects of the enzymes appear mainly to be symptomatic, but have possible neuroprotective effects. Currently, four cholinesterase inhibitors, like tacrine, donepezil, galantamine, and rivastigmine (FDA-approved) are used for the symptomatic treatment of AD. Senile or amyloid-beta (Aβ) plaques and neurofibrillary tangles are the two main hallmarks of AD, as formerly known[5, 7-9]. These two key hypotheses named the amyloid hypothesis and the tau hypothesis was established in light of these characteristics. According to the amyloid hypothesis, Aβ accumulation is a basic cause of AD[5, 7-9]. In normal brain physiology, APP is cleaved by β, γ, and α secretase enzymes yielding 40 soluble amino acid peptides[5, 7-9]. But, in the case of AD, a two-step proteolytic process is initiated by the Swedish double mutation at the BACE1 (β-secretase) followed by γ-secretase (catalytic subunits presenilin 1 and 2) yielding a 42 insoluble amino acid peptide called amyloid-β (Aβ), and consequently forming β-amyloid plaque[5, 7-9]. According to the tau hypothesis, neurofibrillary tangles are formed when tau proteins are hyperphosphorylated[5, 7-9]. This ultimately results in the transport mechanism of the neuron failing due to the disintegration of microtubules[5, 7-9]. Furthermore, dysregulation in several protein kinases, including glycogen synthase kinase 3 (GSK3), mitogen-activated protein kinases (MAPK), casein kinase 1 (CK1), and cyclin-dependent kinase 5 (CDK5), which can act as effective targets, contributes to the hyperphosphorylation of tau proteins[5, 7-9]. According to the myelin hypothesis, AD may potentially be led on by the breakdown of myelin with aging in the brain. It is thought that further harm is caused by the local release of iron and cholesterol during myelin breakdown. According to the oxidative stress theory, the pathophysiology of neuronal death in AD may be caused by the brain experiencing chronically elevated levels of oxidative stress. Anti-oxidants can therefore be used in this situation to eliminate reactive oxygen species (ROS) or stop their synthesis. One of the potential targets is a monoamine oxidase (MAO-B), which is known to be involved in the formation of ROS, which directly destroys neuronal cells. Moreover, the mitochondria-targeted antioxidants have also demonstrated efficacy against Aβ plaques in an animal investigation, while humans (with mild to moderate AD) have shown better cognitive performance and behavioral impairments. Important targets of AD that are based on the above-mentioned hypotheses are shown in **Figure 1.2**.

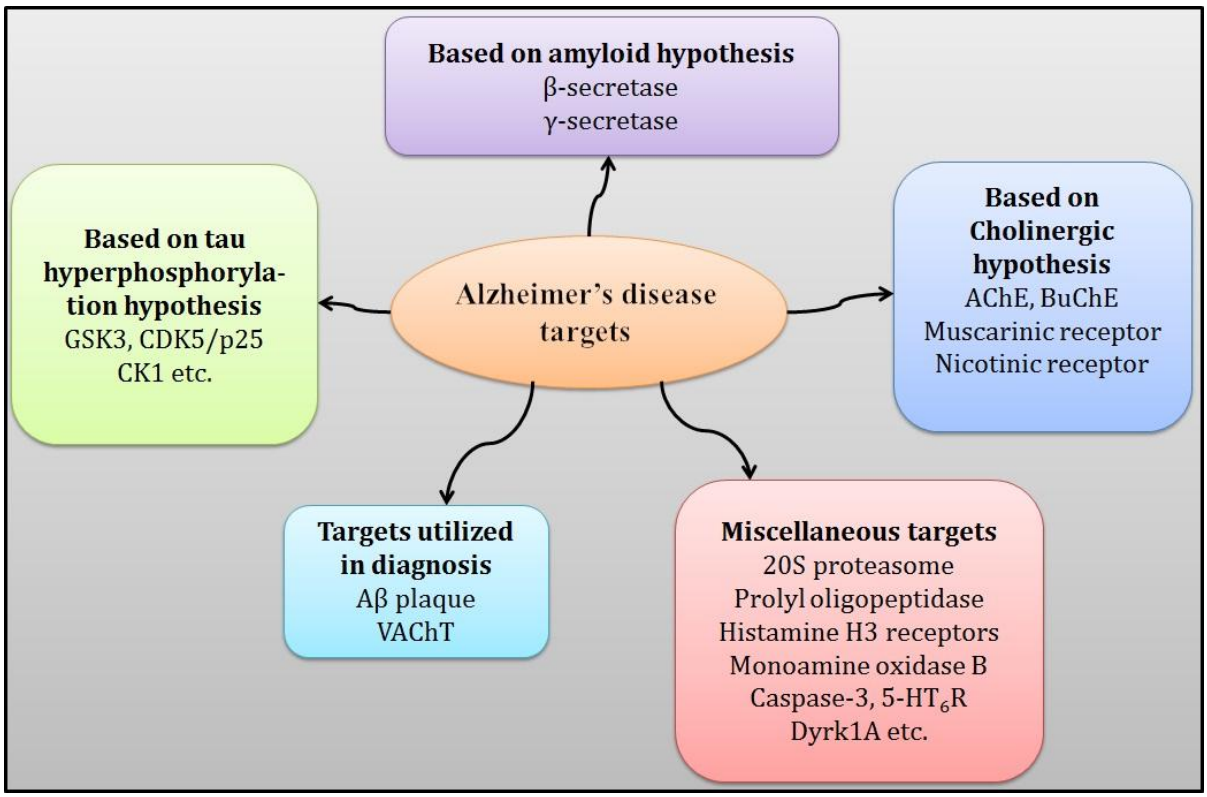


Figure 1.2. Proposed targets for the design and development of drugs against AD. 5-HT6R: 5-hydroxytryptamine-6 receptor; Aβ: beta-amyloid; AChE: Acetylcholinesterase; CDK5/p25: Cyclin-dependant kinase 5/p25; CK1: Casein kinase 1; Dyrk1A: Dual-specificity tyrosine phosphorylation-regulated kinase 1A; GSK3: Glycogen synthase kinase-3; VAChT: Vesicular acetylcholine transporter.

1.1.5. Diagnosis and treatment

It has been shown that the early brain damage associated with AD starts to manifest far before the onset of the clinical symptoms[2, 5]. The diagnosis of AD is based primarily on the patient's medical history, the collateral histories of their family, and clinical observations on the presence or absence of distinctive neurological and cognitive characteristics[2, 5]. Structure imaging, functional imaging, and molecular imaging are among the brain imaging or neuroimaging technologies employed in AD research[2, 5]. Molecular imaging techniques are one of the most active study fields among the aforementioned technologies that could help detect AD early on, that is, even before changes to the structure or activity of the brain. Molecular imaging agents now used in Alzheimer's disease research (**see Table 1.1**) include Pittsburgh Compound-B (PiB, 2-(4-[^{11}C] methylaminophenyl)-6-hydroxybenzothiazole), F-flutemetamol, F-florbetapir, and F-florbetaben[2, 5]. Undeniably, worldwide researchers have made significant efforts to the identification of novel strategies for the treatment of AD. AChE enzyme inhibitors (**Donepezil, Galantamine, Tacrine**, and **Rivastigmine**) and NMDA receptor blockers (**Memantine**) are the only two kinds of drugs available for the treatment of AD (see **Figure 1.3**)[2, 5]. Very recently, **Aducanumab** (Brand name: Aduhelm) was approved by FDA for the treatment of AD (see **Table 1.1**)[10]. It is an amyloid beta-directed monoclonal antibody that targets aggregated forms of amyloid beta found in the brains of people with AD to reduce its buildup[10]. It was developed by Biogen and Eisai[10]. Additionally, the availability of more recent atypical neuroleptics and serotonin-modulating antidepressants has increased hope for effective therapy of AD[10]. The therapeutic efficacy of these substances is still somewhat limited, despite their usage in medicine. As a result, throughout the last

decade, the emphasis has switched to the development of a new therapeutic strategy for the treatment of AD. New hope for the treatment of AD has been provided by the development of a 7-mehtoxytacrine (7-MEOTA) derivative based on the multi-target-directed ligand.

Table 1.1. A list of pharmaceuticals and neuroimaging agents that have received FDA approval for diagnosis and treating symptoms of Alzheimer's disease, respectively.

Serial No.	Name	Brand name (NDA* number)	Molecular target	Approved for disease stage	Year of FDA approval
		Drugs for symptomatic treatment			
1	Donepezil	Aricept (NDA 20-690)	AChE	All stages	1996
2	Rivastigmine	Exelon (NDA 20-823)	AChE	Mild to moderate	2000
3	Galantamine	Razadyne (NDA 21-169)	AChE	Mild to moderate	2001
4	Memantine	Namenda (NDA 21-487)	NMDA receptor	Moderate to severe	2003
5	Donepezil + Memantine	Namzaric (NDA 206-439)	AChE + NMDA receptor	Moderate to severe	2014
6	Aducanumab	Aduhelm (NDA 761178)	β-Amyloid	Moderate to severe	2021
		Imaging agents for diagnosis			
1	^{18}F-florbetapir	Amyvid (NDA 202008)	Aβ plaques	-	2012
2	^{18}F-flutemetamol	Vizamyl (NDA 203137)	Aβ plaques	-	2013
3	^{18}F-florbetaben	Neuraceq (NDA 204677)	Aβ plaques	-	2014

*NDA: New Drug Application

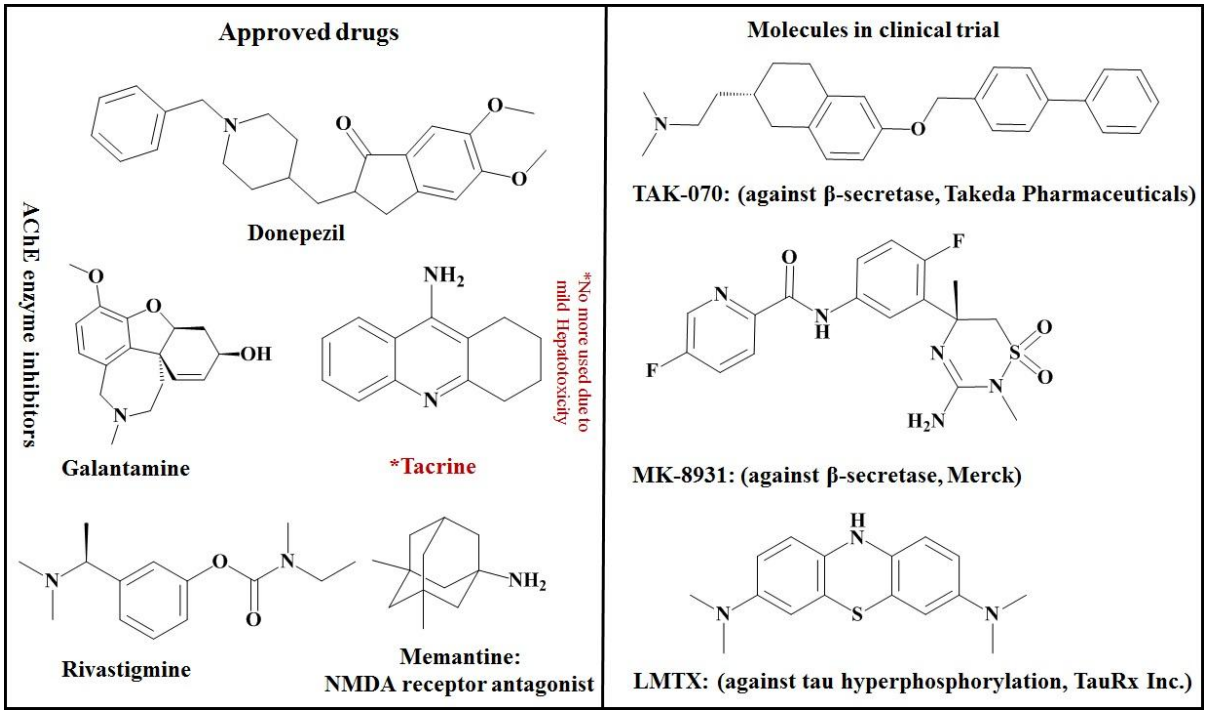


Figure 1.3. List of approved drugs and molecules in clinical trials.

1.1.6. Different biological targets focused in this study

1.1.6.1. Acetylcholinesterase (AChE) and Butyrylcholinesterase (BuChE) enzyme

The well-known "cholinergic hypothesis" relates neuronal degeneration with the loss of cholinergic neurotransmission. This is the oldest hypothesis of AD progression, according to which, a reduced synthesis of neurotransmitters acetylcholine (terminated by acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE))[11-12] results in neuroinflammation and large scale aggregation of β-amyloid. Even though the distinctive features of neurodegeneration in Alzheimer's brains are well known, one of the current difficulties is related to the lack of solid evidence regarding the crucial factors that give rise to the pathogenesis of this disease, creating a great challenge for the efficient treatments of AD. The current treatment strategy for AD patients is the use of AChE enzyme inhibitors, which give only symptomatic relief. However, recent studies indicated a long-lasting effect in a certain percentage of patients. There are accumulating shreds of evidence that AChE and BuChE have secondary non-cholinergic functions including the processing and deposition of β-amyloid (Aβ). BuChE and AChE could play a role in the Aβ metabolism and during an early step in the development of the senile plaque, as revealed by the finding that AChE and BuChE accelerate Aβ deposition. Considering the non-classical BuChE and AChE functions, their relationships with AD hallmarks and the assumed role of the peripheral anionic site in all these functions, the dual binding site ChE inhibitors may acquire importance for the AD treatment. On the other hand, the interference of AChE inhibitors with Aβ processing is not a general rule for this class of compounds with the involvement of other features such as chemical structure and/or genetic regulation[11-12]. The recent development of highly selective BuChE inhibitors (2-Phenylbenzofuran derivatives)[11-12] will allow testing these new agents in patients with AD to find out whether they represent an advantage or not for the treatment of patients with AD as compared with selective (Donepezil) or relatively non-selective (Rivastigmine, and Galantamine) ChE inhibitors presently in use[11-12]. There is a shred of rising evidence that both AChE and BuChE may be important in the development and progression of AD[11-12]. Structural features of both enzymes suggest the differences in their substrate specificity; AChE is highly selective for ACh

neurotransmitter hydrolysis, while BuChE is capable of metabolizing several different neuroactive peptides including ACh etc[11-12]. The substrate diversity is identified by the amino acid sequences of the ChE (AChE and BuChE) enzymes that determine the 3D size and shape of their respective receptors[11-12]. In the case of AChE, the available space for ligand binding is limited by the presence of two large amino acids, phenylalanine (Phe295 and Phe297), but in BuChE, these two amino acids are replaced by two smaller amino acids, valine, and leucine, creating greater space that permits binding of various larger molecules [11-12]. The substituted amino acids must affect the size and hydrophobic nature of the active site[11-12].

1.1.6.2. Amyloid-beta (Aβ) plaques

AD is distinguished by the formation of aberrant neuritic plaques and neurofibrillary tangles in the brain [13-14]. Neuritic plaques are hemispheric minuscule plaques with an extracellular amyloid beta-peptide core surrounded by increased axonal terminals. A transmembrane protein known as an amyloid precursor protein (APP) produces beta-amyloid peptide (β-AP) as a byproduct[13-14]. In the brain, generally, the β-AP is cleaved from APP by the action of three proteases christened alpha (α), beta (β), and gamma (γ)-secretase[13-14]. APP is normally cleaved by alpha or beta-secretase, and the resulting microscopic fragments are not toxic to neurons[13-14]. Though, successive cleavage by beta and then gamma-secretase outcomes in 42 amino acid peptides called beta-amyloid-42 peptide. Raise in the levels of beta-amyloid 42 leads to the accumulation of amyloid which causes neuronal toxicity by the deposition at the synaptic site in the brain[13-14]. Over normal APP breakdown, beta-amyloid 42 promotes the formation of aggregated fibrillary amyloid protein[13-14]. The APP gene is positioned on chromosome 21, which is associated with familial AD[13-14]. In AD, amyloid deposition happens around meningeal and neural vessels, as well as a gray matter[13-14]. Multifocal gray matter deposits aggregate to become milliary structures known as plaques[13-14]. Though some people without dementia had amyloid plaques detected during brain scans, some people with dementia did not.

1.1.6.3. Beta (β)-site amyloid precursor protein cleaving enzyme 1 (BACE1 or β-secretase1)

BACE1 (β-secretase) is the first protein that acts on amyloid precursor protein (APP) in the production of amyloid-β (Aβ)[15-16]. Due to its evident rate-limiting function, BACE1 seems to be a prime target to prevent Aβ generation in AD[15-16]. The BACE1 enzyme has long been observed as an important therapeutic target for AD in the development of inhibitor drugs for the reduction of Aβ[15-16]. The cloning and identification of β-secretase were first reported in 1999 which energized research on both the protease and its inhibitor drugs. Presently, β-secretase is a major drug target for AD, and the development of its inhibitor drugs is being pursued in many research laboratories around the world[15-16]. Heparan sulphite and its derivatives were reported as BACE1 enzyme inhibitors [15-16]. Furthermore, other oligosaccharides and their analogs have also been reported to inhibit the BACE1 enzyme and reduce the Aβ deposition[15-16]. Enoxaparin has been reported to lower Aβ plaque deposition and recover cognitive function in AD-transgenic mice[15-16].

1.1.6.4. 5-hydroxytryptamine receptor

Serotonin (5-hydroxytryptamine, 5-HT), a monoamine neurotransmitter, has been found to have an important function in the CNS in regulating cognitive behavior, sensory and affective processes, autonomic responses, and motor activity[17-18]. The behavioral elements of AD development are influenced by serotonergic neuron degeneration and 5-HT neurotransmitter hypofunction[17-18]. In particular, 5-HT concentration was shown to be significantly lower in the hippocampus area of affected brains. 5-HT receptors have been classified into seven families based on transducer processes, spanning from 5-HT1 to 5-HT7[17-18]. The pathophysiology of AD has been associated with these kinds of

serotonergic receptors, including 5-HT6 receptors, which are implicated in learning and memory. This is corroborated by the finding that the 5-HT6 receptor gene's single nucleotide polymorphism C267T is a risk factor in the genesis of AD. Notably, apolipoprotein E epsilon 4, a crucial factor in the development of AD, is not dependent on the genetic polymorphism C267T, which is involved in the late onset of AD[17-18]. The function of several neurotransmitters, including glutamate and acetylcholine, which are crucial for memory and learning, is regulated (more precisely, down-regulated) by 5-HT6 receptors[17-18]. 5-HT3 receptors have been discovered in the hippocampus[17-18]. Serotonin reduces cholinergic tone via interacting with 5-HT3 receptors located in the hippocampus[17-18]. The reduced cholinergic tone during AD may also contribute to the course of the disease[17-18]. Thus, 5-HT3 receptor antagonism may be advantageous in Alzheimer's patients. Furthermore, stimulation of 5-HT4 receptors has been shown to enhance acetylcholine release simultaneously decreasing Aβ toxicity[17-18]. However, 5-HT1, 5-HT2, and 5-HT7 receptor dysfunction may potentially be involved in the etiology of AD[17-18].

1.1.6.5. Gamma-secretase enzyme

One of the key pathogenic hallmarks of AD is the formation of amyloid-beta (Aβ) peptide, which is mediated by the crucial enzyme named gamma-secretase (γ-secretase)[19-20]. Gamma-Secretase is a protease enzyme that slices the transmembrane domain of the amyloid protein precursor (APP) to produce the amyloid β-peptide (Aβ), an aggregation-prone product that accumulates in the brain throughout AD[19-20]. As data from previous studies revealed that Aβ is a crucial characteristic responsible for Alzheimer's pathogenesis, γ-secretase is regarded as an important target for the development of disease-modifying therapies[19-20]. Peptide aldehyde type calpain and proteasome inhibitors were the first reported γ-secretase inhibitors (GSIs) to be disclosed, as γ-secretase has proven to be a promising target for the prospective treatment of AD[19-20]. Although current research indicates that gamma-secretase plays crucial functions in cellular signaling, medications that control the formation of Aβ by reducing gamma-secretase activity may be an effective treatment for AD[19-20].

1.1.6.6. Glutaminyl Cyclase enzyme (QC)

Glutaminyl cyclase (QC) is one of the zinc-dependent aminoacyltransferase enzymes which transforms the N-terminal glutamate residue of A3-40/42 into the analogous pyroglutamate state (pE-A3-40/42) through intramolecular cyclization[21-22]. Since QC is more abundantly generated in the brains and cerebrospinal fluid (CSF) of AD patients compared to normal brains, it is possible to detect QC in the early stages of AD. QC inhibitors effectively decreased the levels of pE-Aβ and Aβ plaques in the brain, thereby reestablishing cognitive function in an AD mouse model and improving memory problems in AD mice[21-22]. In the phase 2 clinical trial, varoglutamstat is the first small molecule QC inhibitor[21-22]. According to phase 2a investigations, the varoglutamstat therapy group exhibited considerably better working memory, lower levels of synaptotoxicity and neurogranin, as well as improvements in several other scientific biomarkers[21-22]. Another powerful QC inhibitor (diphenyl-conjugated imidazole) shows good blood-brain barrier (BBB) penetrability and enhanced activities in AD model mice[21-22]. Furthermore, N-Methyltriazole-based inhibitor was identified by the pharmacophore and model-based on in silico approaches and in vitro screening with improved cognitive behavior in an animal model[21-22]. These findings support the use of QC inhibition as a feasible disease-modifying treatment for Alzheimer's disease.

1.1.6.7. Phosphodiester enzymes (PDEs)

The primary function of the broad family of enzymes known as PDEs is to hydrolyze the 3′-phosphodiester link in cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP), which are involved in signal-transduction pathways[23-24]. Due to their ability to modify cAMP and cGMP levels, they play a significant role in the pathways that are crucial for many pharmacological processes, including cell function[23-24]. The hippocampus-expressed cAMP response element binding protein (CREB) is regulated by cAMP/cGMP and promotes neuronal survival via pathways related to synapse strengthening and synaptic plasticity[23-24]. CREB positively regulates memory consolidation and performance by upregulating brain-derived neurotrophic factor (BDNF)[23-24]. Acute neuroinflammation has been shown to impact learning and memory-related CREB signaling in the hippocampus of mice via tumor necrosis factor (TNF)-dependent processes [23-24]. Furthermore, Aβ reduces BDNF levels via a mechanism involving transcription factor CREB downregulation. It is widely understood that most of the drugs that interfere with neurodegenerative enhancement, including cognition enhancers, target a specific neurotransmitter[23-24]. PDE inhibition contributes to neurodegeneration by enhancing cGMP and/or cAMP intracellular availability[23-24]. Although PDEs are extensively expressed in the human brain, cAMP and cGMP levels can influence neurodegenerative processes[23-24]. From a broad perspective, PDE-inhibitors could enhance cognition by controlling neural transmission by altering presynaptic neurotransmitter release and postsynaptic intracellular processes after extracellular neurotransmitter interaction[23-24]. As a result, researchers have found the PDE family to be an appealing multipotential target for numerous disease pathologies. PDEs are a large group of enzymes comprised of 11 isoenzyme families (PDE1 to PDE11), which differ from one another in terms of substrate specificity and affinities, kinetic characteristics, tissue and subcellular distributions, regulatory mechanisms, and drug and modulator susceptibility[23-24]. These isoforms catalyze the hydrolysis of cAMP and cGMP, which are involved in the proliferation, differentiation, apoptosis, gene expression, visual transduction, inflammation, and metabolic pathways[23-24]. These isoenzyme families have been discovered to have above 40 PDE isoforms (encoded by 21 genes), with some specific to cAMP (i.e. PDE4, PDE7, and PDE8) and others unique to cGMP (i.e. PDE5, PDE6, and PDE9), while others can operate on both cAMP and cGMP (i.e. PDE1, PDE2, PDE3, and PDE11)[23-24]. The presence of different isoforms in PDE families was discovered in experimental animals in the 1970s[23-24]. As a result of the abundance of isoforms and subtypes in this enzyme family, isoform/subtype-specific inhibitors should be created, which is a difficult task in drug research[23-24]. However, a significant effort has recently been made to develop novel PDEIs as subtype-selective medicines addressing a variety of disorders using experimental and computational approaches[23-24]. Furthermore, phosphodiesterase inhibitors (PDEIs) have a beneficial impact on cognition improvement through information processing, memory, memory, and executive functioning[23-24]. There are few clinical studies on the effects of PDE-inhibitors on cognitive function in AD, and the majority of the available data was obtained from preclinical animal models of AD, such as transgenic mice or central Aβ insertion[23-24]. Clinical experiments revealed that hippocampus mRNA expression of PDE4D and 8B was altered in age-related memory-impaired participants as well as in patients with mild to moderate AD[23-24]. The findings of behavioral preclinical and clinical studies revealed that PDE7 improved memory function in AD patients and altered PDE7A mRNA expression in the AD brain[23-24]. Heckman et al. 2017[25] proposed that inhibiting PDE4D isoform subtypes is an appropriate target for AD treatments. As a potential new target for AD therapy, PDEIs may prove useful. More than ten medications have been given market approval thus far. The most successful PDE5 inhibitor, sildenafil (Viagra), is used to treat ED, demonstrating the therapeutic potential of PDE targeting[23-24]. Furthermore, the PDE4 inhibitor Rolipram has been shown to improve cognitive performance in AD mice, which substantially supports

and promotes the development of PDE inhibitors for the treatment of AD. In addition to pre-clinical investigations, several PDEIs are currently in clinical trials for cognitive enhancement, particularly related to AD, as stated in **Table 1.2**.

Table 1.2. List of the PDEs inhibitors in clinical trials (completed/ recruited/terminated) for the treatment of AD or associated diseases.

Clinical Trial Identifier	Drug	Target	Disease/condition	Phase	Completion Date/year
NCT04854811	Roflumilast	PDE4	Memory and Functional Recovery	II (Recruiting)	April 1, 2024
NCT04658654	Roflumilast	PDE4	Mild Dementia Patients	II (Recruiting)	October 1, 2023
NCT05297201	CPL500036	PDE10A	Parkinson Disease	II (Recruiting)	March 1, 2023
NCT01429740	PF-0999	PDE2	Schizophrenia	I (completed)	2011
NCT01530529	PF-05180999	PDE2	Healthy volunteers	I (completed)	2012
NCT02584569	TAK-915	PDE2	Healthy volunteers	I (completed)	2015
NCT01409564	Cilostazol	PDE3	Alzheimer's disease	IV (completed)	2011
NCT02491268	Cilostazol	PDE3	Mild cognitive impairment	II (Recruiting)	2015
NCT01433666	Roflumilast	PDE4	Dementia	II (completed)	2011
NCT02051335	Roflumilast	PDE4	Memory impairment, Alzheimer's disease	I (completed)	2014
NCT02835716	Roflumilast	PDE4	Alzheimer disease	Preclinical (Recruiting)	2016
NCT02013310	HT-0712	PDE4	Age-associated memory impairment	II (completed)	2013
NCT00880412	Etazolate	PDE4	Alzheimer's disease	II (completed)	2009
NCT03030105	BPN14770	PDE4	Alzheimer's disease	I (not recruiting participants)	2017
NCT02840279	BPN14770	PDE4	Alzheimer's disease	I (completed)	2016
NCT02648672	BPN14770	PDE4	Alzheimer's disease	I (completed)	2016
NCT00455715	Sildenafil	PDE5	Schizophrenia	IV (completed)	2007
NCT01941732	Sildenafil	PDE5	Parkinson's disease	IV (completed)	2013
NCT02450253	Tadalafil	PDE5	Dementia, vascular	II (recruiting)	2017
NCT00930059	PF-04447943	PDE9	Alzheimer's disease	II (completed)	2009

NCT00988598	PF-04447943	PDE9	Alzheimer's disease	I (completed)	2009
NCT01097876	PF-04447943	PDE9	Healthy	I (completed)	2010
NCT02240693	BI-409306	PDE9	Alzheimer's disease	II (recruiting)	2014
NCT02337907	BI-409306	PDE9	Alzheimer's disease	II (recruiting)	2014
NCT02197130	PF-02545920	PDE10	Huntington's disease	II (completed)	2014
NCT02037074	EVP-6308	PDE10	Schizophrenia	I (completed)	2014
NCT01900522	ITI-214	PDE1	Schizophrenia	I (terminated)	2013
NCT00362024	MK-0952	PDE4	Alzheimer's disease	II (terminated)	2006
NCT01215552	HT-0712	PDE4	Healthy elderly volunteers	I (terminated)	2010
NCT02162979	Sildenafil	PDE5	Parkinson's disease	II (terminated)	2007
NCT02342548	PF-02545920	PDE10	Huntington's disease	II (terminated)	2015
NCT02477020	TAK-063	PDE10	Schizophrenia	II (terminated)	2015
NCT01952132	OMS643762	PDE10	Schizophrenia	II (terminated)	2013

**Agents in clinical trials for the treatment of AD and related diseases in 2022 (from clinicaltrials.gov accessed on 05/09/2022)

1.1.6.8. N-methyl-D-aspartate (NMDA) receptor

NMDA receptor-mediated excitotoxicity plays an imperative part in the advancement of AD[26-27]. Excess glutamate-mediated NMDA receptor over-activation appears to result in the formation of amyloid plaques, which leads to neuronal death[26-27]. NMDA receptors become over-activated after Aβ aggregation and NFT formation in AD, resulting in Ca^{2+} overflow into the cytoplasm[26-27]. The vital enzyme CREB (cyclic AMP response element binding protein) is activated by Ca^{2+} influx, which leads to mitochondrial dysfunction and signal suppression that lowers phospho-CREB levels[26-27]. The synthesis of pro-survival molecules like BDNF (brain-derived neurotrophic factor) is decreased when phospho-CREB is downregulated, making cells more susceptible to oxidative stress-induced cellular malfunction and neuronal death[26-27]. All of the aforementioned activities are synergistically mediated by nitric oxide, which stimulates the Aβ protein to produce more glutamate[26-27]. Overexpression of Aβ reduces glutamate absorption by glial cells, which increases glutamatergic excitotoxicity[26-27]. Memantine, a non-competitive NMDA receptor antagonist approved for moderate to severe AD treatment in the United States and Europe under the brand name Namenda, may have the potential to mitigate other neurological disorders such as vascular dementia (VD) and Parkinson's disease (PD)[26-27]. Memantine has been shown to have favorable effects on vascular and neurodegenerative processes in several animal models of neurodegenerative disease. Memantine blocks the NMDA glutamate receptors to restore the glutamatergic system and improve cognitive and memory deficiencies[26-27]. Despite the relatively minor nature of its side effects, Memantine has been demonstrated to produce only a slight decrease in clinical deterioration in AD and VD, and hence efforts are being made to design new and more potent Memantine-based therapeutics to potentially provide higher efficacy.

1.1.6.9. Cyclin-dependent kinases 5 (CDK5)

The set of enzymes known as cyclin-dependant kinases (CDKs) regulates the cell cycle's activities and takes the role of apoptosis[28-29]. Although managing the cell cycle is the primary biological role of CDKs, some CDKs, including CDK5, are also involved in governing cell differentiation in neural cells[28-29]. CDK5 is expressed in post-mitotic cells of the central nervous system and it plays a vital role during neuronal differentiation[28-29]. It is a unique member of the family since it requires association with a p35 activator for activation instead of cyclins[28-29]. Since p35 is a neuronal-specific activator, CDK5 activity is only found in neurons. Deregulation of CDK5 activity has been linked to several neurodegenerative disorders, including Alzheimer's[28-29]. The calcium-dependent cysteine protease calpain cleaves the protein p35 into the active fragment p25, which is then shown to accumulate in the brains of AD patients, which causes this dysregulation[28-29]. Despite the lack of a clear mechanism, it is thought that CDK5/p25 has considerably more catalytic activity than CDK5/p35 and that the hyperactive CDK5/p25 complex causes tau proteins to be phosphorylated excessively, which leads to the formation of neurofibrillary tangles, which are thought to be the cause of neuronal cell toxicity[28-29]. Tau is a microtubule-associated protein that helps to stabilize microtubules, which are crucial for nutrition transport within neurons[28-29]. The hyperphosphorylation of tau proteins by hyperactive kinases such as CDK5/p25 promotes self-aggregation, which leads to microtubule disintegration and, finally, the collapse of the neuron's transport system[28-29] (showed in **Figure 1.4**). Therefore, targeting such kinases is a feasible option for reversing aberrant tau hyperphosphorylation.

1.1.6.10. Glycogen synthase kinase 3β (GSK-3β)

Glycogen synthase kinase 3 (GSK-3β) is a proline-directed serine/threonine protein kinase that has also been linked to tau pathology in AD[30-31]. GSK-3β plays a role in apoptosis, gene expression, and cell architectural maintenance[30-31]. It associates with microtubules and is highly expressed in healthy brain tissue[30-31]. GSK-3β is thought to be important in the aberrant hyperphosphorylation of tau and neurodegeneration (illustrated in **Figure 4**) in AD. It is unclear how GSK-3β functions exactly at the molecular level in AD[30-31]. Recently, Jin et al.[32] proposed that truncation of GSK-3β by Ca^{2+}/calpain-I may contribute to tau hyperphosphorylation and neurofibrillary degeneration in AD.

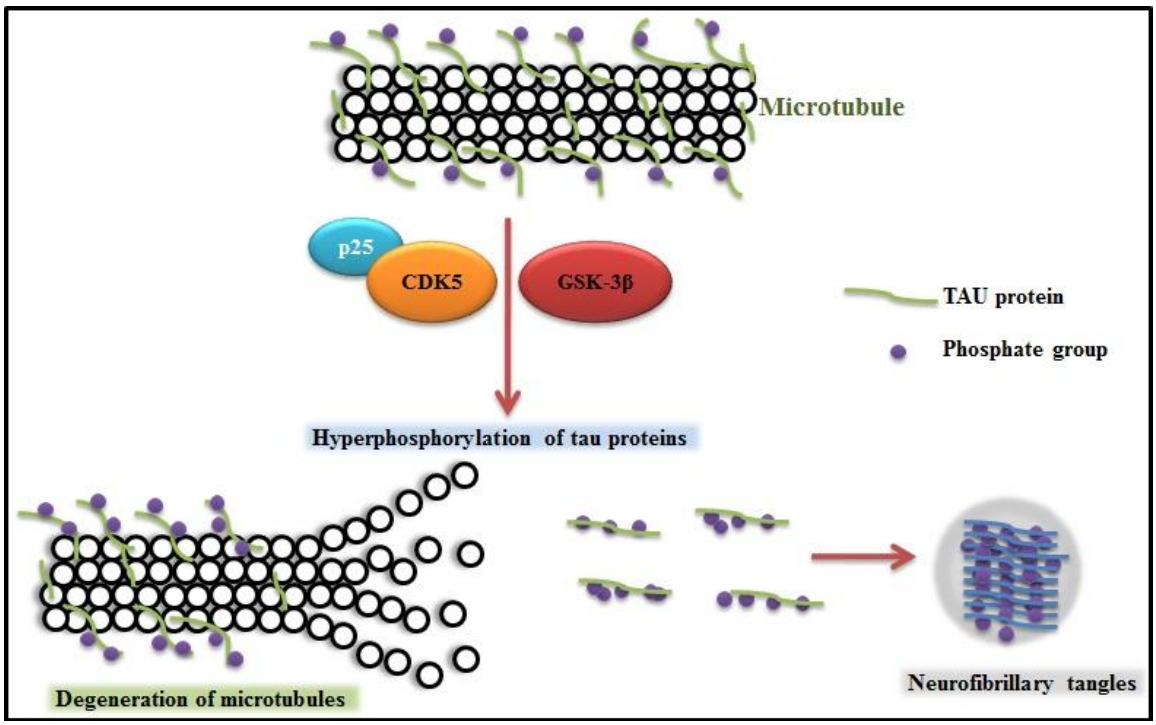


Figure 1.4. Hyperphosphorylation of tau proteins and the formation of neurofibrillary tangles.

1.1.6.11. Monoamine oxidase B (MAO-B)

Different evidence suggests that specific MAO-B inhibition is significant in the treatment of AD and Parkinson's disease[33-35]. In addition to playing a crucial role in the metabolism of neuroactive and vasoactive amines in the peripheral tissues and central nervous system, MAO-B is a protein that is an essential component of the outer mitochondrial membrane, which catalyzes the oxidative deamination of biogenic and xenobiotic amines[33-35]. MAO-B is also responsible for the generation of reactive oxygen species which directly harms nerve cells[33-35]. MAO-B-produced hydrogen peroxide can be transformed into highly hazardous hydroxyl radicals via the Fenton reaction when it comes into contact with iron[33-35]. Additionally, these hydroxyl radicals can release methylene hydrogens from polyunsaturated fats in brain membrane phospholipids, resulting in the initiation of lipid peroxidation and cell death[33-35]. Furthermore, it has been identified that MAO-B levels are rising with age, implying a role in natural aging-related cognitive decline and the possibility of its participation in the development of neurological illnesses such as Alzheimer's[33-35].

1.1.7. Current approaches to design multi-target directed ligands (MTDLs) in AD

1.1.7.1. Dual binding site: 'AChEIs' targeting Aβ aggregation

AChE has been identified as a potential target in recent studies, even though AChE inhibitors (AChEIs) may not be a viable treatment for AD [36]. Under the MTDLs design strategy, the majority of research is currently being directed toward the development of AChEIs with "dual binding sites"[37]. These AChEIs have been proven to dramatically boost cognitive performance by concurrently blocking the catalytic site and peripheral anionic sites (PAS), which are both active sites for the AChE enzyme[38]. It may raise a question on how a drug with two binding sites may be referred to as a multi-target inhibitor[39]. The rationale is that inhibitors with substantial binding at both the PAS and catalytic triad

sites are effective against AChE-induced Aβ aggregation[40]. Accordingly, AChE not only metabolizes ACh but also contributes to the development of Aβ plaque. This discovery sparked an interest in developing hybrid compounds that inhibit both the cholinergic effect and AChE-induced Aβ aggregation at the same time[41]. Numerous in vitro and in silico studies[37-44] have been carried out, and it has been revealed that a number of possible AChE dual-binding site inhibitors exist [42]. Munoz-Ruiz et al. designed a series of indole-tacrine heterodimers as strong dual-binding site inhibitors by combining the indole ring (PAS binding unit) and the 1,2,3,4-tetrahydroacridine fused ring or the tacrine moiety (catalytic triad site binding unit)[43]. More examples of the investigated scaffolds as dual binding site AChE inhibitors include derivatives of 2-(aminoalkyl)-isoindoline-1, 3-dione, 6,7-dimethoxy-2H-2-chromenone, 6,7-dimethoxycoumarin, and many more[44].

1.1.7.2. Dual-target directed ligands (DTDLs) designing

Dual-target-directed ligands (DTDLs) are capable of concurrently binding to two targets, and their development strategy typically entails the generation of hybrid molecules[45]. In most cases, important substructures from two distinct compounds that are acting on two different targets are combined to generate hybrid molecules[46]. These compounds are crucial for their respective activities. Novel Tacrine-Melatonin, Tacrine-Ferulic Acid, Tacrine-8-hydroxyquinoline, and Tacrine-Lipocrine are just a few examples of the hybrid compounds designed to combine AChE inhibition and anti-oxidant effects in a single molecule[45-52]. More research has been done to discover the chemical building blocks that were used to design MAO-B-AChE dual inhibitors, such as carbamate derivatives of hydroxy amino-indans and phenethylamines, coumarin derivatives, N-substituted pyrazoline derivatives, etc[47]. Additionally, research has been done on the synergistic effects of combining the inhibitory characteristics of BACE1 and AChE in a single molecule[48-50]. The dual inhibition of BACE1-AChE was also discovered to be a property of coumarin derivatives and diterpenoids from *Aralia cordata*[51-52].

1.1.7.3. Multi-target-directed ligands (MTDLs) designing

Multi-target-directed ligands (MTDLs), a novel trend in drug design and discovery, have emerged since the year 2000[53-54]. Such a strategy looks to be especially useful in the treatment of complex disorders like AD. Caproctamine, one of the earliest examples of purposefully developed MTD, demonstrated synergistic cholinergic activity against AD by antagonizing presynaptic muscarinic acetylcholine M2 autoreceptors and inhibiting the AChE enzyme[53-54]. This technique has been gradually used in AD drug discovery over the last two decades, and many papers have highlighted the advantages of the MTD approach over classical target-specific medicines (TSMs) and their combinations, while rarely emphasizing its limitations and shortcomings[53-54]. But so far, this strategy has fallen short of expectations and has only produced two clinical candidates (ladostigil and NP-61), both of which were unsuccessful in clinical trials[53-54]. Due to the consistently disappointing findings of clinical studies, novel candidate medications generally seem to be a long way from being approved (**Table 1.3**). The evaluation of clinical trials needs to be changed, thus there is a big demand for it. It is also recommended that continuous outcome measurements and discrete clinical state ratings be integrated to boost statistical power in clinical trials relative to any single endpoint, in addition to a more exact allocation of patients to trial groups. In **Figure 1.5**, we have highlighted nine major targets linked with AD, which are AChE, β-amyloid aggregation, BACE-1, GSK-3β, MAOs, metal ions in the brain, NMDA receptor, 5-hydroxytryptamine (5-HT) receptors, the third subtype of histamine receptor (the H3 receptor), and phosphodiesterase (PDEs). Additionally, eleven multi-target design strategies have been categorized by the involvement of AChE (AChE and BACE-1, AChE and GSK-3β, AChE and MAOs, AChE and

metal ions, AChE and NMDA receptor, AChE and 5-HT receptors, AChE and H3 receptor, AChE and PDEs) and without the involvement of AChE (BACE-1 and GSK-3β, MAO-B and metal ions, PDEs and metal ions), which were reported in recent years[53-54] for improvement of AD therapy.

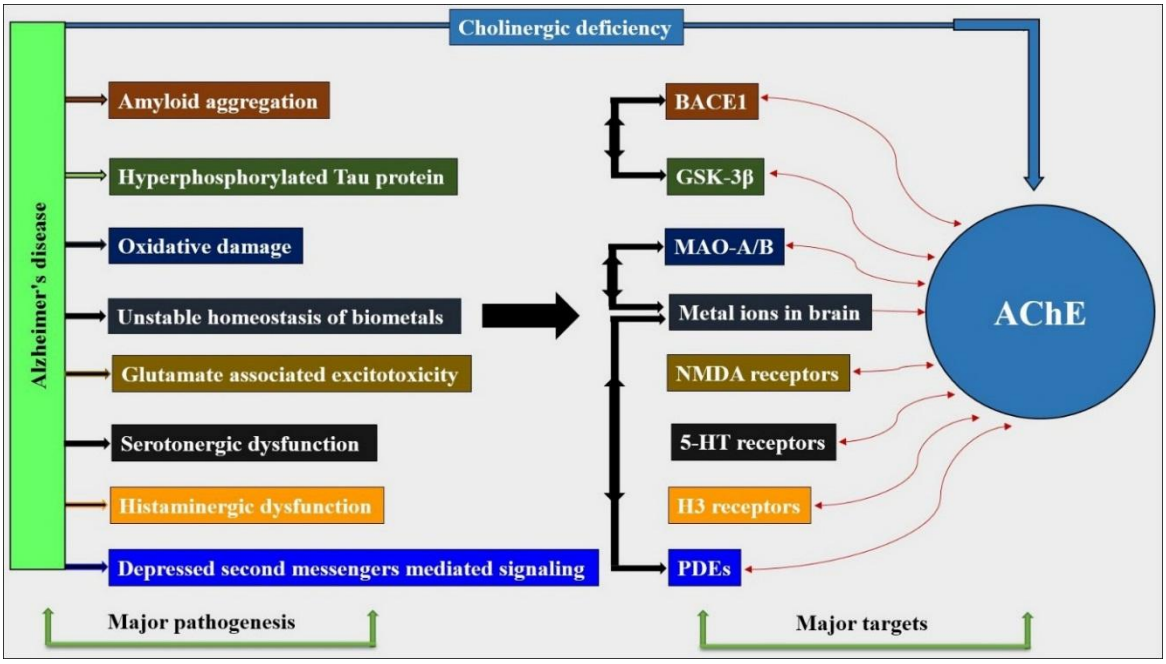


Figure 1.5. Brief connections between AD and nine major targets, and eleven multi-target design strategies based on the targets. (Note: Red arrow denotes the multi-target strategies involving AChE and the Black arrow denotes the multi-target strategies without AChE involvement).

Table 1.3. Multitarget therapies for AD treatment in clinical trials (https://clinicaltrials.gov/, Date of data collection: 06/10/2022).

Clinical Trial Identifier	Drug	Phase	Mechanism of action on targets
NCT02913664	Losartan+amlodipine+atorvastatin	III	Angiotensin II receptor blocker: losartan Calcium channel blocker: amlodipine Cholesterol agent: atorvastatin
NCT03533257	AMX0035 (sodium phenylbutyrate and tauroursodeoxycholic acid combination)	II	Chemical chaperone to inhibit endoplasmic reticulum stress responses. (Sodium phenylbutyrate) Naturally occurring bile acid to tackle mitochondrial dysfunction. (tauroursodeoxycholic acid)
NCT02547818 NCT04570644	ALZT-OP1 (Combination of cromolyn and ibuprofen)	III II	Mast cell stabilizer (cromolyn) Anti-inflammation (ibuprofen)

NCT03790709 NCT04314934 NCT02756858	ANAVEX2-73 (Blarcamesine)	III III II	Sigma-1 receptor agonist M1 receptor agonist and M2 receptor antagonist GSK-3β inhibitor
NCT03393520 NCT02446132 NCT04464564 NCT04408755	AVP786 (Combination of dextromethorphan and quinidine)	III III III III	Sigma-1 receptor agonist (dextromethorphan) NMDA receptor antagonist (dextromethorphan)
NCT03620981 NCT03594123 NCT03548584 NCT03724942	Brexpiprazole	III III III III	D2 receptor agonist 5-HT receptor agonist
NCT02008357	Gantenerumab and solanezumab	III	Monoclonal antibody directed at plaques and oligomers (Ganenerumab) Monoclonal antibody directed at monomers (Solanezumab)
NCT04063124	Dasatinib+Quercetin (Combination therapy)	II	Tyrosine kinase inhibitor (Dasatinib) Flavonoid with antioxidant and anti-Aβ fibrilization properties (Quercetin)
NCT02033941	Grapeseed Extract	II	Polyphenolic compound with antioxidant property Anti-oligomerization
NCT03062449	L-serine	II	Synthesis of sphingolipids and phosphatidylserine The precursor of D-serine, a co-agonist of NMDARs.
NCT03867253	ORY-2001 (Vafidemstat)	II	LSD1 inhibitor MAO-B inhibitor
NCT02085265	Telmisartan+Perindopril	II	Angiotensin Ⅱ receptor blocker (Telmisartan) Angiotensin-converting enzyme inhibitor (Perindopril)
NCT03748303	Allopregnanolone	I	Growth hormones to promote neurogenesis Positive allosteric GABAARs modulators

5-HT: 5-hydroxytryptamine; GABAARs: γ-aminobutyric acid type A receptors; GSK-3β: glycogen synthase kinase 3; LSD1: Lysine-specific histone demethylase 1A; MAO-B: monoamine oxidase B; NMDA: N-methyl-D-aspartate; NMDARs: N-methyl-D-aspartate.

1.2. Computer-aided drug design (CADD)

Although drug research is a time-consuming and expensive process, computational biology and bioinformatics methodologies have simplified the initial identification of potential therapeutic compounds. Such computational biology strategies have become extremely relevant, from lead identification to optimization. In this regard, computational approaches such as quantitative structure-activity relationship (QSAR), chemical Read-Across (RA), pharmacophore modeling, molecular docking, Molecular Dynamic (MD) Simulations, etc. are playing imperative roles in the design and discovery of new compounds with enhanced therapeutic activity[55-60]. For decades, chemoinformatics and molecular modeling approaches have been utilized to identify and optimize novel compounds with improved therapeutic potential in various fields[55-60]. In silico modeling is currently used in the conventional drug discovery process, and such methods are typically used in the search for novel therapeutics or the optimization of the therapeutic action of a chemical series during the early stages of drug development[55-60]. Computational methodologies have given several potential drug molecules against the predictable and promising targets against AD[55-60]. Many molecules have been discovered to be excellent lead compounds (including flavonoids, carbamates, pyridonepezil, and coumarin derivatives) that can be propagated against AD[61]. Numerous molecules have entered various stages of clinical trials, including MK-8931 (against β-secretase, Merck), TAK-070 (against β-secretase, Takeda Pharmaceuticals), and LMTX (against tau hyperphosphorylation, TauRx Inc.)[61].

1.2.1. Quantitative structure-activity relationship (QSAR)

Quantitative structure-activity relationships (QSARs) is a statistical approach to finding the consistent relationship between the biological activity (dependent variable) of compounds and their structural arrangements and chemical property (independent variables)[62]. The chemical domain space refers to the chemical information, which is derived in terms of descriptors (independent variables) utilizing various software tools. The dependent variable acquired from an experiment stands as the endpoint and is modeled using the general ***Equation 1.1*** that is given below. There are different regression and pattern recognition techniques (S-MLR, GA, etc.) that can be used for variable selection and QSAR model development[62]. The developed model for the QSAR is used particularly in chemoinformatics, drug discovery and to evaluate the biological activity of new chemical compounds, apart from these it is also used for toxicological and ecotoxicological evaluations of specific chemicals within the meaning of risk management[62].

$$Biological\ activity/toxicity\ =\ f\ (Structure, Physicochemical\ properties)$$

Equation 1.1

The modeling method may be referred to as quantitative structure-activity relationship (QSAR), quantitative structure-toxicity relationship (QSTR), quantitative structure-property relationship (QSPR), or other terms depending on the type of endpoint modeled. The term "activity" itself may often be employed as additional descriptors, as in the cases of "quantitative structure-activity-activity relationship" (QSAAR), etc. Only the QSAR aspects of modeling have been taken into consideration when writing this chapter. The selection of data (chemical/biological) is one of the most crucial elements in the accomplishment of any. Two separate types of data information, namely biological data (endpoints) and chemical data in terms of molecular descriptors, are needed for the development of QSAR models. Before using any modeling methodology for model development, the feature selection approach is then employed to choose the appropriate number of meaningful and informative descriptors. The flowchart for the development of the QSAR model is given in **Figure 1.6**.

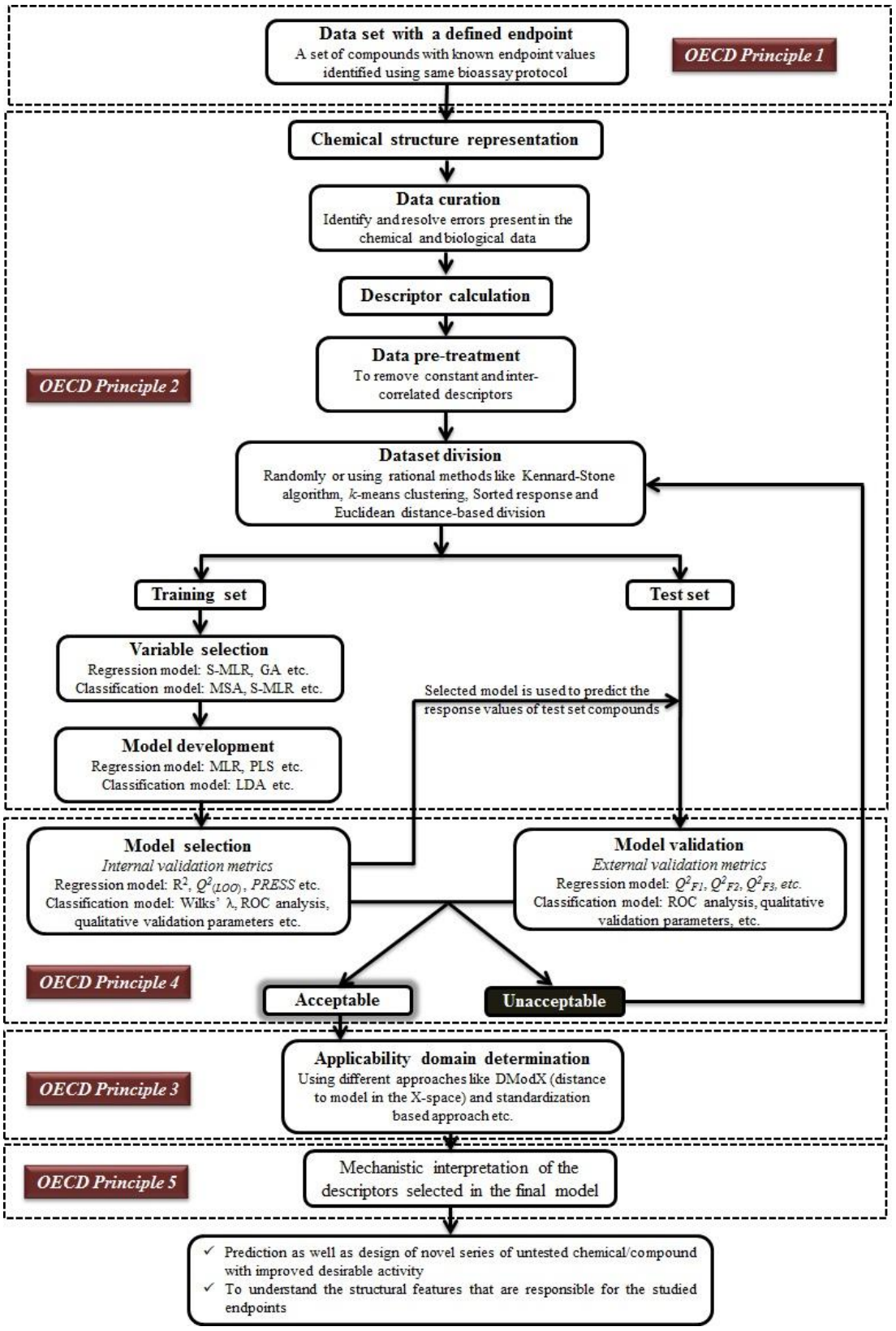


Figure 1.6. Schematic outlining the process for the development of a QSAR model.

1.2.1.1. Biological data

The biological data represent experimental endpoints, such as activity values ($IC_{50,}$ etc.). The dose-response curve yields the activity values. The concentration values IC_{50}, and EC_{50} are effective quantities that inhibit and affect 50% of the test population. The dosage values are presented in molar units and then transformed to a negative logarithmic scale, where a higher number on the positive scale corresponds to more activity, and conversely. The following points are to be noted while choosing any biological data for QSAR modeling:

1. The QSAR model should have a specified endpoint following Organization for **Economic Co-operation and Development (OECD) Guideline No. 1**[63]. The endpoint predicted by the model should be transparent. Only those compounds should be used to develop models that have defined activities. Another crucial feature is that the compounds selected to build the model must have the same mode of action.
2. The dependent variables should fall within a logarithmic scale of at least three to four units. Since the conventional metrics of validation (R^2, Q^2, and R^2pred) are response range-based parameters, effort should be made to ensure that the data points are consistent and there are no gaps[64-66].
3. All molecules should be evaluated under the same experimental conditions using the same bioassay methodology.
4. Before using the dataset, it must be properly curated. The data points containing information concerning activity cliffs should also be carefully considered[67].
5. The researcher should carefully select the data when working with small data sets. Working with limited data sets is a significant challenge for QSAR modeling. The absence of experimental observational activities, such as activities, is the cause of the lack of data points[68].

1.2.1.2. Molecular descriptors

Molecular descriptors are numerical terms that characterize specific information about a studied molecule. They are the "numerical values associated with the chemical constitution for correlation of chemical structure with various physical properties, chemical reactivity, or biological activity". Descriptors can be classified in multiple ways. In general, there are several types of descriptors like structure-explicit descriptors (topological), structure-implicit (hydrophobicity and electronic), and cryptic descriptors (quantum chemical). It is interesting to point out that the majority of QSAR researchers prefer to classify the types of descriptors concerning their dimensions. From a broader perspective, descriptors (specifically, physicochemical descriptors) can be classified into two major groups: (1) substituent constants and (2) whole molecular descriptors. Before the development of a QSAR model, it is necessary to convert the chemical structural information into numerical values. These numerical values are referred to as 'descriptors' in QSAR studies, which are connected with the chemical constitution to correlate chemical structure with various physical characteristics, chemical reactivity, or biological activity[69]. The characteristics of an ideal molecular descriptor have been summarized as follows:

1. For a given end point, a descriptor must be correlated with the structural characteristics of the chemicals and have an inconsiderable association with other descriptors.
2. A descriptor should apply to a diverse class of compounds.
3. A descriptor that can be computed easily and does not rely on investigational outcomes is preferable to one that is computationally intensive and largely reliant on experimental data.

4. Even if the structural variations are minor, a descriptor should produce dissimilar values for structurally diverse compounds. This implies that the descriptor should have a low degree of degeneracy. A descriptor should be continuous, meaning that slight structural changes should result in modest changes in the descriptor's value, along with having low degeneracy.
5. To encode the desired properties of the investigated molecules, the descriptor must be physically interpretable.
6. Another useful feature is the ability to visually represent descriptor values by mapping them directly to the chemical structure[69]. These conceptions are logical only when descriptor values can be linked with chemical structural properties.

In a QSAR analysis, a 'dimension' serves as a constraint that governs the nature of the study. In predictive model development, the term 'dimension' is approximately related to the complexity of the modeling technique which directly explains the degree of descriptors. The molecular descriptors can be categorized based on dimensions as represented in **Table 1.4**[69].

Table 1.4. Discussion on the categorization of the molecular descriptors based on dimensions with suitable examples.

Dimensions of descriptors	Discussion	Examples
0D descriptors	The class of 0D descriptors includes all molecular descriptors for which no information on molecule structure or atom connectivity is required.	Constitutional indices, molecular property, atom, and bond count.
1D descriptors	The 1D descriptors include all molecular descriptors that may be computed from sub-structural information about the molecule. These descriptors are frequently expressed as fingerprints, which are binary vectors in which 1 denotes the existence of a specified substructure and 0 indicates its absence.	Fragment counts fingerprints.
2D descriptors	The 2D descriptors are molecular descriptors based on a graph theoretical representation of the molecule. Various structural and/or physicochemical property indices are also included in this class of molecular descriptors.	Topological, structural, and physicochemical parameters including thermodynamic descriptors.
3D descriptors	3D descriptors, also known as geometrical descriptors are generated from the geometrical representation of the molecules, or the x–y–z Cartesian coordinates of the atoms.	Electronic, spatial parameters, WHIM descriptors, 3D-MoRSE descriptors, GETAWAY descriptors, etc.
4D descriptors	4D-descriptors are generated from the produced energy between molecules which are embedded into the grid and probe.	Volsurf, GRID, Raptor, etc. derived descriptors.
5D descriptors	5D descriptors take into account induced-fit parameters and aim to develop a ligand-based virtual or pseudo-receptor model.	Flexible-protein docking.

6D descriptors	6D descriptors are derived by combining information from 5D descriptors with the representation of various solvation circumstances.	Quasar.
7D descriptors	7D descriptors are derived using real receptor or target-based receptor model data.	-

1.2.1.3. Division of the dataset

A rational splitting of a dataset into a training and a test set is one of the most important steps to develop a QSAR model. Generally, a training set is used to develop a model and the test set, also known as the validation set, is used to check the predictive ability of the developed model. Since the training set is used to develop the model, it is usually given a larger number of compounds than the test set. The complete dataset is divided so that the test set compounds fall within the chemical space of the training set, i.e., the training set represents the test set. The techniques for dividing the dataset may include (1) Euclidean distance (diversity-based)[70], (2) Kennard-Stone[71], (3) k-means clustering[72], (4) sorted response[73], etc.

1.2.1.4. Feature selection

The presence of insignificant descriptors may reduce the model's robustness and interpretability. Therefore, variable selection has become one of the important aspects in developing a QSAR model in which the important descriptors are selected to build the model; conversely, the removal of the insignificant descriptors results in the improvement of the predictive ability of the model. Since it would take too long to run the computations, the entire descriptor pool cannot be used for modeling. For regression-based modeling, approaches such as stepwise selection (S-MLR)[74], Genetic algorithm (GA) method[74], Factor analysis (FA)[74], and so on should be used to select an adequate number of descriptors.

1.2.1.5. Development of QSAR model

The model should be established using a specified algorithm, and the description of the modeling algorithm should be consistent, according to **OECD Guideline No. 2**[74-76]. This includes the formalizations applied to the preprocessing of data, dataset division, feature selection, and data modeling[74-76]. The OECD recommends several commonly used linear modeling algorithms, including principal component analysis (PCA), principal component regression (PCR), ordinary least squares (OLS), multiple linear regression (MLR), and partial least squares (PLS). Using a mechanistic basis or an evolutionary method, such as a genetic algorithm (GA), as well as methods like principal component analysis (PCA) or factor analysis (FA), etc., is also suggested by the OECD guideline for undertaking a priori feature selection[74-76]. There are three basic types of model-building tools in QSAR: regression-based approaches, classification-based approaches, and machine learning[74-76]. The regression-based methodology is used when a response (endpoint) and independent (descriptors) variable values are accessible[74-76]. The classification-based strategy is for ordered response data where the response is accessible in a Boolean form such as active/inactive and positive/negative (as in the case of linear discriminant analysis, logistic regression, and cluster analysis)[74-76]. Since the machine learning method does not explicitly follow programmed instructions, it builds and develops its learning depending on available data (as in the case of an artificial neural network, Bayesian neural network, decision tree, and random forest protocol)[74-76].

1.2.1.6. Validation of QSAR models

The basic objective of the validation of QSAR models is to check the reliability of the developed models in terms of their predictive ability and reproducibility using different validation parameters[77-88]. The robustness of the developed model is judged based on the statistical quality of the model, the corresponding metrics include determination coefficient (R^2), explained variance (R_a^2), standard error of estimate (s), F-value, normality distribution test, variance inflation factor (VIF)[77-88]. The validation of the models is done using two approaches: one is based on the training set, known as internal quality assessment or validation, the corresponding metrics include cross-validated correlation coefficient (Q^2), predicted residual sum of squares (PRESS), and the other is based on the test set, known as external validation, the corresponding metrics include R^2_{pred} (Q^2F_1), Q^2F_2 and Q^2F_3, etc[77-88]. Additional validation parameters include modified r^2_m ($r_m^2{}_{(LOO)}$, $r_m^2{}_{(test))}$ and overall model predictivity ($r_m^2{}_{(overall)}$) (threshold value in all cases of different validation is = 0.5)[77-88] and Y-randomization[77-88]. The reliability of the developed model is based on the prediction ability of the model through various validation parameters for untested compounds[77-88]. One particular model may not be enough to predict the whole test set of compounds, which means one QSAR model may be the best model for the prediction of a test compound while other models may be the best predictor for another test compound[77-88]. In this regard, developed intelligent consensus predictor (ICP) tool to improve the prediction of the external test through the "intelligent" selection of multiple models[77-78]. This software judges the performance of the consensus predictions and compares them with the prediction quality obtained from the individual (MLR/PLS) models based on MAE-based criteria ($MAE_{95\%}$)[77-88]. The applicability domain (AD)[77-88], which is a theoretical region in chemical space defined by the model descriptors and modeled response, is an important criterion to check, which enables judging the reliability of the predictive performance of a model[77-88]. Other validation tests/metrics such as Golbraikh and Tropsha's criteria, concordance correlation coefficient, etc. are also extensively used for model validation purposes[77-88].

1.2.1.7. Checking domain of applicability of developed models

The applicability domain of a QSAR model has been described as the response and chemical structure space that is defined by the nature of the chemicals in the training set. If a new compound falls within the Applicability domain of the developed model, only then the developed model can predict the compound precisely. QSAR developers must have information about the applicability domain of the developed model to identify interpolation (true predictions) or extrapolation (less reliable predictions) (**as per OECD guideline no. 3**)[63]. Three key factors such as structural data, physicochemical characteristics, and response space influence a model's applicability domain. QSAR models can be developed for specific chemical classes working through the same mechanism of action due to the potential involvement of several mechanistic bases in different regulatory endpoints. It is possible for a single general QSAR model to fail to discriminate across chemical classes on occurrence, making it impossible to offer a precise estimate for a given chemical class[63]. To achieve the global applicability, the OECD recommends using (a) several predictive models against the endpoint on diverse domains of applicability shared to provide a global estimation or (b) using a statistical technique that provides global modeling attributes across multiple mechanisms of action concerning the same endpoint. By comparing the chemical domain concerning each specified regulatory endpoint, one can compare the domain of applications of the QSAR models that are currently available and identify the data gaps[63]. In our studies, we have checked the applicability domain of the developed model employing the DModX (distance to model X) approach at a 99% confidence level using SIMCA-P software[79].

1.2.1.8. Mechanistic interpretation of the QSAR model

As per the **OECD guideline 5**, the researcher should provide the mechanistic interpretation of the developed model if possible. Because it is not always possible to provide a mechanistic explanation of a particular QSAR model, the principle advises that the Modeller should publish if such knowledge is available, thus encouraging future study on that endpoint. The precise knowledge of the mechanism of action of chemicals toward a process can influence the design and development of the desired innovative analogs[80].

1.2.1.9. A few significant issues in QSAR

Several conditions must be met before experimental data can be prepared for QSAR analysis. Typically, the concentrations or dosages required for a defined response, such as IC_{50} values, are applied as the response for activity-based QSAR analysis[62-65]. The concentration values should be expressed in a molar unit and a negative logarithmic scale so that a higher value represents higher activity or toxicity[62]. To determine the statistical soundness of a QSAR model, there should be a high degree of freedom[64-65]. As a consequence, the number of observations used to develop a model should be significantly higher than the number of independent variables employed in the modeling. Even if using a suitable number of training compounds is vital even in the case of machine learning approaches, this element is less crucial for more reliable procedures[64-65]. The difficulty of modeling small datasets is a common challenge for QSAR researchers because there may not always be enough experimental observations for all endpoints. Multiple linear regression (MLR) is a popular approach for regression-based QSARs, but it has various drawbacks, including intercorrelation between descriptors, bias in descriptor selection due to a predetermined composition of the training set, inability to handle many descriptors in the model, and so on[62-65]. A more reliable modeling method, such as partial least squares (PLS), which reduces the original set of descriptors into a smaller number of latent variables (LVs) that are functions of the original descriptors, can be used to prevent this problem. Working with a dataset having a small number of data points requires special consideration. In these circumstances, a double cross-validation strategy might be useful. This approach uses two loops to carry out the validation. The training set is further divided into "n" calibration and validation sets in the inner loop, resulting in various compositions that are then utilized for model development and model selection, whilst the test set in the external loop is simply used for model assessment[62-65]. Another strategy involves using consensus predictions, which have been shown in various studies to be more accurate than predictions from individual models since the former account for the contribution of the greatest number of significant descriptors[62-65]. Additionally, this strategy can provide more chemical space coverage. An intelligent consensus modeling technique has recently been proposed in light of the possibility that not all query chemicals will respond predictably to predictions from a single QSAR model. It also is crucial to estimate the accuracy of predictions for unknown molecules, which may not be completely dependent on the applicability domain[62-65].

1.2.1.10. Applications of QSAR

In general, the chemicals modeled using the QSAR can be categorized into three main types that are as follows, 1) chemicals with health benefits, such as drugs, pharmaceuticals, food ingredients, etc., 2) chemicals involved in industrial/laboratory processes like solvents, reagents, etc., and 3) the chemicals posing hazardous outcome, such as persistent organic pollutants (POPs), volatile organic solvents, toxins, xenobiotics, carcinogens, etc[81-83]. The use of QSAR in the rational monitoring of the activity, property, and toxicity of the aforementioned chemicals makes it beneficial in a wide range of applications[81-83]. QSAR applications can be classified into three main categories drug design, materials

science, and predictive toxicity[81-83]. The principal use of QSAR modeling in the field of drug discovery and design is the screening of a large number of potential lead compounds that are active against the target enzyme[81-83]. Moreover, it is also used in lead optimization and the prediction of pharmacological pharmacokinetic profiles i.e. absorption, distribution, metabolism, and excretion (ADME) properties of pharmaceuticals[81-83]. In materials science, the QSAR technique can be used to investigate a variety of properties for different kinds of materials, including metal oxide nanoparticles, ionic liquids, polymers, fullerenes, surfactants, and many more[81-83]. In the field of toxicology, QSAR techniques are used to predict a variety of toxicity endpoints for in vitro cell cultures or in vivo animal experiments[81-83]. Chemical toxicity evaluation entails assessing systemic toxicity as well as monitoring eco-toxicological hazards. Drugs and pharmaceuticals can cause toxicity to certain organ systems, such as hepatotoxicity, nephrotoxicity, and so on, and these drugs can also cause eco-toxicity[81-83]. Additionally, QSAR has been found to be useful in agricultural sciences where chemical toxicity potential is a key component, such as fungicidal activity. QSAR is well-accepted as an alternative to animal testing as well as it is also used for regulatory purposes because of its broad applicability and reliability[81-83].

1.2.2. Other in silico methods employed

1.2.2.1. 3D QSAR Pharmacophore mapping

A pharmacophore model is an assembly of steric and electronic features essential to ensure optimal supramolecular interactions with a particular biological target and to activate or inhibit its biological response[84-86]. A pharmacophore model represents the binding patterns of bioactive molecules with the target binding site, through different 3D arrangements of conceptual interaction features accounting for different types of non-covalent interactions[84-86]. These interaction types could be hydrogen bonding interactions, hydrophobic interactions, metal interactions, aromatic contact, charge transfer interactions, etc[84-86]. Features employed by major different programs for the development of pharmacophore models are as follows: hydrogen bond donor (HBD), hydrogen bond acceptors (HBA), positive and negative charge features, hydrophobic features, ring aromatic, steric constraints features, etc[84-86]. A pharmacophore model can be developed either in a ligand-based (using a set of defined molecules) approach through superimposing a set of active molecules and extracting common chemical features that are important for their bioactivity, or in a structure-based (using the active site of the protein structure) approach, by probing possible interaction points between the macromolecular target and ligands[84-86]. The common steps involved in the development of a significant pharmacophore model are as follows: in the case of ligand-based pharmacophore model generation: selection of molecules with defined activity, conformational search, feature extraction and representation, pattern identification and scoring and validation (internal and external); in case of structure-based pharmacophore model development: active site identification, complementary image construction, generation of queries, searching and hit analysis and validation [84-86]. Pharmacophore models have been used broadly in virtual screening, de novo design, and other applications such as lead optimization and multitarget drug design[84-86]. Various automated pharmacophore generators have been developed, including commercially and free available tools and software such as Biovia Discovery studio[87], Schrodinger software[88], PharmaGIST online tool (https://bioinfo3d.cs.tau.ac.il/PharmaGist/), LigandScout[89], etc.

1.2.2.2. Molecular docking study

Molecular docking is one of the functional methods of structure-based drug discovery (SBDD) that predicts the binding affinities between small molecules and macromolecular targets[90-92]. Molecular docking is generally carried out in two major stages: first, to predict the stable conformation and orientation of the ligand, and second, to evaluate the binding affinity and binding orientation of the ligand within active sites[90-92]. The information obtained from the docking analysis can be used to suggest the binding orientation, binding energy, free energy, interaction energy, and stability of complexes for the discovery of novel compounds[90-92]. Molecular docking has a broad range of uses and applications in drug discovery, such as lead optimization, chemical mechanism studies, structure-activity studies, finding potential leads by virtual screening, assisting X-ray crystallography in the fitting of substrates and inhibitors to electron density, providing binding hypotheses to make easy predictions and combinatorial library design[90-92]. According to the degrees of flexibility of the molecules, docking methods are divided into three classes such as rigid docking, semi-flexible docking, and flexible docking[90-92]. There are four basic steps involved the molecular docking, which is as follows, 1) target selection and preparation, 2) ligand selection and preparation, 3) molecular docking, and 4) molecular docking analysis. There are some tools and online servers such as AutoDock Vina[93], Schrodinger software [88], Molegro virtual docker software 6.0 (MVD)[94], Biovia Discovery Studio[87], idock (https://github.com/HongjianLi/idock), "Achilles" Blind Docking Server (https://bio-hpc.ucam.edu/achilles/), FlexX[95] and Smina software (https://sourceforge.net/projects/smina/), which are mostly used for high throughput docking simulations.

1.2.2.3. Virtual screening

There are many publicly or commercially available databases that can be used for computational drug discovery applications since high throughput technologies for biological screening and compound synthesis have recently become available. Virtual screening techniques are increasingly being recognized as the most cost-effective and time-saving approach to introducing new chemical entities into the pharmaceutical market to address the economic pressure on the pharma industry. A ligand-based method[96] and a receptor-based approach[96] can both be used as the framework for virtual chemical database screening. Pharmacophore mapping is a ligand-based virtual screening technique that can be used to efficiently identify novel potential lead compounds. A pharmacophore model identifies the key chemical properties underlying the bioactivities of the compounds under investigation. Furthermore, in receptor-based virtual screening, molecular docking is used to identify compounds based on their binding energy and interaction patterns with the target protein. Additionally, one can perform virtual screening simply based on some criteria, such as Lipinski's rule of 5, central nervous system (CNS) drug-like properties, or any other user-defined physicochemical properties.

1.3. Review: QSAR studies performed on anti-Alzheimer's compounds

Since 1996, QSAR analyses of anti-AD drugs have been published. Before 2000, almost all QSAR investigations were conducted against the AChE enzyme, with only a few reports on muscarinic or nicotinic agonists. This could be because the FDA approved two AChE inhibitors, tacrine in 1993 and donepezil in 1996 (**Figure 1.3 and Table 1.1**), where donepezil was discovered through chemical modeling and QSAR research. This might have encouraged numerous researchers to conduct QSAR on AChE inhibitors. The discovery of additional AD complexity led to the identification of numerous other related targets. **Table 1.5** contains the representative set of QSAR findings[97-130] that have been examined against numerous AD targets. As previously stated, the first QSAR investigations were only reported against targets based on the cholinergic hypothesis. But since the last ten years, many new

targets have been identified and are currently being investigated. Most of the QSAR investigations are focused on targets with cholinergic (AChE), amyloid (β and γ secretase), and tau hyperphosphorylation (GSK3β, CDK5) bases. Other targets such as the 20S proteasome, prolyl oligopeptidase, glutaminyl cyclase, histamine H_3 receptor, monoamine oxidase B, caspase 3, 5-HT_6R, Dyrk1A, etc. have also been subjected to QSAR analysis, although the number of studies on these targets is limited, maybe because there is insufficient information or it is unclear how these targets specifically contribute to AD. The appropriate diagnosis and comprehensive treatment of AD are now the two most important concerns that must be overcome in the near future, and the QSAR technique has enormous promise in doing so. As previously stated, early detection is very important since brain deterioration occurs before clinical symptoms manifest. It is advised to conduct more QSAR studies to investigate biomarkers or imaging agents since there have only been a few QSAR studies on biomarkers reported to date. Additionally, the QSAR method can be effectively used to analyze selectivity problems like those seen with protein kinases. It assists in understanding the structural requirements for achieving or improving activity against the desired target enzyme as well as selectivity over undesirable enzymes with comparable active sites. Additionally, it has been noted that the majority of QSAR investigations have been for a single biological target, even though AD is a complex disease that involves the concurrent dysfunction of multiple biological targets. Consequently, one of the main objectives is the rational design of novel leads as versatile inhibitors for various targets related to AD.

Table 1.5. QSAR studies performed on anti-Alzheimer's compounds since 1996.

Author name and year of publication	QSAR method performed	Chemical scaffold	Name of the Target	Importance of the study
Based on the cholinergic hypothesis				
Tong *et al.* (1996)[97]	CoMFA	1-benzyl-4-[2-(*N*-benzoylamino) ethyl]piperidine derivatives; *N*-benzylpiperidine benzisoxazoles derivatives	AChE	The CoMFA model revealed a close relationship between the activity of these N-benzylpiperidines and the steric and electronic variables that influence their biological activity.
Recanatini *et al.* (1997)[98]	2D-QSAR	Physostigmine analogues; 1,2,3,4 tetrahydroacridines; Benzylamine analogue	AChE	(1) Hydrophobicity is important in both the physostigmine and benzylamine-derived groups. (2) Electronic effects are key in the interactions carried out by the variable part of benzylamine analogs.
Kaur *et al.* (2000)[99]	2D-QSAR	Derivatives of physostigmine, Tacrine, donepezil, huperzine A etc.	AChE	The presence of log P in the majority of QSAR models revealed that all derivatives were hydrophobic.
Nicolotti *et al.* (2004)[100]	2D-QSAR CoMFA MoQSAR	300 nicotinic agonists from diverse chemical classes	Nicotinic receptor (nAChR)	The descriptors log P, MR, and low inter-correlated WHIM indices were used to describe highly active molecules.
Shen *et al.* (2007)[101]	CoMFA CoMSIA	2-substituted 1-indanone derivatives	AChE	The contour map revealed that the binding affinity might be increased by replacing the small protonated nitrogen moiety with a more hydrophobic and bulky group with a highly partial positive charge.
Solomon *et al.* (2009)[102]	2D-QSAR	*N*-aryl derivatives	AChE; BChE	Thermodynamic descriptors and charge descriptors are key contributors to potency, according to developed models.

Gupta *et al.* (2011)[103]	2D-QSAR	Dual binding inhibitors *via* high-throughput *in vitro* screening of a library consisting of 56,000 compounds	AChE	The potency is largely influenced by electrotopological, two thermodynamic, and an electronic descriptor.
Decembrino de Souza *et al.* (2012)[104]	HQSAR	4-(diethylamino)methyl]-phenol derivatives	AChE; BChE	The significance of numerous structural fragments to the activity levels of this group.
Kumar et al. (2020)[105]	2D-QSAR	Diverse scaffold	BuChE	Hydrophobic, ring aromatic, and hydrogen bond acceptors/donors responsible for the enhancement of the activity were identified.
Kumar et al. (2020)[106]	2D-QSAR GQSAR	Structurally Diverse Carbamates	AChE	Structural features appearing in the models are responsible for the enhancement of the inhibitory activity against the AChE enzyme
Based on the amyloid hypothesis				
Keerti *et al.* (2005)[107]	2D-QSAR	Benzodiazepine derivatives	γ-secretase	A high lipophilicity and low molar refractivity of a compound allows for the maintenance of robust activity.
Al-Nadaf *et al.* (2010)[108]	2D-QSAR	129 compounds with diverse scaffolds	β-secretase or BACE1	This research resulted in the identification and synthesis of novel pyridinium-based compounds with low μM BACE inhibitors.
Meek *et al.* (2012)[109]	Binary QSAR	3-hydroxyanthranilic acid derivatives	Aβ aggregation	Whether a substance will be active or inactive can be predicted using a binary QSAR model.
Valasani *et al.* (2013)[110]	2D-QSAR	Frentizole, benzothiazole-urea derivatives	ABAD	Novel small compounds with improved BBB crossing capacity, such as benzothiazole phosphonate and frentizole phosphonate derivatives, were developed.
Kumar et al (2019)[111]	2D-QSAR	Diverse scaffold	BACE1	Heteroatoms (nitrogen, oxygen, etc.) present within an aromatic nucleus and the structural features such as hydrophobic, ring aromatic, and hydrogen bond acceptor/donor are responsible for the enhancement of the BACE1 enzyme inhibitory activity

Kumar et al. (2020)[112]	2D-QSAR	Diverse scaffold	Aβ aggregation	Features obtained from the developed models (2D-QSAR and 3D-pharmacophore) and molecular docking can be helpful for the design of novel inhibitors against β-amyloid aggregation.
Based on the tau hyperphosphorylation hypothesis				
Martinez *et al.* (2005)[113]	CoMFA	2,4-disubstituted thiadiazolidinones (TDZDs)	GSK-3	The interaction of TDZDs with GSK-3β was revealed to be dependent on molecular electrostatic field interaction.
Lather *et al.* (2008)[114]	2D-QSAR, 3D-QSAR	Indirubin derivatives	GSK-3β	The hydrophobic groups could be positioned at specific locations on the phenyl ring of the indirubin-like molecules to increase their bioactivity.
Park *et al.* (2010)[115]	3D-QSAR	neutral phenylthiazolylhydrazide (PTH) derivatives	Tau aggregation-n inhibitor	The PTH structure is shaped like a tweezer, and the biological activity of the PTH depends on the relative orientation of the two aromatic rings linked to both ends.
Fang *et al.* (2011)[116]	3D-QSAR	benzofuran-3-yl-(indol-3-yl) maleimides derivatives	GSK-3β	A new lead identification procedure was suggested.
Haq *et al.* (2011)[117]	CoMFA, CoMSIA	Thienyl triazoles derivatives	CDK5/p25	The steric complementarity of the area close to the phenyl ring linked with the thiadiazine ring could be fine-tuned to increase the selectivity of the ligands.
QSAR studies of imaging agents utilized in the diagnosis				
Wang *et al.* (2005)[118]	2D-QSAR	Benzothiazole (BTA) derivatives	Aβ plaques	Due to its smallest dipole moment, 3'-^{125}I-BTA had the highest initial brain uptake among the four Aβ probes.
Kim *et al.* (2007)[119]	CoMFA CoMSIA	ThioT analogs; stilbene derivatives	Aβ plaques	It has been found that positive electrostatic interaction and positively charged or electron-donating substituents improve the binding affinity of ThioT derivatives around the sulfur atom of the benzothiazole ring system.
Kovac *et al.* (2010)[120]	3D-QSAR	Vesamicol derivatives; benzovesamicol derivatives	VAChT	Further research should be done on (+)-(S, S)-5-FBVM as a possible radio ligand.

Cisek *et al.* (2012)[121]	2D-QSAR	Benzothiazole derivatives	Aβ plaques	Binding selectivity can be modulated by polarizability and hydrophobicity.
QSAR studies on miscellaneous targets				
Zhu *et al.* (2006)[122]	CoMFA, CoMSIA	Tripeptide aldehyde inhibitors	20S proteasome	The developed 3D-QSAR models can be utilized to guide the development of new 20S proteasome inhibitors.
Pripp (2006)[123]	2D-QSAR	POP inhibitory peptides derived from β-casein	Prolyl oligopeptidase (POP)	This research is important for the advancement of functional foods as a supplement to pharmacological drugs.
Dastmalchi *et al.* (2008)[124]	2D-QSAR	Arylbenzofuran derivatives	Histamine H_3 receptors	The role of charge transfer interactions in the ligand-receptor interaction was suggested by developed QSAR models.
Firoozpour *et al.* (2012)[125]	2D-QSAR	Diverse scaffolds	Caspase-3	The identified independent variables influencing the caspase-3 inhibitory activity were the atom-centered fragment type CR2X2, electronegativity, polarizability, atomic radius, and lipophilicity of the molecule.
Hajjo *et al.* (2012)[126]	2D-QSAR	Diverse scaffolds	5-HT_6R	QSAR models were used for virtual screening to discover potential 5-HT6R actives.
Bharate *et al.* (2013)[127]	2D-QSAR	Meridianin analogs	Dyrk1A	According to the findings, the kier Chi4 path/cluster, total lipole, VAMP polarisation ZZ component, dipole moment Z component, and log P all play significant roles in inhibiting Dyrk1A.
Multi-target QSAR studies				
Prado-Prado *et al.* (2012)[128]	3D mt-QSAR	Diverse scaffold	More than 500 FDA-approved targets	The obtained model will assist with the prediction of novel drugs that have activity against several AD targets.

Speck-Planche *et al.* (2013)[129]	mt-QSAR (LDA)	Diverse scaffold	5 proteins targets associated with AD	New molecular entities were suggested and a model was used to quickly and efficiently identify the fragments in charge of the activity against the five targets.
Kumar et al. (2020)[130]	2D-QSAR	Diverse scaffold	AChE; BuChE	Ring size, −CH2- groups, secondary aromatic amines, and aromatic ketones contribute to AChE inhibition. Distances between nitrogens, X-C(=X)-X and R--CR-X, secondary aromatic amides contribute to BuChE inhibition.

5-HT_6R: 5-hydroxytryptamine-6 receptor; Aβ: beta-amyloid; AChE: Acetylcholinesterase; ABAD:Aβ-binding alcohol dehydrogenase; BChE: Butylcholinesterase; CDK5/p25: Cyclin-dependant kinase 5/p25; CK1: Casein kinase 1; Dyrk1A:Dual-specificity tyrosine phosphorylation-regulated kinase 1A; GSK3: Glycogen synthase kinase-3; VAChT: Vesicular acetylcholine transporter.

1.4. Activity data sources, and freely available CADD software and tools

One of the main elements of the study needed to perform the various tasks involved in in-silico studies is a software tool. A publicly available software tool allows any researcher to execute, copy, and share the tool with the scientific community. **Table 1.6** contains a collection of freely available databases and in silico software applications.

Table 1.6. The Weblinks of tools and servers in alphabetical order.

Acronyms	Weblinks
Achilles Blind Docking Server	https://bio-hpc.ucam.edu/achilles/
ACD-Chemsketch	https://www.acdlabs.com/resources/freeware/chemsketch/download.php
AMBER18	https://ambermd.org/GetAmber.php
AutoDock Vina	http://vina.scripps.edu/
Avogadro version 1.2	https://avogadro.cc/news/avogadro-1-2-0-released/
Biovia discovery studio	https://discover.3ds.com/discovery-studio-visualizer-download
Binding Database	https://www.bindingdb.org/bind/index.jsp
CHARMM-GUI Glycolipid Modeler	http://www.charmm-gui.org/?doc=input/glycan
Chiron online server	https://dokhlab.med.psu.edu/chiron/login.php
ChemDraw Ultra12v	https://en.freedownloadmanager.org/users-choice/Chemdraw_Ultra_7.0_Free_Download.html
ClustalW	https://www.genome.jp/tools-bin/clustalw
ClustalX	http://www.clustal.org/clustal2/
DISPHOS 1.3 server	https://dabi.temple.edu/disphos/pred/predict
ERRAT	https://servicesn.mbi.ucla.edu/ERRAT/
ChEMBL database	https://www.ebi.ac.uk/chembl/
FlexX	https://www.biosolveit.de/FlexX/
Gaussview program	https://gaussian.com/gaussview6/
GROMACS 5.1.2	http://manual.gromacs.org/documentation/5.1.2/download.html
GlycoEP server	https://bio.tools/glycoep
HMMER tool	http://hmmer.org/
HHsearch tool	https://github.com/soedinglab/hh-suite
Hyperchem	http://www.hyper.com/
Idock tool	https://github.com/HongjianLi/idock

I-TASSER server web server	https://zhanglab.ccmb.med.umich.edu/I-TASSER/
LigandScout tool	https://ligandscout.software.informer.com/
MODELLER 9.23	https://salilab.org/modeller/9.23/release.html
Molegro Molecular viewer	http://molexus.io/molegro-molecular-viewer/
Molegro virtual docker software	http://molexus.io/molegro-virtual-docker/
MolProbity web server	http://molprobity.biochem.duke.edu/
MMTK software	http://dirac.cnrs-orleans.fr/MMTK.html
NAMD 2.13	https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=NAMD
NCBI database	https://www.ncbi.nlm.nih.gov/
NetPhos 3.1 sever	http://www.cbs.dtu.dk/services/NetPhos/
NetSurfP-2.0 sever	http://www.cbs.dtu.dk/services/NetSurfP/
OGTSITE server	http://csb.cse.yzu.edu.tw/OGTSite/
PaleAle 5.0 server	http://distilldeep.ucd.ie/paleale/quickhelp.html
PeptideMass online server	https://web.expasy.org/peptide_mass/
PEPFOLD3 server	https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/
PharmaGIST online tool	https://bioinfo3d.cs.tau.ac.il/PharmaGist/
PileUp tool	http://www.dbbm.fiocruz.br/cgc/pileup.html
pmemd.cuda module in AMBER18 tool	https://ambermd.org/GPUHowTo.php
PROCHECK	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
Profilescan	http://130.88.97.239/bioactivity/newpfscan.html
profile-profile alignment tool FFAS03	http://ffas.godziklab.org/ffas-cgi/cgi/ffas.pl
PubChem	https://pubchem.ncbi.nlm.nih.gov/
PyMol tool	https://pymol.org/dsc/
PyRx 0.8	https://pyrx.sourceforge.io/

QSAR model	https://dtclab.webs.com/software-tools
SAM tool	https://compbio.soe.ucsc.edu/sam.html
Schrodinger software	https://www.schrodinger.com/
Sima software	https://sourceforge.net/projects/smina/
SANCDB	https://sancdb.rubi.ru.ac.za/
SwissDock server	http://www.swissdock.ch/
Swiss model web server	https://swissmodel.expasy.org/
UCSF Chimera	https://www.cgl.ucsf.edu/chimera/
Verify3D	https://servicesn.mbi.ucla.edu/Verify3D/
Verification Server	https://servicesn.mbi.ucla.edu/SAVES/
vROCS (OpenEye)	https://docs.eyesopen.com/applications/rocs/vrocs/vrocs.html
Yasara server	http://www.yasara.org/minimizationserver.htm
YinOYang 1.2 server	http://www.cbs.dtu.dk/services/YinOYang/
ZINC15 database	https://zinc15.docking.org/

CHAPTER - 2

PRESENT WORK

Chapter 2: Present work

Alzheimer's disease (AD) is a progressive neuropathological disorder, found in the most common form of dementia, which causes severe brain deterioration and cognitive function loss[1-2]. AD is a degenerative ailment that is thought to begin decades before symptoms appear. Clinicians are only able to identify even the first signs of AD after significant damage has already been done to crucial biological components[3]. Despite the challenge of researchers to definitively identify the initial trigger that leads apart a series of harmful processes, a lot of studies have identified essential components in AD pathogenesis[3-4]. According to evidence from autosomal dominant and sporadic types of AD, amyloid plaques and tau protein-based neurofibrillary tangles can develop for up to 20 years before the onset of clinical dementia[5]. The staging of AD pathological abnormalities during the preclinical stage of the disease is facilitated by the recent growth of imaging and fluid biomarkers for AD pathogenesis[3-5]. AD is a developing healthcare concern, with increased life expectancy as the primary risk factor[3-5]. Disease prevalence is expected to more than double over the next several decades in the absence of adequate prevention and treatment alternatives[3-5]. About 200 clinical studies have been conducted to date to identify disease-modifying treatments for AD, but these efforts have generally failed, with many failures being attributable to ineffectiveness or excessive toxicity[3-5]. Every failed clinical study of a novel molecular entity (NME) takes a significant amount of time and money. Repurposing medications that have already been approved by the Food and Drug Administration (FDA) for a different indication, however, is less expensive, involves known potential toxicities, and has a greater success rate (30%) than developing an NME[3-6]. Significant effort has been devoted in recent years[5-7] to identify therapies that halt neurodegeneration in AD, but we are still far from finding exact treatment techniques[6-7]. The early diagnosis and treatment of AD is now a fast-developing field of both scientific and clinical research because current treatments only help with the symptoms of the disease. There are now only five approved drugs for the treatment of cognitive symptoms of Alzheimer`s disease. Among them, four drugs are acetylcholinesterase enzyme (AChE) inhibitors (Tacrine, Rivastigmine, Galantamine, and Donepezil), and the remaining one drug is non-competitive glutamate (NMDA) receptor antagonist (Memantine "FDA approved")[6-7]. Their use is only symptomatic, and no treatment has been proven to slow or stop the progression of the disease[6-7]. The long-term effects of AChE inhibitors have recently been postulated to be due to these medications interfering with the metabolism of amyloid precursor protein (APP)[6-7]. The cause and progression of AD are still not well understood. The search for treatments in the field of neurodegenerative disorders is extremely active, yet there is still no cure for AD.

Furthermore, the development of inhibitors against AD is a challenging and difficult procedure due to the complication of the molecular pathways involved in the progression of the disease[83-85]. Computer-Aided Drug Design (CADD) uses computer power, three-dimensional graphics, mathematics, and statistics to understand and predict the binding mode and energy of small molecule inhibitors with potential targets[83-85]. The most common in-silico techniques employed by medicinal chemists to help them rationalize the selection of hit compounds and to perform hit-to-lead optimization include structure-based design like molecular docking and dynamics and ligand-based design like quantitative structure-activity relationships (QSAR), chemical Read-Across and pharmacophore mapping[83-85]. Among these methodologies, the quantitative structure-activity relationship (QSAR) and molecular docking have great applications in the area of in-silico search. The QSAR methods are essential for the exploration of important structural features and prediction of the biological activity of novel compounds based on mathematical and statistical relations[83-85]. The idea of QSAR is based on the concept that endpoint values of compounds change systematically with modification of the structural attributes[83].

In the present thesis work, numerous *in silico* techniques were employed to study the potential leads against AD. The main objective was to use different in silico approaches to find and improve potential anti-Alzheimer's leads against several crucial targets involved in AD. Along with the single-target drug designing approach, we have also focused on identifying or designing dual-binding site AChE inhibitors, as well as multi-target inhibitors. Further, we have explored the selectivity issue of inhibitors against AChE over BuChE, which is a commonly observed issue while designing molecules against enzymes. Although we used a variety of *in silico* methods, such as QSAR, molecular docking, pharmacophore modeling, virtual screening, and so on, the majority of our work is focused on developing predictive and statistically robust QSAR models. The QSAR approach is used extensively in the lead optimization step of any drug development effort to reduce time, money, and, most importantly, animal sacrifice. A QSAR model is used to identify the structural features responsible for the activity as well as to achieve selectivity. Additionally, we have also developed the quantitative structure activity-activity relationship (QSAAR) and selectivity-based models to explore the most important features contributing to the dual inhibition against the respective targets. Furthermore, the model provides significant information for designing new compounds with improved activity, and it is used to predict the activity of a query or newly designed compound.

2.1. Datasets employed

For performing the requisite *in silico* studies, several datasets were collected from various reliable sources (i.e., *literature and online database*) as mentioned in **Table 2.1**.

Table 2.1. Datasets employed in the present work. *Ncompds: Number of compounds in the dataset

Datasets	Alzheimer's disease targets	N_{compds}*	Class of compounds	Reference
I	Butyrylcholinesterase (BuChE)	1130	Diverse classes	Journal Article[131-184]
II	Beta-secretase 1 (BACE1)	98	Diverse classes	Online Database[185]
III	Amyloid-beta (Aβ) plaques	314	Diverse classes	Online Database[185]
IV	Acetylcholinesterase (AChE)	78	Carbamate derivatives	Journal Articles[186]
V	AChE	997	Diverse classes	Journal Articles[187-299]
	BuChE	716		
	Selectivity (AChE- BuChE)	198		
VI-A	5-hydroxytryptamine 6 (5-HT6)	80	Diverse classes	Online Database[185]
VI-B	AChE	1733		
VI-C	BuChE	2507		

VI-D	BACE1	905
VI-E	Amyloid-beta (Aβ) plaques	262
VI-F	Cyclin Dependent Kinase 5 (CDK-5)	225
VI-G	Gamma-secretase enzyme	217
VI-H	Glutaminyl Cyclase (QCs)	132
VI-I	Glycogen synthase kinase-3β	159
VI-J	Monoamine oxidase B (MAO-B)	170
VI-K	N-methyl-D-aspartate (NMDA)	356
VI-L	Phosphodiester 10A (PDE 10A)	289
VI-M	Dual (AChE and BACE1)	43
VI-N	Dual (AChE and β-amyloid)	83
VI-O	Dual (AChE and BuChE)	113
VI-P	Dual **(**AChE and MAO-B)	52
VI-Q	Dual **(**BACE1 and GSK-3β)	20
VI-R	Dual **(**BuChE and BACE1)	51
VI-S	Dual **(**BuChE and β-amyloid)	23
VI-T	Dual **(**BuChE and MAO-B)	48
VI-U	Dual **(**AChE and GSK-3β)	21
VI-V	Dual **(**BuChE and GSK-3β)	21

2.1.1. Dataset I (study 1)

In this study, the activity values of a set of 1130 diverse classes of compounds against the BuChE enzyme were collected from previously published papers[131-184] for the development of the QSAR model. The experimental activity values of the dataset compounds were expressed as IC_{50} values (nM) and converted to pIC_{50}values for model development purposes. The data taken from the above-mentioned sources were checked and filtered by the criteria of a defined endpoint, and the same experimental procedures, following the OECD guidelines. The main of this study was to identify the structural requirements which are essential for BuChE enzyme inhibitory activity. Furthermore, we have also performed a molecular docking study with the most active, moderately active, and least active compounds from the whole dataset.

2.1.2. Dataset II (study 2)

In this study, a set of 98 heterocyclic compounds (BACE1 enzyme inhibitors) were collected from the BindingDB database[185] for the development of the QSAR model. The experimental IC_{50} values (nM) of the dataset compounds were converted into pIC_{50} values for model development purposes. These molecules employed similar bioassay protocols and conditions (FRET bioassay) for determining the biological activity against the BACE1 enzyme. The main purpose of this study was to identify the structural requirements which are essential for BACE1 enzyme inhibitory activity and to predict the activity of unknown compounds against the BACE1 enzyme. Additionally, we have performed pharmacophore mapping using the above dataset to reveal the structural requirements for the inhibitory activity and to categorize the compounds into more active and less active classes against the BACE1 enzyme. Furthermore, we have performed a molecular docking study with the most active and least active compounds from the whole dataset and tried to justify the contributions of different descriptors/features as evident in the QSAR/pharmacophore model.

2.1.3. Dataset III (study 3)

In this study, a set of 314 heterocyclic compounds (β-amyloid aggregation inhibitors) were collected from the BindingDB database[185] with defined β-amyloid aggregation inhibitory activity for QSAR model development. The experimental IC_{50} values (nM) of the dataset compounds were converted into pIC_{50} values for model development purposes. The dataset compounds utilized in this study followed the same experimental protocol (Thioflavin T (ThT) spectrofluorometric assay method). The main purpose of this study is to determine the essential structural features which are responsible for the inhibition of β-amyloid aggregation. Moreover, we also performed pharmacophore modeling using the above dataset to reveal the structural requirements for inhibitory activity against β-amyloid aggregation and also to categorize the compounds into active and less active classes. Furthermore, we performed a molecular docking study with the most active and least active compounds from the dataset and tried to justify the contributions of different descriptors/features obtained from QSAR/pharmacophore models.

2.1.4. Dataset IV (study 4)

In this study, the activity values of a congeneric series of 78 carbamate derivatives against the AChE enzyme were collected from the previously published literature[186]. The activity values for all the compounds were measured using the same experimental method (Ellman assay) by the same research group. The activity values of all the dataset compounds expressed as IC_{50} (μM) values were converted to the negative logarithm of IC_{50} (i.e., pIC_{50}) values for model development. In GQSAR model development, the designation of common scaffold and substitution sites is a prerequisite step. There are three substitution sites R_1, R_2, and R_3 in the congeneric series used in the G-QSAR study. In the GQSAR methodology, every dataset molecule is considered a set of fragments, and the fragmentation scheme is either template-based or user-defined. Once the common scaffold and substitution sites are defined, various descriptors are calculated for each fragment of the molecule. Now using an appropriate variable selection method, the significant descriptors representing the particular substituent sites are selected. The main purpose of this study was to identify the structural features responsible for the enhancement of the inhibitory activity against the AChE enzyme. Furthermore, we have performed the pharmacophore mapping using the above dataset to unveil the structural requirements for the inhibitory activity. Additionally, molecular docking studies were performed to understand the molecular interactions involved in binding, and the results are then correlated with the requisite structural features obtained from the QSAR and pharmacophore models.

2.1.5. Dataset V (study 5)

In the current work, we have used two different datasets against two important targets, namely AChE (number of compounds = 997) and BuChE (number of compounds = 761) enzymes collected from previously published literature[187-299]. In both the datasets, some compounds have inhibitory activities both against AChE and BuChE enzymes (number of compounds 198), therefore, we have used the difference of activities against AChE and BuChE enzymes as the dependent variable (AChE-BuChE) for modeling selectivity, and the third dataset has been prepared. The datasets comprise diverse classes of heterocyclic compounds, and the experimental activity of each compound is expressed in IC_{50} (nM) value, derived following the same bioassay protocol (modified colorimetric Ellman assay. For the model's development, we have converted the IC_{50} values to pIC_{50} values as customary in QSAR analysis. Before descriptor calculation, we carefully checked all structures in the datasets for the development of significant 2D-QSAR models. Additionally, we have also implemented the molecular docking studies using the most and least active compounds from the datasets and tried to rationalize the influences of different descriptors/features as apparent from the 2D-QSAR models.

2.1.6. Dataset VIA-V (study 6)

The activity data against twelve major targets of AD were collected from the BindingDB database[185]. Initially, 80 inhibitors against 5-hydroxytryptamine receptor 6 (5-HT6) following cell-based Radio ligand binding assay, 1733 compounds against acetylcholinesterase (AChE) enzyme following modified colorimetric Ellman assay, 2507 compounds against butyrylcholinesterase (BuChE) enzyme following modified colorimetric Ellman assay, 905 inhibitors against beta-secretase 1 (BACE1) enzyme following FRET (fluorescence resonance energy transfer) assay, 262 β-amyloid aggregation inhibitors following Thioflavin T-based fluorometric assay, 225 compounds against Cyclin Dependent Kinase 5 (CDK-5) protein following Scintillation proximity assay, 217 inhibitors against gamma-secretase enzyme following cell-based sandwich ELISA assay, 132 compounds against Glutaminyl Cyclase (QCs) enzyme following Continuous Spectrometric Assay, 159 inhibitors against glycogen synthase kinase-3 beta (GSK-3β) enzyme following Kinase-Glo reagent based luminescence assay, 170 compounds against Monoamine oxidase B (MAO-B) enzyme following Fluorometric method, 356 compounds against N-methyl-D-aspartate (NMDA) receptor following Fluorescence-based assay, 289 compounds against Phosphodiester 10A (PDE 10A) enzyme following TR-FRET assay. The datasets comprise diverse classes of heterocyclic compounds, and the experimental activity values are quantified in IC_{50} (nM). Before proceeding with the development of the regression models, we executed preliminary dataset preparation and data curation (chemical and biological) strategy using a KNIME workflow (available from https://dtclab.webs.com/software-tools). After dataset curation, screening of the activity datasets was performed to find the common compounds having dual inhibitory activity against the listed targets. Accordingly, we have found that the 43 compounds with dual inhibitory activities both against AChE and BACE1 enzymes, 83 compounds against AChE and β-amyloid, 113 compounds against AChE and BuChE enzymes, 52 compounds against AChE and MAO-B enzymes, 20 compounds against BACE1 and GSK-3β enzymes, 51 compounds against BuChE and BACE1enzymes, 23 compounds against BuChEI and β-amyloid, 48 compounds against BuChE and MAO-B enzymes, 21 compounds against AChE and GSK-3β enzymes and 21 compounds against BuChE and GSK-3β enzymes were retained and used for the development of the respective QSAAR and selectively based models. The activity endpoint values (IC_{50}) were converted to the negative logarithmic scale, pIC_{50}, as customary in QSAR modeling. In addition, we also executed molecular docking analyses with the most and least active molecules from the datasets, attempting to explain the influences of different properties as seen in the 2D-QSAR models.

CHAPTER - 3

MATERIALS AND METHODS

Chapter 3: Materials and methods

In this section, details about the all the datasets and procedures have been described that were employed to carry out several in silico studies, namely, QSAR, molecular docking, pharmacophore mapping, and virtual screening. Along with the methodology, all the software tools (including in-house developed) have been mentioned that were employed in our studies. Note that all the in-house developed QSAR relevant software tools and in-house designed Konstanz Information Miner (KNIME) workflows mentioned in this thesis are available at http://teqip.jdvu.ac.in/QSAR_Tools/ and http://dtclab.webs.com/software-tools.

3.1. Study 1-A Multi-layered Variable Selection Strategy for QSAR Modeling of Butyrylcholinesterase Inhibitors

3.1.1. QSAR methodology

3.1.1.1. Dataset selection

In this study, a PLS-regression-based QSAR model was developed for BuChEI using diverse classes of compounds (n =1130) collected from previously published papers[131-184] to identify the structural requirements which are essential for BuChE enzyme inhibitory activity. The experimental activity values of the dataset compounds were expressed as IC_{50} values (nM) and converted to pIC_{50}values for model development purposes. The data taken from the above-mentioned sources were checked and filtered by the criteria of a defined endpoint, and the same experimental procedures, following the OECD guidelines[300]. All the structures were drawn using the ChemDraw ultra 12.0 software (Available from: https://chemistry.com.pk/software/free-download-chemdraw-ultra-12/). Then each molecular structure was cleaned and hydrogens were added using Marvin view ChemAxon tool (Available from: https://chemaxon.com/products/marvin) and the structures were saved as MDL.mol format.

3.1.1.2. Descriptor calculation and data pretreatment

Molecular descriptors are mathematical representations of molecular structure information obtained by a well-specified algorithm. The descriptors were calculated using two software tools, namely, Dragon software version 7[301] and PaDEL-descriptor 2.20 software[302]. In this work, only 2D descriptors were calculated covering constitutional, ring descriptors, connectivity index, functional group counts, atom-centered fragments, 2D atom pairs, atom type E-states, molecular properties(using Dragon software version 7), and extended topochemical atom (ETA) indices (using PaDEL-Descriptor software). Additionally, data pretreatment was performed to remove inter-correlated descriptors from the dataset using the tools Pretreatment V-WSP version 1.2 (available at http://dtclab.webs.com/software-tools).

3.1.1.3. Dataset division

In this present study, the aim was to develop a QSAR model which is statistically robust and capable of making accurate and reliable predictions. Therefore, the developed QSAR model was validated using new chemical entities, i.e., a test set to check the predictive capacity of the developed models. The whole data set was divided into an internal set (training set) and an external set (test set) using the "Modified *k*-medoid" clustering technique (available at http://dtclab.webs.com/software-tools). The clustering technique categorizes a set of compounds into clusters so that the compounds present in the same cluster are similar

to each other. On the other hand, when two compounds belong to two different clusters, they are expected to be dissimilar. The representative compounds within a cluster are called medoids. This technique tends to select *k* from most middle compounds as the initial medoid. Eight clusters were generated for the BuChEI dataset containing 1130 compounds. We have selected approximately 25% of compounds from each cluster for the test set and the remaining 75% of compounds were selected for the training set. The training set was used for model development and the test set was used for model validation purposes.

3.1.1.4. Multi-layered variable selection and QSAR model development

The selection of important and meaningful descriptors from a large pool is a crucial step in QSAR model development. The selection of significant descriptors from a large initial pool is important to reduce the noise in the input. Thus, a multi-layered variable selection strategy was employed before the development of the final model using stepwise regression (using a suitable stepping criterion, e.g., 'F-for-inclusion 'and 'F-for-exclusion' based on partial F-statistic) followed by genetic algorithm (GA) followed by again stepwise regression and finally, best subset selection. For this purpose, first, we have run stepwise regression using the whole pool of descriptors and kept the model descriptors aside. Next, we have run again stepwise regression using the remaining pool (after removing the descriptors obtained from the first run stepwise regression) of descriptors and selected the model descriptors. In this way, we have selected 80 descriptors from the initial pool of 600 descriptors (Layer-I). After the first layer of descriptor selection, we developed some models using a genetic algorithm (GA) (available at http://dtclab.webs.com/software-tools) and selected 57 descriptors (Layer-II) from 80 descriptors. After that, top 20 descriptors were selected using the stepwise regression technique (Layer-III) again. Using these 20 descriptors, we have run the best subset selection using a tool developed in our laboratory (available at http://dtclab.webs.com/software-tools) to develop a 15 descriptor model which was selected based on Mean Absolute Error (MAE) based criteria[303-304]. Although many groups of authors reported different variable selection strategies, we have followed here stepwise regression followed by GA followed by stepwise regression, and finally, the best subset selection method as reported previously also[305-308]. The final model was developed by employing PLS-regression methodology to avoid intercorrelation among the modeled descriptors using Minitab software (Available from: http://www.minitab.com/en-US/default.aspx). Multi-layered variable selection strategy is schematically represented in **Figure 3.1**.

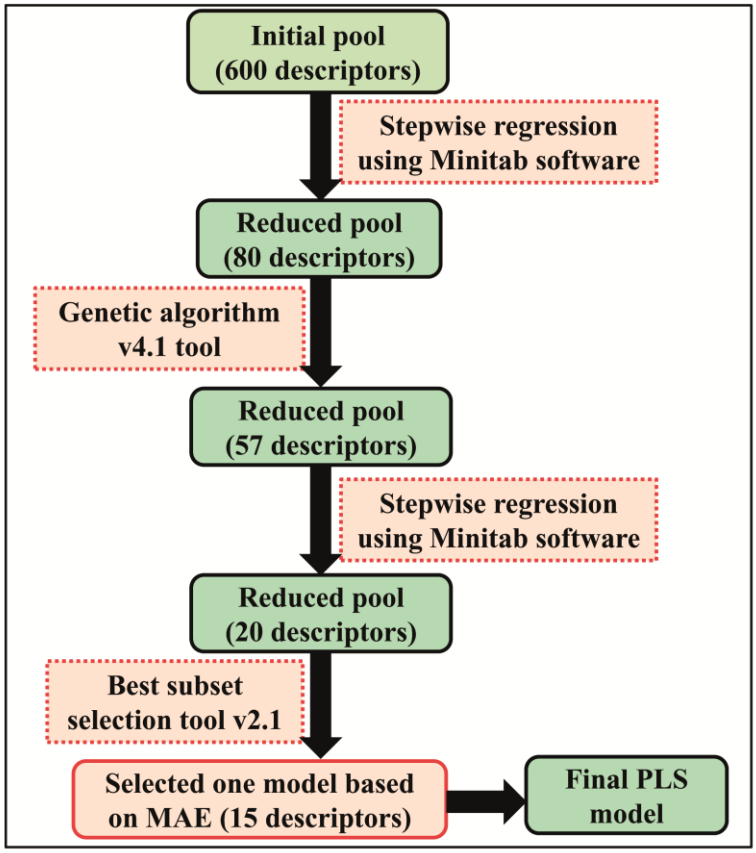


Figure 3.1. Schematic representation of multi-layered variable selection strategy.

3.1.1.5. Statistical validation metrics

In the present work, a PLS-regression-based QSAR model was developed for the inhibitory activity of the BuChE enzyme. The developed model was validated using both internal and external validation parameters. The internal statistical parameters used in this study are the determination coefficient (R^2) and leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$). The determination coefficient (R^2) represents how much variability of a factor can be explained by its relationship to another factor, it is computed as a value between 0 (0 percent) and 1 (100 percent). The higher value of this parameter indicates a better fit of the model. But these parameters are not good enough to evaluate the robustness and predictivity of a significant model. Thus, we have employed additional statistical validation parameters such as R^2_{Pred} (external prediction variance) or Q^2F_1 and Q^2F_2 to assure the significance of the developed model. Additionally, a Y-randomization test, checking applicability domain criteria, etc. were performed to investigate the robustness of the developed model. The main objective of the Y-randomization test is to ascertain whether the developed model is obtained by chance or not. The Y-randomization test was performed using the Simca-P software (Available from: https://umetrics.com/kb/getting-started-simca-p) by randomly reordering (100 permutations) the dependent variable. The validation parameter of the model obtained under such conditions should be of poor quality. The value of the R^2_{yrand} intercept should not exceed 0.3 and the value of the Q^2_{yrand}intercept should not exceed 0.05[309].

3.1.1.6. Applicability domain (AD) assessment

The Applicability domain of a QSAR model has been described as the response and chemical structure space that is defined by the nature of the chemicals in the training set. If a new compound falls within the Applicability domain of the developed model, only then the developed model can predict the compound precisely. It is extremely useful for QSAR developers to have information about the applicability domain of the developed model to identify interpolation (true predictions) or extrapolation (less reliable predictions)[310-311]. Here, we have checked the applicability domain of the developed model employing the DModX (distance to model X) approach at a 99% confidence level using SIMCA-P software (Available from: https://umetrics.com/kb/getting-started-simca-p).

3.1.1.7. Randomization of the PLS model

The purpose of the Y-randomization test is to identify and quantify chance correlations between the dependent variable and the descriptors[309]. Here, the term chance correlation means that the real model may contain descriptors that are statistically well correlated to Y, but in reality, there is no cause-effect relationship encoded in the respective correlations with Y because they are not related to the mechanism of action [309]. The Y-randomization test consists of several runs for which the original descriptor matrix X is kept fixed, and only the vector Y is randomized[309]. The validation parameter of the model obtained under such conditions should be of poor quality and without real meaning[309-310]. The value of the R^2_{yrand} intercept should not exceed 0.3 and the value of the Q^2_{yrand} intercept should not exceed 0.05[309]. In the present study, for the training set, the X data remained constant and the data Y were shuffled randomly. Here, we have generated the randomized model using 100 permutations. The schematic work flow for the development of the QSAR model against BuChE inhibitors is shown in **Figure 3.2**.

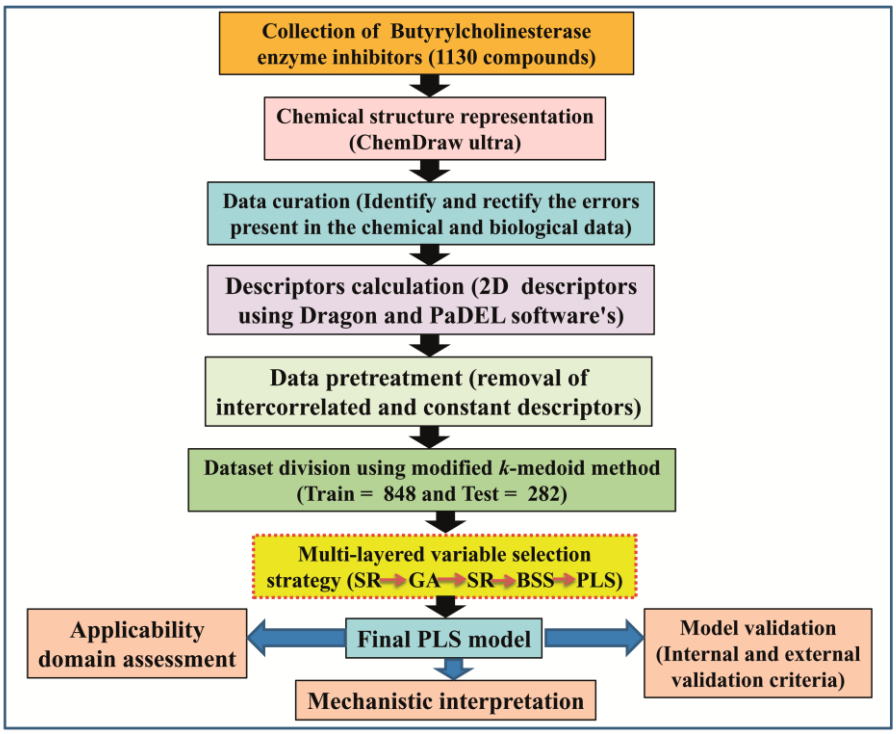


Figure 3.2. Schematic work flow of QSAR model development against BuChE inhibitors (BuChEI). [PLS = Partial least squares, SR = Stepwise regression, BSS = Best subset selection].

3.1.2. Docking studies

Molecular docking is a technique to understand drug-biomolecular interactions for the rational drug design and discovery as well as for the mechanistic study by placing a molecule (ligand) into the desired binding site of the target definite region of the protein/enzyme (receptor) mainly in a non-covalent manner to form a stable complex of potential efficacy and specificity[312]. The docking process includes two basic stages: prediction of the ligand conformation as well as its position and orientation within these sites (usually referred to as pose) and assessment of the binding affinity. The evidence obtained from the docking study can be used to suggest the binding orientation, binding energy, free energy, interaction energy, and stability of complexes. In the current study, we have employed molecular docking studies to comprehend the interactions between the BuChE enzyme (the structure of the protein was retrieved from Protein Data Bank with PDB ID: 6EZ2, available from: https://www.rcsb.org/structure/6EZ2) and the selected BuChE enzyme inhibitors. In this context, we have applied the CDOCKER module of receptor-ligand interaction available in BIOVIA Discovery Studio client 4.1 (available from: https://3dsbiovia.com/resource-center/downloads/). Before the docking experiment, we defined the active site of the enzyme using the protocol Receptor-Ligand Interaction section using the option “define site from receptor cavities” in the BIOVIA Discovery Studio client 4.1 platform. The selected inhibitors were subjected to ligand preparation to find a series of ligand conformers. Each orientation was used in the CDOCKER module for molecular docking using CHARMm-based interaction energy using a rigid receptor[313]. The poses are sorted according to CHARMm interaction energy, and the top scoring (most negative, thus favorable to binding) poses are kept.

3.2. Study 2- Exploring 2D-QSAR for prediction of Beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease

3.2.1. QSAR methodology

3.2.1.1. The Dataset

In this study, a **98** heterocyclic compounds (BACE1 enzyme inhibitors) (**Table 3.1**) were collected from the BindingDB database[185] for the development of the QSAR model. The experimental IC_{50} values (nM) of the dataset compounds were converted into pIC_{50} values for model development purposes. The reported assay (FRET bioassay) procedure of all the compounds used in this study followed the same protocol[314-320]. All the structures were drawn using the Marvin ChemAxon tool (Available from https://chemaxon.com/products/marvin) followed by cleaning of molecules, and finally, saved as MDL .mol format. All the compound structures were properly checked before the calculation of descriptors. The main purpose of this study was to identify the structural requirements which are essential for BACE1 enzyme inhibitory activity and to predict the activity of unknown compounds against the BACE1 enzyme.

3.2.1.2. Preliminary dataset preparation and data curation

Before the development of QSAR models, we performed dataset preparation and data curation (chemical and biological) steps. The dataset which we downloaded in a structural data format (SDF) had both structural and biological activity information. The identifiers were given to all compounds present in the dataset that carry the following information, i.e., the name of the respective enzyme and a serial number. The activity values from the dataset file was extracted to classify the compounds in four orders of magnitude. At last, we have submitted the dataset to chemical and biological curation. In this study, we have employed the most common steps **(**as detailed below**)** to perform the chemical curation that has been implemented in the in-house designed Konstanz Information Miner (KNIME) workflow (available at https://dtclab.webs.com/software-tools).

3.2.1.2.1. Reading and storing the information present in the downloaded SDF file from the BindingDB database

The dataset file (SDF) contains all the structural and topological information like molecule name, coordinates, bond counts, etc. Along with this necessary information, it also contains biological information. For chemical data curation, structural information is enough to identify the correct chemical structure and remove duplicates. Later, the biological data curation was also performed to store other information related to the biological property, which is necessary for correct biological data curation. In the KNIME workflow, we have used the "SDF Reader" node for reading the input SDF file and storing the structural information. We have discarded the molecules with incorrect and incomplete information, and the molecules which have correct structural and biological information were stored as the revised SDF file for further use.

3.2.1.2.2. Removal of salts, mixtures, inorganics, and organo-metallics

The molecular descriptors are generally computed for organic compounds and thus the majority of software tools can only consider organic compounds. Thus, the presence of salts, mixtures, inorganics, and organo-metallics in the dataset may lead to incorrect descriptor values, or such compounds are simply rejected by

the software. So, in the present study, we have removed all the salts, mixtures, in-organics, and organo-metallics compounds before calculating the descriptors using 'RDKit Salt Stripper', 'Connectivity' node and 'Element Filter' nodes[59] respectively.

3.2.1.2.3. Normalization of chemical structures

There is a possibility of demonstrating the same functional group using different structural forms in the dataset. Therefore, normalization of chemical structure is necessary to remove different structural patterns. For example, nitro groups can be represented using two double bonds between nitrogen and oxygens, one single bond linking the nitrogen and the protonated oxygen, or linking both nitrogen and oxygen atoms that are oppositely charged[59]. Different representations of the same chemical structure may create serious problems in a QSAR study because molecular descriptors calculated for these different representations of the same functional group could be significantly different. Thus, the transformation of all such functional groups to standard forms is highly essential. In this study, the 'RDKit Structure Normalizer node was used for standardizing the chemical structures.

3.2.1.2.4. Biological curation

After chemical curation, biological curation was performed using the screened compounds, and it includes two important steps, i.e., duplicate analysis and activity cliff analysis. In the case of biological data curation, we have again employed the KNIME workflow (available at http://dtclab.webs.com/software-tools) to perform duplicate identification and activity cliff determination. First, we simply performed the duplicate analysis based on the BindingDB Monomer ID and using a distance similarity index[59] (calculated using a 3D D-Similarity node available in KNIME), where the two compounds were considered identical or duplicates, only if both the BindingDB Monomer ID and distance similarity values are identical for both the compounds. Although the BindingDB Monomer ID might have been sufficient to identify duplicates, to confirm that there is no error in the BindingDB Monomer ID itself, distance similarity index was computed using KNIME workflow. The list of molecules present in the dataset with their names, structures, and activity against the BACE1 enzyme is depicted in **Table 3.1**.

Table 3.1. The list of molecules present in the dataset with their names, structures, and activity against the BACE1 enzyme.

Name	SMILES notation	Observed value (pIC_{50} nM)
1	C1(=N[C@](C(=O)N1C)(c1cc(ccc1)NC(=O)c1occc1)C1CCCCC1)N	-1.301
2	C1(=N[C@](C(=O)N1C)(C1CCCCC1)c1cc(ccc1)NC(=O)COC)N	-1.602
3	C1(=N[C@](C(=O)N1C[C@H]1CC[C@H](CC1)C(=O)O)(c1ccccc1) C1CCCCC1)N	-2.079
4	C1(=N[C@](C(=O)N1C)(c1cccc(c1)NC(=O)c1n(ccc1)C)C1CCCCC1)N	-2.255
5	C1(=N[C@@](C(=O)N1CCCCCC)(C1CCCCC1)c1ccccc1)N	-2.431
6	C1(=N[C@](C(=O)N1CCCCCC(=O)O)(c1ccccc1)C1CCCCC1)N	-2.491
7	C1(=N[C@](C(=O)N1C)(c1cccc(c1)NC(=O)c1ccc2c(c1)OCO2)C1CCCCC1)N	-2.579
8	C1(=N[C@](C(=O)N1Cc1cc(cc(c1)F)F)(c1ccccc1)C1CCCCC1)N	-2.612
9	C1(=N[C@](C(=O)N1C)(C1CCCCC1)c1cc(ccc1)NC(=O)CCC)N	-2.690
10	C1(=N[C@@](C(=O)N1C)(c1ccccc1)C1CCCCC1)N	-2.707
11	c1(ccccc1)c1ccc(c2ccc(cc2)NC(=O)c2cc(ccc2)Br)n1CC(=O)NC(=N)N	-2.778
12	C1(=N[C@](C(=O)N1C)(c1cccc(c1)NC(=O)c1n(ccc1)C)C1CCCCC1)N	-2.934

13	C1(=N[C@@](C(=O)N1C)(c1ccccc1)C1CCCCC1)N	-2.995
14	C1(=NC(C(=O)N1C)(C1CCCCC1)C1CCCCC1)N	-3.033
15	C1(=N[C@](C(=O)N1C)(c1cccc(c1)NC(=O)c1ccc2c(c1)OCO2)C1CCCCC1)N	-3.060
16	C1(=N[C@@](C(=O)N1C)(c1cccc(c1)NC(=O)c1cccc(c1)OC)C1CCCCC1)N	-3.133
17	C1(=N[C@](C(=O)N1C)(c1cc(ccc1)NC(=O)CCOC)C1CCCCC1)N	-3.267
18	C1(c2ccccc2)(c2ccccc2)N=C(N(C1=O)C)N	-3.531
19	C1(=N[C@](C(=O)N1C)(c1cc(ccc1)NC(=O)CCOC)C1CCCCC1)N	-3.895
20	C1(=N[C@@](C(=O)N1C)(c1ccccc1)C1CCCCC1)N	-4.380
21	C1(c2ccccc2)(N=C(N2C1=NCCC2)N)c1ccccc1	-4.579
22	c12cc(ccc1c(c(cn2)C(=O)NCc1ccc2OCOc2c1)O)S(=O)(=O)Nc1cccc(c1)C(F)(F)F	-2.491
23	c1(csc(n1)NC(=O)CSc1nc(c(c(=O)[nH]1)C#N)c1ccc(cc1)OC)c1ccccc1	-3.677
24	c1(n(c(cc1)c1c(cccc1)Cl)CC(=O)NC(=N)NCCCO)c1ccc(cc1)OCCC	-2.041
25	c1(n(c(cc1)c1c(cccc1)Cl)CC(=O)NC(=N)N)c1ccc(cc1)Oc1ccc(cc1)C(=O)C	-2.380
26	c1(n(c(cc1)C12CC3CC(C1)CC(C3)C2)CC(=O)NC(=N)NCCCO)c1ccccc1	-3.505
27	c1(n(c(cc1)C12CC3CC(C1)CC(C3)C2)CC(=O)NC(=N)N)c1ccccc1	-3.770

28	c1(n(c(cc1)c1ccccc1)CC(=O)NC(=N)N)c1ccccc1	-4.287
29	C1[C@@H]2[C@@H]3[C@H](N[C@@H]2CN[C@H]1Cn1cc(nn1)c1ccc2c(c1)ccc(c2)OC)CCCC3	-3.173
30	[C@H]1(C(=O)[C@H](SCc2cc(ccc2)C(F)(F)F)C(=O)N1)[C@@H](C)CC	-4.778
31	C1(=O)[C@@H](OC(=O)[C@H]1Sc1ccccc1)c1ccccc1	-5.021
32	C1(=O)[C@H](NC(=O)[C@@H]1SCc1occc1)Cc1ccccc1	-5.161
33	C1(=O)[C@@H](NC(=O)[C@@H]1SCc1cccc(c1)C(F)(F)F)CC(C)C	-5.287
34	C1(=O)[C@@H](NC(=O)[C@H]1SCC(=O)NCc1cc2OCOc2cc1)CC(C)C	-5.301
35	C1(=O)[C@@H](NC(=O)[C@H]1SCc1occc1)[C@@H](CC)C	-5.301
36	C1(=O)CN(C(=O)[C@H]1SCc1cccc(c1)C(F)(F)F)Cc1ccccc1	-5.326
37	C1(=O)CN(C(=O)[C@H]1SCCc1ccccc1)Cc1cc2OCOc2cc1	-8.143
38	C1(=O)CN(C(=O)[C@H]1SCc1cccc(c1)C(F)(F)F)CCc1ccccc1	-8.264
39	C1(=O)CN(C(=O)[C@H]1SCCc1ccccc1)Cc1cccc(c1)C(F)(F)F	-8.326
40	C1(=O)CN(C(=O)[C@@H]1SCc1cccc(c1)C(F)(F)F)Cc1ccc(cc1)OC	-8.382
41	C1(=O)CN(C(=O)[C@@H]1SCc1cccc(c1)C(F)(F)F)Cc1cc(ccc1)C(F)(F)F	-8.411

42	C1(=O)CN(C(=O)[C@H]1SCc1cccc(c1)C(F)(F)F)Cc1cc2OCOc2cc1	-8.423
43	C1(=O)CN(C(=O)[C@H]1SCCc1ccccc1)CCc1ccccc1	-8.627
44	C1(=O)CN(C(=O)[C@@H]1SCCc1ccccc1)Cc1ccccc1	-8.661
45	C1(=N[C@@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1ccc(cc1)OC(F)(F)F)N	-1.477
46	C1(=N[C@](c2cccc(c2)c2c(nccc2)F)(c2ccc(OC(F)(F)F)cc2)C2=NOCCN12)N	-1.778
47	C1(=N[C@](C2=NCCCN12)(c1cccc(c1)c1cncnc1)c1ccc(cc1)OC(F)(F)F)N	-1.903
48	C1(=N[C@@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1ccc(cc1)OC(F)(F)F)N	-1.903
49	C1(=N[C@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1ccc(cc1)OC)N	-2
50	C1(=N[C@](c2cccc(c2)c2c(nccc2)F)(c2ccc(OC(F)(F)F)cc2)C2=NCCCCN12)N	-2.113
51	C1(=N[C@](C2=NCCCN12)(c1cc(c2c(nccc2)F)ccc1)c1cc(c(cc1)OCC)OCC)N	-2.278
52	C1(=N[C@@](c2cccc(c3c(nccc3)F)c2)(c2ccc(OC(F)(F)F)cc2)C2=NCCN12)N	-2.897
53	C1(=N[C@](C2=NCCCN12)(c1cccc(c2cncnc2)c1)c1ccc(cc1)OC)N	-3.309
54	C1(=N[C@](C2=NCCCN12)(c1cccc(c1)c1cncnc1)c1ccc(cc1)OC(F)(F)F)N	-3.380
55	c1(n(c(cc1)c1c(cccc1)Cl)CC(=O)NC(=N)NCCCO)c1ccc(cc1)OCCC	-3.577

56	C1(=N[C@](c2cccc(c2)c2c(nccc2)F)(c2ccc(OC(F)(F)F)cc2)C2=NCCCCCN12)N	-3.819
57	C1(=N[C@@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1cc2c(cc1)OCO2)N	-3.845
58	C1(=N[C@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1cc2c(cc1)OCCO2)N	-3.982
59	C1(=N[C@](c2cccc(c2)c2c(nccc2)F)(c2ccc(OC(F)(F)F)cc2)C2=NC(CN12)(C)C)N	-4.029
60	C1(=N[C@@](C2=NCCCN12)(c1cccc(c1)CCC1CC1)c1ccc(cc1)OC(F)(F)F)N	-4.037
61	C1(=N[C@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1cc2c(cc1)OC(O2)(F)F)N	-4.086
62	C1(=N[C@](C2=NCCCN12)(c1cccc(c1)c1cncnc1)c1ccc(cc1)OC(F)(F)F)N	-4.215
63	C1(=N[C@@](C2=NCCCN12)(c1cccc(c1)CCC)c1ccc(cc1)OC(F)(F)F)N	-4.238
64	C1(=N[C@](c2cccc(c2)Br)(c2ccc(cc2)OC(F)(F)F)C2=NCCCN12)N	-4.318
65	C1(=N[C@](C2=NCCCN12)(c1cccc(Cc2ccccc2)c1)c1ccccc1)N	-4.387
66	C1(=N[C@](C2=NCCCN12)(c1cccc(Cc2ccc(cc2)F)c1)c1ccccc1)N	-4.480
67	C1(=N[C@](C2=NCCCN12)(c1cccc(Cc2ccc(cc2)OC)c1)c1ccccc1)N	-4.579
68	C1(=N[C@@](c2cccc(c2)C(F)(F)F)(c2ccccc2)C2=NCCCN12)N	-4.583
69	C1(=N[C@@](C2=NCCCN12)(c1cc(ccc1)OC)c1ccccc1)N	-4.755

70	C1(=N[C@@](c2cccc(c2)C(C)(C)C)(c2ccccc2)C2=NCCCN12)N	-4.895
71	C1(c2ccccc2)(N=C(N2C1=NCCC2)N)c1ccccc1	-5.176
72	c1c(n(Cc2ccc(c(n2)N)OCCO)c(c2ccccc2Cl)c1)c1ccc(Oc2cncnc2)cc1	-1.602
73	c1c(n(Cc2ccc(c(n2)N)NCCO)c(c2ccccc2Cl)c1)c1ccc(Oc2cncnc2)cc1	-1.845
74	c1c(c2ccc(Oc3cncnc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2
75	c1c(c2ccc(OCCCCC)cc2)n(Cc2ccc(c(n2)N)OCCO)c(c2ccccc2Cl)c1	-2.041
76	c1c(c2ccc(OCCCCC)cc2)n(Cc2ccc(c(n2)N)NCCO)c(c2ccccc2Cl)c1	-2.230
77	c1c(n(Cc2ccc(c(n2)N)NCCCO)c(c2ccccc2Cl)c1)c1ccc(OCCCCC)cc1	-2.342
78	c1c(c2ccc(Oc3cccs3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.462
79	c1c(c2ccc(Oc3cccnn3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.612
80	c1c(c2ccc(Oc3cccnc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.612
81	c1c(c2ccc(OCCCC#N)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.623
82	c1c(c2ccc(Oc3cnccn3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.672
83	c1c(c2ccc(OCCCCC)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.740
84	c1c(c2ccc(Oc3ccncc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.770

85	c1c(c2ccc(Oc3ccccn3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.770
86	c1c(c2ccc(Oc3ccccc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.908
87	c1c(c2ccc(Nc3cncnc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.991
88	c1c(c2ccc(cc2)OCCC)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-3
89	c1c(c2ccc(C(=O)NC3CC3)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.060
90	c1c(c2ccc(OCCCC)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-3.079
91	c1c(c2ccc(C(=O)NC(C)C)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.217
92	c1c(c2ccc(cc2)C(=O)NCC)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.250
93	c1c(c2ccc(cc2)OC)n(c(c1)c1ccccc1Cl)Cc1cccc(n1)N	-3.352
94	c1c(c2ccc(C(=O)NCCC)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.389
95	c1c(c2ccc(C(=O)NCCCC)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.477
96	c1c(c2ccc(C(=O)NC3CCC3)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.477
97	c1c(c2ccccc2)n(c(c1)c1ccccc1)Cc1cccc(n1)N	-3.531
98	c1c(c2ccc(C(=O)NCC=C)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.851

3.2.1.3. Descriptor calculation and data pretreatment

The molecular descriptors were calculated using two software tools, namely, Dragon software version 7[301] (covering constitutional, ring descriptors, connectivity index, functional group counts, atom centered fragments, 2D atom pairs, atom type E-states, and molecular properties) and PaDEL-descriptor 2.20 software[302] (for extended topochemical atom or ETA indices). Molecular descriptors may be defined as the way of mathematical representation of a molecule by values associated with the chemical constitution for correlation of chemical structures with various chemical reactivity, biological activity, or physical property[69]. After the calculation of descriptors, data pretreatment was implemented using the tool Pretreatment V-WSP version 1.2 (available at http://dtclab.webs.com/software-tools) to remove descriptors with at least one missing value, variables with constant or near constant values (standard deviation less than 0.0001), descriptors with all missing values and descriptors with (absolute) pair correlation larger than or equal to 0.95 from the initial pool of descriptors[69].

3.2.1.4. Dataset division

The whole dataset was divided into training (n_{train}=**76**) and test (n_{test}=**22**) sets based on the k-Medoids clustering technique. We aimed to develop a QSAR model which is statistically robust and capable of making accurate and reliable predictions. For this, we have employed a software tool "Modified k-Medoids" (version 1.2) developed in our laboratory (available at http://dtclab.webs.com/software-tools). The clustering method classifies a set of compounds into clusters so that compounds belonging to the same cluster are similar to each other, while when two compounds belonging to two different clusters are expected to be dissimilar[321]. The representative compounds within a cluster are called medoids. This technique tends to select k from most middle objects or compounds as the initial medoid. After clustering, we sorted the whole dataset according to the cluster number followed by activity values. We have selected around 22% of compounds from each cluster as test set compounds (n_{test}= **22**) and the remaining 78% as training set (n_{train}= **76**) compounds. Consequently, the developed QSAR model was validated using new chemical entities, i.e., the test set to check the predictive ability of the developed model.

3.2.1.5. Variable selection and QSAR model development

In this study, a QSAR model was developed for inhibitory activity against the BACE1 enzyme using pIC_{50} values as the response variable. The selection of important and meaningful descriptors from a large descriptors pool is a crucial step in QSAR model development. Thus, we employed a variable selection strategy before the development of the final model using stepwise regression (using a suitable stepping criterion, e.g., 'F-for-inclusion 'and 'F-for-exclusion' based on partial F-statistic) followed by the best subset selection. For this purpose, stepwise regression was run using the whole pool of descriptors (393) to develop a stepwise regression model. In this case, stepwise regression has been applied using the initial pool of 393 descriptors, and the selected model descriptors were removed from the initial pool of descriptors and kept aside. Further, stepwise regression was run using the remaining pool of descriptors, and so on. Finally, we have clubbed the selected model descriptors in different cycles. In this way, we have reduced the initial pool of descriptors from 393 to 60 descriptors. Using these 60 descriptors, we ran the best subset selection tool v2.1 developed in our laboratory (available at http://dtclab.webs.com/software-tools) to generate a 5-descriptor model. Best subset regression is an investigative model-building regression analysis technique. This technique compares all possible models using a specified set of predictors and displays the best-fitting models that contain one predictor, two predictors, and so on. A result is several models and their

summary statistics. It is up to us to compare them and choose one. Sometimes the results do not point to one best model and so our judgment is required to choose the best significant model. When selecting the best subset model, we are looking for the highest R^2, Q^2, R^2pred, and lowest MAE. Among the equations generated from the best subset selection, we selected the best model, based on the highest R^2, Q^2, R^2pred, and lowest MAE criteria. Among these models, we have selected one model based on Mean Absolute Error (MAE) based criteria[303]. The final model was developed by employing PLS-regression methodology to avoid intercorrelation among the modeled descriptors using the Minitab software (Available from http://www.minitab.com/en-us/products/minitab/). Information about the original variables is stored in latent variables (LV) generated by PLS. The final PLS model was developed with 5 selected descriptors using 3 latent variables (LV).

3.2.1.6. Statistical validation metrics

The developed models were validated using both internal and external validation parameters for measurement of the fitness, stability, robustness, and predictivity of the developed PLS model. Various internal statistical parameters like determination coefficient (R^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), and some r_m^2 metrics like average $r_m^2{}_{(LOO)}$ and $\Delta r_m^2{}_{(LOO)}$ were used to measure the robustness of the model. The determination coefficient (R^2) represents how much variability of a factor can be caused or explained by its relationship to another factor. It is computed as a value between 0 (0 percent) and 1 (100 percent)[322]. The higher value of this parameter indicated a better fit of the model. But these parameters are not good enough to evaluate the robustness and predictivity of a significant model. Thus, we have employed some other statistical validation parameters (external prediction) such as R^2_{Pred} or Q^2F_1, Q^2F_2 to assure the significance of the developed model. Besides these parameters, r_m^2 metrics like average r_m^2 and Δr_m^2 and concordance correlation coefficient (CCC) were also calculated for external validation. The basic application of a predictive QSAR model is to judge the prediction errors for an external set, which should be within the chemical and response-based domain of the internal set (i.e., training set). The Q^2_{ext}-based metrics (i.e., R^2_{pred} and Q^2_{F2}) are not always able to provide a correct indication of the prediction quality because of the influence of the response range as well as the distribution of the values of response in both the training and test set compounds. Thus, we have also validated the model using the mean absolute error (MAE) based criteria for both external and internal validation tests. The error-based metrics were used to determine the true indication of the prediction quality in terms of prediction error since they do not evaluate the performance of the model in comparison with the mean response[322]. Additionally, we have performed a Y-randomization test and DModX approach (applicability domain criteria) using the Simca-P software (Available from https://umetrics.com/products/simca). The Y-randomization test was performed to check whether the model was obtained by any chance or not, and the DModX approach was performed to check whether the test set compounds lie within the applicability domain or outside the applicability domain of training set compounds. The Y-randomization test was performed using the Simca-P software (Available from https://umetrics.com/products/simca) through randomly reordering (100 permutations) the dependent variable values. The validation parameter of the model obtained under such conditions should be of poor quality as compared to the selected model. The value of the R^2_{yrand} intercept should not exceed 0.3 and the value of the Q^2_{yrand} intercept should not exceed 0.05. The detailed methodology of work for the development of the 2D-QSAR model for BACE1 enzyme inhibitors is shown in **Figure 3.3**.

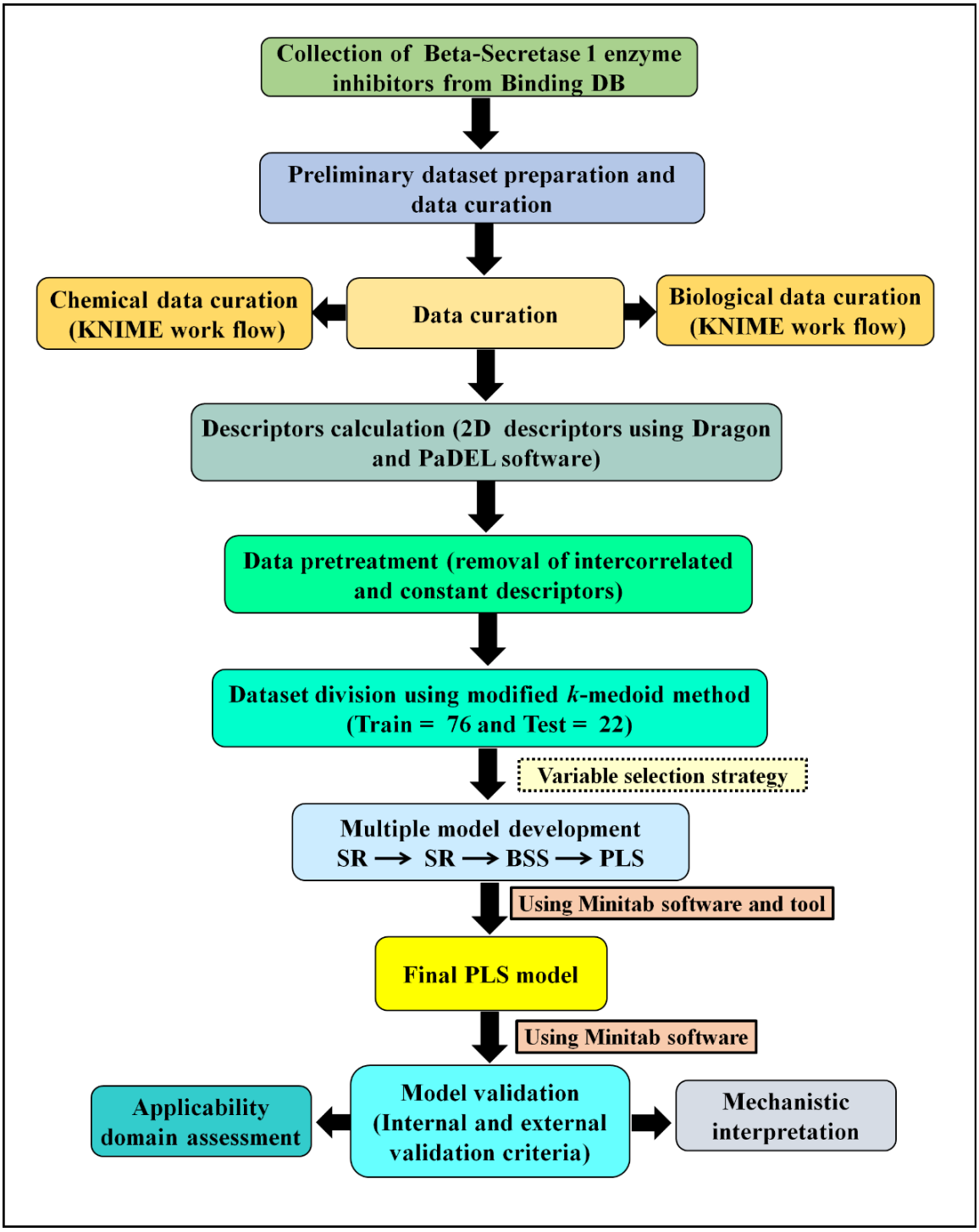


Figure 3.3. Schematic work flow of 2D QSAR model development against BACE1 inhibitors, [PLS = Partial least squares, SR = Stepwise regression, BSS = Best subset selection].

3.2.2. 3D-QSAR pharmacophore modeling

In the current study, 3D-QSAR pharmacophore modeling was for the identification of essential pharmacophoric features, which are necessary for the inhibitory activity against the BACE1 enzyme. The inhibitory activity against the BACE1 enzyme expressed in terms of IC_{50} values was used as the dependent variable for the pharmacophore model development. For the development of the pharmacophore model, we have utilized the compounds (training set) which were selected as the test set compounds in the case of the 2D QSAR model whereas the training compounds in the 2D QSAR model were used as test compounds here [81]. The development of the 3D QSAR pharmacophore model was carried out using the training set compounds and validated using the test set compounds. Before the development of pharmacophore models, we performed conformation generation using training set compounds. After conformation generation, we performed a features mapping protocol for identifying the meaningful pharmacophoric features from the training set compounds was carried out in the BIOVIA Discovery Studio client 4.1 (Available from https://www.3dsbiovia.com/) platform and it resulted in hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrophobic (HYD), hydrophobic (aromatic and aliphatic) and ring aromatic (RA) features. In the development of the pharmacophore model, different parameters were adjusted such as activity, uncertainty value was kept 2, maximum five features containing hydrophobic (HYD), hydrophobic aliphatic, aromatic and ring aromatic (RA), and hydrogen bond acceptors (HBA) were selected and the final models were developed using the FAST method of poling algorithm[323].

3.2.2.1. Validation of developed pharmacophore model

For to validate the best pharmacophore model, we have used different validation parameters such as quantitative and qualitative methods. In terms of test analysis, validation of developed models was performed by mapping the whole test set molecules on the pharmacophore model. It was carried out in the BIOVIA Discovery Studio client 4.1 (Available from https://www.3dsbiovia.com/) platform with the same setting as we have used in pharmacophore model development. The predictive ability of a model to categorize both active and less active compounds has been determined by organizing the molecules with an activity threshold of 1000 nM. In the whole dataset, the training set (n = 22) consists of 12 most active and 10 least active compounds, whereas the test set (n = 76) consists of 32 most active and 44 least active compounds. To judge the quality of the developed pharmacophore model, we have performed the qualitative validation test calculating the confusion matrix (validation parameters namely true positives, true negatives, false positives, and false negatives) based on the observed and predicted activity values obtained from test set analysis[323]. The validation parameters in terms of qualitative analysis used for the pharmacophore model are sensitivity, specificity, accuracy, precision, F-measure, and G-means. According to Aher et al.[323], the selected model is measured to be robust, if all the validation parameter values are more than 60% for both the sets (training and test set). In terms of internal validation, we have performed a cost analysis and selected the model based on the RMSD, correlation, fit values, and cost difference values. We have also performed the Fischer randomization test (F-test) at a confidence level of 95%, to check whether the obtained model is by chance or not. It was carried out by randomly reordering the activity data of training set molecules and developing the model with the same settings as used for the actual pharmacophore model development[323]. According to F-test, the actual model is considered to be a better model, if the results obtained from randomized models are bad quality than the actual model[323]. The validated pharmacophore model could be further utilized for the prediction of the inhibitory activity of the new compound against the BACE1 enzyme.

3.2.2. Molecular Docking studies

Aβ is the main component of pathophysiology in Alzheimer's disease, and the BACE1 enzyme is responsible for the amyloidogenic cleavage of APP-generating Aβ. Consequently, controlling the BACE1 enzyme activity to decrease Aβ is a rational therapeutic goal[324]. Molecular docking is a key tool in structural molecular biology and computer-assisted drug design[313]. The goal of ligand-protein docking is to predict the predominant binding mode of a ligand with a protein of known three-dimensional structure [68]. In the present study, molecular docking studies was employed to understand the interactions pattern of BACE1 enzyme inhibitors within the active site of BACE1 enzyme (the structure of the protein was retrieved from Protein Data Bank[325] with PDB ID: 4ivt (Available from https://www.rcsb.org/structure/4IVT). Molecular docking was performed by using the CDOCKER module of receptor-ligand interaction available in BIOVIA Discovery Studio client 4.1. Before the docking, the active site of the enzyme was defined using the protocol Receptor-ligand Interaction section employing the option "define site from receptor cavities" available in the BIOVIA Discovery Studio client 4.1 platform (Available from https://www.3dsbiovia.com/). The selected inhibitors were subjected to ligand preparation to find a series of ligand conformers (maximum 255). Each orientation was used in the CDOCKER module for molecular docking using CHARMm-based interaction energy using a rigid receptor[326]. The poses are sorted according to CHARMm interaction energy and the top scoring (most negative, thus favorable to binding) poses are kept.

3.3. Study 3- Cheminformatic modeling of β-amyloid aggregation inhibitory activity against Alzheimer's disease

3.3.1. QSAR methodology

3.3.1.1. The Dataset

In this study, 314 heterocyclic compounds (β-amyloid aggregation inhibitors) were collected from the BindingDB database[185] with β-amyloid aggregation inhibitory activity for QSAR model development. The experimental IC_{50} values (nM) of the dataset compounds were converted to pIC_{50} (=-logIC_{50}) values for model development purposes. The dataset compounds utilized in this study followed the same experimental protocol (Thioflavin T (ThT) spectrofluorometric assay method)[327-346]. All the compounds were carefully checked and filtered using different software and tools like KNIME (https://dtclab.webs.com/software-tools), MarvinView, and MarvinSketch (Available from http://www.chemaxon.com). The compounds were drawn by using the MarvinView and saved in the MDL.mol format.

3.3.1.2. Preliminary dataset preparation and data curation

Chemical curation is very important when researchers collect data from different sources[347]. In this work, before the development of the regression models, preliminary dataset preparation and data curation (chemical and biological) strategy were implemented using KNIME work flow (https://dtclab.webs.com/software-tools). The accuracy of the KNIME workflow was confirmed by Mariana et al. 2017[348], Domenico et al. 2018[349], and Fabian P et al. 2015[350]. The dataset was downloaded from BindingDB[185] in a structural data format (SDF) containing important information related to the structure and endpoint values against the β-amyloid aggregation. An identifier was given to every compound present in the dataset, characterizing the name of a respective protein and a serial number. The endpoint values from the dataset file was extracted to classify the compounds in four orders of magnitude. In the end, we incorporated the dataset into chemical and biological curation.

3.3.1.2.1. Reading and storing the information obtained from the Binding database

In this methodology, the "SDF reader" was utilized using the KNIME workflow (https://dtclab.webs.com/software-tools) to read the input file and store the important structural features of the compounds. The downloaded dataset contained all of the essential information related to the compounds such as the molecule name, coordinates, bond counts, bond order, number of rings, end point, biological assay, etc. The compounds with incorrect information were deleted from the source file and the compounds with correct information were saved for further use.

3.3.1.2.2. Elimination of salts, mixtures, inorganics, and organo-metallics from the dataset

In this study, we removed all of the salts, mixtures, in-organics, and organo-metallic compounds before calculating the descriptors using 'RDKit Salt Stripper', 'Connectivity' node, and 'Element Filter node[351], respectively.

3.3.1.2.3. Standardization of chemical structures

Normalization is an important step in QSAR modeling to correct the structural pattern before the calculation of the molecular descriptors. In the current work, the 'RDKit Structure Normalizer' node was applied for

correction of the geometry of the chemical structure using the KNIME workflow (https://dtclab.webs.com/software-tools).

3.3.1.2.4. Biological curation

The molecules obtained from chemical curation were subjected to biological curation. Biological data curation is one of the important steps in QSAR modeling. In the current work, we implemented biological data curation as described by Ambure et al.[59] using the KNIME workflow (http://dtclab.webs.com/software-tools) for duplicate identification and activity cliff determination.

3.3.1.3. Descriptor calculation and data pre-treatment

A pool of 457 descriptors was computed by applying two software tools, namely, Dragon version 7[301] and PaDEL-Descriptor software version 2.20[302]. In this work, only 2D descriptors were computed including constitutional, ring descriptors, connectivity index, functional group counts, atom-centered fragments, 2D atom pairs, atom type E-states, molecular properties, and extended topochemical atom (ETA) indices. After descriptor calculation, we performed the data pre-treatment using the tool Pre-treatment V-WSP version 1.2 (http://dtclab.webs.com/software-tools) to discard the descriptors with incomplete information or with nearly constant values.

3.3.1.4. Dataset division

In this study, we aimed to develop QSAR models having good reliable prediction ability. Therefore, QSAR models were developed by using a training set and validated using new chemical entities, i.e., a test set to check the predictive capacity of the developed models. In this study, the whole data set (n = 314) was divided into a training set (n = 252, 80% of the total number of compounds), and a test set (n = 62, 20% of the total number of compounds) based on a Euclidean distance-based algorithm using the "Dataset Division GUI" developed by our group (https://dtclab.webs.com/software-tools).

3.3.1.5. Multilayered variable selection strategy and model development

Before the development of the final model, we tried to extract the important descriptors from the large pool of initial descriptors using various variable selection strategies. For this purpose, we applied a multilayered variable selection strategy before the development of the final model using multistage stepwise regression (using a suitable stepping criterion, e.g., 'F-for-inclusion 'and 'F-for-exclusion' based on partial F-statistic) followed by a genetic algorithm (GA) followed by the best subset selection, and the final models were built by using partial least squares (PLS) regression techniques. The detailed multi-layered variable selection strategy is schematically represented in **Figure 3.4**.

3.3.1.6. Statistical validation metrics

Validation of the robustness and predictive ability of the developed models is an important step in a QSAR study[77]. In this study, we employed different statistical approaches such as internal and external validation metrics to justify the robustness and predictive quality of the developed models. In the case of internal validation, we determined various statistical metrics such as determination coefficient (R^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), Avg $r_m^2{}_{(LOO)}$, and $\Delta r_m 2$[77]. Higher values of the metrics R^2, $Q^2_{(LOO)}$) and Avg $rm^2_{(LOO)}$ indicated a better fit of the model, but all of these parameters are not sufficient to evaluate the robustness and predictivity of significant models[77]. Thus, we determined other statistical validation parameters (external validation parameters) such as Q^2F_1, Q^2F_2, and r_m^2 parameters like average

r_m^2(test) and Δrm^2 and concordance correlation coefficient (CCC) to assure the significance of the developed models[77]. Moreover, we also performed a Y-randomization test, checked applicability domain criteria, etc. to investigate the robustness of the developed models. The Y-randomization test was performed using the Simca-P software (Available from https://umetrics.com/products/simca) by randomly reordering (100 permutations) the dependent variable[310]. The details of the methodology are depicted in **Figure 3.4**.

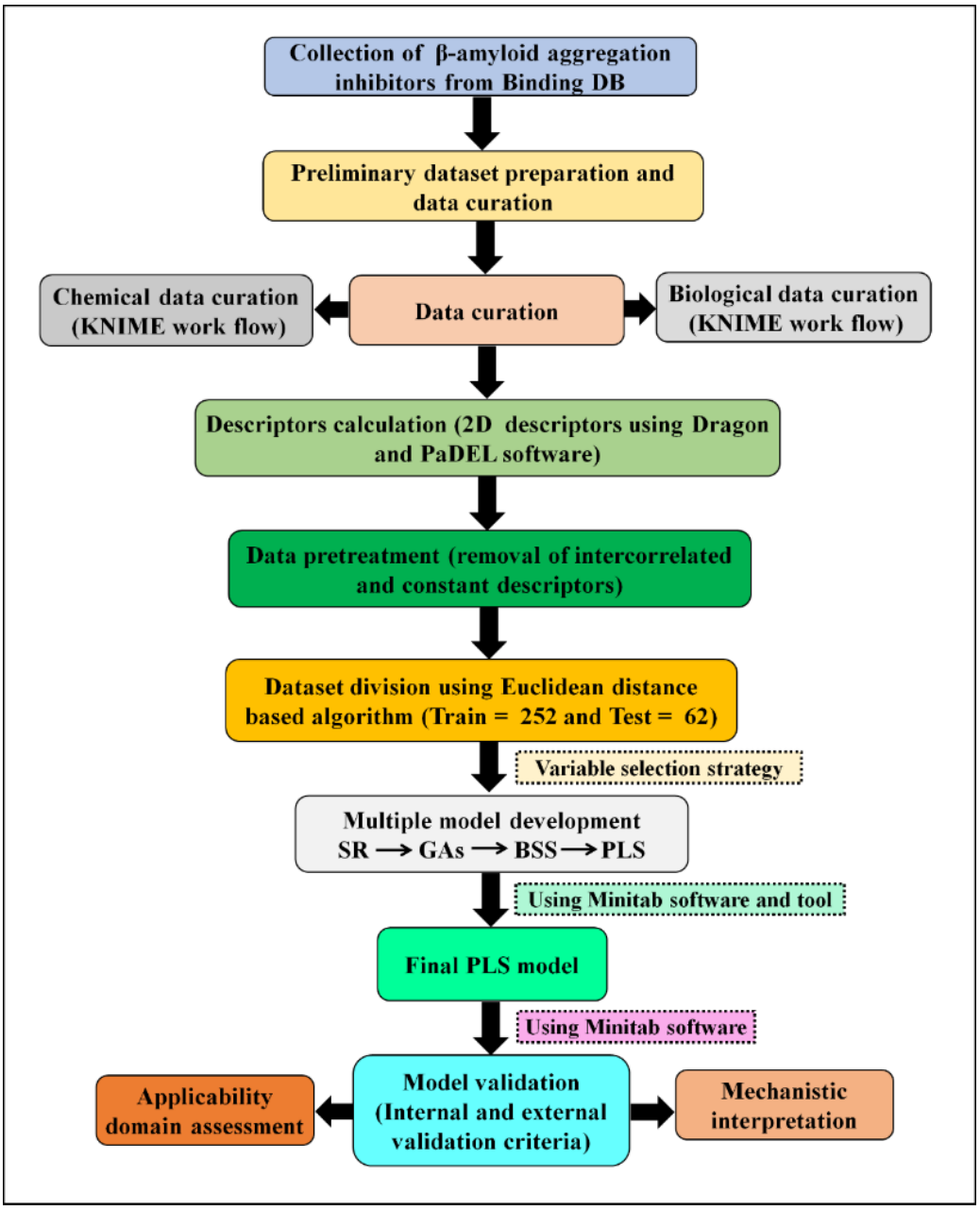


Figure 3.4. Schematic work flow of QSAR model development against β-amyloid aggregation [PLS = Partial least squares, SR = Stepwise regression, BSS = Best subset selection, GAs = Genetic algorithms].

3.3.2. Development and validation of the 3D-pharmacophore model

In the current work, a pharmacophore modeling study was performed to reveal the required features which are essential for β-amyloid aggregation inhibitory activity. The β-amyloid aggregation inhibitory activity stated in terms of IC_{50} (nM) was used as the dependent variable for the development of pharmacophore models. Previously prepared compound structures were used for this study. The dataset was rationally distributed into training (62 compounds) (for model development) and test sets (252 compounds for validation) based on the biological activity values spanned over four orders of magnitude[313]. The BIOVIA Discovery Studio Client 4.1 (Available from https://www.3dsbiovia.com/) platform was used to build the pharmacophore models. The details of the methodology for the development of the pharmacophore model are described by Aher et al.[323]. Validation of the developed models was performed using different parameters like cost analysis, the Fischer randomization test (F-test), and test set prediction to judge the robustness and predictive quality of models as described by Aher et al.[323].

3.3.3. Molecular Docking studies

Investigation of important structural features that will be helpful for the development of novel inhibitors which control the aggregation of β-amyloid was the goal of this study. Here, a molecular docking study was performed to identify the interaction pattern between the β-amyloid peptide (PDB ID: 1IYT, Available from https://www.rcsb.org/structure/1IYT) and selected β-amyloid aggregation inhibitors from the dataset. Molecular docking studies were performed by using the BIOVIA Discovery Studio client 4.1 platform using the CDOCKER module of receptor-ligand interactions[326]. After docking, the generated poses were sorted according to CDOCKER interaction energy, and the top-scoring poses were kept for further analysis.

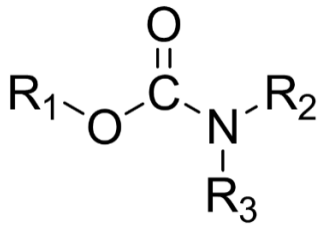
3.4. Study 4- Chemometric modeling of structurally diverse carbamates for the inhibition of acetylcholinesterase enzyme (AChE) in Alzheimer's disease

3.4.1. QSAR modeling

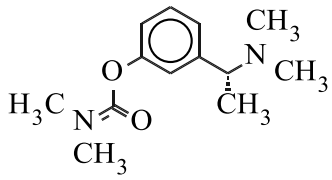
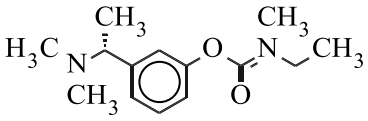
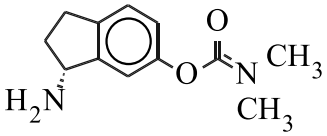
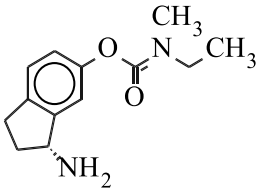
3.4.1.1. Dataset

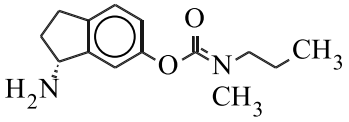
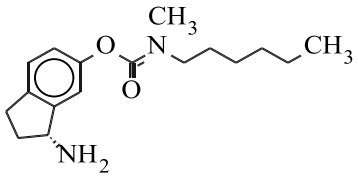
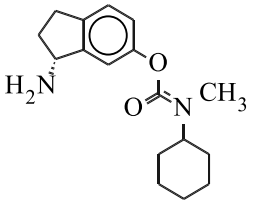
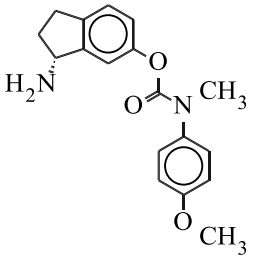
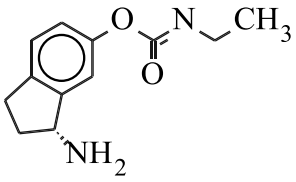
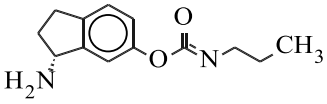
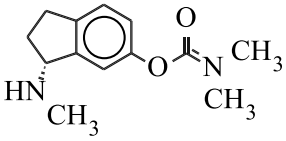
In the current study, a congeneric series of 78 carbamate derivatives[186] with inhibitory activity against the AChE enzyme were collected from the previous published literature. All 78 structures of the AChEI (carbamate derivatives) present in the dataset with their names, structures, and activity (observed and predicted) against the AChE enzyme are depicted in **Table 3.2**. The activity values for all the compounds were measured using the same experimental method (Ellman assay)[186] by the same research group. The activity values of all the dataset compounds expressed as IC_{50} (μM) values were converted to the negative logarithm of IC_{50} (i.e., pIC_{50}) values for model development. The structures were drawn using Marvin Sketch software version 5.9.4 (Available from https://chemaxon.com/products/marvin) and ChemDraw Ultra software version 12.0 (Available from https://www.perkinelmer.com/category/chemdraw). In GQSAR model development, the designation of common scaffold and substitution sites is a prerequisite step as shown in **Table 3.2**[186]. It can be seen that there are three substitution sites R1, R2, and R3 in the congeneric series used in the G-QSAR study. In the GQSAR methodology, every dataset molecule is considered a set of fragments, and the fragmentation scheme is either template-based or user-defined. Once the common scaffold and substitution sites are defined, various descriptors are calculated for each fragment of the molecule. Now using an appropriate variable selection method, the significant descriptors representing the particular substituent sites are selected.

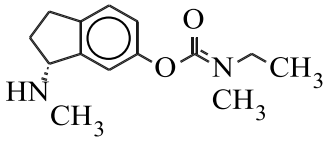
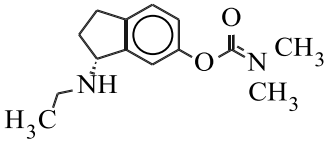
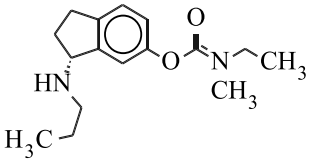
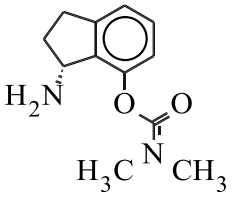
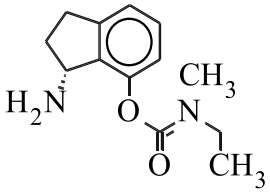
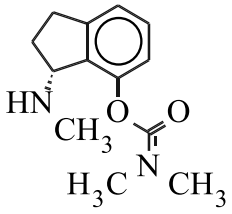
Table 3.2. The list of Carbamate derivatives present in the dataset with their name structure and activity value against the AChE enzyme.

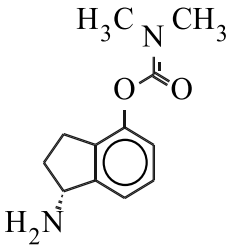
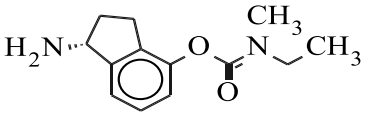
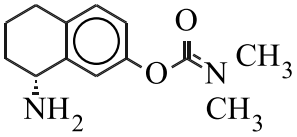
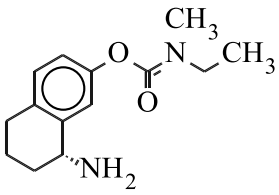
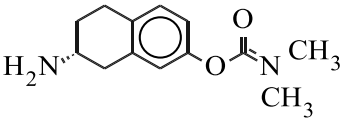
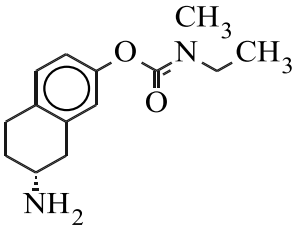


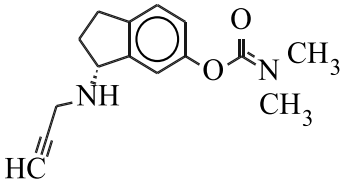
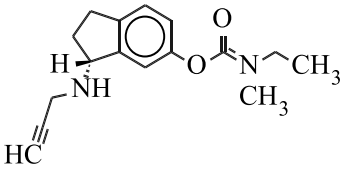
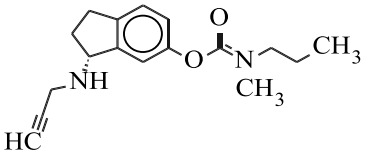
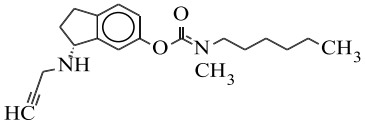
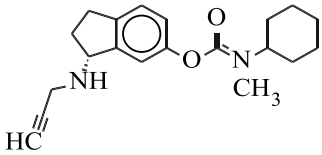
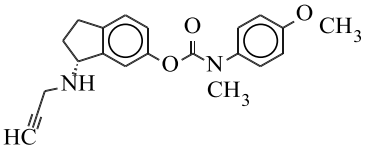
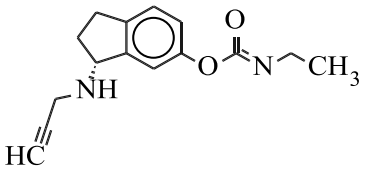
Common scaffold and the substitution sites

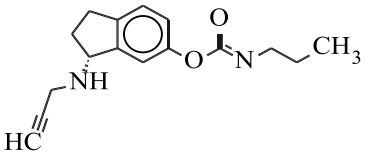
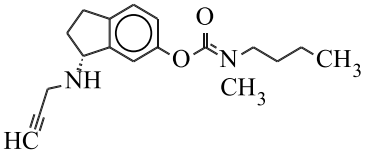
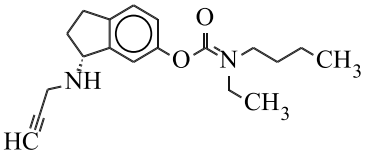
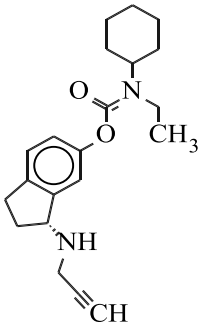
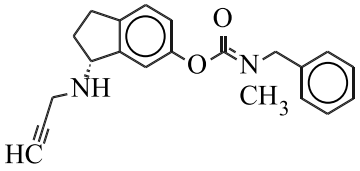
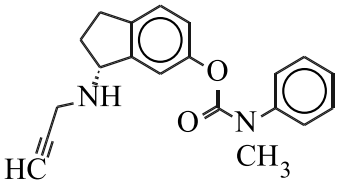
Name	Molecular structures	Observed value (pIC_{50} µM)
1		1.522
2		0.036
3		0.119
4		-1.278

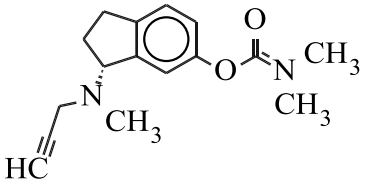
5		-0.863
6		0.275
7		-0.597
8		0.522
9		-1.247
10		-0.170
11		-0.029

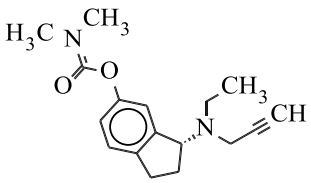
12		-1.586
13		-1.389
14		-0.556
15		0.337
16		-1.021
17		0.292

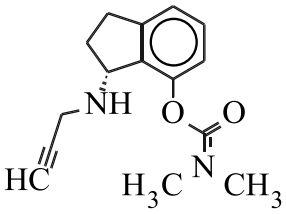
18		2.045
19		1.585
20		-0.170
21		-0.793
22		-0.510
23		-1.900

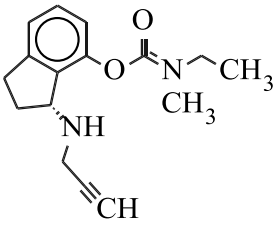
24		-0.462
25		-1.672
26		-1.164
27		-1.195
28		-1.614
29		0.065
30		-1.136

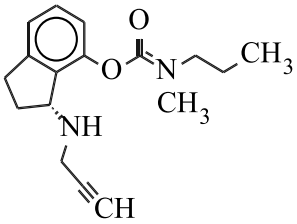
31		-0.376
32		-1.053
33		-1.859
34		-1.252
35		-0.250
36		0.259

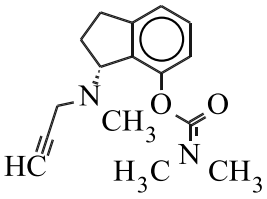
37  -1.110

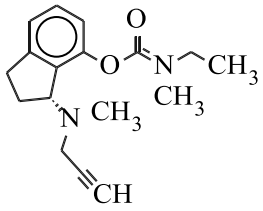
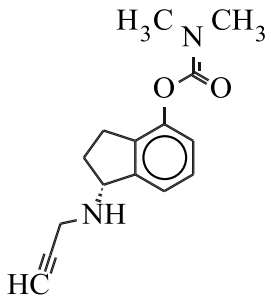
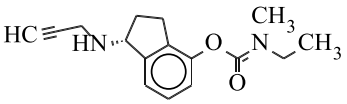
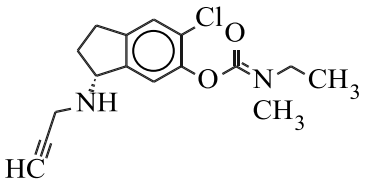
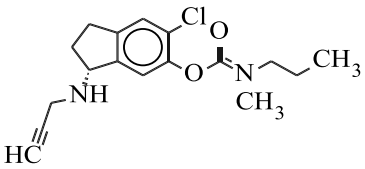
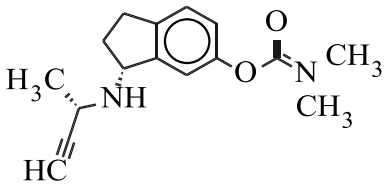
38  -1.252

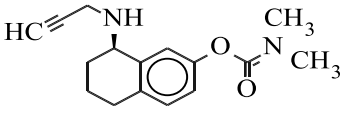
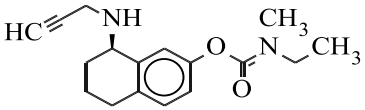
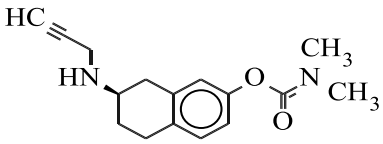
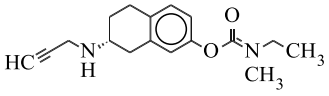
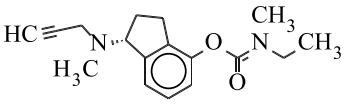
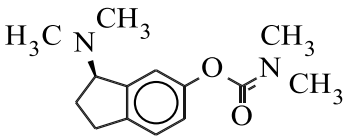
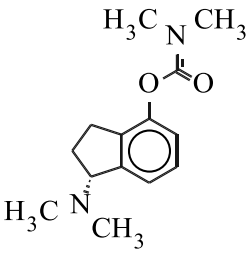
39  -0.397

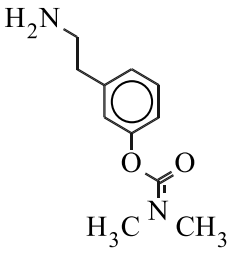
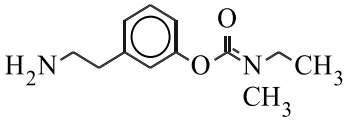
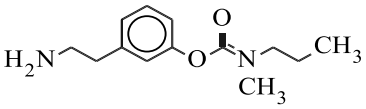
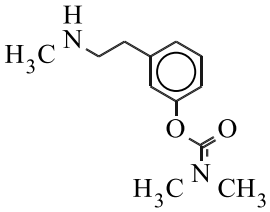
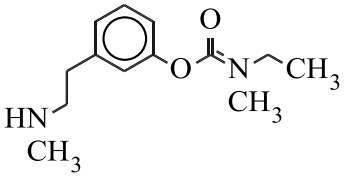
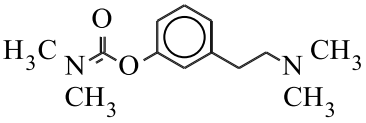
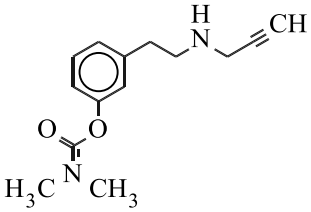
40  -2.720

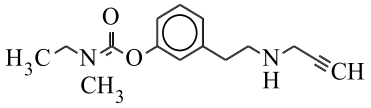
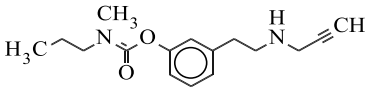
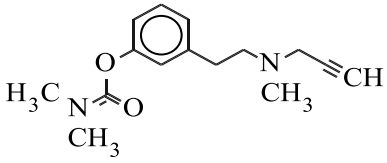
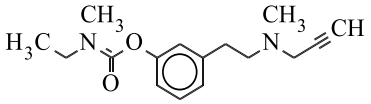
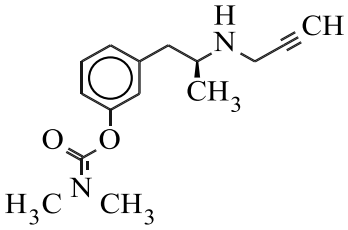
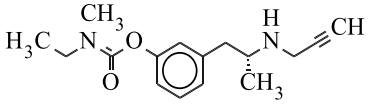
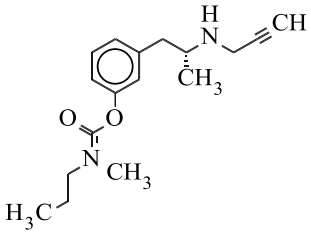
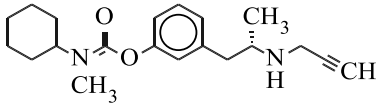
41  -1.653

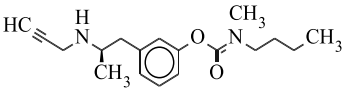
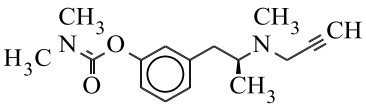
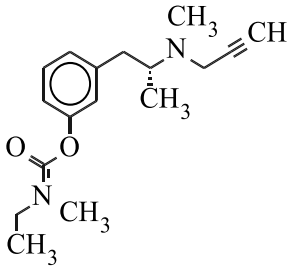
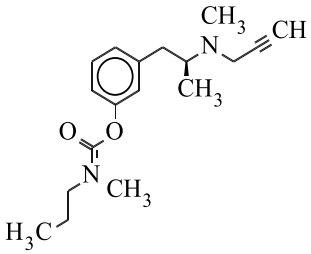
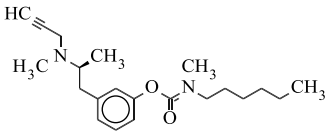
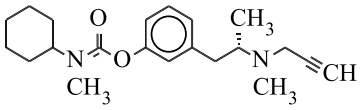
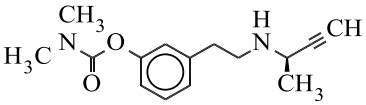
42  -0.845

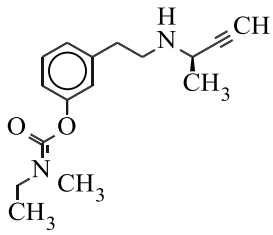
43		-2.642
44		1.275
45		-0.332
46		-1.406
47		-1.642
48		-0.255

49		-0.595
50		-1.719
51		-0.612
52		-2.309
53		-1.173
54		-0.332
55		1.886

56		0.638
57		-1.550
58		-0.892
59		0.552
60		-1.315
61		0.795
62		0.657

63		-1.480
64		-1.187
65		0.070
66		-1.220
67		0.267
68		-1.530
69		-1.281
70		-0.556

71		-1.086
72		-0.214
73		-2.369
74		-1.519
75		-0.485
76		-0.938
77		0.769

78		-1.143

3.4.1.2. Descriptor calculation and data pretreatment

In the case of 2D-QSAR model development, a pool of 174 descriptors was computed from the congeneric series of 78 carbamate derivatives[186] against AChE enzyme using two software tools, namely, Dragon version 7[301] and PaDEL-Descriptor version 2.20 software[302]. We have calculated only 2D descriptors covering constitutional, ring descriptors, connectivity index, functional group counts, atom-centered fragments, 2D atom pairs, atom type E-states, and molecular properties. In the case of GQSAR modeling, a pool of 213 descriptors was calculated for every fragment at the substitution sites (R1, R2, and R3) using the tool VLifeMDS version 3 (Available from https://www.vlifesciences.com/support/request_demo.php) software covering physiochemical and atom type count descriptors. In current investigation, data pretreatment was performed for both sets of descriptors to remove inter-correlated descriptors from the datasets using the tool Pretreatment V-WSP version 1.2 (available from http://dtclab.webs.com/software-tools).

3.4.1.3. Dataset division

The primary aim of this study was to develop QSAR models that are robust enough and capable of making accurate and reliable predictions. Therefore, QSAR models were developed by a training set and validated using new chemical entities, i.e., a test set to check the predictive capacity of the developed models. In this study, the whole data set was divided into a training set and a test set based on activity/property algorithm using the "Dataset Division GUI" developed by our group (available from http://dtclab.webs.com/software-tools). The same strategy was also applied in the case G-QSAR study.

3.4.1.4. Multilayered variable selection and QSAR model development

Before the development of the final models, authors tried to extract the important descriptors from the large pool of initial descriptors using various variable selection strategies[111]. For this purpose, we have applied a multilayered variable selection strategy before the development of the final models using stepwise regression (using a suitable stepping criterion, e.g., 'F-for-inclusion 'and 'F-for-exclusion' based on partial F-statistic)[352] followed by genetic algorithm (GA) followed by best subset selection, and the final models were built using partial least squares (PLS) regression techniques. The same variable selection strategy was applied in both cases (2D- QSAR, and GQSAR). The detail multi-layered variable selection strategy is schematically represented in **Figure 3.5**.

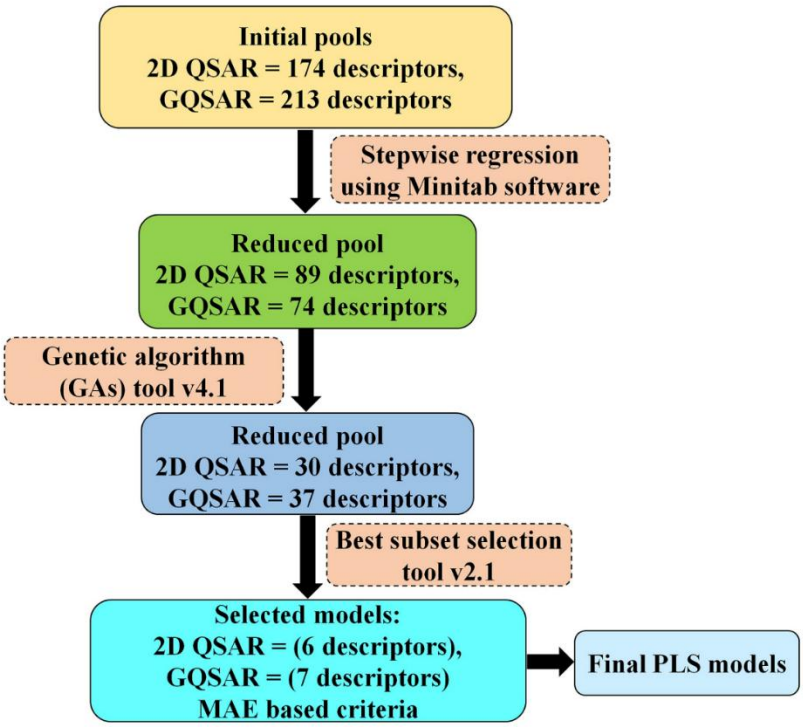


Figure 3.5. Schematic representation of multi-layered variable selection strategy for the development of the model.

3.4.1.5. Statistical validation metrics

To judge the robustness and predictive quality of the developed models is a critical step in the QSAR study[322]. In this study, different statistical approaches such as internal and external validation metrics were employed to justify the robustness and predictive quality of developed models. In the case of internal validation, we have determined various statistical metrics such as determination coefficient (R^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), Avg $rm^2_{(LOO)}$, and Δrm^2 etc.[322]. Higher values of the metrics R^2, $Q^2_{(LOO)}$), and Avg $rm^2_{(LOO}$ indicated a better fit of the model, but all these parameters are not sufficient to evaluate the robustness and predictivity of significant models[322]. Thus, authors have determined other statistical validation parameters (external validation parameters) such as Q^2F_1, Q^2_{F2}, and r^2m parameters like average rm^2 (test) and Δrm^2 and concordance correlation coefficient (CCC) to assure the significance of developed models[322]. Moreover, we have also performed a Y-randomization test, checked applicability domain criteria, etc to investigate the robustness of developed models. The Y-randomization test was performed using the Simca-P software (Available at http://www.minitab.com/en-us/products/minitab/) through random reordering (100 permutations) of the dependent variable [65]. The details of the methodology are depicted in **Figure 3.6**.

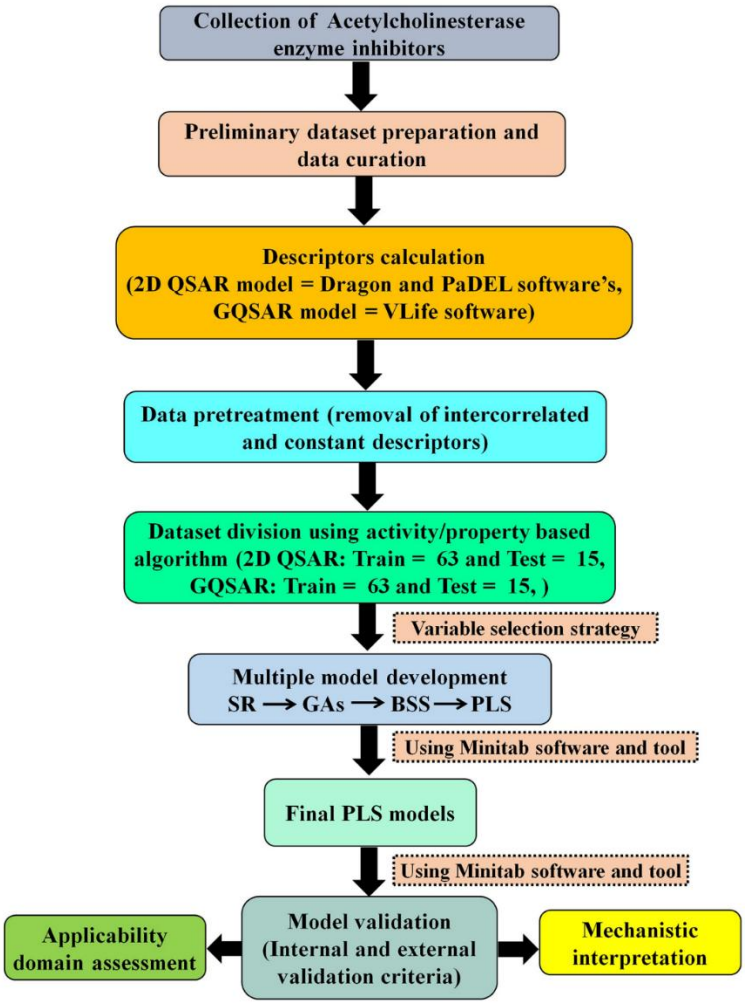


Figure 3.6. Schematic work flow of QSAR model development against AChE enzyme [PLS = Partial least squares, SR = Stepwise regression, BSS = Best subset selection, GAs = Genetic algorithms].

3.4.2. Development and validation of 3D pharmacophore model

In the present study, a pharmacophore modeling study was applied to reveal the required features essential for the inhibitory activity against the AChE enzyme. The AChE enzyme inhibitory activity expressed in terms of IC_{50} (µM) was used as the dependent variable for the development of pharmacophore models. Previously prepared compound structures were used for this study. The dataset was rationally distributed into training (23) (for model development) and test set (55) (For validation) compounds based on the span over four orders of magnitude[313]. BIOVIA Discovery Studio client 4.1 (Available from https://www.3dsbiovia.com/) platform was used to build the pharmacophore models. The details of the methodology for the development of the pharmacophore model are described by Aher et al.[323]. Validation of the developed models was performed using different parameters like cost analysis, Fischer randomization

test (F-test), and test set prediction, to judge the robustness and predictive quality of models as described by Aher et al[323].

3.4.3. Molecular docking studies

In this study, molecular docking studies was performed to identify the interactions of the AChE enzyme (the structure of the protein was retrieved from Protein Data Bank[325] with PDB ID: 4M0E available from https://www.rcsb.org/structure/4M0E) with the most and least active AChE enzyme inhibitors from the dataset. Molecular docking studies were carried out using BIOVIA Discovery Studio client 4.1 (Available from https://www.3dsbiovia.com/) platform using the CDOCKER module of receptor-ligand interaction[326]. After docking, the generated poses were sorted according to CDOCKER interaction energy, and the top scoring (most negative, thus favorable to binding) poses are kept. Schematic workflow for the methodologies adopted in this study given in **Figure 3.7**.

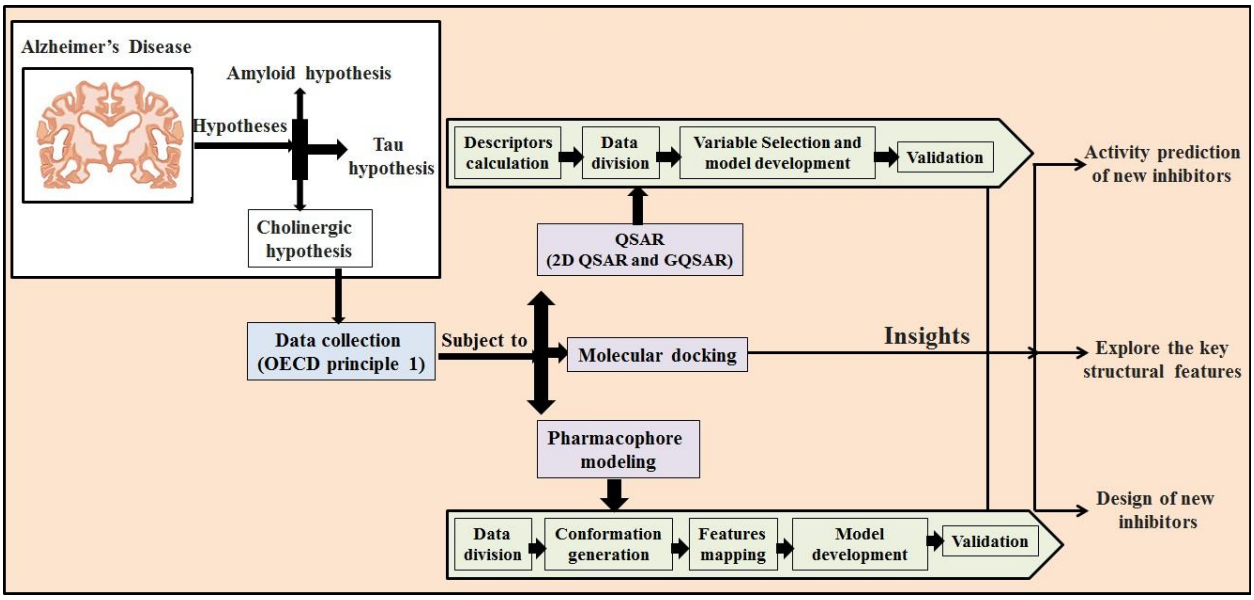


Figure 3.7. Schematic workflow for the methodologies adopted in this study.

3.5. Study 5- In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease

3.5.1. 2D-QSAR analysis

3.5.1.1. Dataset selection

In the current work, authors have used two different datasets against two important targets, namely AChE (number of compounds = 997) and BuChE (umber of compounds = 761) enzymes collected from previously published literature[187-299]. In both the datasets, some compounds have inhibitory activities both against AChE and BuChE enzymes (number of compounds 198), therefore, we have used the difference of activities against AChE and BuChE enzymes as the dependent variable (AChE-BuChE) for modeling selectivity, and the third dataset has been prepared. The datasets comprise diverse classes of heterocyclic compounds, and the experimental activity of each compound is expressed in IC_{50} (nM) value, derived following the same bioassay protocol (modified colorimetric Ellman assay. For the model's development, we have converted the IC_{50} values to pIC_{50} (pIC_{50} = -logIC_{50}) values as customary in QSAR analysis, and all the compounds from each dataset set were drawn using MarvienSketch (Available from https://chemaxon.com/products/marvin), followed by cleaning of molecules. Then, hydrogen was added, and the file was saved in SDF format. Before descriptor calculation, authors have carefully checked all structures in the datasets for the development of significant 2D-QSAR models.

3.5.1.2. Descriptor calculation and pretreatment

In this section, only 2D descriptors were calculated using two software, namely Dragon 7[301] (covering functional group counts, constitutional, ring descriptors, connectivity index, atom centered fragments, 2D atom pairs, atom type E-states, and molecular properties) and PaDEL descriptor 2.20[302] (for extended topochemical atom indices). After descriptor calculation, we performed data curation utilizing the tool Pretreatment V-WSP version 1.2 (available at http://dtclab.webs.com/software-tools) to eliminate the descriptors with missing or near-constant values.

3.5.1.3. Dataset division

In this method, authors have split the whole dataset into training and test sets based on the sorted activity-based algorithm using the "Dataset Division GUI" developed by our group (available at http://dtclab.webs.com/software-tools). Initially, the dataset was divided into training and test sets randomly[353] for 30 trials. Then the whole range of activities was sorted in ascending order, and every fourth compound was assigned to the test set. Finally, an attempt was made to rationalize the division process, in which the division was performed so that points representing both training and test sets were distributed within the whole descriptor space occupied by the entire dataset, and each point of the test set was close total east one point of the training set[353]. The training set was used for the development of models, and the test set compounds for the validation of the developed models.

3.5.1.4. Multi-layered variable selection strategy and 2D QSAR model development

Selection of important descriptors from the large pool of initial descriptors using different variable selection methodologies is an important task in QSAR modeling. Based on this concept, a multi-layered variable selection strategy was adopted before the development of the final models. In the multi-layered variable selection, first, we have applied stepwise regression in successive iterations using the Minitab software

(Available from http://www.minitab.com/en-us/products/minitab/) followed by genetic algorithm (GA) using the GA software (available from http://dtclab.webs.com/software-tools), and afterward, we applied double cross-validation (DCV) (available from http://dtclab.webs.com/software-tools) in case AChE and BuChE enzyme inhibitor models. On the other hand, in the case of the selectivity-based models we have applied best subset selection (BSS) (available from http://dtclab.webs.com/software-tools), and partial least squares (PLS) regression technique was used for the development of final models in all cases. Additionally, we have applied additional selection strategies to check the statistical quality of the developed models. Thus, to address the above said situation, we have performed (1) PLS without stepwise regression (SR) + GA, (2) PLS + SR, (3) PLS + GA, (4) PLS + SR + GA to find out the optimal combinations of predictors. From the developed models, we have found that the models reported in this study are more robust than models obtained from the above strategies. The details of the steps used in the multi-layered variable selection strategy are schematically represented in **Figure 3.8**.

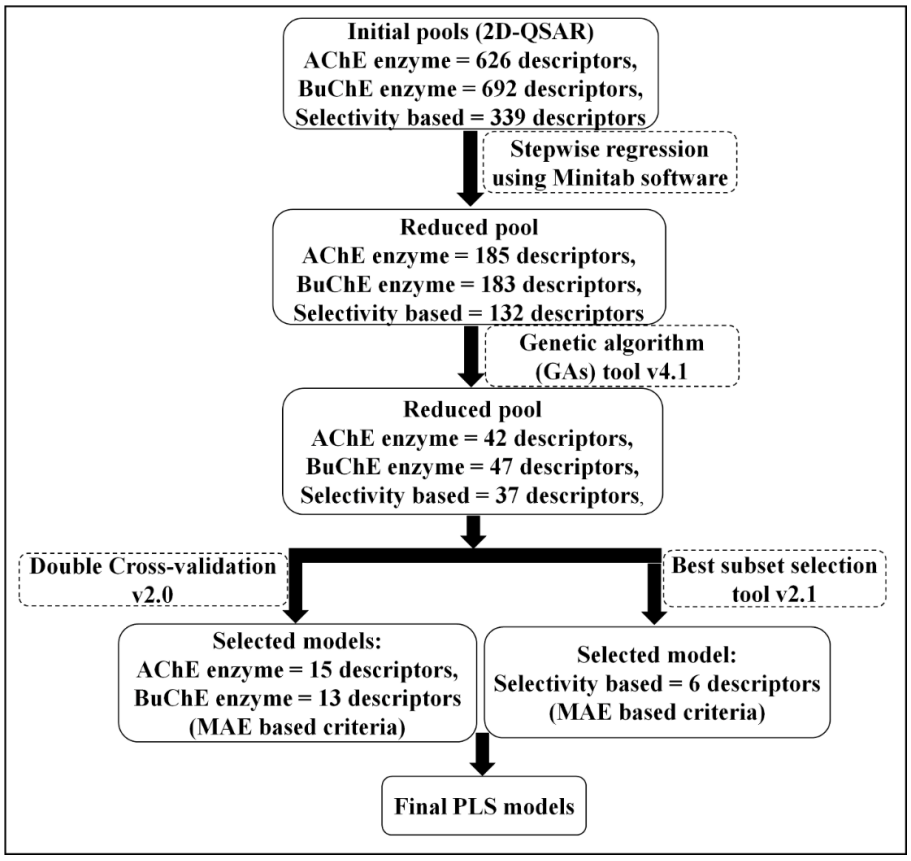


Figure 3.8. Schematic representation of multi-layered variable selection strategy.

3.5.1.5. Statistical validation of the generated2D-QSAR models

In the present work, authors have applied various statistical methodologies like internal (determination coefficient (r^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), Avg $r_m^2{}_{(LOO)}$ and Δrm^2) and external (Q^2F_1, Q^2_{F2}, r_m^2 parameters like average r_m^2 (test) and Δr_m^2 and concordance correlation coefficient (CCC)) validation methods to assure the significant level of the generated models[77, 310]. Additionally, we

have also implemented the Y-randomization test[309], checked applicability domain criteria, etc using Simca-P 10.0 software (available from https://umetrics.com/products/simca). The detailed methodologies are depicted in **Figure 3.9**.

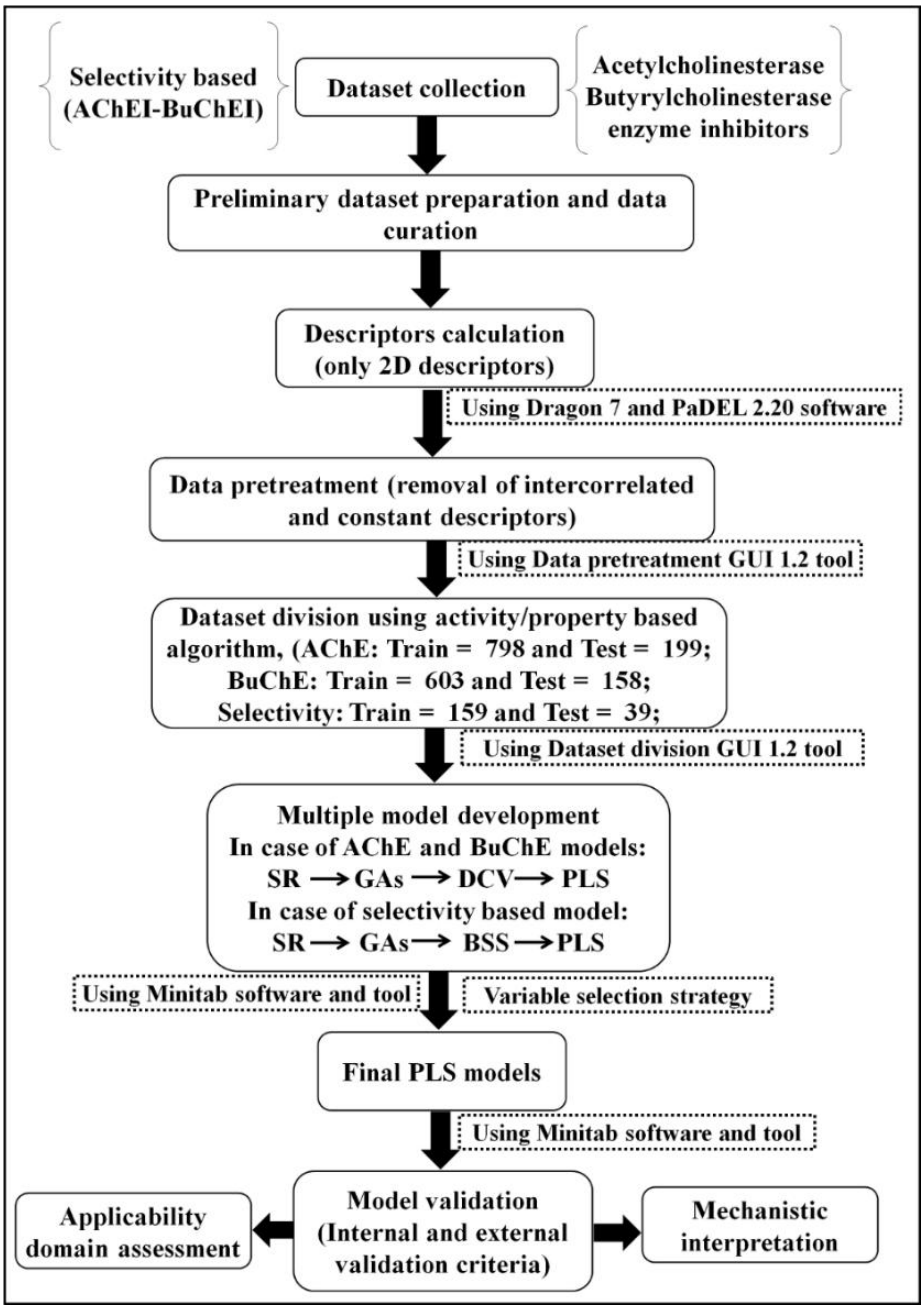


Figure 3.9. Schematic work flow of QSAR model development against AChE, BuChE and based on selectivity [PLS = Partial least squares, SR = Stepwise regression, BSS = Best subset selection, GAs = Genetic algorithms, DCV = double cross validation].

3.5.2. Molecular docking studies

In this analysis, molecular docking studies was performed to investigate the binding pattern of molecules (most and least active compounds from the dataset) with the respective enzymes, such as AChE and BuChE. The enzyme crystal structures were extracted from the protein databank[325] with the PDB id: 4M0E (structure of human acetylcholinesterase bound to Dihydrotanshinone I, available at https://www.rcsb.org/structure/4M0E) and 4BDS (crystal structure of human butyrylcholinesterase in Complex with Tacrine, available at https://www.rcsb.org/structure/4BDS) (AChE and BuChE enzymes, respectively). The molecular docking study was performed by using BIOVIA discovery studio client 4.1 (Available from https://www.3dsbiovia.com/) platform with the CDOCKER module of receptor-ligand interaction as discussed by Pal S *et al* and Kumar V *et al*[313, 112]. Before the docking analysis, we prepared the target enzyme and selected inhibitors using the protein and ligand preparation protocol available in BIOVIA discovery studio client 4.1. The active site in the enzyme was defined by the "define and edit binding site" protocol available in BIOVIA discovery studio client 4.1. After docking analysis, we sorted the generated poses as per the CDOCKER interaction energy, and the poses with top-scoring values were used for further analysis. The obtained poses were validated using the bound ligand present in the crystal structure of the enzyme. Based on the number of interactions and the active residues interacting with the bound ligand, we have selected the final pose for further study. From the ligplot (**Figure 3.10**), we can see the number of interactions and active residues responsible for the significant interaction in the crystal structure of AChE and BuChE enzymes with their bound ligand.

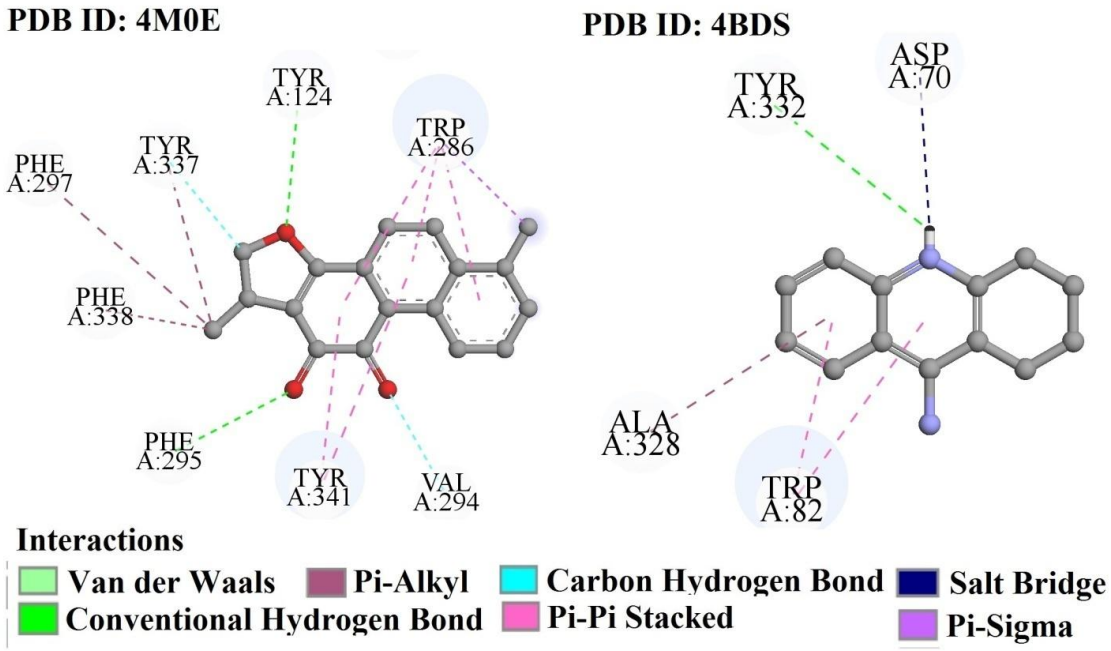


Figure 3.10. Ligplot of AChE (4M0E) and BuChE (4BDS) enzyme and with their bound ligands.

3.6. Study 6- Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease

3.6.1. 2D-QSAR modeling

3.6.1.1. Data collection, curation, and dataset preparation

The activity data against twelve major targets of AD were collected from the BindingDB database[185] (available from www.bindingdb.org). Initially, 80 inhibitors against 5-hydroxytryptamine receptor 6 (5-HT6) following cell-based Radio ligand binding assay, 1733 compounds against acetylcholinesterase (AChE) enzyme following modified colorimetric Ellman assay, 2507 compounds against butyrylcholinesterase (BuChE) enzyme following modified colorimetric Ellman assay, 905 inhibitors against beta-secretase 1 (BACE1) enzyme following FRET (fluorescence resonance energy transfer) assay, 262 β-amyloid aggregation inhibitors following Thioflavin T-based fluorometric assay, 225 compounds against Cyclin Dependent Kinase 5 (CDK-5) protein following Scintillation proximity assay, 217 inhibitors against gamma-secretase enzyme following cell-based sandwich ELISA assay, 132 compounds against Glutaminyl Cyclase (QCs) enzyme following Continuous Spectrometric Assay, 159 inhibitors against glycogen synthase kinase-3 beta (GSK-3β) enzyme following Kinase-Glo reagent based luminescence assay, 170 compounds against Monoamine oxidase B (MAO-B) enzyme following Fluorometric method, 356 compounds against N-methyl-D-aspartate (NMDA) receptor following Fluorescence-based assay, 289 compounds against Phosphodiester 10A (PDE 10A) enzyme following TR-FRET assay were collected from the BindingDB database[185] (available from www.bindingdb.org). The datasets comprise diverse classes of heterocyclic compounds, and the experimental activity values are quantified in IC_{50} (nM). Before proceeding with the development of the regression models, we executed preliminary dataset preparation and data curation (chemical and biological) strategy using a KNIME workflow (available from https://dtclab.webs.com/software-tools) following the protocol as discussed by Kumar et al.[112]. The precision of the KNIME workflow was confirmed by Mariana et al., 2017[348], Domenico et al., 2018[349], and Fabian et al., 2015[350]. After dataset curation, screening of the activity datasets was performed to find the common compounds having dual inhibitory activity against the listed targets. Accordingly, we have found that the 43 compounds with dual inhibitory activities both against AChE and BACE1 enzymes, 83 compounds against AChE and β-amyloid, 113 compounds against AChE and BuChE enzymes, 52 compounds against AChE and MAO-B enzymes, 20 compounds against BACE1 and GSK-3β enzymes, 51 compounds against BuChE and BACE1enzymes, 23 compounds against BuChEI and β-amyloid, 48 compounds against BuChE and MAO-B enzymes, 21 compounds against AChE and GSK-3β enzymes and 21 compounds against BuChE and GSK-3β enzymes were retained and used for the development of the respective QSAAR and selectively based models. Marvin Sketch software version 5.5.0.1 (available from https://chemaxon.com) was used to draw the chemical structures of all compounds, followed by the addition of explicit hydrogens in the structures. The activity end point values (IC_{50}) were converted to the negative logarithmic scale, pIC_{50}, as customary in QSAR modeling.

3.6.1.2. Computation of the molecular descriptors and data pretreatment

In this section, authors have calculated only 2D descriptors using software, namely the alvaDesc (v2.0.12) tool (available from https://www.alvascience.com/alvaDesc/) covering atom-type E-state indices, 2D Atom Pairs, 2D autocorrelations, 2D matrix-based descriptors, atom-centered fragments, 2D Autocorrelation,

connectivity indices, constitutional indices, ETA indices, functional group counts, information indices, MDE descriptors, molecular properties, P_VSA-like descriptors Rotatable Bonds Count Descriptor, Pharmacophore descriptors, Ring descriptors, Rule of Five Descriptor and Topological indices. After descriptor calculation, we have executed data pretreatment employing the tool Pretreatment V-WSP version 1.2 (available from http://dtclab.webs.com/software-tools) to remove the descriptors with missing or near constant values.

3.6.1.3. Dataset division

After data pretreatment, all the datasets were divided into training and test sets. In this work, the division of the data sets was implemented following three different dataset division methods, namely, activity-property, Euclidean distance-based, and modified k-medoid clustering techniques using "Dataset Division GUI" version 1.2 and "Modified k-Medoid" version 1.3 software tools, respectively (Available from http://teqip.jdvu.ac.in/QSAR_Tools/). The training set was used for the development of models, and the test set compounds for the validation of the obtained models. The datasets containing less than 20 molecules in the whole dataset have been modeled by the application of the "small dataset modeler_beta version" (available from http://teqip.jdvu.ac.in/ QSAR_Tools/) without dividing the dataset into training and test sets. Instead of dividing the small dataset into training and test sets, the double cross-validation (DCV) technique is utilized here to model for small data sets[354-356]. Therefore, all potential combinations (k) of the validation set, which contains r compounds, and the calibration set, which contains n-r compounds, are calculated[354-356]. This is because the inner loop does not produce the "modeling set" (containing *n* compounds)[354-356]. The software enables the user to define the number of compounds to be kept in the validation set (*r*) depending on how the calibration and validation sets are made[354-356]. Genetic algorithm-multiple linear regression (GA-MLR) models are generated using calibration set chemicals[354-356]. Several internal and external validation metrics are generated for each of the chosen models during the thorough double cross-validation process. Furthermore, for each MLR model, the software generates partial least squares (PLS) regression models.

3.6.1.4. Multi-layered variable selection strategy and model development

In the current investigation, authors have adopted a multi-layered variable selection strategy to extract the meaningful and important descriptors before developing the final model. In this approach, initially, we have applied stepwise regression in successive iterations using the Minitab software (Available from https://www.minitab.com/en-us/products/minitab/) using with the whole pool of descriptors, followed by a genetic algorithm using the GeneticAlgorithm_v4.1 software (available from http://dtclab.webs.com/software-tools), with a reduced pool of descriptors. Finally, we have implemented the best subset selection (available from http://dtclab.webs.com/software-tools) on the reduced pool of descriptors obtained from the genetic algorithm step. Finally, the acquired pool of descriptors was used to develop the final model. All of the final QSAR, selectivity, and QSAAR models were developed using the partial least squares (PLS) regression method, except the QSAAR models (between β-amyloid and BuChE enzyme inhibitory activity, and BACE1 enzyme and BuChE enzyme inhibitory activity), and selectivity based models (between BACE1 and GSK-3β enzyme, BuChE and BACE1 enzyme, AChE and GSK-3β enzyme and BuChE and GSK-3β enzyme inhibitors), which were developed by using the multiple linear regression (MLR) techniques.

3.6.1.5. Statistical validation of the developed 2D-QSAR models

To establish a model's significance and reliability in terms of robustness and prediction accuracy, statistical validation is one of the most important steps in the model development process. In the current work, authors have calculated different internal and external validation metrics to establish that the developed models are robust and predictive enough to satisfy the acceptability criteria. In terms of internal validation metrics, we have calculated metrics like determination coefficient (R^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), Avg $r_m^2{}_{(LOO)}$, and $\Delta r_m^2{}_{(LOO)}$) to perform the validation of the training set compounds. Better values of these metrics (R^2, $Q^2_{(LOO)}$), Avg $r_m^2{}_{(LOO)}$, and $\Delta r_m^2{}_{(LOO)}$) indicate a better fit and robustness of the model[77, 310]. Since the internal validation metrics are insufficient to assess the predictive accuracy and robustness of the developed model, comprehensive validation of test set compounds using various external validation parameters is required like Q^2F_1, Q^2F_2, r_m^2 parameters like average r_m^2 (test), and Δr_m^2 (test) and concordance correlation coefficient (CCC)) validation methods to guarantee the predictive nature of the developed models[77, 310]. Furthermore, we have also performed the Y-randomization test[309], applicability domain criteria (DModX (distance to model) in the X-space), etc using Simca-P 10.0 software (Available from https://umetrics.com/products/simca).

3.6.1.6. Database preparation and activity prediction using developed 2D-QSAR models

To predict the inhibitory activity using developed models, authors have used four chemical drug-like databases, namely, Asinex database (338604 compounds) (available from https://www.asinex.com/screening-libraries-(all-libraries), InterBioscreen (IBS) database (552793 compounds) (available from https://www.ibscreen.com/), NCI Open Database (265242 compounds) (available from https://cactus.nci.nih.gov/download/nci/), and Zinc[12] Database (17900742 compounds) (available from https://zinc12.docking.org/subsets/drug-like). Before the prediction, we developed the alvaModel by establishing 2D QSAR model descriptors against each listed target and then converted the alvaModel into the alvaRunner project file using software, namely the alvaModel v2.0.4 tool (available from https://www.alvascience.com/alvaModel/). The established alvaRunner project files were individually used to compute the predicted values of the above databases' compounds; the validated models were capable of precisely predicting the inhibitory activity of the majority of the compounds, as suggested by 'alvaRunner version 2.0.4' tool (https://www.alvascience.com/alvarunner/) product of Alvascience solution. We have predicted the inhibitory activity of these compounds considering the applicability domain of our PLS-based 2D QSAR models against the respective targets.

3.6.2. Similarity-based Read-Across prediction

Read-across prediction is a similarity-based *in silico* technique that predicts the biological response of unknown compounds based on known activity values[357-360]. In this study, the chemical read-across approach based on machine learning was employed to estimate the activity of the test set chemicals using the modeled descriptors. For a successful prediction with Laplacian kernel similarity-based (LK), Gaussian kernel similarity-based (GK), and Euclidean distance-based (ED) estimations, we have optimized the hyperparameter sigma (σ) and gamma (γ) respectively using validation sets. For the optimization, the initial training set is randomly divided into sub-training and sub-test sets in a 3:1 proportion. The sub-training and sub-test sets were then subjected to 'Read-across v4.1' (Available from https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home) with different σ and γ values. During optimization, the other tool parameters, including the number of nearby training compounds, the

distance threshold, and the similarity threshold were kept constant. The optimized setting has been selected by checking the external validation metrics (Q^2F_1 and Q^2F_2). Finally, the optimized setting was combined with the original training and test sets to get the final prediction. To obtain the best predictions, we gradually reduced the number of similar training compounds from 10 to 2.

3.6.3. Molecular docking study

In this investigation, the molecular docking study was performed using the most and least active compounds from the initial datasets and also the top predicted compounds from the chemical databases to identify the interaction pattern with the respective targets. The crystal structure of the targets such as AChE enzyme (PDB ID: 4M0E), BACE1 enzyme (PDB ID: 4ivt), β-amyloid aggregation (PDB ID: 1IYT), BuChE enzyme (PDB ID: 6EZ2), Cyclin-dependent kinase 5 (PDB ID: 3O0G), Gamma-secretase enzyme (PDB ID: 6IYC), Glutaminyl Cyclase (QC) enzyme (PDB ID: 3PBB), GSK-3β enzyme (PDB ID: 5F94), MAO-B enzyme (PDB ID: 2V5Z), NMDA receptor (PDB ID: 1PBQ), and PDE 10A (PDB ID: 6MSA) were extracted from the protein databank (available from https://www.rcsb.org/). In the case of 5-HT6 protein, there are no experimental structures available in the protein data bank, so we have retrieved the predicted protein structure from the AlphaFold Protein Structure Database (Available from https://alphafold.ebi.ac.uk/entry/P50406) with the UniProt: P50406, Source organism: Homo sapiens (Human), and AlphaFold id: AF-P50406-F1-model_v2. To confirm the reliability of the predicted structure, we have validated the structure by Ramachandran plot server (Available from https://swift.cmbi.umcn.nl/servers/html/ramchk.html) and found Ramachandran Z-score: -5.259 which represents the good quality of the model (see **Figure 3.11**). The molecular docking study was executed using the Biovia Discovery Studio client 4.1 (Available from https://www.3dsbiovia.com/) platform following the protocol discussed by Robertson et al.[326] and Kumar et al.[361]. Before molecular docking, the protein was prepared by checking for any missing residues, having explicit hydrogen added, and generating the active site. The active site was generated using the Biovia Discovery Studio client 4.1 platform from the ligand binding domain of the bound ligand and generating the site 'from the current selection' program in the 'receptor-ligand interaction module' of the software. The bound ligand was taken out after active site generation for new molecule docking. In the case of the 5-HT6 protein, we have predicted the multiple active sites at the surface of the protein using the Biovia discovery studio 4.1 client platform from the "define and edit binding site" using the module "generate active site from receptor cavities", and docked the ligand in each site to identify the favorable binding site (identified most favorable active site coordinate X: -18.945, Y: -0.896, Z: 11.313, the radius of sphere 19.299). To prepare ligands, the selected compounds were run through the Discovery Studio platform's 'small-molecule module', where several ligand conformers were formed. Each of these generated conformers was subsequently employed in the CDOCKER module for molecular docking using a CHARMm-based molecular dynamic scheme. The CDOCKER interaction energy parameter (kcal/mol) was examined for all receptor-ligand complexes, and the highest-scoring (more negative; hence favorable to binding) poses with only non-covalent interactions (ionic bonds, hydrophobic interactions, hydrogen bonds, etc.) were kept for future investigation. A graphic representation of the complete methodologies implemented in this study is shown in **Figure 3.12**.

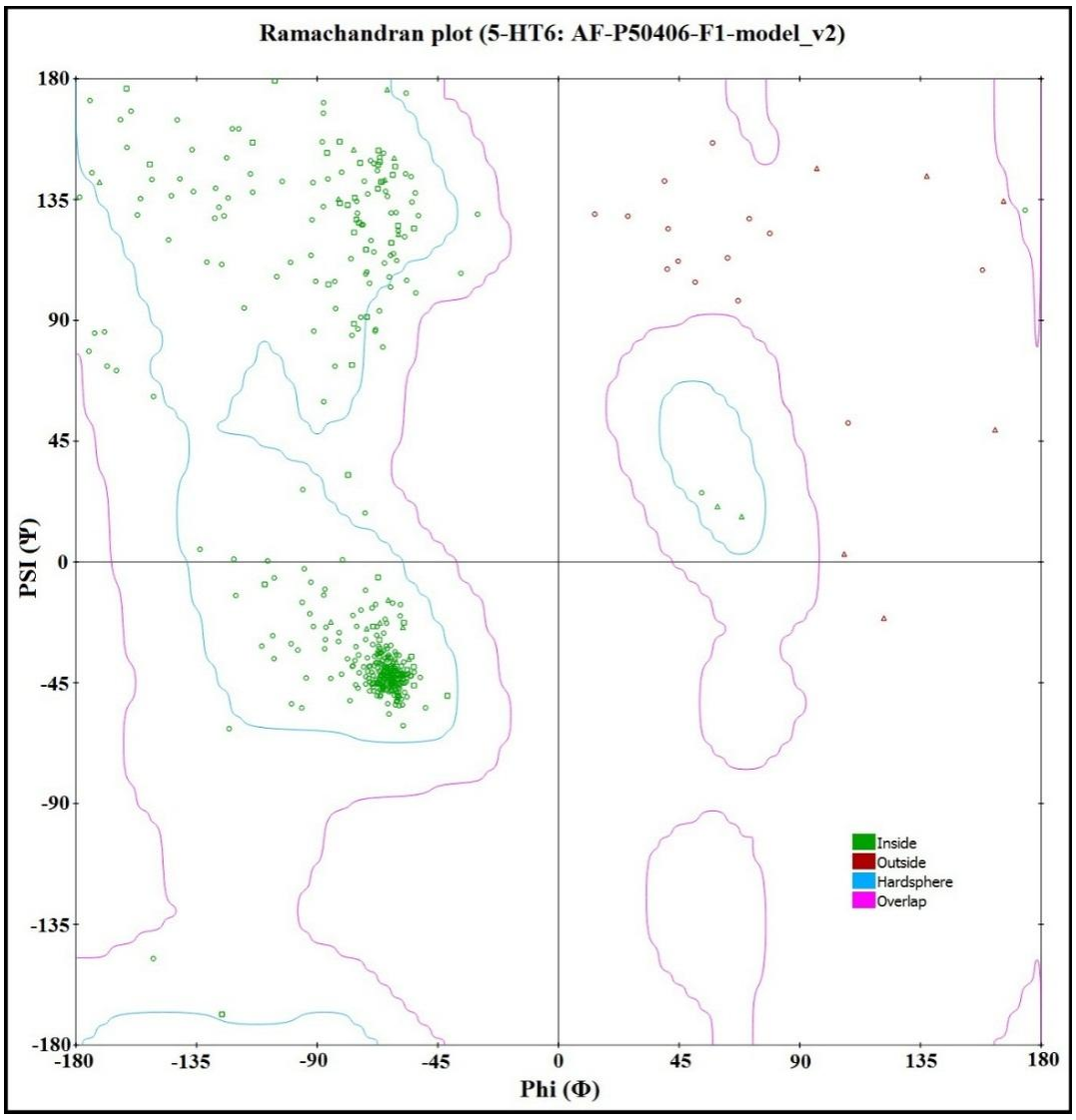


Figure 3.11. Ramachandran plot for 5-HT6 protein model (UniProt: P50406, Source organism: Homo sapiens (Human), and AlphaFold id: AF-P50406-F1-model_v2).

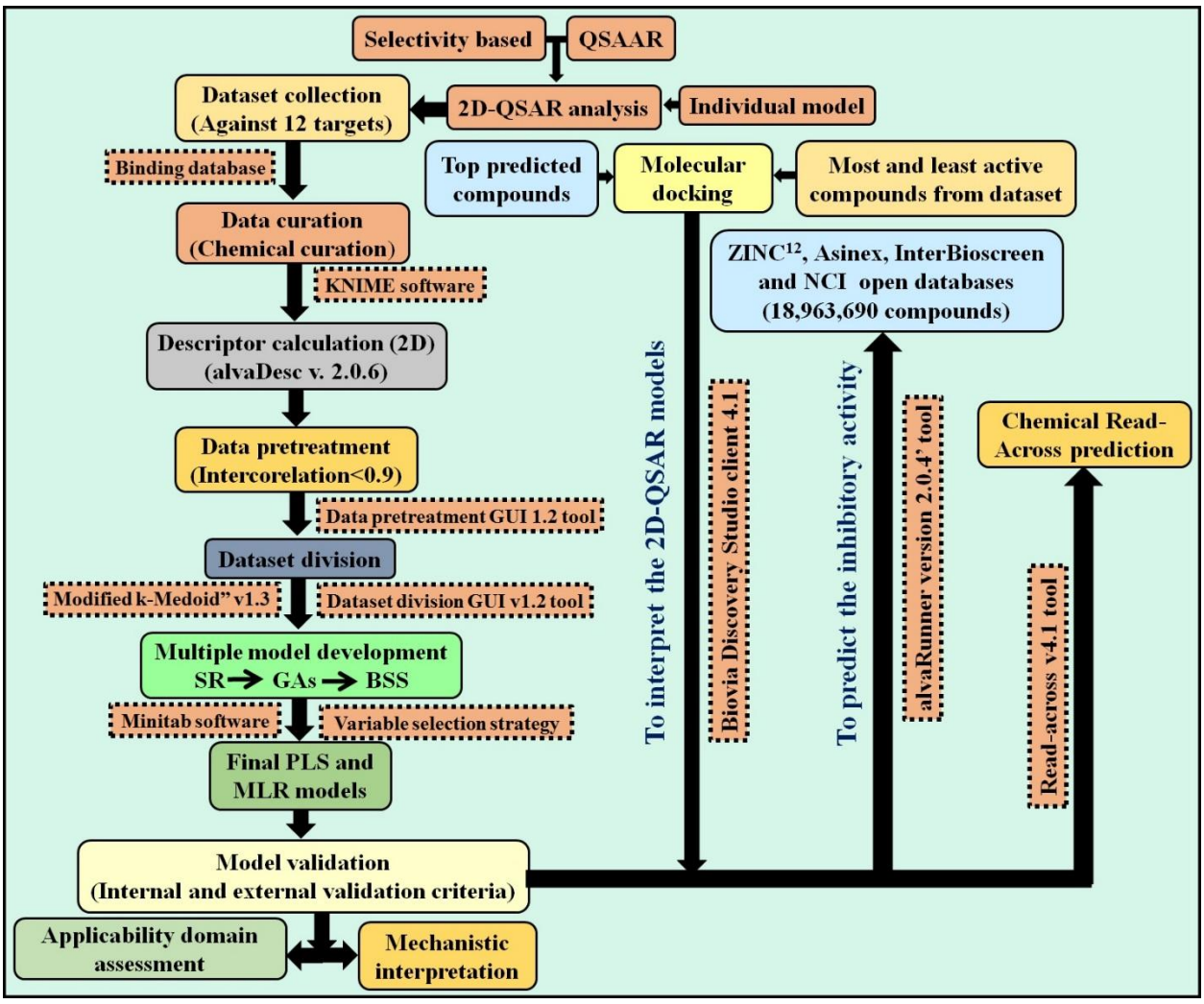


Figure 3.12. The framework of the methodologies implemented in this investigation.

CHAPTER - 4

RESULTS AND DISCUSSIONS

Chapter 4: Results and discussions

4.1. Study 1- A Multi-layered Variable Selection Strategy for QSAR Modeling of Butyrylcholinesterase Inhibitors

In this study, 2D QSAR modeling and molecular docking study were performed to identify the important structural features responsible for the inhibition of the BuChE enzyme. Here, we will discuss the results obtained from this study, which will include the selected 2D QSAR model and its validation, molecular docking analysis, and finally, the identified important structural features against the BuChE enzyme.

4.1.1. 2D QSAR modeling analysis

4.1.1.1. Mechanistic interpretation of modeled descriptors

The reported PLS model was developed by using 15 descriptors with corresponding latent variables of 6. The R^2 (0.664), Q^2 (0.650), and R^2_{pred} values (0.657) of the PLS model were higher than 0.6 (**Equation 4.1**) which indicated the acceptability and predictive ability of the model. Thus, the results obtained from the PLS model (**Equation 4.1**) suggested that the models are acceptable in terms of fitness, stability, and classical predictivity measures. The descriptors appearing in the model (see **Table 4.1**) define the structural and functional requirements which can improve the inhibitory activity of molecules against the BuChE enzyme. The proximity of the observed and predicted values for the BuChE enzyme inhibitors in the data set can be further established from the scatter plot as shown in **Figure 4.1.** The quantitative contributions of similar/dissimilar descriptors are given in the loading plot (similar descriptors are placed in close proximity), and the interrelationships between the X-variables and the Y-response are depicted in the loading plot **(Figure 4.2)**. Additionally, we have also performed a Y-Randomization test to check whether the model was obtained by any chance or not. The results (R^2_{rand} = -0.000805 and Q^2_{rand} = -0.147) obtained from the randomized model suggested that the developed model was not obtained by any chance correlation (**Figure 4.2**).

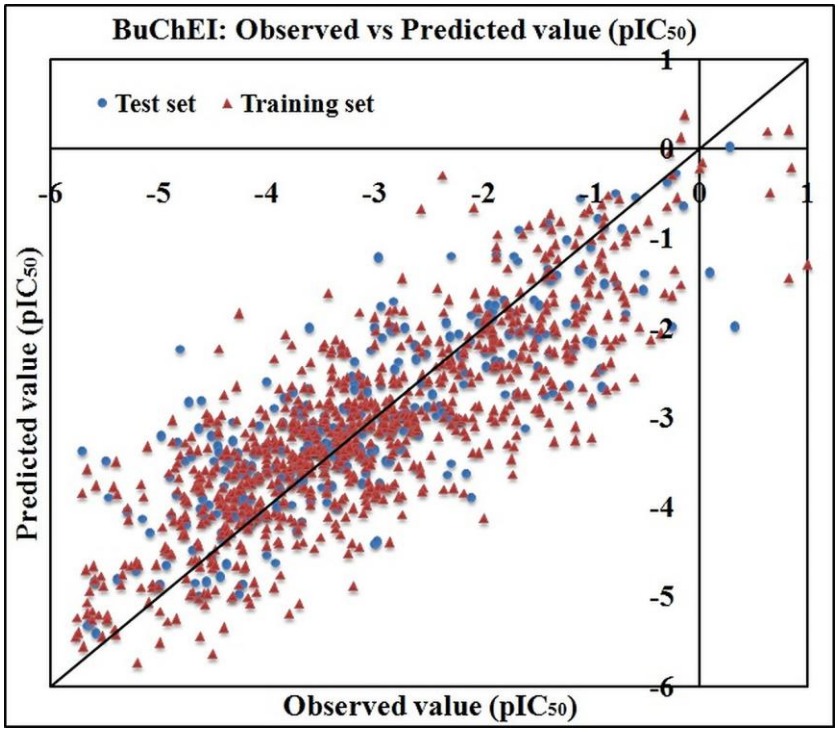


Figure 4.1. The scatter plot of observed and predicted values of the final PLS model against the BuChE enzyme.

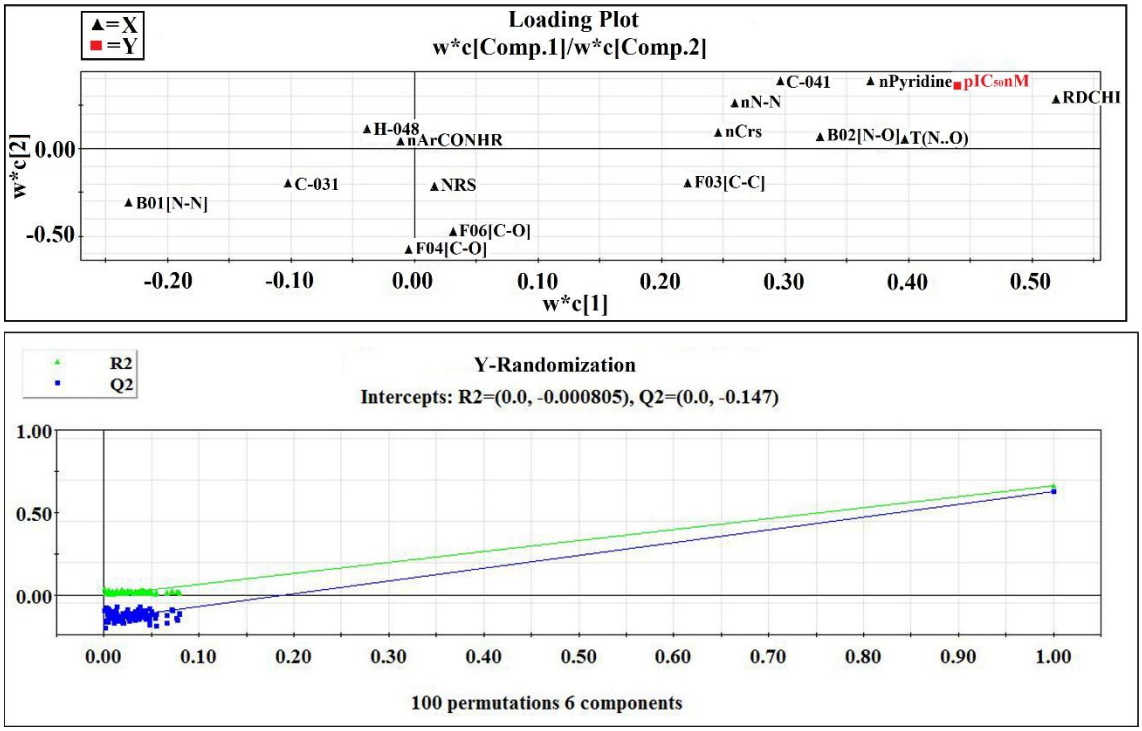


Figure 4.2. Loading plot and Randomization plot for final PLS model against the BuChE enzyme.

The developed PLS regression model is shown below:

$$pIC50(nM) = -6.41 + 1.15 \times RDCHI - 0.08 \times F04[C-O] + 0.46 \times nPyridines + 0.004 \times T(N..O) - 0.07 \times F06[C-O] + 1.38 \times C-041 + 0.12 \times B02[N-O] - 0.015 \times F03[C-C] + 0.06 \times nCrs + 3.10 \times nN-N - 1.34 \times B01[N-N] + 0.18 \times H-048 - 0.43 \times NRS - 1.33 \times C-031 + 1.05 \times nArCONHR$$ **Equation 4.1**

$n_{training}$ = 848, R^2 = 0.664, R^2_{adj} = 0.662, Q^2_{LOO} = 0.650, LV=6, n_{test} = 282, Q^2F_1 = 0.657, Q^2F_2 = 0.657

The descriptors in the PLS model are arranged accordingly to their importance and then described separately. The significance level of the modeled descriptors towards the BuChE inhibitory activity is computed based on a variable importance plot (VIP) (**Figure 4.3**). The VIP defines the importance of each variable obtained from the final PLS model that is responsible to regulate the BuChE inhibitory activity. As per the VIP plot, the significance level of the modeled descriptors is established to be in the following manner: RDCHI, F04[C-O], nPyridines (N..O), F06[C-O], C-041, B02[N-O], F03[C-C], nCrs, nN-N, B01[N-N], H-048, NRS, C-031, and nArCONHR. According to the regression coefficient plot, among these descriptors, RDCHI, C-041, nN-N, T(N..O), nPyridines, nArCONHR, nCrs, H-048 and B02[N-O] contributed positively but C-031, F03[C-C], F06[C-O], B01[N-N], F04[C-O] and NRS descriptors contributed negatively towards the BuChE inhibitory activity as shown in **Figure 4.3**.

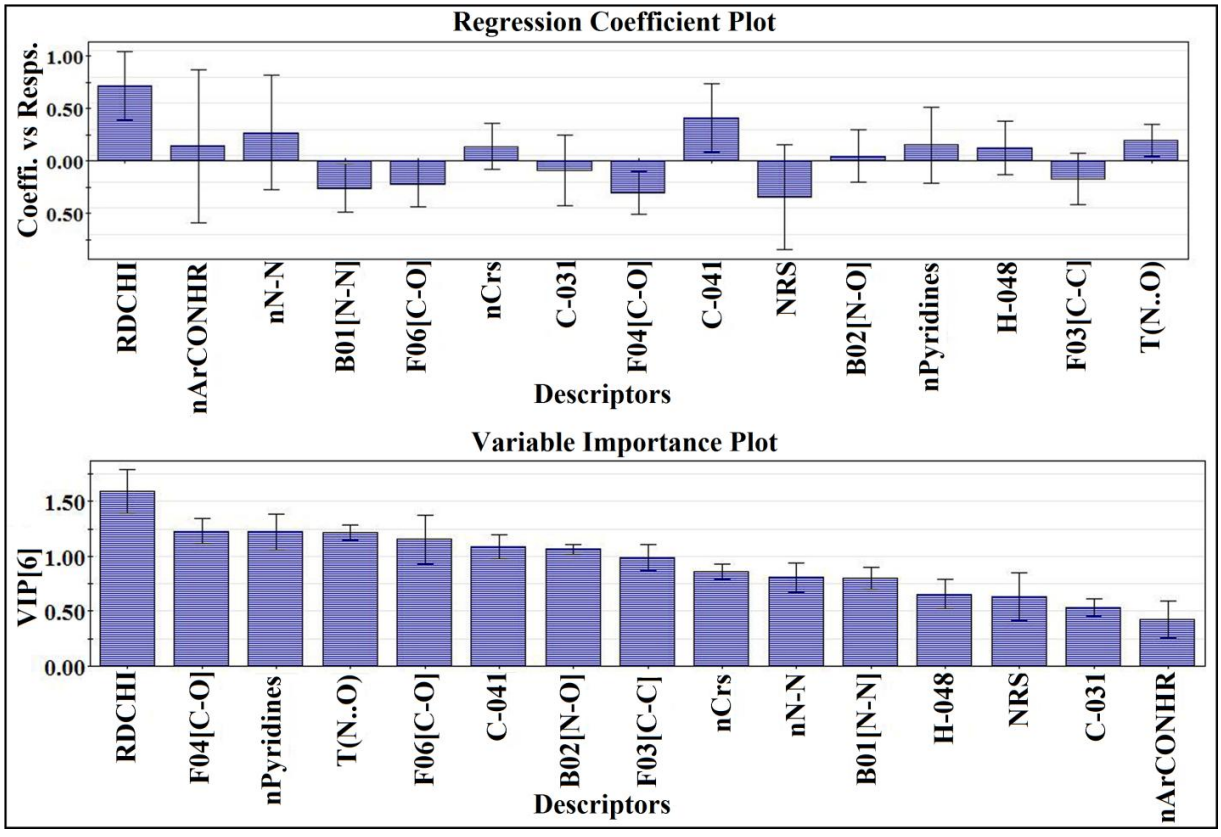


Figure 4.3. Regression coefficient plot and variable importance plot (VIP) of final PLS model against BuChE enzyme.

Table 4.1. Contribution, definition, and mechanism of all the descriptors obtained from the PLS model.

S. No.	Name of Descriptors	Contribution	Family/Short description	Mechanism
1	RDCHI	+ve	Connectivity indices: Reciprocal distance sum Randic-like index	Hydrophobic interactions
2	nArCONHR	+ve	Functional group counts: Number of secondary amides (aromatic)	Hydrogen bonding and electrostatic interactions
3	nN-N	+ve	Functional group counts: Number of N hydrazines	Hydrogen bonding interaction
4	B01[N-N]	-ve	2D Atom Pairs: Presence/absence of N - N at topological distance 1	Hydrogen bonding interaction
5	F06[C-O]	-ve	2D Atom Pairs:- Frequency of C - O at topological distance 6	Hydrogen bonding interaction
6	nCrs	+ve	Functional group counts:- Number of ring secondary C(sp3)	π-π interactions
7	C-031	-ve	Atom-centred fragments:- X--CR--X	Electrostatic interactions
8	F04[C-O]	-ve	2D Atom Pairs:- Frequency of C - O at topological distance 4	Hydrogen bonding interaction
9	C-041	+ve	Atom-centred fragments:- X-C(=X)-X	Electrostatic interactions
10	NRS	-ve	Ring descriptors:-Number of ring systems	Hydrophobic interactions
11	B02[N-O]	+ve	2D Atom Pairs:- Presence/absence of N - O at topological distance 2	Hydrogen bonding and electrostatic interactions
12	nPyridines	+ve	Functional group counts:- Number of Pyridines	Hydrogen bonding and electrostatic interactions
13	H-048	+ve	Atom-centred fragments:- H attached to C2(sp3)/C1(sp2)/C0(sp)	Electrostatic interactions
14	F03[C-C]	-ve	2D Atom Pairs:- Frequency of C - C at topological distance 3	Hydrophobic interactions
15	T(N..O)	+ve	2D Atom Pairs:- Sum of topological distances between N..O	Electrostatics and π -π Interaction

The descriptor, RDCHI, simply characterizes the size and branching of molecules. Its value increases with molecular size but decreases with molecular branching. It can be calculated through the Randic-like formula (**Equation 4.2)** as shown below.

$$RDCHIindex = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} aij.(RDSi.RDSj)^{-1/2} \qquad \textbf{Equation 4.2}$$

Here, A is the number of vertices and a_{ij} is equal to 1 only for pairs of adjacent vertices and zero otherwise, RDS is the sum of reciprocal distance[362]. The descriptor contributes positively towards the BuChE inhibitory activity as suggested by the positive regression coefficient, as shown (in **Figure 4.4**) in compounds **264** (pIC_{50}: 1), **266** (pIC_{50}: 0.823), and **690** (pIC_{50}: 0.629) (containing descriptor values 5.482, 5.277 and 5.246 respectively)**.** These compounds have a large molecular size and less molecular branching. Conversely, compounds **292** (pIC_{50}: -5.63), **293** (pIC_{50}: -4.62), and **835** (pIC_{50}: -5.13) (containing descriptor values 2.351, 2.608 and 2.727 respectively) are low active and have higher branching **(Figure 4.4).** The 2D atom pair descriptor, F04[C-O], denotes the frequency of C-O at the topological distance 4. The BuChE enzyme inhibitory activity is inversely correlated to the numerical value of this descriptor as indicated by its negative regression coefficient. The frequency of the C-O fragment at the topological distance 4 may reduce the inhibitory activity of the BuChE enzyme. The higher number of C-O fragments correlates to lower inhibitory activity as observed in (**Figure 4.4**) compounds **883** (pIC_{50}: -4.41) and **642** (pIC_{50}: -5.53) (containing descriptor values 8 and 15 respectively), while a lower numerical value of this descriptor correlates to higher inhibitory activity as observed in **(Figure 4.4)** compounds **757** (pIC_{50}: -0.819) and **752** (pIC_{50}: -0.991).

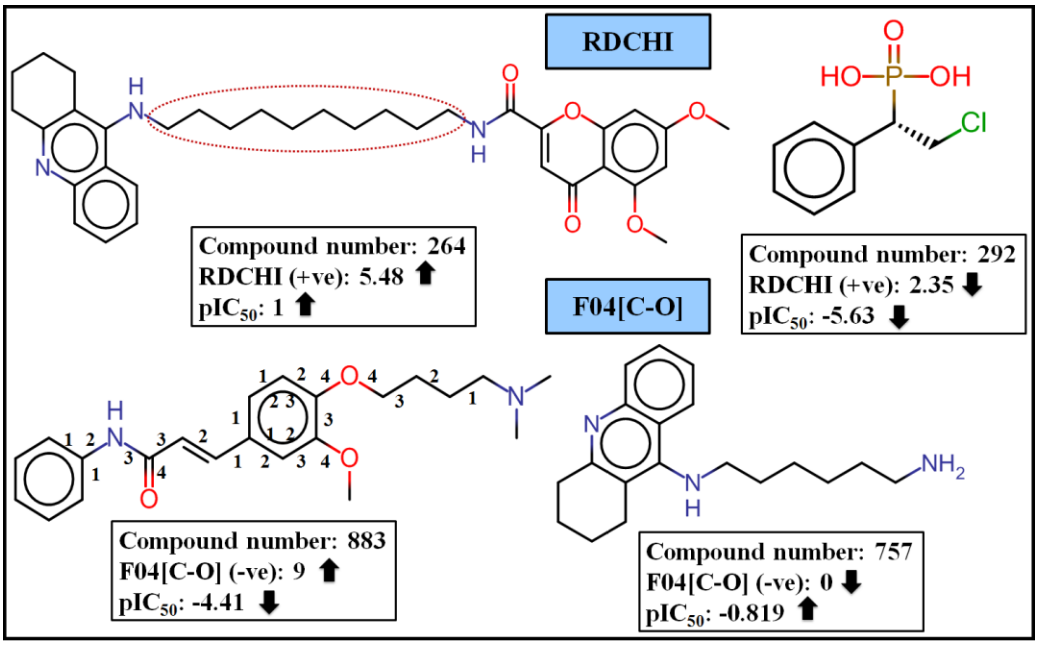


Figure 4.4. Contribution of RDCHI and F04[C-O] descriptors on BuChE enzyme inhibition.

The functional group count descriptor, nPyridines, describes the number of Pyridine rings present in the compounds. The positive regression coefficient of this descriptor indicates that the presence of Pyridine rings in the compounds may enhance the inhibitory activity against BuChE enzyme as found in **(Figure 4.5)** compounds **690** (pIC_{50}: 0.621), **688** (pIC_{50}: 0.017), **697** (pIC_{50}: 0.85) and **695** (pIC_{50}: 0.645) (containing

1 pyridine ring in each case) and vice versa in case of compounds **190** (pIC_{50}: -5.74), **204** (pIC_{50}: -5.02) and **849** (pIC_{50}: -4.91) (containing no such fragments) as given in **Figure 4.5**. The pyridine ring with free electron-pairs and having no active atoms are "passive hydrophilic" moieties and are capable of forming "hydrogen-bridges" with other polar molecules. The structural polarity of pyridine makes it hydrophilic. Another 2D atom pair descriptor, T(N..O), stands for the sum of topological distances between N..O. This descriptor contributes positively towards the BuChE inhibitory activity as indicated by the positive regression coefficient. Thus, the molecules bearing higher topological distance between **(N..O)** fragment may have higher BuChE inhibitory activity as shown in **(Figure 4.5)** compounds **264** (pIC_{50}: 1) and **690** (pIC_{50}: 0.621) (containing descriptor values 202 and 237 respectively) whereas in contrary, compounds **190** (pIC_{50}: -5.74) and **841** (pIC_{50}: -2.23) which do not contain any such fragment shows less BuChE inhibitory activity **(Figure 4.5)**. From this observation, it can be concluded that the topological distances between nitrogen and oxygen atoms should be higher for better inhibitory activity against BuChE.

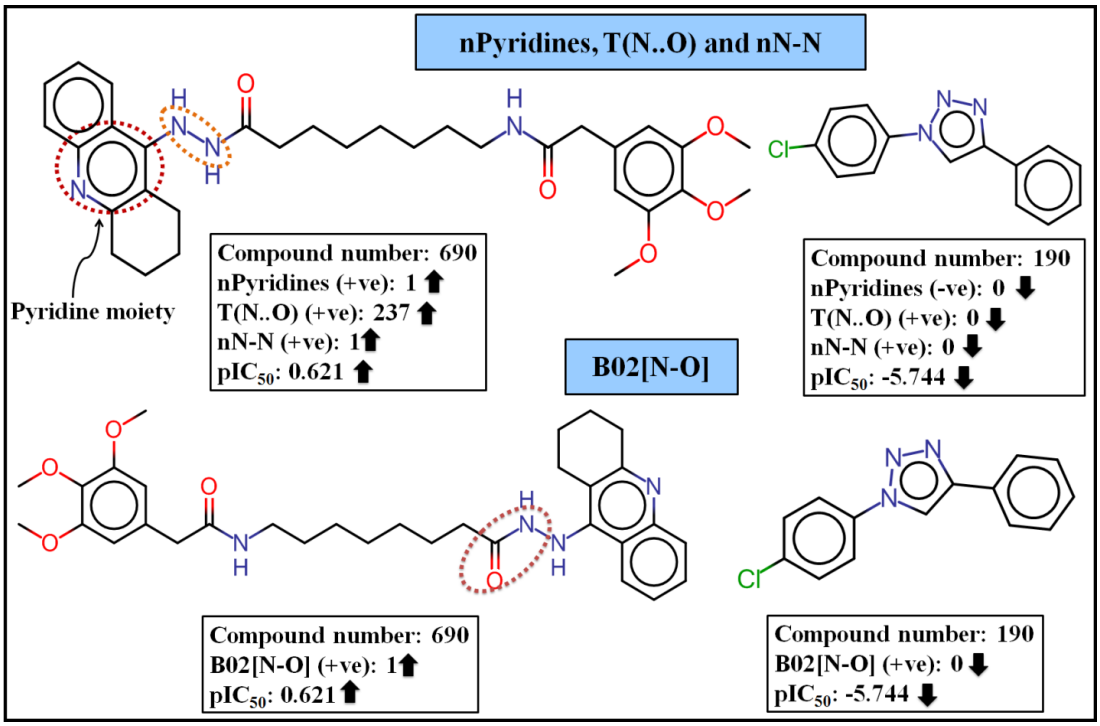


Figure 4.5. Contribution of nPyridines, T(N..O), nN-N and B02[N-O] descriptors on BuChE enzyme inhibition.

Another 2D atom pair descriptor, F06[C-O], indicates the frequency of the C-O fragment at the topological distance 6. The negative regression coefficient of this descriptor suggests that the descriptor is inversely proportional to the BuChE inhibitory activity as observed in the case of compounds **757** (pIC_{50}: -0.819) and **758** (pIC_{50}: -0.44) (having higher enzyme inhibitory activity as the corresponding numerical descriptor value is in the lower range) whereas the reverse is observed in case of compounds **880** and **642** having lower enzyme inhibitory activity (pIC_{50}= -4.22 and -5.53 respectively) **(Figure 4.6)**. The atom-centered fragments descriptor**,** C-041, represents the number of fragments containing C(sp2) atoms that are attached with two electronegative atoms (O, N, S, Se, and halogens), i.e., one by a single bond and another by a double bond. The positive regression coefficient suggests the influential effect of the feature containing

C(sp2) atoms directly attached to two electronegative atoms toward BuChE inhibitory activity. This is observed in the case of compound **269** (pIC_{50}: 0.823) (descriptor value 1), and the opposite is seen in compound **84** (pIC_{50}: -4.39) as depicted in **Figure 4.6**. Thus, this descriptor provides us with an assumption that these fragments might involve in polar interaction with the binding pocket amino acid residues. This electrostatic bond may also help in stabilizing the π-π stacking binding and increase the affinity between the BuChE enzyme and its inhibitors[363].

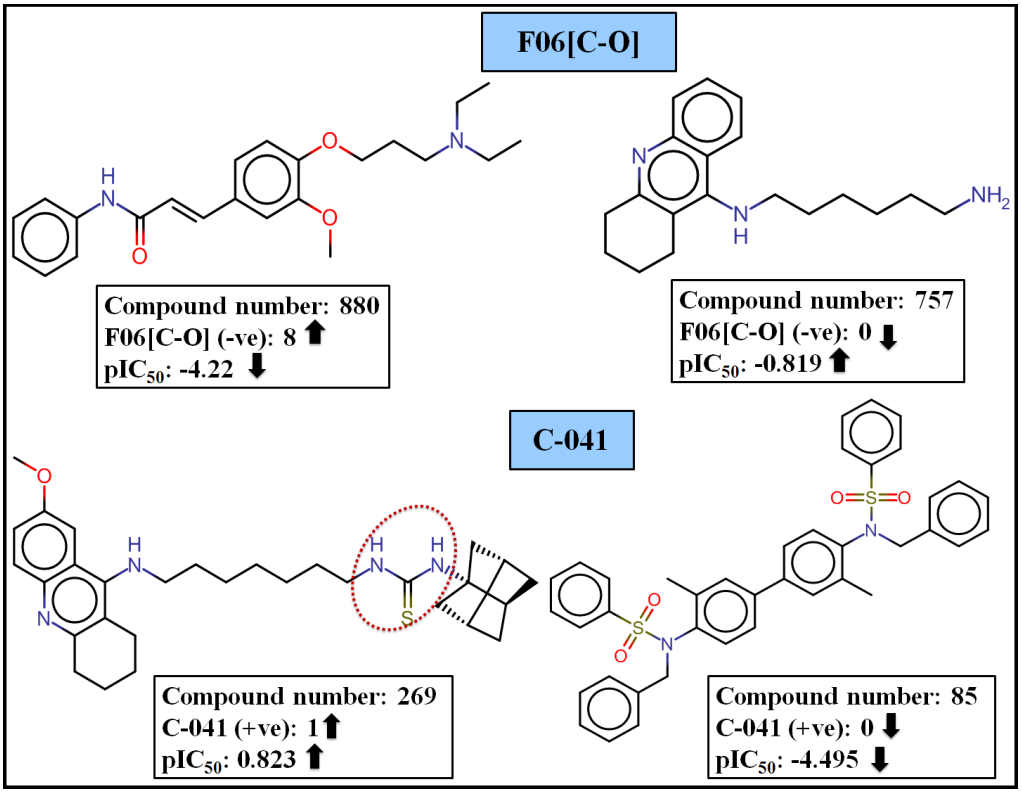


Figure 4.6. Contribution of F06[C-O] and C-041 descriptors on BuChE enzyme inhibition.

Another functional group count descriptor, B02[N-O], stands for the presence/absence of the N-O fragment at the topological distance 2. The positive regression coefficient of this descriptor indicates that the presence of the N-O fragment at topological distance 2 may favor the inhibitory activity of inhibitors against the BuChE enzyme as found in **(Figure 4.5)** compounds **690** (pIC_{50}: 0.621), **688** (pIC_{50}: 0.017), **697** (pIC_{50}: 0.850) and **695** (pIC_{50}: 0.645) (containing descriptor value of 1 for all the cases). On the other hand, compounds with the lower numerical value of this descriptor show lower inhibitory activity as observed in **(Figure 4.5)** compounds **190** (pIC_{50}: -5.74), **204** (pIC_{50}: -5.02), and **849** (pIC_{50}: -4.91). In the latter case, the compounds do not have such N-O fragments at the topological distance 2. Thus, N-O fragments at topological distance 2 are influential for the BuChE inhibitory activity. Another 2D atom pair descriptor, F03[C-C], indicates the frequency of the C-C fragment at the topological distance 3. The negative regression coefficient of this descriptor suggests that the presence of the C-C fragment at the topological distance 3 inversely affects the BuChE inhibitory activity. This is observed in compounds **996** (pIC_{50}: -4.25) and **987** (pIC_{50}: -5.37) (containing higher descriptor values 90 and 81 respectively) **(Figure 4.7)**. The opposite is observed in compounds **675** (pIC_{50}: -1.83) and **430** (pIC_{50}: -1.91) (containing descriptor values

9) show higher enzyme inhibitory activity due to lower numerical values of this descriptor. The functional group count descriptor, nCrs, represents the number of ring secondary C (sp3) atoms present in the compounds. This descriptor positivity influences the activity of BuChE inhibitors as indicated by its positive regression coefficient. Thus, the compounds containing a higher number of ring secondary C (sp3) atoms may have high inhibitory activity against BuChE, as shown in **(Figure 4.7)** compounds **385** (pIC_{50}: -0.477), **386** (pIC_{50}: 0.9085) and **269** (pIC_{50}: 0.823) (containing 10, 10 and 9 ring secondary C (sp3) atoms respectively), whereas the compounds containing no such ring secondary C (sp3) atom have low inhibitory activity against BuChE as shown in **(Figure 4.7)** compounds **1041** (pIC_{50}: -5.33) and **1098** (pIC_{50}: -4.60). From this descriptor, it can be inferred that the cyclic ring containing carbon atoms without any unsaturation may favor the BuChE inhibitory activity.

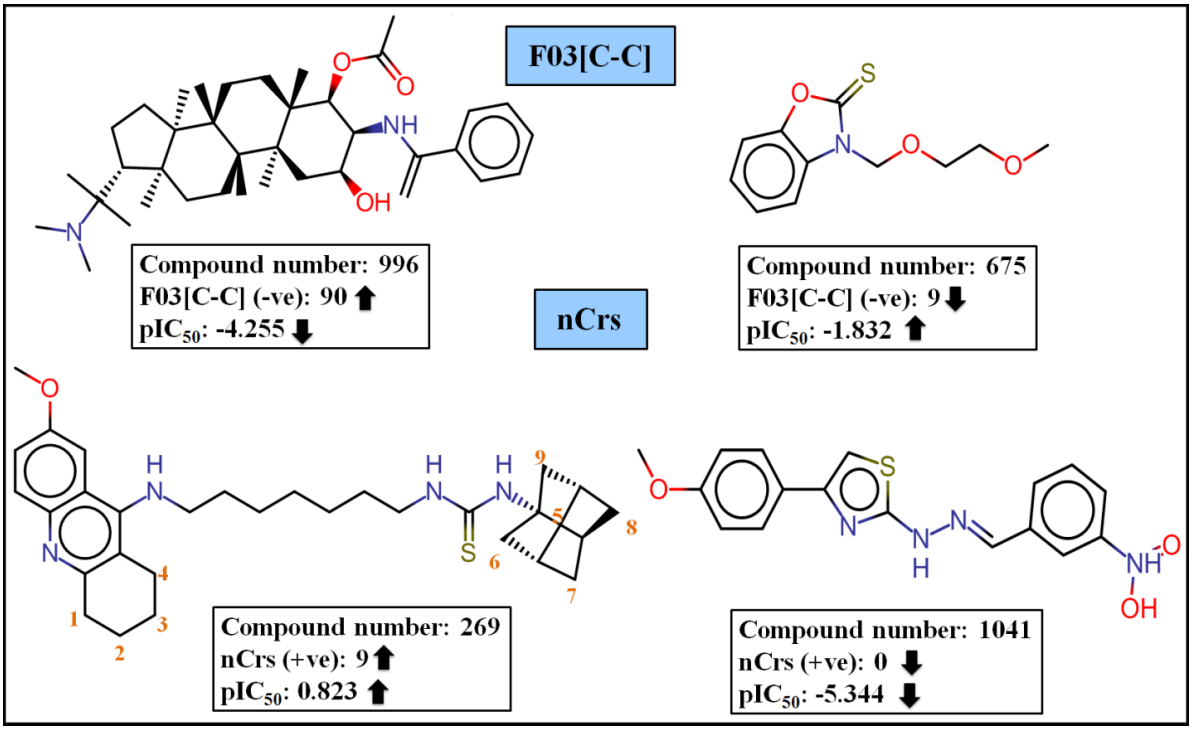


Figure 4.7. Contribution of F03[C-C] and nCrs descriptors on BuChE enzyme inhibition.

Another functional group count descriptor, nN-N, denotes the presence of the number of hydrazine moieties in the compounds. The positive regression coefficient of this descriptor indicates that the activity of inhibitors is directly proportional to the numerical value of the nN-N descriptor. Thus, the compounds having a higher number of hydrazine moiety may have higher BuChE enzyme inhibitory activity as shown in **(Figure 4.5)** compounds **690** (pIC_{50}: 0.621), **688** (pIC_{50}: 0.017), **697** (pIC_{50}: 0.85) and **695** (pIC_{50}: 0.645) (containing descriptor values 1), whereas the compounds such as **190** (pIC_{50}: -5.74), **204** (pIC_{50}: -5.02) and **849** (pIC_{50}: -4.91) have less BuChE enzyme inhibitory activity due to the absence of such fragment as shown in **Figure 4.5**. The hydrazine fragment may be involved in hydrogen bonding interactions with the surrounding amino acid residues in the binding pocket of the BuChE enzyme. We have observed from docking studies (discussed later) that N-N fragments in the molecules form hydrogen bonds along with electrostatic interaction with their surrounding amino acid residues.

The 2D atom pair descriptor, B01[N-N], describes the presence/absence of N-N at topological distance 1. The negative regression coefficient of this descriptor indicates that compounds containing the lower number of such fragments have good inhibitory activity against the BuChE enzyme as shown in **(Figure 4.8)** compounds **264** (pIC_{50}: 1) and **266** (pIC_{50}: 0.823) (containing descriptor value of 1), while a higher number of this fragment shows the lower inhibitory activity as observed in **(Figure 4.8)** compounds **187** (pIC_{50}: -5.52) and **175** (pIC_{50}: -5.39). We have found that descriptors nN-N and B01[N-N] are showing opposite effects on the inhibitory activity against BuChE. The descriptor nN-N representing the presence of the number of hydrazine moiety in the compounds has a positive regression coefficient. In a hydrazine group, two adjacent nitrogen atoms are attached with a single bond; there is no presence of unsaturation. On the other hand, the descriptor B01[N-N] denotes the presence/absence of N-N at topological distance 1 and has a negative regression coefficient. Here, the descriptor signifies the presence or absence of N-N at topological distance 1 without considering the unsaturation effect (i.e., unsaturation might be present or absent). In compounds **187** and **175**, the hydrazine group is absent (nN-N descriptor) but the presence of unsaturation between two adjacent nitrogen atoms can be explained by B01[N-N] but not by then N-N descriptor.

The atom centered fragment, H-048, denotes the number of H attached to C2(sp3)/C1(sp2)/C0(sp). This descriptor is defined as the number of specific atom types in a molecule and can be calculated by knowing only molecular composition and atom connectivity. The count of hydrogen atoms of type H-048 discloses the importance of hydrogen bond interaction. The positive regression coefficient of this descriptor indicates that compounds containing a higher number of such hydrogen atoms have good inhibitory activity against BuChE enzyme as shown in **(Figure 4.8)** compounds **795** (pIC_{50}: -0.77) and **461** (pIC_{50}:-0.699) (containing descriptor values 1) while the compounds **927** (pIC_{50}: -5.29) and **243** (pIC_{50}:-5.70) show lower inhibitory activity due to the absence of such H atom **(Figure 4.8)**.

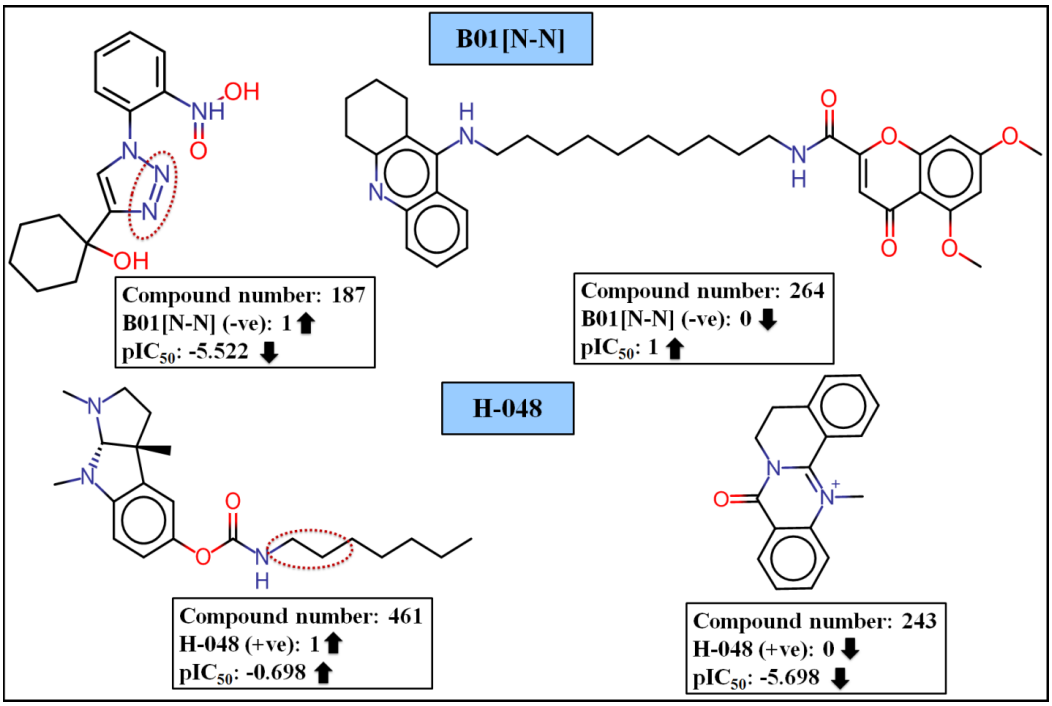


Figure 4.8. Contribution of B01[N-N] and H-048 descriptors on BuChE enzyme inhibition.

Another functional group count descriptor, NRS, indicates the number of ring systems present in the compounds, which contributes negatively towards the BuChE enzyme inhibitory activity. As observed in the docking study (discussed later), the results properly corroborate this observation. Hydrophobicity plays an important role in better BuChE inhibitory activity as we have observed in compounds such as **264** (pIC_{50}: 1) and **697** (pIC_{50}: 0.850) containing descriptor value 2 showing higher inhibitory activity, but according to the regression coefficient plot, this descriptor contributes negatively. Thus, we have concluded that BuChE inhibitory activity decreases with increasing the descriptor value which could be due to some other features present in the molecules as shown in (**Figure 4.9**) Compounds **996** (pIC_{50}: -4.25) and **987** (pIC_{50}: -5.37) (containing descriptor values 2 and 1 respectively) and vice versa in case of compounds **675** (pIC_{50}: -1.83) and **430** (pIC_{50}: -1.91) **(Figure 4.9)**. The atom-centered fragment descriptor, C-031, simply refers to X-CR-X, where, R represents any group linked through a carbon atom; X represents any heteroatom (O, N, S, P, Se, and halogens). This descriptor contributes negatively towards the BuChE enzyme as indicated by the negative regression coefficient. For example, compounds **306** (pIC_{50}: -4) and **303** (pIC_{50}: -4.22) (containing descriptor values 1) have lower BuChE inhibitory activity. On the contrary, the molecules which do not contain such features have higher inhibitory activity as shown in compounds **264** (pIC_{50}: 1) and **697** (pIC_{50}: 0.850) as mentioned in **Figure 4.9**. The functional group count descriptor, nArCONHR, represents the presence of the number of secondary amides (aromatic) in the compounds. There are only 32 compounds **(261, 270, 272, 1061, 1068, 227, 229, and 231, etc)** out of 1130 molecules in the whole data set, which contain such fragments, and the frequency of these fragments in compounds is 1 **(Figure 4.9)**. It may be assumed that this fragment contributes significantly to increasing the intermolecular interactions by forming strong H-bonds. This descriptor contributed positively towards the BuChE inhibitory activity as indicated by the positive regression coefficient. Thus, the molecules bearing this fragment may enhance the BuChE inhibitory activity as shown in **(Figure 4.9)** compounds **261** (pIC_{50}: -1.04) and **272** (pIC_{50}: -1.25) (containing descriptor values 1). On the other hand, the compounds containing no such fragments have lower inhibitory activity as shown in compounds **179** (pIC_{50}: -5.61) and **175** (pIC_{50}: -5.39). It may be mentioned here that the entire dataset may not follow the exact pattern of correlation for all data points concerning a single descriptor, since it is obvious that the property of any molecules is a function of multiple features. Here, we have given appropriate representative examples to understand the role of different features and descriptors in controlling the response values.

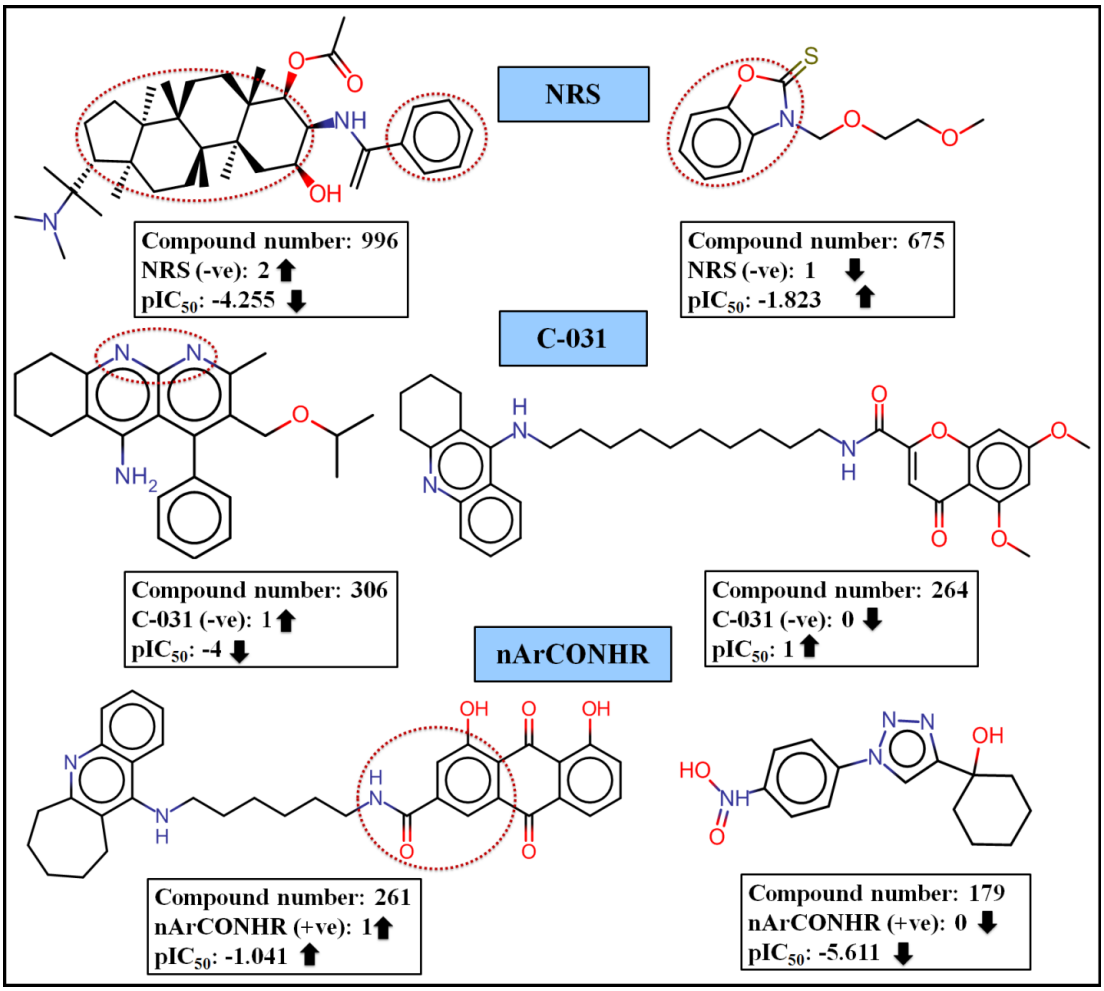


Figure 4.9. Contribution of NRS, C-031, and nArCONHR descriptors on BuChE enzyme inhibition.

4.1.1.2. Applicability domain of the PLS model

The applicability domain of a QSAR model is the structural, biological, and Physico-chemical information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds[310]. The applicability domain of a QSAR model should be described in terms of the most significant parameters that appeared in the developed model. Ideally, the QSAR should only be used to make predictions within that domain by interpolation, not extrapolation. The proposed PLS model was checked using the applicability domain at a confidence level of 99% according to the DModX (distance to model in the X-space) approach using SIMCA-P 10.0 software (available from https://umetrics.com/kb/getting-started-simca-p). In the case of the proposed model, **(Figure 4.10)** we found that 59 compounds (i.e. compounds number 99, 189, 221, 222, 225-231, 233, 261, 270, 272, 300-306, 512, 683, 684, 686, 688, 690, 691, 692, 693, 695, 696, 697, 778, 960, 975-981, 984, 988, 990, 995, 996, 999, 1003, 1006, 1011, 1057, 1059, 1060, 1061, 1062, 1065, 1066, 1068 and 1069) in the training set are located outside the critical DModX value (D-Crit=1.62) and in case of the test set, 16 compounds (i.e., compounds number 223, 228, 301, 685, 687, 689, 694, 959, 989, 993, 997, 1002, 1058, 1063, 1064 and 1067) are located outside the critical DModX value (D-Crit=1.62).

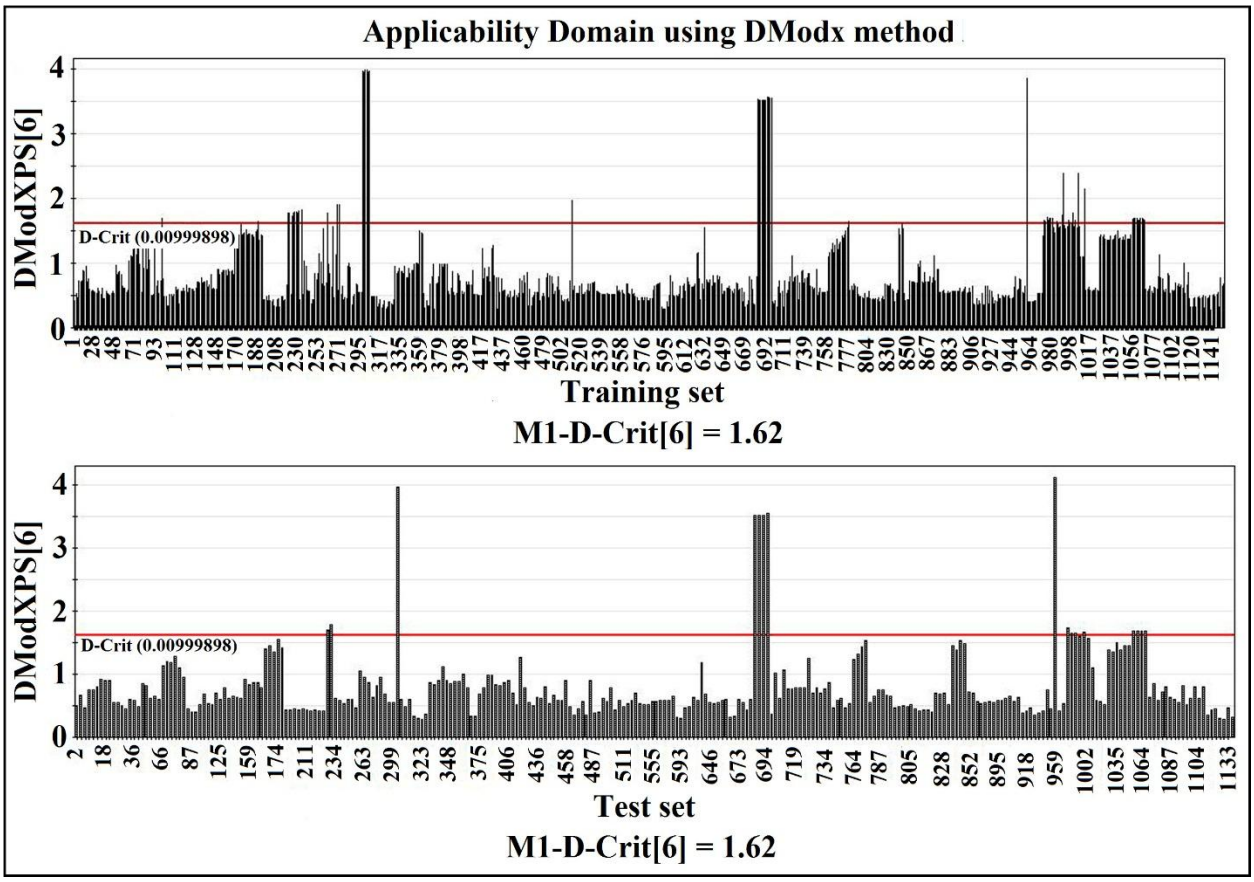


Figure 4.10. Applicability domain DModX values of the training and test set compounds at 99% confidence level of the developed PLS model against BuChE enzyme.

4.1.2. Molecular Docking analysis

4.1.2.1. Molecular Docking for the most active compounds from the dataset

In the case of compounds **264**, **688**, **697**, **695,** and **690** having higher activity values (pIC_{50} = 1, 0.01, 0.850, 0.645, and 0.621 respectively), the interaction forces include hydrogen bond interactions (carbon-hydrogen bonds and conventional hydrogen bonds), π-interactions (π-sigma, π-anion, π-cation, π-alkyl bonds and alkyl hydrophobic) and others attractive forces. The amino acid residues involved in the interaction are ALA A:101, ASP A: 3, ASN A:17, ARG A:14, ILE A:99, ILE A:5, ILE A:4, ASP A:3, THR A:26, MET A:16 and LYS A:103 (**Figure 4.11, 4.12, 4.13, 4.14 and 4.15**). **Table 4.2** contains the docking results and correlation with the final QSAR model.

From compound **264 (Figure 4.11)**, we can see the interacting residues include ALA A: 101, ASP A: 3, MET A: 16, ASN A: 17, ILE A: 4, LYS A: 103, and THR A: 26 of which ALA A: 101 is bound with the ligand via alkyl bonding, ASP A: 3, ASN A: 17, MET A: 16 and THR A: 26 interact with the ligand with hydrogen bonding, while ILE A:4 and LYS A:103 show π-alkyl interactions.

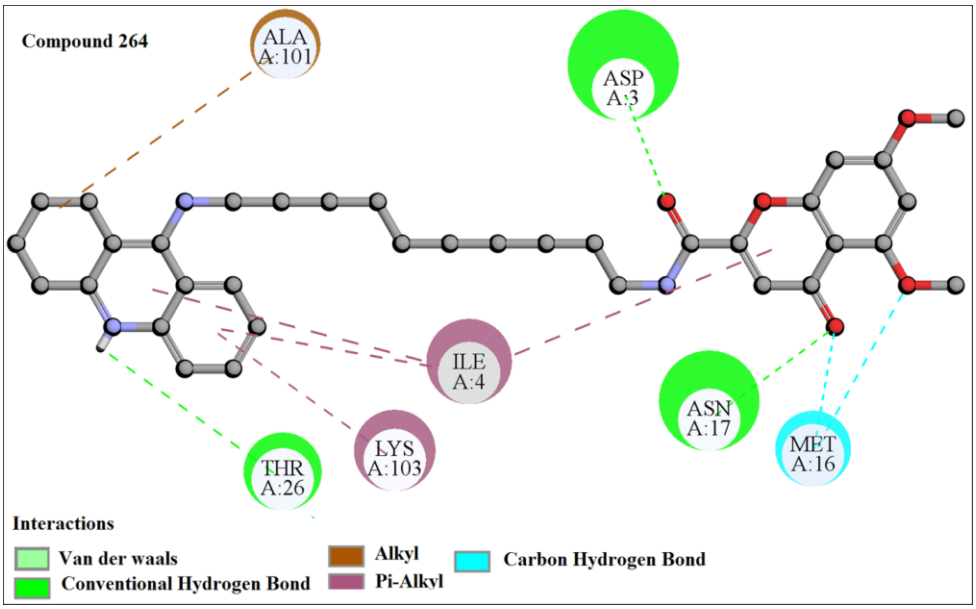


Figure 4.11. Docking interactions of most active compound (compound **264**).

In the case of compound **688**, the interacting amino acids include ILE A: 5, ASP A: 3, and ILE A: 4. The different interactions are shown in **Figure 4.12.** Hydrogen bonding interactions are exhibited by ILE A: 5, and ASP A: 3 while ILE A: 4 shows π-sigma interaction. It is observed from compound **697 (Figure 4.12)** that amino acids ILE A: 4 and ASP A: 3 interact with the ligand via hydrogen bonding interaction, ASP A: 3 and ARG A: 14 interact with ligand via π-anion, π-cation and some other attractive charges whereas LYS A: 103, MET A: 16 show π-alkyl and alkyl interaction with the ligand.

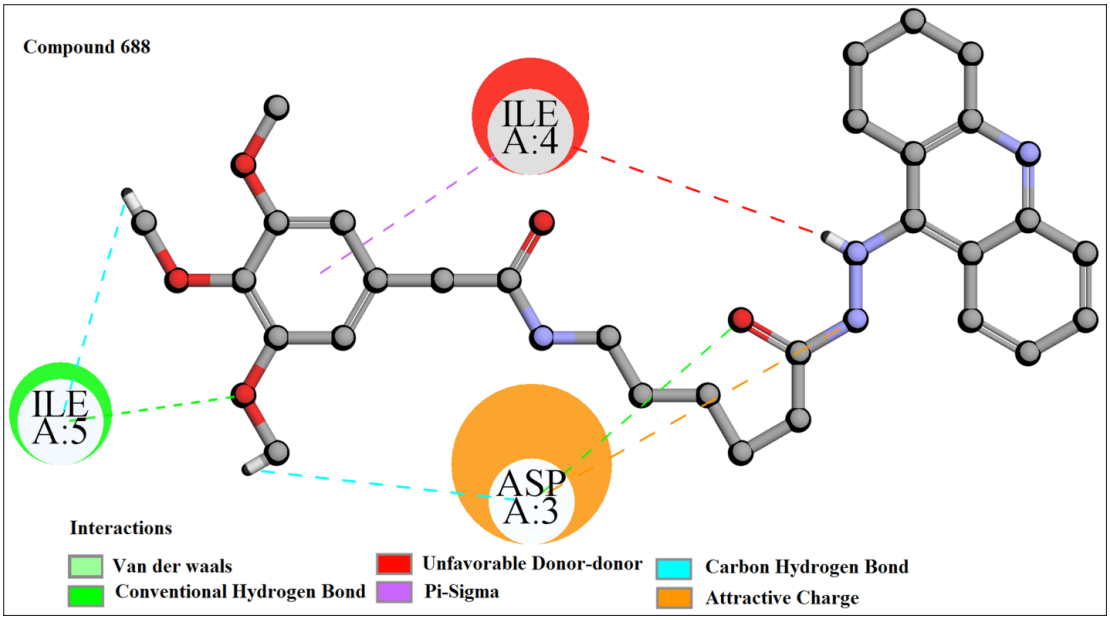


Figure 4.12. Docking interactions of most active compound (compound **688**).

In the case of compound **697**, which interacted with the active site pocket of the enzyme (**Figure 4.13**) through hydrogen bonding with the amino acid residues ILE A: 4, and ASP A: 33, via pi-cation with the amino acid residue ARG A:14, through attractive charge and pi-anion with the amino acid residue ASP A: 3, via pi-alkyl with the amino acid residue LYS A: 103 and alkyl bond with the amino acid residue MET A: 16.

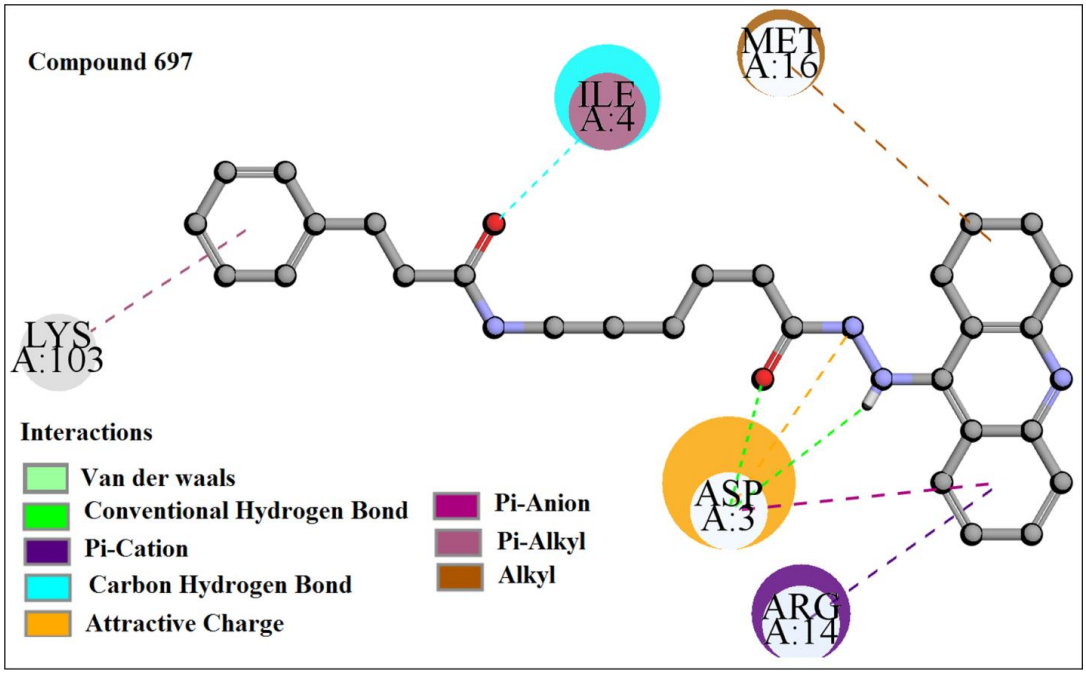


Figure 4.13. Docking interactions of most active compound (compound **697**).

In the case of compound **695**, the interacting amino acids include LYS A:103, ILE A:99, ASP A:3, ILE A:4, and ASN A:17. The different interactions are shown in **Figure 4.14**. Conventional hydrogen bond interactions are formed with ASN A:17 and ASP A:3 while LYS A:103, ILE A:99, and ILE A:4 form π-alkyl and alkyl bonding.

It is observed from compound **690** (**Figure 4.15**) that amino acids ILE A:5, ILE A:4, ASP A:3, and THR A:26 interact with the ligand via hydrogen bonding interaction whereas MET A:16 and LYS A:103 show π-alkyl interaction with the ligand.

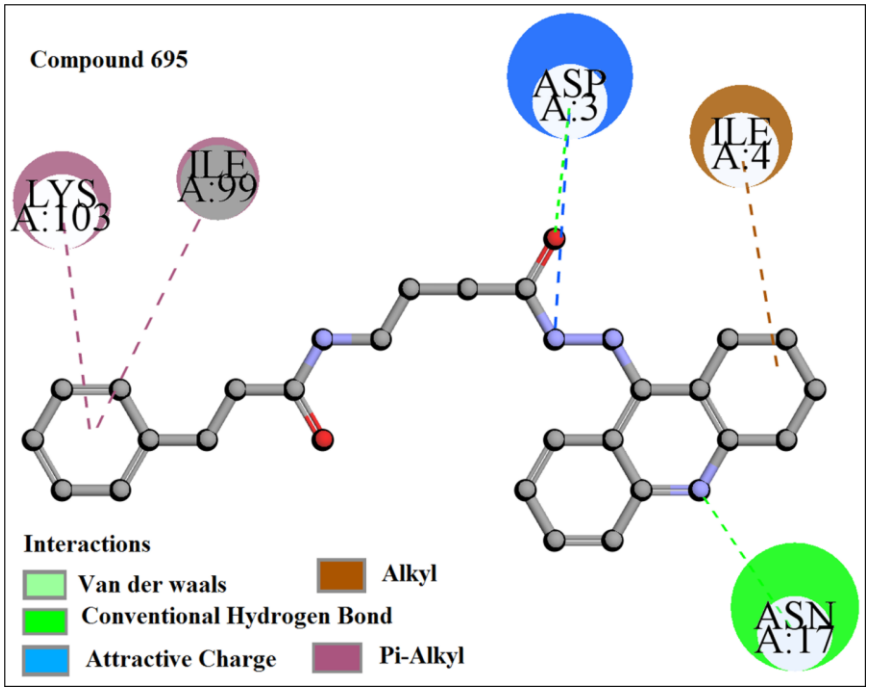


Figure 4.14. Docking interactions of most active compound (compound **695**).

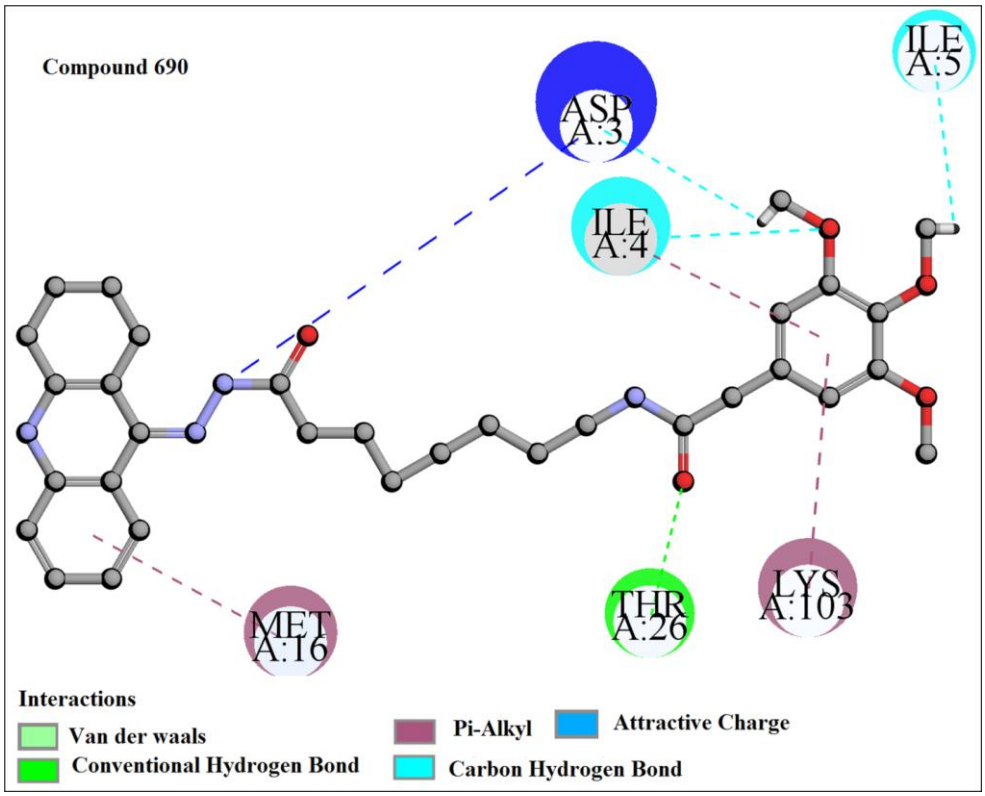


Figure 4.15. Docking interactions of most active compound (compound **690**).

4.1.2.2. Moderately active molecules from the dataset

In the case of compounds **35**, **118**, **867**,**45 and 300** (pIC_{50} = -2.99, -2.96, -2.91, -3.770, and -3.959 respectively) which are moderately active against the BuChE enzyme, the interaction forces include hydrogen bond (carbon-hydrogen bonds and conventional hydrogen bonds), pi-interaction (π-cation, π-alkyl, π-lone pair, π-cation, π-anion, and π-π stacking) and interacting amino acids residues include such as LYS A: 103, ASN A: 17, ILE A: 4, ASP A: 3, HIS A:438, TRP A: 82, TRP A: 430, TYR A: 332, THR A: 120, MET A: 16, THR A: 59, ILE A:99, LEU A:125, ALA A:328, and GLY A:115.

It is observed from compound **35 (Figure 4.16)** that amino acids ASN A: 17 and ASP A: 3 interact with the ligand via hydrogen bonding interaction, and LYS A: 103 and ILE A: 4 interact with ligand via π-alkyl and alkyl bonding.

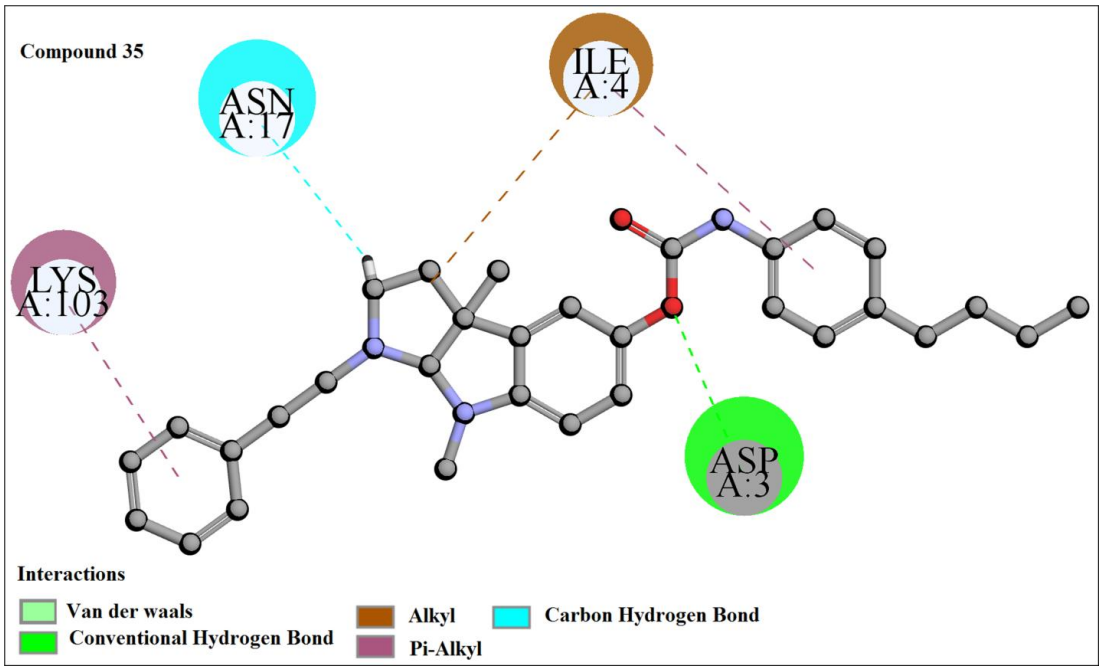


Figure 4.16. Docking interactions of moderately active compound (compound **35**).

In the case of compound **118**, the interacting amino acids include HIS A:438, TRP A: 82, TRP A: 430, TYR A: 332, and THR A: 120. The different interactions are shown in **Figure 4.17**. The amino acid residues HIS A: 438 interact with ligands through hydrogen bonding, whereas HIS A; 438, TRP A: 82, TRP A: 430, TYR A: 332, and THR A: 120 share their hydrophobic feature through π-interactions (π-alkyl, pi-lone pair, and π-cation).

In the case of compound **867 (Figure 4.18)**, the interacting amino acid residues are ASP A: 3 and THR A: 59 which interact with the ligand making hydrogen bonding interaction while ILE A: 4, MET A: 16 bind with ligand via π-alkyl and alkyl bonds.

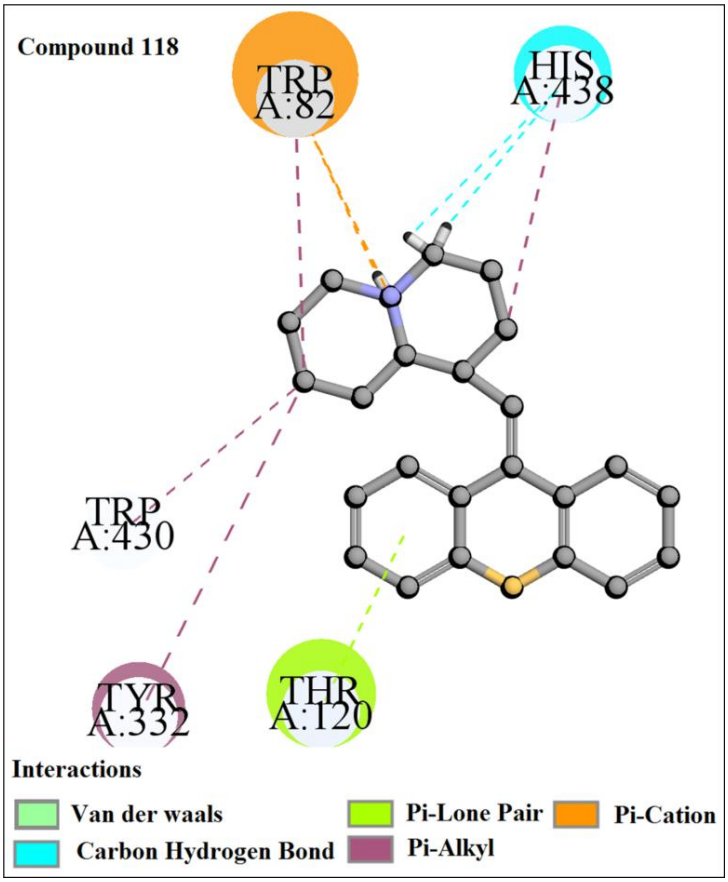


Figure 4.17. Docking interactions of moderately active compound (compound **118**).

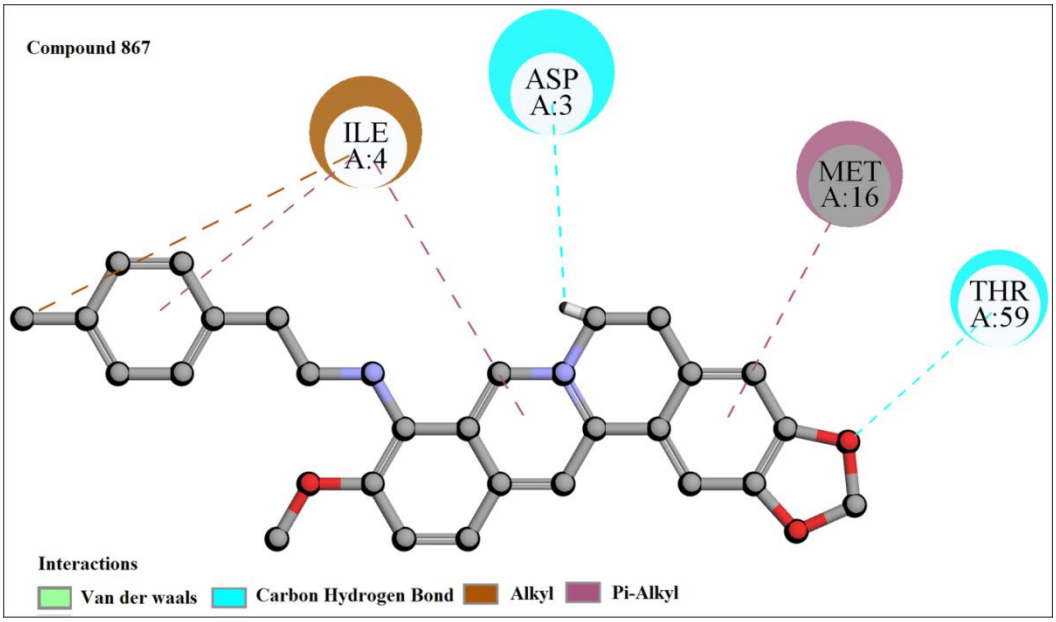


Figure 4.18. Docking interactions of moderately active compound (compound **867**).

From compound **45**, we can see **(Figure 4.19)** the interacting residues include ILE A:99, LYS A:103, ILE A:4, and ASP A:3, of which ASP A:3 is bound with the ligand via hydrogen bonding while ILE A:4, ILE A:99, and LYS A:103 share their hydrophobic feature via π-alkyl interactions.

In compound **300 (Figure 4.20)**, the interacting amino acids include LEU A:125, TYR A:332, THR A:120, TRP A:82, ALA A:328, TRY A:332, GLY A:115, TRP A:430, and HIS A:438. The different interactions are shown in **Figure 4.20**. The amino acid residues HIS A;438, GLY A:115, and THR A:120 share their hydrophilic feature through hydrogen bonding interaction whereas TRY A:332, ALA A:328, TRP A:82, TRY A:128, LEU A:125 and TRP A:430 are bound to the ligand via π-π interaction.

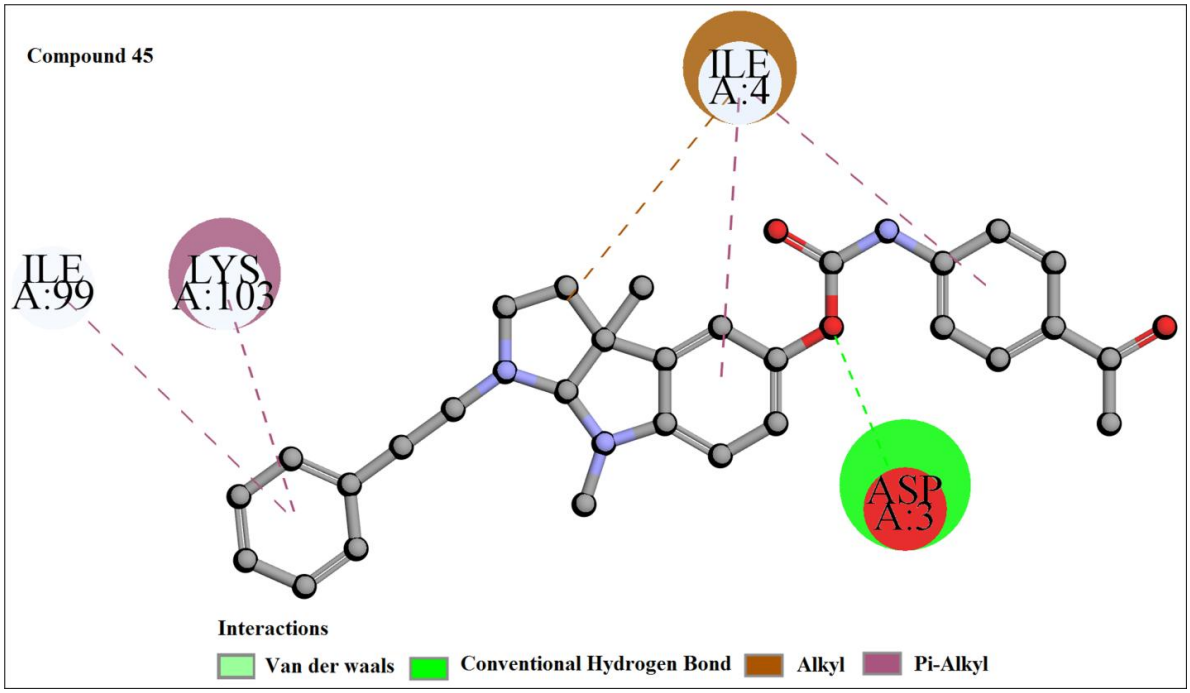


Figure 4.19. Docking interactions of moderately active compound (compound **45**).

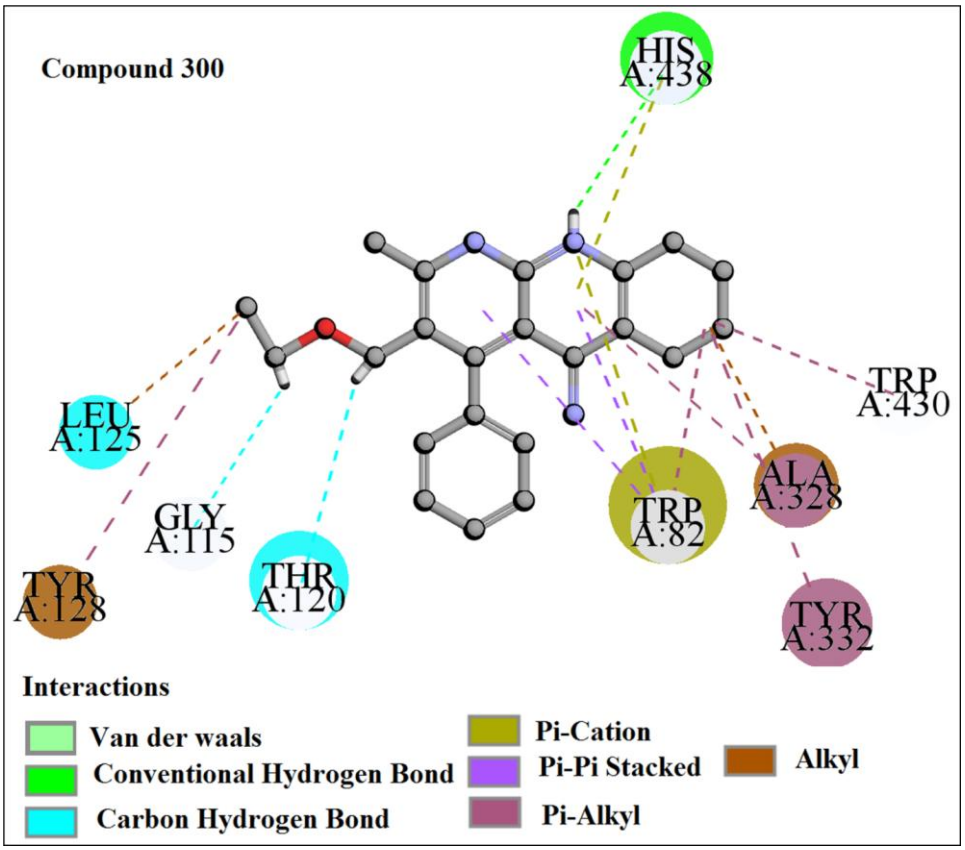


Figure 4.20. Docking interactions of moderately active compound (compound **300**).

4.1.2.3. Least active molecules from the dataset

The BuChE enzyme inhibitors **185**, **243**, **277**, **204** and **835** having a lower inhibitory activity (pIC_{50} = -5.65, -5.69, -5.60, -5.021, and -5.136 respectively) show similar kinds of interactions (alkyl interaction, hydrogen and π interactions) as in case of higher inhibitory activity compounds, but the number of interacting amino acid residues are much less as shown in **Figure 4.21, 4.22, 4.23, 4.24 and 4.25**.

From compound **185 (Figure 4.21),** we can see the interacting residues include TYR A: 332, TRP A: 82, TYR A: 128, and GLU A: 197, of which GLU A: 197 and TYR A: 128 are bound with the ligand via hydrogen bonding while TYR A: 332 and TRP A: 82show π-sigma and π-π stacked interactions.

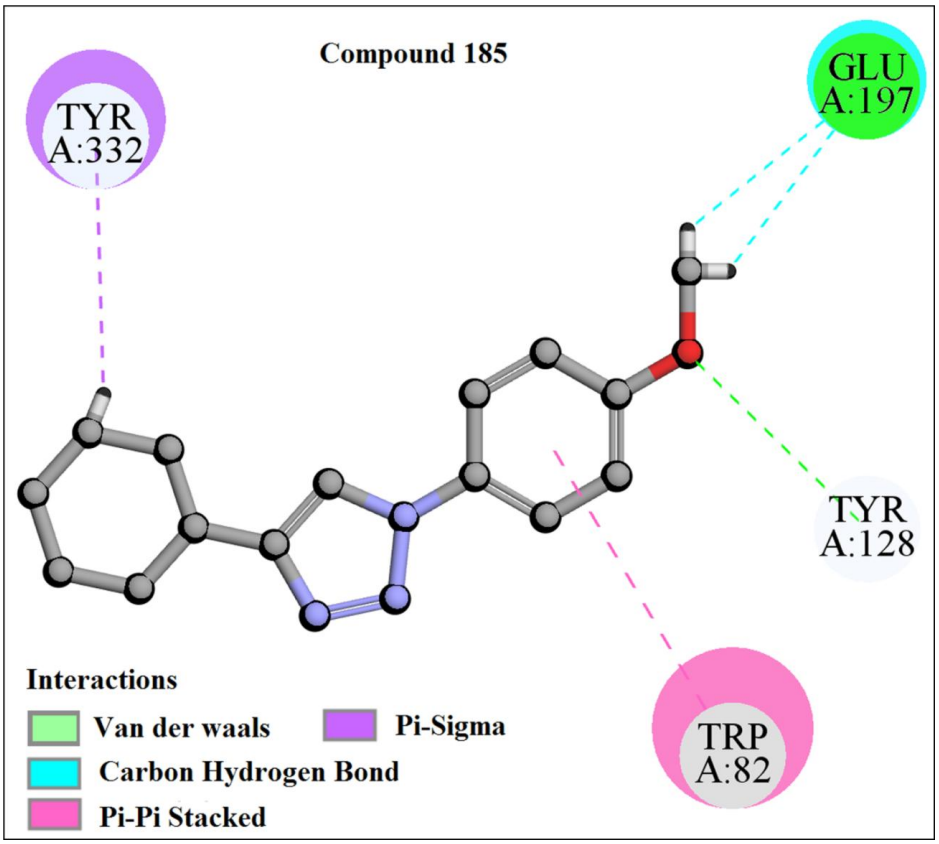


Figure 4.21. Docking interactions of least active compound (compound **185**).

In the case of compound **243 (Figure 4.22)**, the amino acids involved in the interaction are GLU A: 197, TRP A: 82, GLY A: 439, HIS A: 438, ALA A: 32. The amino acid residues HIS A:438 and GLU A:197 interact with the ligand via hydrogen bonding whereas TRP A:82, GLY A:439, ALA A:32 interact with the ligand through π interaction (π-alkyl, π-cation, and π-π stacked).

From compound **277 (Figure 4.23)**, we can see the interacting residues include HIS A: 438, TYR A: 332, ALA A: 199, and GLY A: 117, in which GLY A: 117 and ALA A: 199 are bound with the ligand via hydrogen bonding while HIS A: 438 and TYR A: 332 show π-π T shaped interactions.

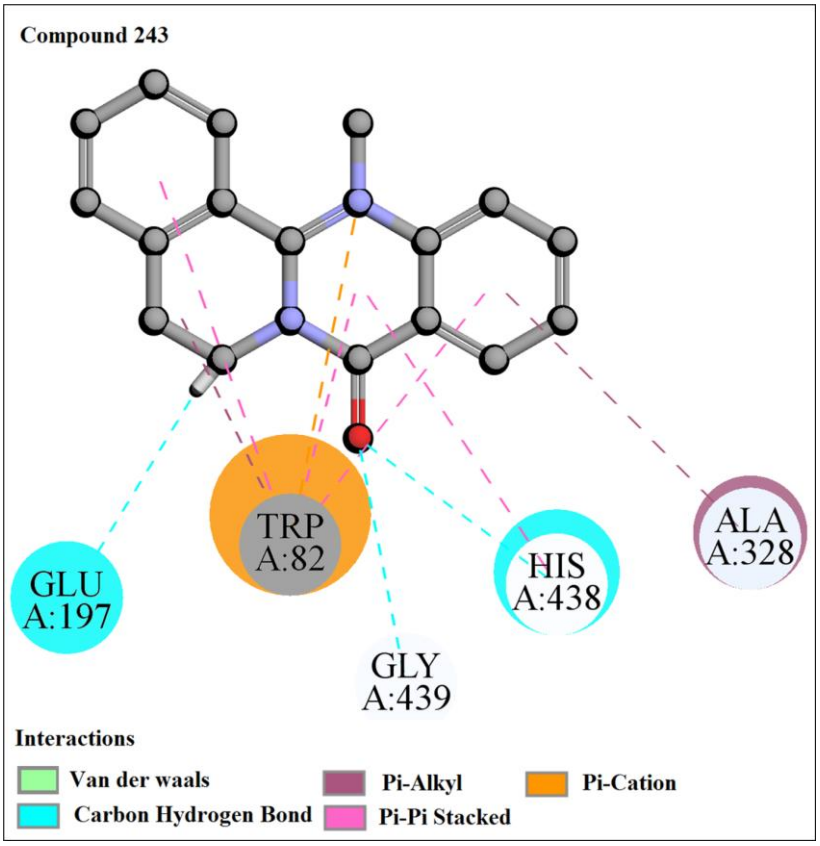


Figure 4.22. Docking interactions of least active compound (compound **243**).

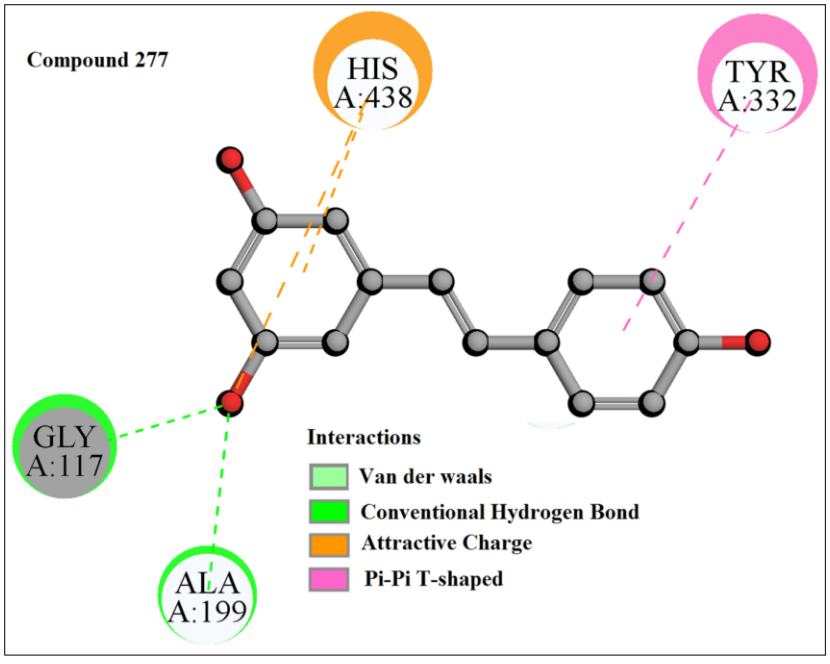


Figure 4.23. Docking interactions of least active compound (compound **277**).

In the case of compound **204 (Figure 4.24)**, the interacting amino acid residues are TRP A: 231, GLU A: 197, TRP A: 128, and GLY A: 115. GLU A: 197 and TRP A: 128 interact with the ligand by hydrogen bonding while TRP A: 231 binds with the ligand via T-shaped π-π interaction, and GLY A: 115 binds with the ligand through amide-π stacking interaction.

In the case of compound **835 (Figure 4.25)**, the amino acids involved in the interaction are TRP A: 82 and ALA A: 328. The amino acid TRP A: 82 interacts with the ligand through π-π stacking and π-sulfur interaction whereas ALA A: 328 binds with the ligand via π-alkyl bonding interaction.

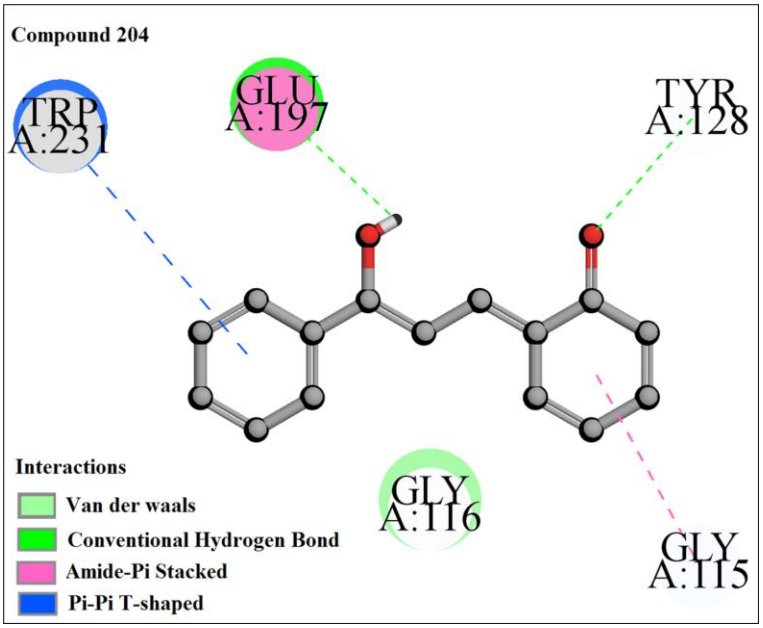


Figure 4.24. Docking interactions of least active compound (compound **204**).

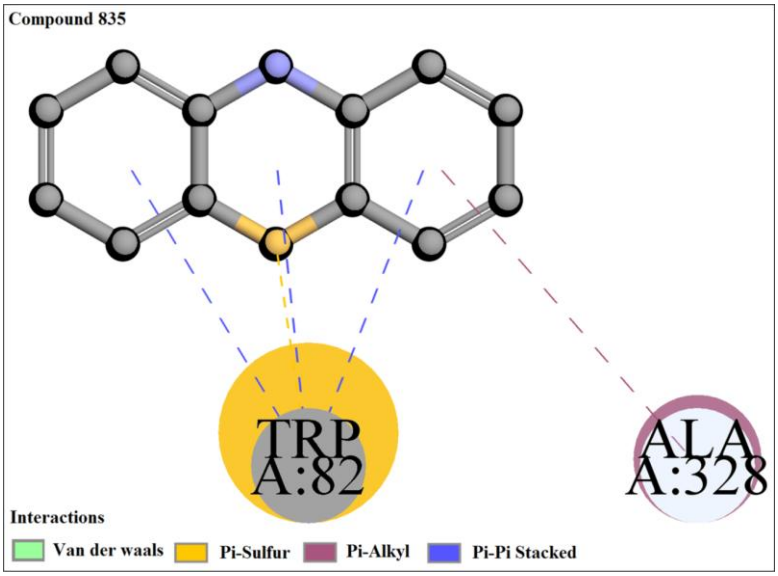


Figure 4.25. Docking interactions of least active compound (compound **835**).

Table 4.2. Docking results and correlation with the final QSAR model.

S. No.	Compound Number	- CDocker interaction energy	Interacting residues	Interactions	Correlation with QSAR model
1	**264** (high pIC_{50})	45.1	ALA A: 101, ASP A: 3, MET A: 16, ASN A: 17, ILE A: 4, LYS A: 103, THR A: 26	Vdw, Hydrogen bonding, alkyl and pi-alkyl	nPyridines, nCrs and T(N..O)
2	**688** (high pIC_{50})	47.20	ILE A: 5, ASP A: 3, ILE A: 4	Vdw, Hydrogen bonding, attractive charges and pi-sigma	nPyridines, nN-N, and T(N..O)
3	**697** (high pIC_{50})	42.67	Met A: 16, ILE A: 4, LYS A: 103, ASP A: 3, ARG A: 14	Vdw, Hydrogen bonding, attractive charges, pi-cation, pi-anion, alkyl and pi-alkyl	nPyridines, RDCHI and nN-N, nCrs
4	**695** (high pIC_{50})	40.56	LYS A:103, ILE A:99, ASP A:3, ILE A:4 and ASN A:17	Vdw, Attractive charges, Hydrogen bonding, Pi-alkyl and alkyl	nPyridines, nN-N, nCrs and T(N..O)
5	**690** (high pIC_{50})	48.38	Met A:16, ILE A: 4, ILE A: 5, LYS A:103, THR A: 26, ASP A:3	Vdw,Hydrogen bonding, Attractive charge, pi-alkyl	nPyridines, RDCHI and nN-N
6	**35** (moderate pIC_{50})	42.46	LYS A: 103, ASN A: 17, ILE A: 4, ASP A: 3	Vdw, Hydrogen bonding, alkyl and pi-alkyl	RDCHI
7	**118** (moderate pIC_{50})	42.09	HIS A:438, TRP A: 82, TRP A: 430, TYR A: 332, THR A: 120	Vdw, Hydrogen bonding, Pi-cation, pi-alkyl, pi-lone pair	NRS and nCrs
8	**867** (moderate pIC_{50})	28	ILE A: 4, ASP A: 3, MET A: 16, THR A: 59	Vdw, Hydrogen bonding, alkyl, pi-alkyl	T(N..O)
9	**45** (moderate pIC_{50})	40.59	ILE A:99, LYS A:103, ILE A:4 and ASP A:3	Vdw, Hydrogen bonding, Pi-alkyl and alkyl	RDCHI, nCrs and C-041
10	**300** (moderate pIC_{50})	44.43	LEU A:125, TYR A:332, THR A:120, TRP A:82, ALA A:328, TRY A:332, GLY A:115, TRP A:430 and HIS A:438	Vdw, Hydrogen bonding, Pi-alkyl, alkyl, pi-cation and pi-pi staking	nPyridines and nCrs
11	**185** (low pIC_{50})	29.55	TYR A: 332, TRP A: 82, TYR A: 128, GLU A: 197	Vdw, Hydrogen bonding, pi-pi stacked, pi-sigma,	B01[N-N] and NRS
12	**243** (low pIC_{50})	37.65	GLU A: 197, TRP A: 82, GLY A: 439, HIS A: 438, ALA A: 328	Vdw, Hydrogen bonding, pi-cation, pi-pi- stacked, pi alkyl	NRS

13	**277** (low pIC_{50})	40	HIS A: 438, TYR A: 332, ALA A: 199, GLY A: 117	Vdw, Hydrogen bonding, attractive charges, pi-cation, pi-pi-T- shaped	NRS
14	**204** (low pIC_{50})	35.193	TRP A:231, GLU A:197, TRP A:128 and GLY A:115	Vdw, Hydrogen bonding, pi-pi T-shaped and amide pi-stacked	NRS
15	**835** (low pIC_{50})	25.128	TRP A:82 and ALA A:328	Vdw, Pi-alkyl, pi-sulphur, pi-pi-stacked	NRS

4.1.3. Relation of the docking results with the QSAR model

In the docking study, it was demonstrated that the formation of hydrogen bonds and π-π stacking between the ligand and the target play a vital role in binding. Hydrogen bonding and π-π-interactions can be correlated with T(N..O) (sum of topological distances between N...O) and nCrs (number of ring secondary C(sp3)) descriptors in the QSAR model. T(N..O)is related to hydrogen bonding, electrostatic and π-donor hydrogen bonding interactions between protein and ligand. The descriptor, nCrs, gives evidence of π-π interaction. Furthermore, nPyridines(number of Pyridines) descriptor supports the evidence of π-π interaction (π-cation, π-anion, π-π stacking, and π-alkyl) along with hydrogen bonding interaction as we have observed in compound **264**, **688**, **697**, **695, 690** and **300 (Figure 4.11, 4.12, 4.13, 4.14, 4.15 and 4.20)**. The RDCHI (simply characterizes the size and branching of molecules) descriptor also supports the π-π interaction (π-alkyl and alkyl) while C-041 (X-C(=X)-X) descriptor signifies hydrogen bonding interaction. The H-048 (C2(sp3)/C1(sp2)/C0(sp)) descriptor characterizes both the hydrogen bonding and π-π (π-alkyl and alkyl) interactions in the QSAR model. Thus, from the above-mentioned information, we can conclude that hydrogen bonding, hydrophobicity, electrostatic interactions, and unsaturation (π-π interaction) feature obtained from both QSAR and docking study are essential for the inhibitory activity against the BuChE enzyme.

4.1.4 Comparisons of the performance of the proposed study with previously published studies

There are many previous QSAR models reported for the prediction of the bioactivity of BuChE enzyme inhibitors. Here, we have performed a comparison of the best model currently derived with some previous models. (**Table** 4.3). The previously reported models were developed by Multilinear Regression (MLR) analysis, partial least squares (PLS), Genetic function approximation (GFA), Multilayer perceptron (MLP), and Artificial neural network (ANN) method, which gave reliable predictions of bioactivity of BuChE enzyme inhibitors. However, the models reported previously were developed using a very low number of compounds covering a very narrow range of chemical diversity. But in the current study, we have employed an extended list of compounds covering a wide range of chemicals and offering a larger chemical domain. We can see from **Table 4.3**, Fang et al.[364] developed a PLS model against the BuChE enzyme by using only **66** compounds; the model quality was good but the equation length (12) was quite high compared to the number of data points. In this study, we have utilized a wide range of compounds and developed the model with 15 selected descriptors and 6 latent variables. We can also see from **Table 4.3** that Zheng et al.[365] and Bitam et al.[366] developed QSAR models using very narrow groups of samples (151 and 93 compounds respectively) and developed MLR, ANN, and MLP-based models. Solomon et al.[367] reported GFA models utilizing only 59 compounds. The details of different internal and external validation parameters obtained from our model and obtained from the previously reported models are given in **Table 4.3**. The best model

presented in this work is based on a larger group of samples and the validation parameters (both internal and external) of the training and test sets qualified the requisite thresholds. Before the development of the final model, we performed a multilayered variable selection strategy from a large pool of descriptors. The best model was selected based on different validation parameters and low equation length. The final model was built by using the PLS algorithm with latent variables of 6. The 15 selected descriptors reflect the fundamental structural characteristics of molecules that are important in modeling the bioactivity of BuChE enzyme inhibitors. The docking results in this study also well collaborate with the descriptors obtained from the developed QSAR model and justify the significance of the developed model. In comparison with the previously reported models concerning acceptability and reliability, the present work deals with diverse classes of compounds. Due to the wide applicability domain, the model reported in the present study may be used as a screening tool for the discovery and development of leads against the BuChE enzyme.

Table 4.3. Comparison of the developed model with the previously published models.

Sources	Equation Length	Model	Training set			Test set	
			Train	R^2	Q^2	Test	R^2pred
Model in this study	15	PLS	848	0.664	0.650	282	0.657
Fang et al. 2016[364]	12	MLR	48	0.883	0.726	18	0.731
Fang et al. 2016[364]	12	PLS	48	0.883	0.777	18	0.775
Zheng et al. 2014[365]	10	MLR	62	0.89	0.85	31	-
Zheng et al. 2014[365]	10	ANN	62	0.950	0.900	31	-
Solomon et al. 2009[366]	5	GFA	39	0.884	0.857	20	0.820
Bitam et al. 2018[367]	8	MLR	121	0.879	0.857	30	0.847
Bitam et al. 2018[367]	8	MLP	121	0.888	0.895	30	-

4.2. Study 2- Exploring 2D-QSAR for prediction of Beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease

In this study, 2D QSAR modeling, pharmacophore mapping, and molecular docking study were performed to unveil the structural requirements for the inhibitory activity against the BACE1 enzyme. Here, we will discuss the results obtained from this study, which will include the selected 2D QSAR model, pharmacophore mapping, and its validation, molecular docking analysis, and finally, the identified important structural features against the BACE1 enzyme.

4.2.1. 2D QSAR modeling analysis

4.2.1.1. Mechanistic interpretation of modeled descriptors

The reported PLS model was developed with 5 descriptors using 3 LVs (latent variables). The statistical validation parameters like coefficient of determination R^2 (0.826) and cross-validated correlation coefficient $Q^2_{(LOO)}$ (0.795) signifying the reliability of the model, and R^2_{Pred} or Q^2F_1 (0.846), Q^2F_2 (0.846) judge the good predictivity of the model. The descriptors appearing in the model describe the structural and functional requirements which can improve the inhibitory activity of molecules against the BACE1 enzyme. The closeness of the observed and predicted values for the BACE1 enzyme inhibitors in the data set can be further recognized from the scatter plot as shown in **Figure 4.26**. The final PLS model with its validation parameters is depicted below (**Equation 4.3)**:

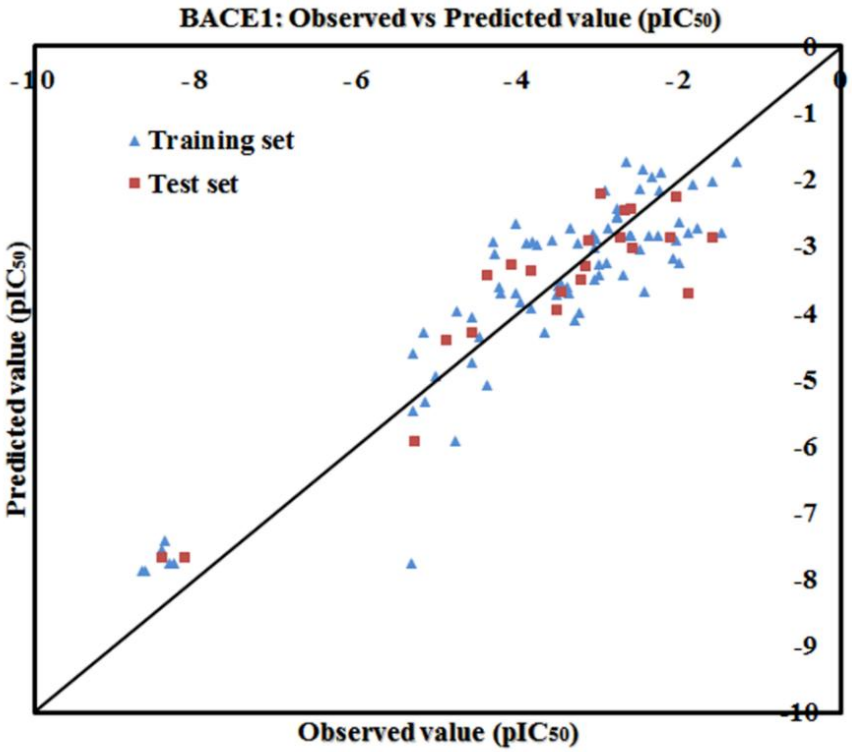


Figure 4.26. The scatter plot of observed and predicted values of the final PLS model against the BACE1 enzyme.

BACE1 enzyme inhibitors: PLS Model

$$pIC_{50}(nM) = -4.235 + 36.67 \times "ETA\_dEpsilon\_D" - 1.82 \times "B03[N-S]" + 1.106 \times "ETA\_Beta\_ns\_d" - 0.107 \times "H-047" + 1 \times "C-033"$$ **Equation 4.3**

Internal validation parameters: $n_{training}$= 76, R^2= 0.826, Q^2= 0.795, Average r_m^2= 0.711, Δr_m^2= 0.139, MAE= 0.625, **Fitting quality**=Moderate, **External validation parameters**: n_{test}= 22, Q^2F_1= 0.846, Q^2F_2= 0.846, Average r_m^2= 0.731, Δr_m^2= 0.140, MAE= 0.544, CCC= 0.911, **Fitting quality** =Good, LV= 3, No. of descriptors= 5

In the present analysis, a simple but statistically robust PLS regression-based QSAR model was developed against the BACE1 enzyme. The regression coefficient plot (see **Figure 4.27**) provides information about the contribution of descriptors in the model toward the activity of the compounds. The positive regression coefficients of the descriptors indicated that the BACE1 inhibitory activity will increase with increasing their descriptors values as shown in the case of ETA_Beta_ns_d, ETA_dEpsilon_D, and C-033 descriptors. In contrast, the negative regression coefficients of the descriptors suggested that the BACE1 enzyme inhibitory activity of the compounds will decrease with increasing the descriptor values as shown in the case of H-047 and B03[N-S] descriptors. The significance level of the modeled descriptors towards the inhibitory activity against the BACE1 enzyme is computed based on the variable importance plot (VIP). The variable importance plot (VIP)[368] defines the order of significance level among the model variables which are responsible to regulate the inhibitory activity towards the BACE1 enzyme. The descriptors contributing most (ETA_dEpsilon_D, B03[N-S] and ETA_Beta_ns_d) and least (H-047 and C-033) to the BACE1 inhibition can be identified with the help of this plot (**Figure 4.27**). The variables show higher statistical significance with a VIP score >1 as compared to one with a low VIP score of 0.45. As suggested by the VIP plot[368], the significance level of the modeled descriptors is found to be in the following order: ETA_dEpsilon_D, B03[N-S], ETA_Beta_ns_d, H-047, and C-033. The details of statistical validation parameters in terms of both internal and external validation parameters are depicted in the model (**Equation 3**). The list of molecules present in the dataset with their names, SMILES notation of respective compounds and observed and predicted activities (2D-QSAR) against the BACE1 enzyme are depicted in **Table 4.4**.

Table 4.4. The list of molecules present in the dataset with their names, SMILES notation of respective compounds, and observed and predicted activities (2D-QSAR) against the BACE1 enzyme.

Name	SMILES notation	Activity (pIC_{50})	
		Observed	Predicted
1	C1(=N[C@](C(=O)N1C)(c1cc(ccc1)NC(=O)c1occc1)C1CCCCC1)N	-1.301	-1.729
2*	C1(=N[C@](C(=O)N1C)(C1CCCCC1)c1cc(ccc1)NC(=O)COC)N	-1.602	-2.877
3	C1(=N[C@](C(=O)N1C[C@H]1CC[C@H](CC1)C(=O)O)(c1ccccc1)C1CCCCC1)N	-2.079	-3.193
4	C1(=N[C@](C(=O)N1C)(c1cccc(c1)NC(=O)c1n(ccc1)C)C1CCCCC1)N	-2.255	-2.169
5	C1(=N[C@@](C(=O)N1CCCCCC)(C1CCCCC1)c1ccccc1)N	-2.431	-3.682
6	C1(=N[C@](C(=O)N1CCCCCC(=O)O)(c1ccccc1)C1CCCCC1)N	-2.491	-3.041

7*	C1(=N[C@](C(=O)N1C)(c1cccc(c1)NC(=O)c1ccc2c(c1)OCO2)C1CCCCC1)N	-2.579	-3.034
8	C1(=N[C@](C(=O)N1Cc1cc(cc(c1)F)F)(c1ccccc1)C1CCCCC1)N	-2.612	-2.853
9*	C1(=N[C@](C(=O)N1C)(C1CCCCC1)c1cc(ccc1)NC(=O)CCC)N	-2.690	-2.455
10	C1(=N[C@@](C(=O)N1C)(c1ccccc1)C1CCCCC1)N	-2.707	-3.428
11	c1(ccccc1)c1ccc(c2ccc(cc2)NC(=O)c2cc(ccc2)Br)n1CC(=O)NC(=N)N	-2.778	-2.566
12	C1(=N[C@](C(=O)N1C)(c1cccc(c1)NC(=O)c1n(ccc1)C)C1CCCCC1)N	-2.934	-2.169
13	C1(=N[C@@](C(=O)N1C)(c1ccccc1)C1CCCCC1)N	-2.995	-3.428
14	C1(=NC(C(=O)N1C)(C1CCCCC1)C1CCCCC1)N	-3.033	-2.894
15	C1(=N[C@](C(=O)N1C)(c1cccc(c1)NC(=O)c1ccc2c(c1)OCO2)C1CCCCC1)N	-3.060	-3.034
16*	C1(=N[C@@](C(=O)N1C)(c1cccc(c1)NC(=O)c1cccc(c1)OC)C1CCCCC1)N	-3.133	-2.923
17	C1(=N[C@](C(=O)N1C)(c1cc(ccc1)NC(=O)CCOC)C1CCCCC1)N	-3.267	-2.967
18*	C1(c2ccccc2)(c2ccccc2)N=C(N(C1=O)C)N	-3.531	-3.962
19	C1(=N[C@](C(=O)N1C)(c1cc(ccc1)NC(=O)CCOC)C1CCCCC1)N	-3.895	-2.967
20*	C1(=N[C@@](C(=O)N1C)(c1ccccc1)C1CCCCC1)N	-4.380	-3.428
21*	C1(c2ccccc2)(N=C(N2C1=NCCC2)N)c1ccccc1	-4.579	-4.286
22	c12cc(ccc1c(c(cn2)C(=O)NCc1ccc2OCOc2c1)O)S(=O)(=O)Nc1cccc(c1)C(F)(F)F	-2.491	-2.149
23	c1(csc(n1)NC(=O)CSc1nc(c(c(=O)[nH]1)C#N)c1ccc(cc1)OC)c1ccccc1	-3.677	-4.303
24	c1(n(c(cc1)c1c(cccc1)Cl)CC(=O)NC(=N)NCCCO)c1ccc(cc1)OCCC	-2.041	-2.915
25	c1(n(c(cc1)c1c(cccc1)Cl)CC(=O)NC(=N)N)c1ccc(cc1)Oc1ccc(cc1)C(=O)C	-2.380	-2.841
26	c1(n(c(cc1)C12CC3CC(C1)CC(C3)C2)CC(=O)NC(=N)NCCCO)c1ccccc1	-3.505	-3.595
27	c1(n(c(cc1)C12CC3CC(C1)CC(C3)C2)CC(=O)NC(=N)N)c1ccccc1	-3.770	-2.975
28	c1(n(c(cc1)c1ccccc1)CC(=O)NC(=N)N)c1ccccc1	-4.287	-3.118
29*	C1[C@@H]2[C@@H]3[C@H](N[C@@H]2CN[C@H]1Cn1cc(nn1)c1ccc2c(c1)ccc(c2)OC)CCCC3	-3.173	-3.303
30	[C@H]1(C(=O)[C@H](SCc2cc(ccc2)C(F)(F)F)C(=O)N1)[C@@H](C)CC	-4.778	-5.935
31	C1(=O)[C@@H](OC(=O)[C@H]1Sc1ccccc1)c1ccccc1	-5.021	-4.964
32	C1(=O)[C@H](NC(=O)[C@@H]1SCc1occc1)Cc1ccccc1	-5.161	-5.332

33*	C1(=O)[C@@H](NC(=O)[C@@H]1SCc1cccc(c1)C(F)(F)F)CC(C)C	-5.287	-5.935
34	C1(=O)[C@@H](NC(=O)[C@H]1SCC(=O)NCc1cc2OCOc2cc1)CC(C)C	-5.301	-5.480
35	C1(=O)[C@@H](NC(=O)[C@H]1SCc1occc1)[C@@H](CC)C	-5.301	-4.609
36	C1(=O)CN(C(=O)[C@H]1SCc1cccc(c1)C(F)(F)F)Cc1ccccc1	-5.326	-7.769
37*	C1(=O)CN(C(=O)[C@H]1SCCc1ccccc1)Cc1cc2OCOc2cc1	-8.143	-7.662
38	C1(=O)CN(C(=O)[C@H]1SCc1cccc(c1)C(F)(F)F)CCc1ccccc1	-8.264	-7.769
39	C1(=O)CN(C(=O)[C@H]1SCCc1ccccc1)Cc1cccc(c1)C(F)(F)F	-8.326	-7.769
40	C1(=O)CN(C(=O)[C@@H]1SCc1cccc(c1)C(F)(F)F)Cc1ccc(cc1)OC	-8.382	-7.429
41*	C1(=O)CN(C(=O)[C@@H]1SCc1cccc(c1)C(F)(F)F)Cc1cc(ccc1)C(F)(F)F	-8.411	-7.662
42	C1(=O)CN(C(=O)[C@H]1SCc1cccc(c1)C(F)(F)F)Cc1cc2OCOc2cc1	-8.423	-7.555
43	C1(=O)CN(C(=O)[C@H]1SCCc1ccccc1)CCc1ccccc1	-8.627	-7.876
44	C1(=O)CN(C(=O)[C@@H]1SCCc1ccccc1)Cc1ccccc1	-8.661	-7.876
45	C1(=N[C@@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1ccc(cc1)OC(F)(F)F)N	-1.477	-2.803
46	C1(=N[C@](c2cccc(c2)c2c(nccc2)F)(c2ccc(OC(F)(F)F)cc2)C2=NOCCN12)N	-1.778	-2.736
47*	C1(=N[C@](C2=NCCCN12)(c1cccc(c1)c1cncnc1)c1ccc(cc1)OC(F)(F)F)N	-1.903	-3.712
48	C1(=N[C@@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1ccc(cc1)OC(F)(F)F)N	-1.903	-2.803
49	C1(=N[C@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1ccc(cc1)OC)N	-2	-3.246
50*	C1(=N[C@](c2cccc(c2)c2c(nccc2)F)(c2ccc(OC(F)(F)F)cc2)C2=NCCCCN12)N	-2.113	-2.869
51	C1(=N[C@](C2=NCCCN12)(c1cc(c2c(nccc2)F)ccc1)c1cc(c(cc1)OCC)OCC)N	-2.278	-2.850
52	C1(=N[C@@](c2cccc(c3c(nccc3)F)c2)(c2ccc(OC(F)(F)F)cc2)C2=NCCN12)N	-2.897	-2.733
53	C1(=N[C@](C2=NCCCN12)(c1cccc(c2cncnc2)c1)c1ccc(cc1)OC)N	-3.309	-4.125
54	C1(=N[C@](C2=NCCCN12)(c1cccc(c1)c1cncnc1)c1ccc(cc1)OC(F)(F)F)N	-3.380	-3.712
55	c1(n(c(cc1)c1c(cccc1)Cl)CC(=O)NC(=N)NCCCO)c1ccc(cc1)OCCC	-3.577	-2.915
56	C1(=N[C@](c2cccc(c2)c2c(nccc2)F)(c2ccc(OC(F)(F)F)cc2)C2=NCCCCCN12)N	-3.819	-2.930

57*	C1(=N[C@@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1cc2c(cc1)OCO2)N	-3.845	-3.357
58	C1(=N[C@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1cc2c(cc1)OCCO2)N	-3.982	-3.845
59	C1(=N[C@](c2cccc(c2)c2c(nccc2)F)(c2ccc(OC(F)(F)F)cc2)C2=NC(CN12)(C)C)N	-4.029	-2.655
60	C1(=N[C@@](C2=NCCCN12)(c1cccc(c1)CCC1CC1)c1ccc(cc1)OC(F)(F)F)N	-4.037	-3.716
61*	C1(=N[C@](C2=NCCCN12)(c1cccc(c2c(nccc2)F)c1)c1cc2c(cc1)OC(O2)(F)F)N	-4.086	-3.280
62	C1(=N[C@](C2=NCCCN12)(c1cccc(c1)c1cncnc1)c1ccc(cc1)OC(F)(F)F)N	-4.215	-3.712
63	C1(=N[C@@](C2=NCCCN12)(c1cccc(c1)CCC)c1ccc(cc1)OC(F)(F)F)N	-4.238	-3.611
64	C1(=N[C@](c2cccc(c2)Br)(c2ccc(cc2)OC(F)(F)F)C2=NCCCN12)N	-4.318	-2.929
65	C1(=N[C@](C2=NCCCN12)(c1cccc(Cc2ccccc2)c1)c1ccccc1)N	-4.387	-5.081
66	C1(=N[C@](C2=NCCCN12)(c1cccc(Cc2ccc(cc2)F)c1)c1ccccc1)N	-4.480	-4.374
67	C1(=N[C@](C2=NCCCN12)(c1cccc(Cc2ccc(cc2)OC)c1)c1ccccc1)N	-4.579	-4.760
68	C1(=N[C@@](c2cccc(c2)C(F)(F)F)(c2ccccc2)C2=NCCCN12)N	-4.583	-4.078
69	C1(=N[C@@](C2=NCCCN12)(c1cc(ccc1)OC)c1ccccc1)N	-4.755	-3.985
70*	C1(=N[C@@](c2cccc(c2)C(C)(C)C)(c2ccccc2)C2=NCCCN12)N	-4.895	-4.415
71	C1(c2ccccc2)(N=C(N2C1=NCCC2)N)c1ccccc1	-5.176	-4.286
72	c1c(n(Cc2ccc(c(n2)N)OCCO)c(c2ccccc2Cl)c1)c1ccc(Oc2cncnc2)cc1	-1.602	-2.033
73	c1c(n(Cc2ccc(c(n2)N)NCCO)c(c2ccccc2Cl)c1)c1ccc(Oc2cncnc2)cc1	-1.845	-2.068
74	c1c(c2ccc(Oc3cncnc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2	-2.636
75*	c1c(c2ccc(OCCCCC)cc2)n(Cc2ccc(c(n2)N)OCCO)c(c2ccccc2Cl)c1	-2.041	-2.266
76	c1c(c2ccc(OCCCCC)cc2)n(Cc2ccc(c(n2)N)NCCO)c(c2ccccc2Cl)c1	-2.230	-1.896
77	c1c(n(Cc2ccc(c(n2)N)NCCCO)c(c2ccccc2Cl)c1)c1ccc(OCCCCC)cc1	-2.342	-1.962
78	c1c(c2ccc(Oc3cccs3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.462	-1.847
79	c1c(c2ccc(Oc3cccnn3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.612	-2.850
80*	c1c(c2ccc(Oc3cccnc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.612	-2.440
81	c1c(c2ccc(OCCCC#N)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.623	-2.835
82	c1c(c2ccc(Oc3cnccn3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.672	-1.743

83*	c1c(c2ccc(OCCCCC)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.740	-2.874
84	c1c(c2ccc(Oc3ccncc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.770	-2.440
85	c1c(c2ccc(Oc3ccccn3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.770	-2.546
86	c1c(c2ccc(Oc3ccccc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.908	-3.244
87*	c1c(c2ccc(Nc3cncnc3)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-2.991	-2.222
88	c1c(c2ccc(cc2)OCCC)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-3	-3.272
89	c1c(c2ccc(C(=O)NC3CC3)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.060	-3.501
90	c1c(c2ccc(OCCCC)cc2)n(c(c2ccccc2Cl)c1)Cc1cccc(n1)N	-3.079	-2.821
91*	c1c(c2ccc(C(=O)NC(C)C)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.217	-3.501
92	c1c(c2ccc(cc2)C(=O)NCC)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.250	-4.008
93	c1c(c2ccc(cc2)OC)n(c(c1)c1ccccc1Cl)Cc1cccc(n1)N	-3.352	-2.743
94	c1c(c2ccc(C(=O)NCCC)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.389	-3.608
95*	c1c(c2ccc(C(=O)NCCCC)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.477	-3.673
96	c1c(c2ccc(C(=O)NC3CCC3)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.477	-3.567
97	c1c(c2ccccc2)n(c(c1)c1ccccc1)Cc1cccc(n1)N	-3.531	-3.738
98	c1c(c2ccc(C(=O)NCC=C)cc2)n(c(c2ccccc2)c1)Cc1cccc(n1)N	-3.851	-3.929

Note: *Test set compounds

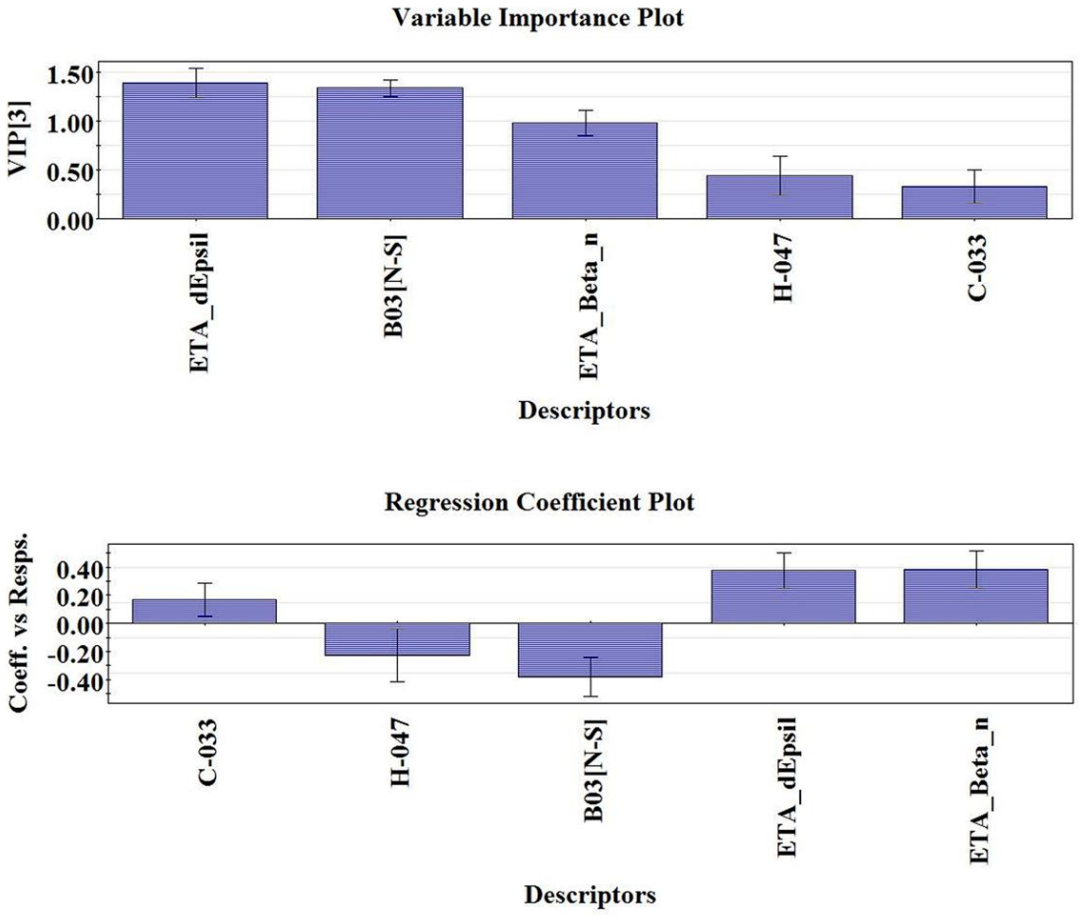


Figure 4.27. Regression coefficient plot and variable importance plot (VIP) of the final PLS model against BACE1 enzyme.

The most contributing descriptor as per the VIP plot is ETA_dEpsilon_D, an extended topochemical atom descriptor, denoting the measure of the contribution of hydrogen bond donor atoms, i.e., the presence of groups such as –OH, –NH_2, –SH, etc[369]. The positive regression coefficients of this descriptor indicated that the activity of inhibitors is directly proportional to the numerical value of ETA_dEpsilon_D. Thus, the compounds having a higher number of hydrogen bond donor atoms may enhance the BACE1 enzyme inhibitory activity as shown in (**Figure 4.28**) compounds like **28** (N-carbamimidoyl-2-(2,5-diphenyl-1H-pyrrol-1-yl)acetamide) (pIC_{50}: -4.28), **82** (6-((2-(2-chlorophenyl)-5-(4-(pyrazin-2-yloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-2-amine) (pIC_{50}: -2.67) and **11** (3-bromo-N-(4-(1-(2-guanidino-2-oxoethyl)-5-phenyl-1H-pyrrol-2-yl)phenyl)benzamide) (pIC_{50}: -2.77) and their corresponding descriptor values are 0.071, 0.066 and 0.0648 respectively. In contrast, compounds like **31** ((3S,5S)-5-phenyl-3-(phenylthio)furan-2,4(3H,5H)-dione), (pIC_{50}: -5.02) **36** ((S)-1-benzyl-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: -5.32) and **38** ((S)-1-phenethyl-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: -8.26) have no such atom to form a hydrogen bond, leading to lower inhibitory activity (**Figure 4.28**). From these observations, we have concluded that a hydrogen bond donor group is important for BACE1 inhibitory activity.

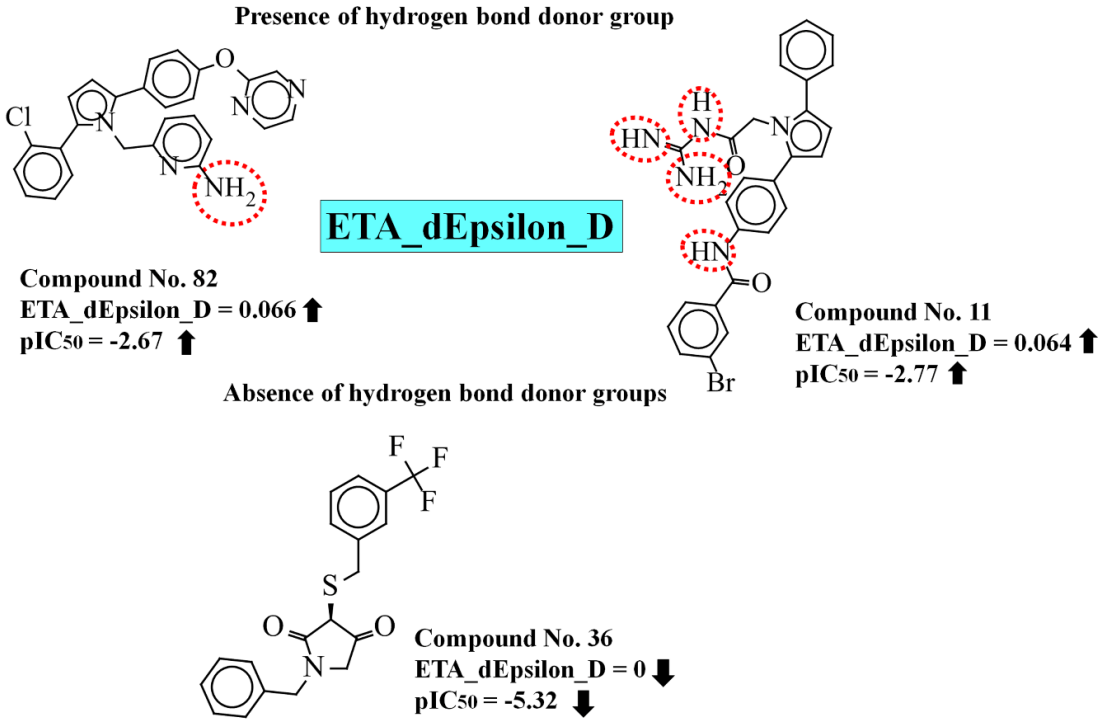


Figure 4.28. Impact of ETA_dEpsilon_D descriptor on pIC_{50} of the compounds.

The next significant descriptor, **B03[N-S]** is an 2D atom pair descriptor that accounts for the presence/absence of N-S fragment at the topological distance 3[69]. It contributes negatively toward the endpoint value which suggested that the numerical values of the descriptor are inversely proportional to the inhibitory activity. Thus, the compounds bearing such fragments show lower values of inhibitory activity as evidenced by (**Figure 4.29**) compounds **40** ((R)-1-(4-methoxybenzyl)-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: -8.38), **38** ((S)-1-phenethyl-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: -8.26) and **39** ((S)-3-(phenethylthio)-1-(3-(trifluoromethyl)benzyl)pyrrolidine-2,4-dione) (pIC_{50}: -8.32)) whereas, compounds having no such fragments show higher BACE1 inhibitory activity as shown in compounds **1** ((S)-N-(3-(2-amino-4-cyclohexyl-1-methyl-5-oxo-4,5-dihydro-1H-imidazol-4-yl)phenyl)furan-2-carboxamide) (pIC_{50}: -1.301) **45** (R)-8-(3-(2-fluoropyridin-3-yl)phenyl)-8-(4-(trifluoromethoxy)phenyl)-2,3,4,8-tetrahydroimidazo[1,5-a]pyrimidin-6-amine) (pIC_{50}: -1.477) and **72** (2-((2-amino-6-((2-(2-chlorophenyl)-5-(4-(pyrimidin-5-yloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-3-yl)oxy)ethanol)) (pIC_{50}: -1.602) (**Figure 4.29**).

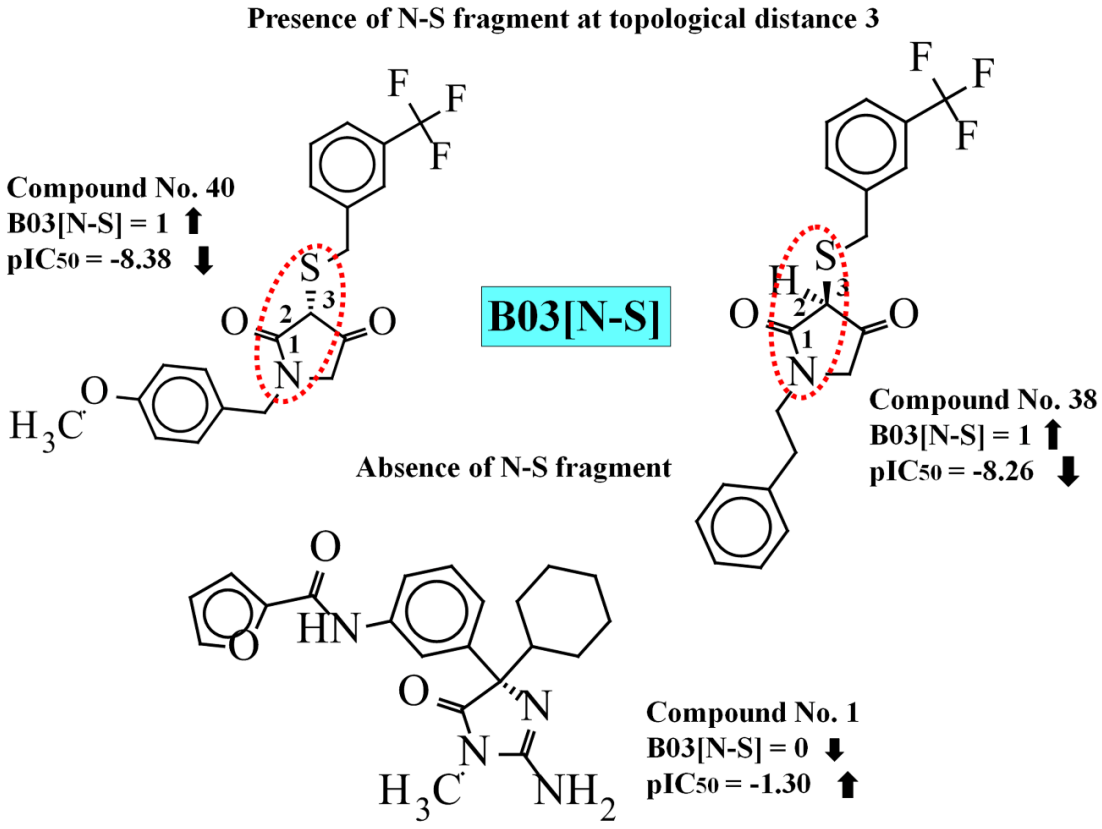


Figure 4.29. Impact of B03[N-S]descriptor on pIC_{50} of the compounds.

Another extended topochemical atom (ETA) descriptor, **ETA_Beta_ns_d,** represents the summed contribution of lone electron pairs capable of forming resonance interaction with an aromatic system. It is defined as the sum of all $\beta_{ns(\delta)}$ values of all vertices[369]. The positive regression coefficient of this descriptor indicates that heteroatoms with a lone pair of electrons capable of resonance with an aromatic system are beneficial for the enzyme inhibitory activity as shown in compounds (**Figure 4.30**) **72** (2-((2-amino-6-((2-(2-chlorophenyl)-5-(4-(pyrimidin-5-yloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-3-yl)oxy)ethanol) (pIC_{50}: -1.60), **73** (2-((2-amino-6-((2-(2-chlorophenyl)-5-(4-(pyrimidin-5-yloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-3-yl)amino)ethanol) (pIC_{50}: -1.84) and **76** (2-((2-amino-6-((2-(2-chlorophenyl)-5-(4-(pentyloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-3-yl)amino)ethanol)) (pIC_{50}: -2.23) (all these compounds have descriptor value 2). In contrast, a low number of heteroatoms is detrimental to the enzyme inhibitory activity as we have observed in (**Figure 4.30**) compounds **42** ((S)-1-(benzo[d][1,3]dioxol-5-ylmethyl)-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione**)** (pIC_{50}: -8.42), **43** ((S)-1-phenethyl-3-(phenethylthio)pyrrolidine-2,4-dione) (pIC_{50}: -8.62) and **44** ((R)-1-benzyl-3-(phenethylthio)pyrrolidine-2,4-dione)) (pIC_{50}: -8.62).

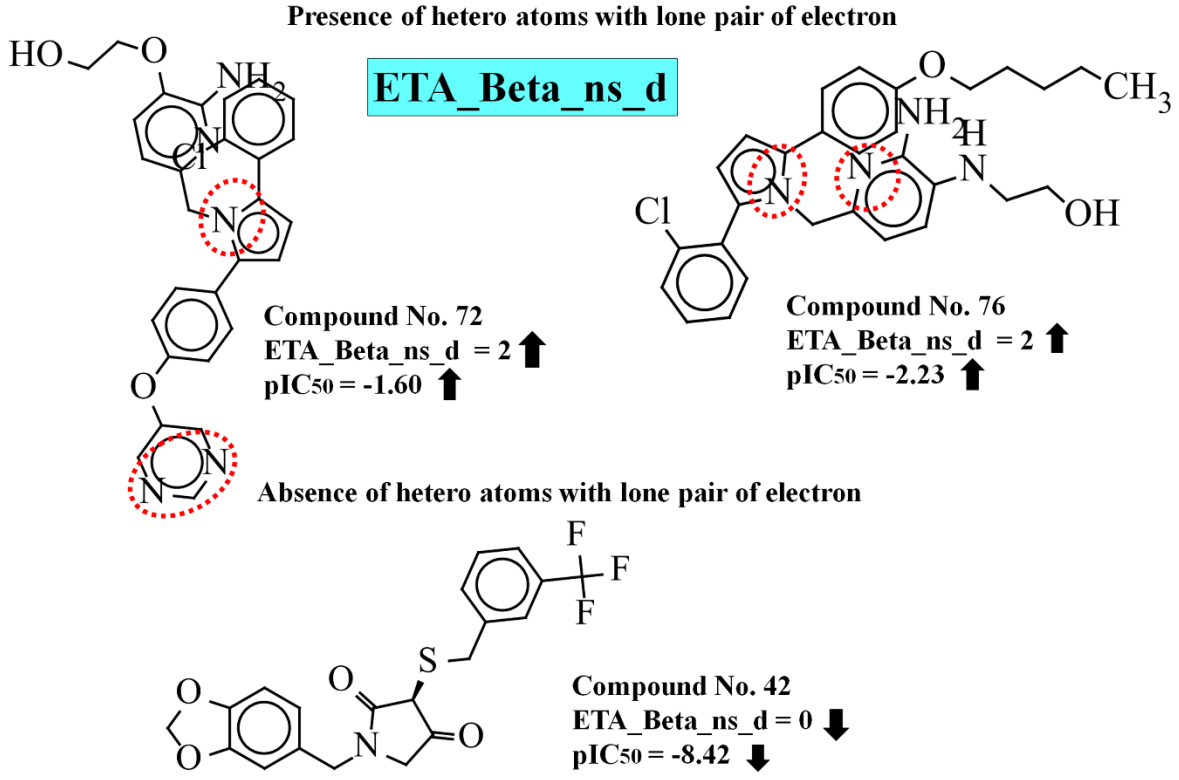


Figure 4.30. Contribution of ETA_Beta_ns_d to pIC_{50} of the compounds.

An atom-centered fragment descriptor, **H-047**, stands for the number of H atoms attached to $C^1(sp3)/C^0(sp2)$; where the superscript represents the formal oxidation number (the formal oxidation number of a carbon atom which is equal to the sum of the conventional bond orders with electronegative atoms)[370]. This descriptor is defined as the number of specific atom types in a molecule which is calculated by knowing only the molecular composition and atom connectivity[371]. The negative regression coefficient of this descriptor suggests that a higher numerical value of this descriptor leads to lower inhibitory activity as evidenced by the compounds **67** ((S)-8-(3-(4-methoxybenzyl)phenyl)-8-phenyl-2,3,4,8-tetrahydroimidazo[1,5-a]pyrimidin-6-amine) (pIC_{50}: -4.57), **40** ((R)-1-(4-methoxybenzyl)-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: -8.38), and **43** ((S)-1-phenethyl-3-(phenethylthio)pyrrolidine-2,4-dione)) (pIC_{50}: -8.62) (corresponding descriptor values are 20, 18 and 17 respectively) (**Figure 4.31**). On the contrary, the compounds with lower descriptor values show higher BACE1 enzyme inhibitory activity as observed in the case of compounds **1** ((S)-N-(3-(2-amino-4-cyclohexyl-1-methyl-5-oxo-4,5-dihydro-1H-imidazol-4-yl)phenyl)furan-2-carboxamide) (pIC_{50}: -1.30), **5** ((S)-2-amino-4-cyclohexyl-1-hexyl-4-phenyl-1H-imidazol-5(4H)-one) (pIC_{50}: -2.43) and **6** ((S)-6-(2-amino-4-cyclohexyl-5-oxo-4-phenyl-4,5-dihydro-1H-imidazol-1-yl)hexanoic acid) (pIC_{50}: -2.49) (their corresponding descriptor values are 9, 7 and 7 respectively) (**Figure 4.31**).

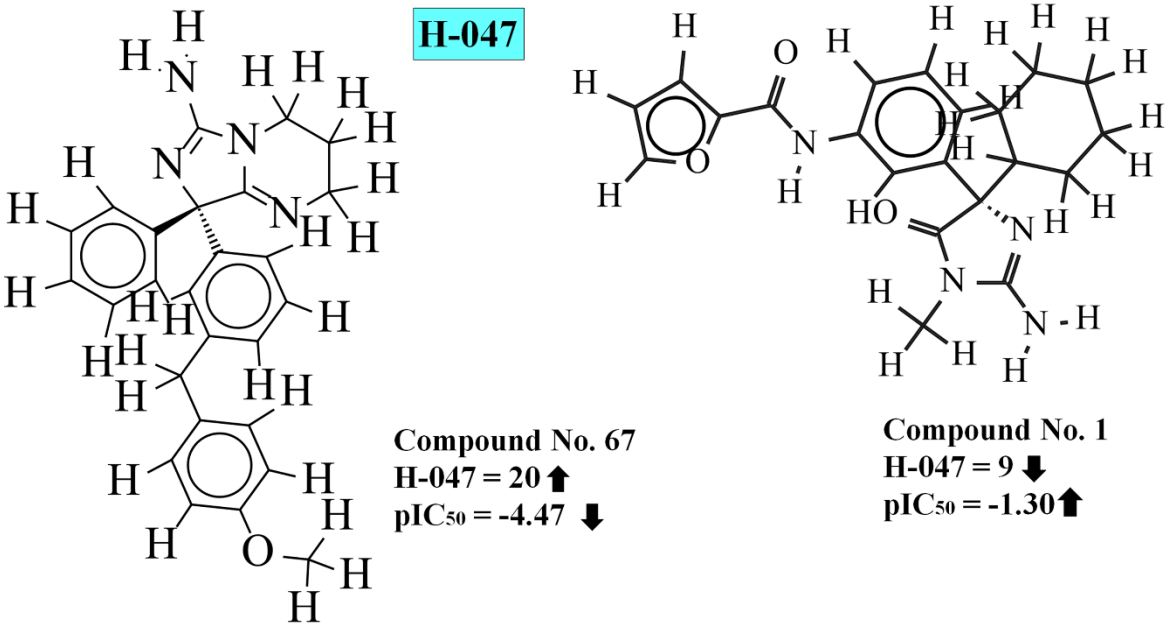


Figure 4.31. Impact of H-047descriptor on pIC_{50} of the compounds.

Another atom-centered fragment descriptor, **C-033,** stands for the fragment R--CH..X. It represents the number of the R--CH…X fragments in a molecule which mean a central carbon atom (C) on an aromatic ring has a carbon neighbor (R), a heteroatom neighbor (X-any heteroatom (O, N, S, P, Se, and halogens)) and the third hydrogen (H) neighbor outside the ring. "--" and "… " stand for aromatic and aromatic single bonds, respectively[69, 371]. For these β-secretase enzyme inhibitors, this fragment indeed plays an important role in the binding process and may influence the inhibitory activity prominently. The positive impact of this descriptor towards the inhibitory activity against β-secretase enzyme was indicated by their positive regression coefficient. Thus, the information obtained from this descriptor suggested that the molecules containing R--CH..X fragment showed higher inhibitory activity to the β-secretase enzyme as shown in compounds **1** ((S)-N-(3-(2-amino-4-cyclohexyl-1-methyl-5-oxo-4,5-dihydro-1H-imidazol-4-yl)phenyl)furan-2-carboxamide) (pIC_{50}: -1.30), **35** ((3S,5S)-5-((R)-sec-butyl)-3-((furan-2-ylmethyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: -5.30) and **23** (2-((5-cyano-4-(4-methoxyphenyl)-6-oxo-1,6-dihydropyrimidin-2-yl)thio)-N-(4-phenylthiazol-2-yl)acetamide)) (pIC_{50}: -3.67) (Figure 7) while compounds **42** ((S)-1-(benzo[d][1,3]dioxol-5-ylmethyl)-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: -8.42), **43** ((S)-1-phenethyl-3-(phenethylthio)pyrrolidine-2,4-dione) (pIC_{50}: -8.62) and **44** ((R)-1-benzyl-3-(phenethylthio)pyrrolidine-2,4-dione**)** (pIC_{50}: -8.66) show lower inhibitory activity due to the absence of this fragment (**Figure 4.32**).

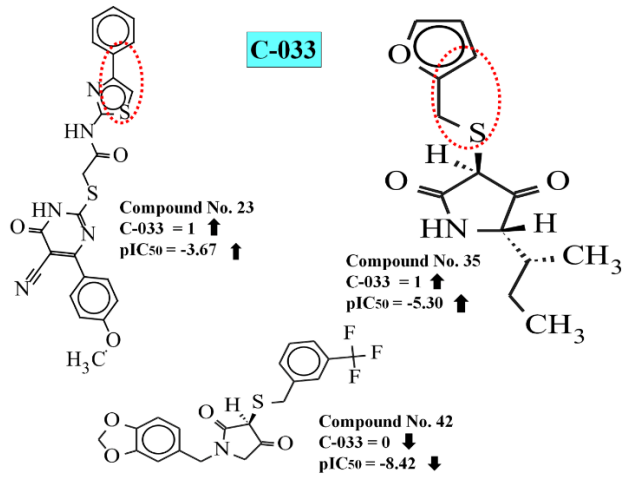


Figure 4.32. Impact of C-033 descriptor on pIC_{50} of the compounds.

4.2.1.2. Randomization model of the PLS model

The predictive quality of the developed model will be poor until the observations are not appropriately independent of each other. The randomization method is a way to test the robustness of the developed model[310]. The purpose of the development of a randomization plot is to identify that the selected descriptors are appropriate and the reported model is not due to chance correlation. In the randomization method, many numbers of models are developed by several runs for which the original descriptor matrix X is kept fixed, and only the vector Y is randomized. The validation metrics of the developed model under such conditions should be poor and the value of the R^2_{yrand} intercept should not be more than 0.3 and the value of the Q^2_{yrand} intercept should not exceed 0.05[310]. In the present study, for the training set, the X data were kept constant and the Y data were scrambled randomly using 100 permutations. The model obtained from such condition shows the intercepts as follows: (see **Figure 4.33**) R^2_{yrand} = -0.0128 and Q^2_{yrand} = -0.26, which signify the validity of the model and confirm that the reported model was not obtained by any chance. The above results suggest that the developed model is non-random and robust, and suitable for the prediction of the inhibitory activity against the BACE1 enzyme.

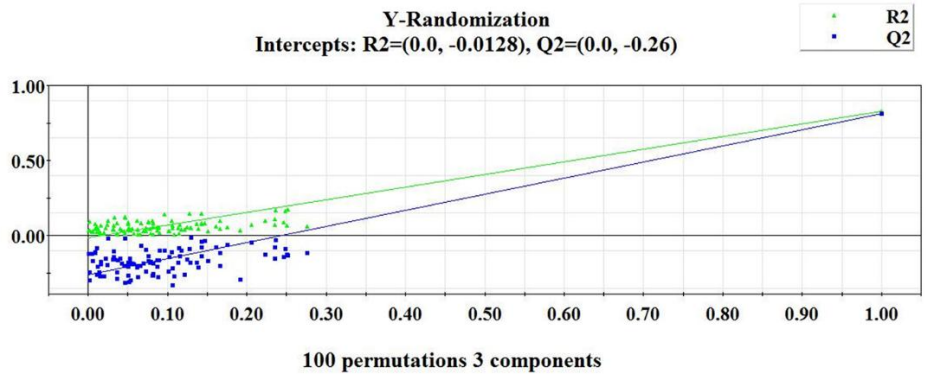


Figure 4.33. Model randomization plots for the final PLS model against the BACE1 enzyme.

4.2.1.3. Applicability domain of the PLS model

Prediction of the activity of the entire space of chemicals is not possible by a robust and validated QSAR model until the compounds are predicted within the applicability domain of the model. The applicability domain (AD) gives a theoretical province in chemical space well-defined by the respective model descriptors and responses in which the predictions of activity are reliable[372]. In this study, we have checked the applicability domain of test set compounds at a 95% confidence level using the DModX (distance to model in X-space) approach available within SIMCA-P 10.0 software (Available from http://www.minitab.com/en-us/products/minitab/). In the plot (see **Figure 4.34**), it was observed that all the test set compounds are within the critical DModX value (D-Crit= 2.584) except compound **29** ((3R,4aR,4bR,8aR,9aS)-3-((4-(6-methoxynaphthalen-2-yl)-1H-1,2,3-triazol-1-yl)methyl)dodecahydro-1H-pyrido[3,4-b]indole).

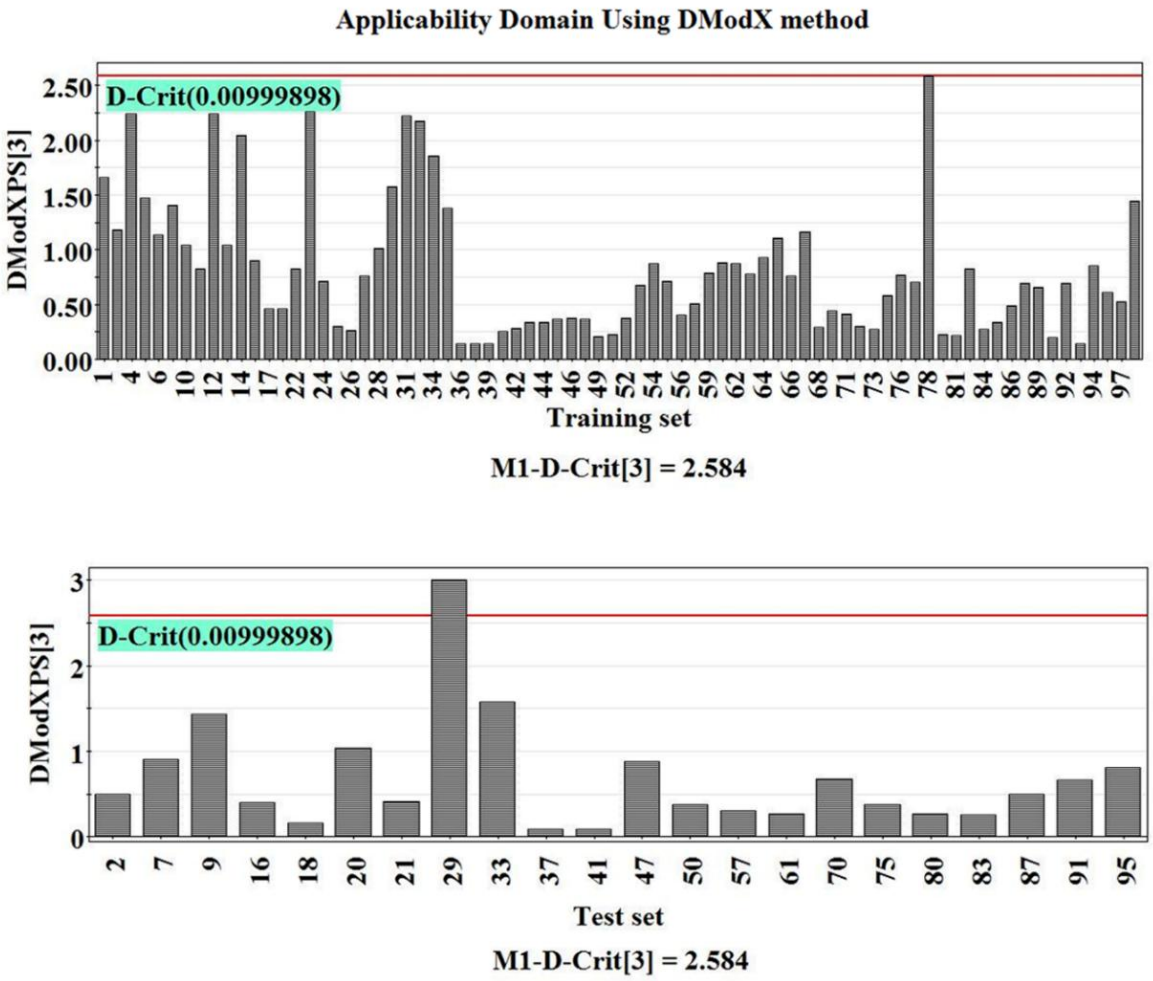


Figure 4.34. Applicability domain DModX values of the training and test set compounds at 99% confidence level of the developed final PLS model against BACE1 enzyme.

4.2.1.4. Loading plot of the PLS model

A loading plot of a PLS model (see **Figure 4.35**) gives information about the relationship between the X-variables and Y-variables. In the loading plot, a descriptor that is close to zero is not well associated with the trends contained in the related scores[373]. In the loading plot (see **Figure 4.35**)**,** we can observe that the X-variable **ETA_dEpsilon_D** is significant for the Y-variable (pIC_{50}) because it is very much close to the Y-variable. It has also been observed that this descriptor is situated on the same side of the Y-variable. From this observation, it can be concluded that this descriptor is directly proportional to the activity. Thus, the BACE1 enzyme inhibitory activity may increase with increasing the numerical value of this descriptor. On the other hand, the variable **B03[N-S]**, which is situated on the opposite side of the plot origin concerning the activity (Y-variable) contributes negatively towards the BACE1 enzyme inhibitory activity. The algebraic sign of the PLS loading is also taken into account, which gives important information about the correlation between the variables.

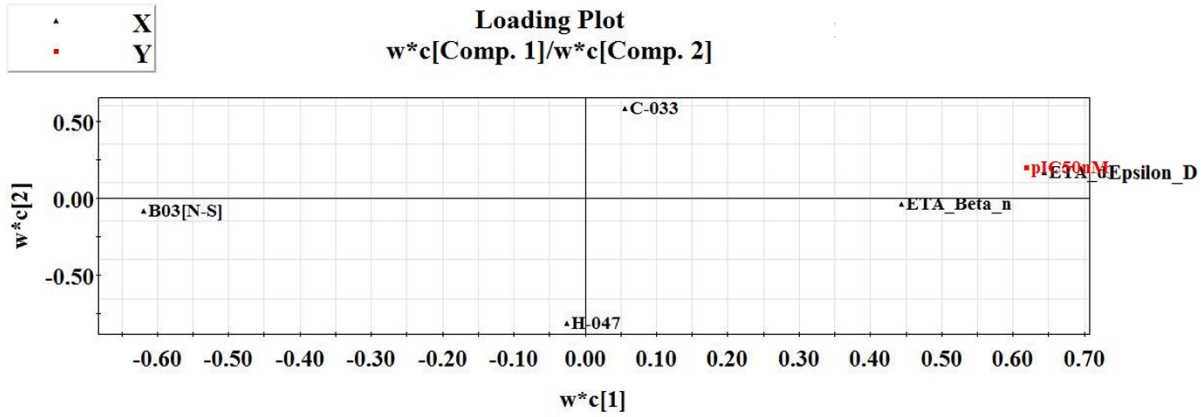


Figure 4.35. Loading plot for final PLS model against BACE1 enzyme.

4.2.2. 3D pharmacophore modeling

In this investigation, ten different pharmacophore hypotheses were developed using a training set (22) of compounds. The best pharmacophore model (Hypo-1) was selected based on the different internal validation parameters such as high correlation coefficient (r: 0.912), lower root means square deviation (rmsd: 1.320), Maximum Fit (8.393), total cost (100.355), configuration cost (16.122), error (83.054) and weight (1.177) was found to be acceptable. In terms of the actual cost for Hypo-1, it is found much closer to the fixed cost with only a difference of 19.240 bits (mentioned in **Table 4.5**) which specifies an accurate correlation of the dataset. From **Table 4.5**, we can see that in Hypo 1 there is a large difference of 78.229 bits between the actual cost and the null cost. Based on all validation matrices Hypo-1 was found to be the best one among the ten hypotheses with one hydrogen bond acceptor (HBA), two hydrophobic (HYD), and one ring aromatic (RA) features (**Figure 4.36**). The external validation of the model has been performed by mapping the test set molecules on the Hypo-1 using the same parameters as we have used in the development of the pharmacophore model. After mapping, we observed that 64 molecules from the data set of 76 compounds were mapped, and only 10 compounds failed in absence of the features that appeared in the developed pharmacophore model. To judge the predictive quality of the selected model to categorize

the compounds into active and less active BACE1 enzyme inhibitors were analyzed by comparing the observed activity with predicted activity by the classification-based technique. The values of different validation parameters for training and test sets are given in **Table 4.5** (qualitative validation parameters). From the observation of activity predicted by the selected model, we have found that the model correctly classified 11 out of 12 compounds as more actives and 9 out of 10 compounds as fewer actives for the training set. For the test set, the model correctly classified 18 out of 25 compounds as most active and 30 out of 39 compounds as less active. Aher et al.[323] suggested that if the values of different validation parameters for both the training and test sets are greater than 60 %, it means that the model is following acceptability criteria and is good enough to predict the activity of the new compound of the same chemical domain. From the above observation, we have concluded that Hypo-1 is best appropriate for the classification of more active BACE1 enzyme inhibitors.

Table 4.5. Different quantitative and qualitative validation parameters of the **Hypo-1** model were obtained for the training and test sets generated by the HypoGen algorithm.

Quantitative validation parameters						
Hypo.	**Total cost**	**ΔCost[a]**	**ΔCost[b]**	**RMS**	**Correlation**	**Features**
1	100.355	78.229	19.240	1.320	0.912563	HBA, HYD, HYD, RA

Dataset	**No. of compounds**	**Qualitative validation parameters**					
		Sensitivity	**Specificity**	**Accuracy**	**Precision**	**F-measure**	**G-Means**
Train	22	91.66	90	90.90	91.66	91.66	90.82
Test	76	68	76.92	73.43	65.38	66.66	72.32

Cost difference[a]= Null cost - total cost, Cost difference[b]= Total cost - fixed cost, Null cost = 178.584, Fixed Cost = 81.1148, Best records in pass: 5, Config. Cost=16.1229, c= Best Hypothesis, Note – RA: Ring aromatic, HYD: Hydrophobic, HBA: Hydrogen bond acceptor. *Compounds with IC_{50} <1000 nM: more active (H) and IC_{50}>1000 nM: less active (L).

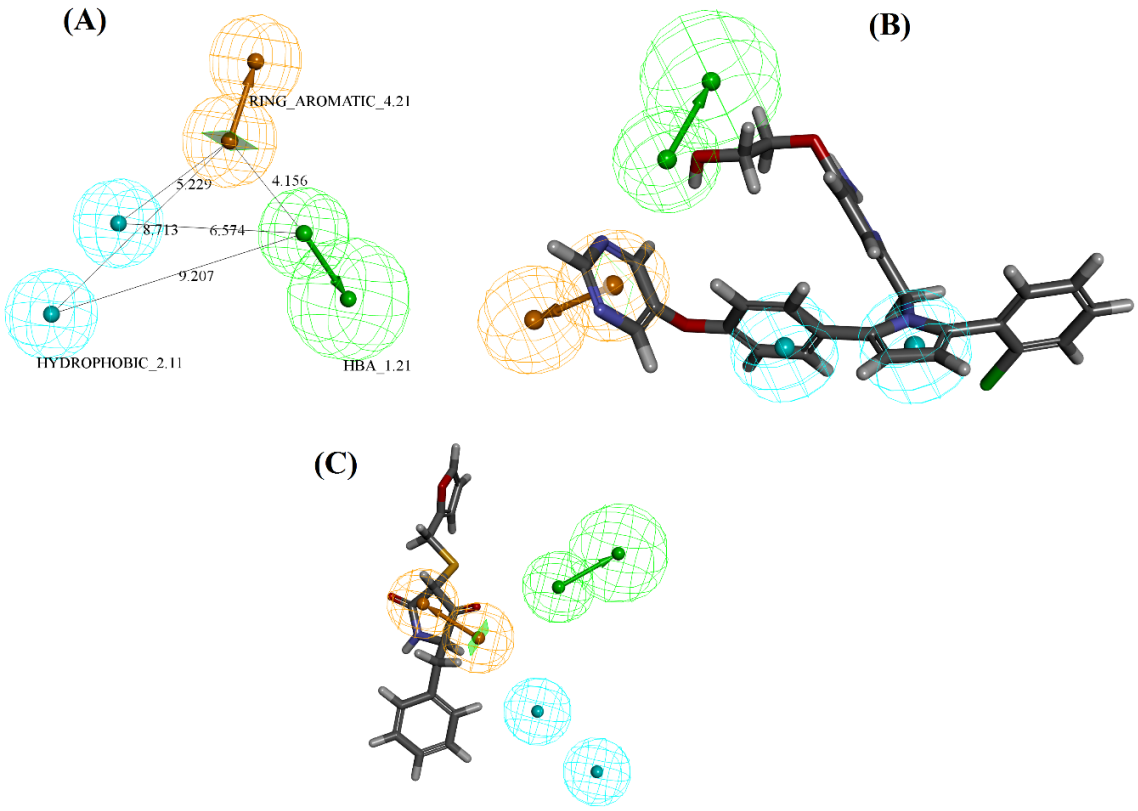


Figure 4.36. The best pharmacophore model (Hypo1) of BACE1 enzyme inhibitors generated by the HypoGen module: (A) the best pharmacophore model Hypo1 represented with distance constraints (Å), (B) Hypo1 mapping with one of the most active compounds **72** of test set compounds and (C) Hypo1 mapping with one of the least active compounds **32** of test set compounds. Pharmacophoric features are colored as follows: hydrogen bond acceptor (green), hydrophobic (cyan), and ring aromatic (orange).

4.2.2.1. Relation of the 3D- pharmacophore model with the 2D QSAR model

In the dataset, all compounds have at least one ring aromatic feature as observed from both QSAR models (2D QSAR and 3D QSAR pharmacophore models). The RA feature is an initial necessity for the inhibitory activity against the BACE1 enzyme. The RA feature from the pharmacophore model is well corroborated with the **ETA_Beta_ns_d** and **C-033** descriptors of the 2D-QSAR model (**Equation 4.3**). **Figure 4.37** has demonstrated that the most active compound of the training set (**45**, IC_{50}: 30nM) mapped correctly with all features appearing in Hypo 1 (**Figure 4.37**). One benzene ring lies in the RA region, a nitro group in the hydrogen bond acceptor region, and a halogen atom of the aromatic ring lies in the hydrophobic region. On the other hand, the least active compound (**70**, IC_{50}: 78700 nM) of the training set does not map correctly with Hypo 1 because of the absence of RA and HBA features in the molecules (**Figure 4.37**). The hydrophobic feature from the developed pharmacophore model is well corroborated with the C-033 descriptor of the 2D-QSAR model (**Equation 4.3**). For the most active compound of the training set (**45**), the hydrophobic feature of Hypo 1 was mapped completely. The most active compound of the test set (**72**, IC_{50}: 40nM) mapped correctly on Hypo-1 with all three features (**Figure 4.36**) appeared in the developed

pharmacophore model. The least active compound (**32**, IC_{50}: 145000 nM) of the test set was mapped partially with Hyo-1 (**Figure 4.36**). From the above observation, it was concluded that the absence of these three features that appeared in the developed pharmacophore model decreases the inhibitory activity of compounds against the BACE1 enzyme. The results obtained from the F-test suggest that the selected pharmacophore model (**Hypo-1**) is not due to a chance. This observation was confirmed by a lower cost value (100.355) of the selected pharmacophore model than the average cost of randomized pharmacophore models (155.91) and a higher correlation coefficient (R: 0.912) of the selected pharmacophore model than the average correlation coefficient of random models (Rr: 0.568). The actual and randomized total cost and correlation values of hypotheses for the F-test are **given in Figure 4.38**.

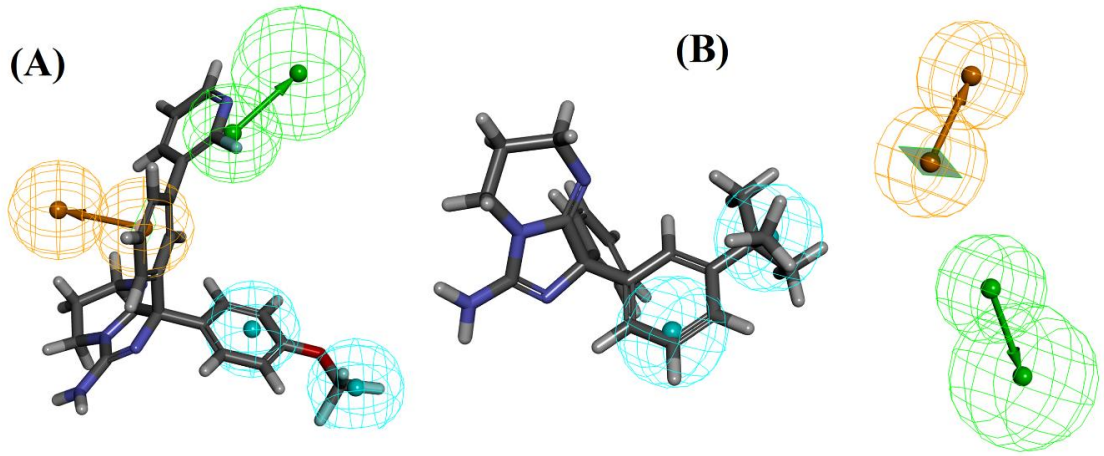


Figure 4.37. Pharmacophore mapping with training set compounds: (A) Hypo1 mapping with one of the most active compounds **45** of training set compounds and (B) Hypo1 mapping with one of the least active compounds **70** of training set compounds.

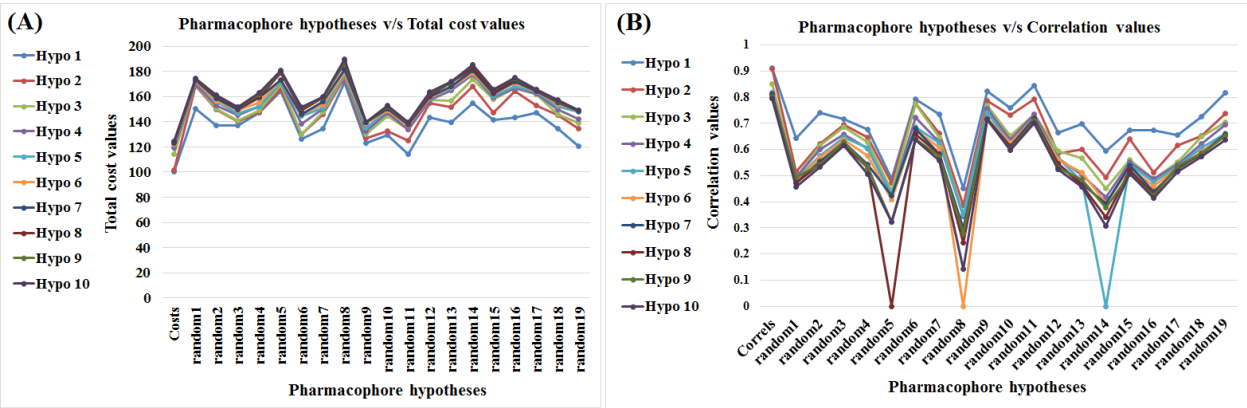


Figure 4.38. The actual and randomized total cost (A) and (B) correlation values of hypotheses for the F-test.

4.2.3. Molecular Docking analysis

In the present study, a molecular docking study was performed of the most active and least active compounds of the dataset. The molecular docking study suggests that the molecules interacted with a pocket containing (ILE A:110, PHE A:108, VAL A:332, GLY A:11, GLY A:13, GLY A:230, VAL A:332, ILE A:226, LEU A:30 and GLY A:34 (hydrophobic nature), GLN A:73, SER A:229, THR A:231, THR A:329, THR A:71, GLN A:73, GLN A:12 and THR A:232 (hydrophilic nature), ARG A:235, ASP A:32, ASP A:228 and LYS A:107 (charged) and TYR A:71 (amphipathic nature)) amino acid residues. Docking results and correlation with the 2D-QSAR model are depicted in **Table 4.6**.

4.2.3.1. Molecular Docking for the most active compounds from the dataset

Three most active compounds from the dataset (pIC_{50} = -1.301, -1.47, and -1.77 respectively) namely **1**, **45,** and **16** interacted with the active site amino acid residues through different interactions forces like hydrogen bonding interactions (carbon-hydrogen bonds conventional hydrogen bonds, and π-donor hydrogen bond), π-interactions (π-alkyl bonds, alkyl hydrophobic, π-lone pair and π-π-T-shaped), salt bridge interaction and halogen bonding (halogen bonding is an attractive, non-covalent interaction that can form between an electrophilic region of a halogen atom (fluorine) in a molecule and a nucleophilic region of a molecule). The amino acid residues involved in interaction with these compounds such as THR A:231, GLY A:11, GLN A:73, THR A:71, ASP A:32, ASP A:228, VAL A:332, GLY A:13, THR A:232, GLY A:230, SER A:229, GLY A:34, LYS A:107, ILE A:110, PHE A:108 and ARG A:235 (shown in **Figure 4.39**).

Figure 4.39 shows that compound **1** (one of the most active compounds in the dataset) interacts with GLN A:73, ASP A:32, and ASP A:228 amino acid residues through hydrogen bonding interaction, with VAL A:332, THR A:71 and THR A:231 amino acid residues through alkyl, π-alkyl and π-lone pair interactions respectively, with ASP A:228 amino acid through salt bridge formation and with ASP A:32 amino acid through attractive charges (interaction between two oppositely charged atoms).

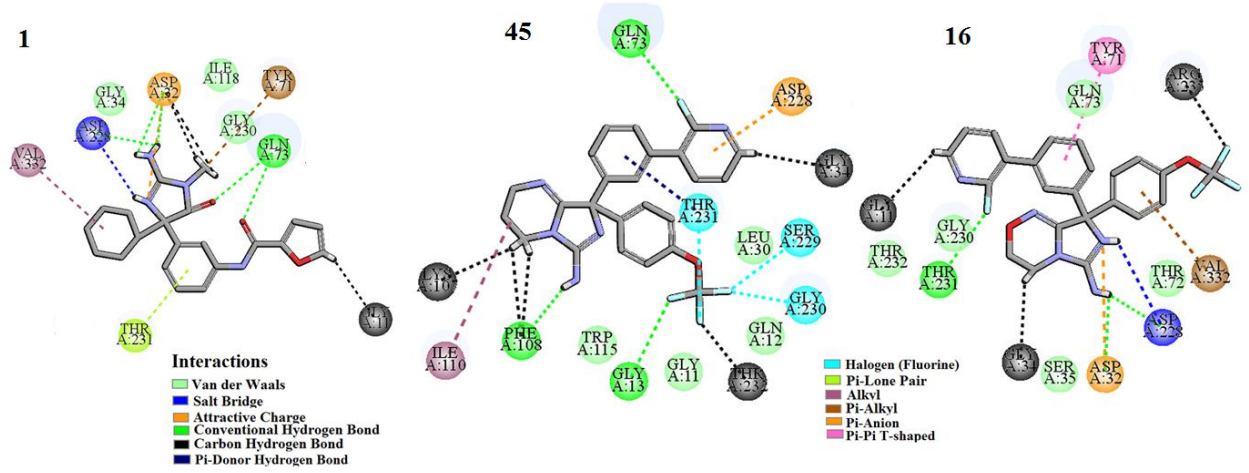


Figure 4.39. Docking interaction of the most active compound **1, 45, and 16.**

Another most active compound, **45,** interacts with the amino acid residues through hydrogen bonding **(**GLNA:73, GLY A:13, PHE A:108, LYS A:107, THR A:232, and GLY A:34**)**, halogen bonding (THR A:231, SER A:229, and GLY A:230), π-anion (ASP A:228), π-donor hydrogen bond (THR A:231) and alkyl bonding (and ILE A:110) interactions (see **Figure 4.39**). **Figure 4.39** shows that compound **16** interacts with GLY A:11, ARG A:235, GLY A:34, THR A:231, ASP A:32, ASP A:228 (through hydrogen bonding), THR A:71 (π-π-T-shaped), VAL A:332 (π-alkyl bonds) ASP A:228 (salt bridge interaction) and ASP A:32 (attractive charges) amino acid residues.

4.2.3.2. Molecular Docking for the least active compounds from the dataset

The three least active compounds from the dataset **(**pIC_{50} = -2.70, -2.77, and -2.770 respectively), namely **84**, **10, and 85,** interact with the active site amino acid residues through different interactions forces like hydrogen bonds (carbon-hydrogen bonds and conventional hydrogen bonds), pi-interaction (π-anion, π-π-T-stacking π-alkyl, and alkyl), salt bridge attractive charge interactions. The amino acid residues involved in interaction with these compounds are VAL A:332, GLN A:73, GLY A:230, ASP A:32, TYR A:71, ASP A:228, GLY A:34, ILE A:226, THR A:329, GLN A:12 and LEU A:30 (shown in **Figure 4.40)**.

Figure 4.40 shows that compound **10** (one of the least active compounds from the dataset) interacts with GLN A:73, GLY A:230, ASP A:32, ASP A:228, and GLY A:34 amino acid residues through hydrogen bonding interaction, TYRA:71 and VAL A:332 amino acid residues through π-alkyl, π-π -T-stacking and alkyl, ASP A:228 amino acid through salt bridge formation and ASP A:32 amino acid through attractive charges interaction. Another least active compound from the dataset, **85,** interacts with the amino acid residues through hydrogen bonding **(**THR A:326 and GLN A:12**)** and hydrophobic interaction such as π-π-T-stacking, π-alkyl and π anion bonding (TYR A:71, LEU A:30, and ASP A:228) (**Figure 4.40**). **Figure 4.40** shows that compound **84** interacts with GLN A:73 (through hydrogen bonding), ASP A:228 (π-anion), and ILE A:226 and VAL A:332 (π-alkyl bonds) amino acid residues.

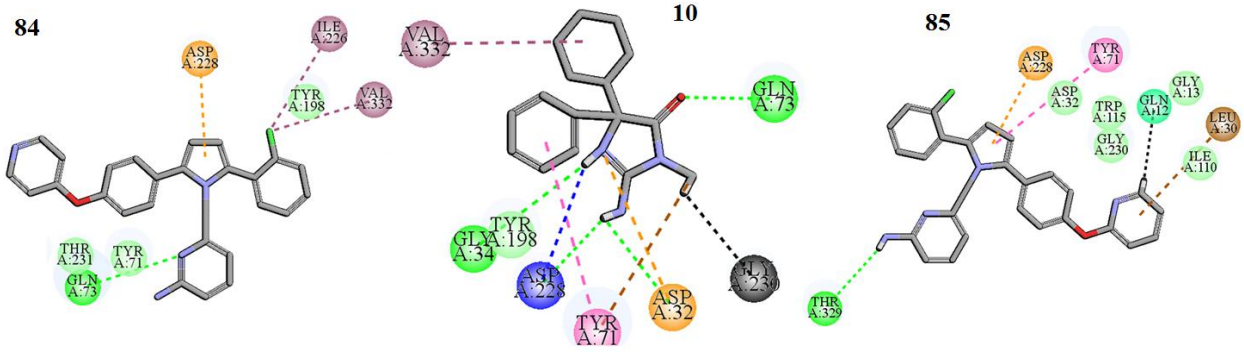


Figure 4.40. Docking interaction of the least active compound **84, 10, and 85.**

Finally, docking analysis has demonstrated that the most active compounds from the dataset such as **1, 45** and **16** (shown in **Figure 4.39**) interacted with the maximum number of active amino acid residues with the higher number of interacting forces (non-covalent forces) in comparison to the least active compounds from the data set like **10, 85** and **84** (shown in **Figure 4.40**). In pharmacophore mapping, we have also observed that the most active compounds from the data set correctly mapped with all features that appeared in the model, whereas the least active compounds partially mapped with the model. It is possible that the least

active compounds from the data set failed to map in the absence of the features appearing in the developed models, which are most important for inhibitory activity against the BACE1 enzyme.

Table 4.6. Docking results and correlation with the 2D-QSAR model.

S. No.	Compound Number	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
1	**1** (high pIC_{50})	45.340	THR A:231, GLY A:11, GLN A:73, TYR A:71, ASP A:32, ASP A:228 and VAL A:332	Vdw, Hydrogen bonding, Pi-alkyl, alkyl, salt bridge, Attractive charge and π-lone pair	ETA_dEpsilon_D, ETA_Beta_ns_d, C-033 and H-047
2	**45** (high pIC_{50})	39.503	GLY A:13, THR A:232, GLY A:230, SER A:229, GLY A:34, ASP A:228, GLN A:73, THR A:231, LYS A:107, ILE A:110 and PHE A:108	Vdw, Hydrogen bonding, halogen (fluorine), π-anion, π-donor hydrogen bond and alkyl	ETA_dEpsilon_D, ETA_Beta_ns_d and H-047
3	**16** (high pIC_{50})	51.278	TYR A:71, ARG A:235, VAL A:332, ASP A:228, ASP A:32, GLY A:34, THR A:231 and GLY A:11	Vdw, salt bridge, Attractive charge, hydrogen bonding, π-π-T-shaped and π-alkyl	ETA_dEpsilon_D, ETA_Beta_ns_d and H-047
4	**10** (low pIC_{50})	46.118	VAL A:332, GLN A:73, GLY A:230, ASP A:32, TYR A:71, ASP A:228 and GLY A:34	Vdw, salt bridge, Attractive charge, hydrogen bonding, π-π-T-shaped, alkyl and π-alkyl	H-047 and ETA_dEpsilon_D
5	**84** (low pIC_{50})	43.144	GLN A:73, ASP A:228, ILE A:226 and VAL A:332	Vdw, hydrogen bonding, π-anion and alkyl	H-047, ETA_dEpsilon_D and ETA_Beta_ns_d
6	**85** (low pIC_{50})	44.710	THR A:329, GLN A:12, ASP A:228, TYR A:71 and LEU A:30	Vdw, hydrogen bonding, π-anion, π - π stacked and π-alkyl	H-047, ETA_dEpsilon_D and ETA_Beta_ns_d

4.2.3.3. Relation with the 2D-QSAR model

Molecular docking analysis has demonstrated that the most active compounds from the dataset such as **1, 45,** and **16** (**see Table 4.6**) interacted with the maximum number of amino acid residues with the higher number of interacting forces with the lower range of CDocker interaction energy (-45.340, -39.503 and -51.278 respectively). On the other hand, the least active compounds from the dataset such as **84** and **85** interacted with less number of active amino acid residues with a lower number of interacting forces, in comparison to the most active compounds, with CDocker interaction energy of -43.144 and -44.710 respectively. It was noted that the interaction energy (-CDocker interaction energy) depends on the number of interactions and the forces involved in the docking. It is not obvious that the most active compounds will always show the highest interaction energy and vice versa. Many compounds show a higher range of interaction energy (for example, compound **10**) due to some insignificant interactions with other elements or amino acids in the active site which have no contribution to the biological activity. Thus, the formation of hydrogen bonding and π-π stacking between the ligands and receptor plays a crucial role in enzyme inhibitory activity. These observations were also observed from the descriptors ETA_dEpsilon_D (a measure of the contribution of hydrogen bond donor atoms) and ETA_Beta_ns_d (a measure of lone electrons entering into resonance with an aromatic system) in the 2D-QSAR model. The information obtained from the descriptor C-033 (R--CH...X) is correlated with hydrogen bonding, attractive charges, and π-donor hydrogen bonding interactions as observed from the docking study (shown in **Figure 4.39**). Thus, from above said information, it was concluded that the hydrogen bonding effect, hydrophobicity, electrostatic interactions, and unsaturation (π-π interaction) feature as obtained from both the 2D QSAR model and docking study are essential for the inhibitory activity against the BACE1 enzyme.

4.2.4. Comparisons of the performance of the present model with previously published models

A comparison of the statistical results obtained from the present QSAR model and previously published models is depicted in **Table 4.7**. Based on the statistical quality in terms of both internal and external validation criteria, the model reported in this work is statistically significant and robust enough as compared to the previously reported models (**Table 4.7**). We have used 2D descriptors only for model development. Many researchers reported QSAR models for the prediction of bioactivity of BACE1 enzyme inhibitors previously using various techniques such as Multiple Linear Regression (MLR) analysis, Partial least squares (PLS), Comparative Molecular Field Analysis (CoMFA), Comparative molecular similarity index analysis (CoMSIA) and linear heuristic method (LHM). In the present study, before the development of the final model, we performed a variable selection strategy using a stepwise regression technique followed by the best subset selection method. The final model was developed by PLS regression technique with 5 selected descriptors using 3 LVs. The 5 selected descriptors reflect the fundamental structural characteristics of molecules that are important in modeling the bioactivity of BACE1 enzyme inhibitors. In comparison with other models, it may be noted the model developed in this study are superior in terms of statistical quality, equation length, LVs, etc. We can see from **Table 4.7** that Ambure et.al.[374] developed PLS and MLR models against BACE1 enzyme inhibitors using only **74** compounds, and the model quality was good. In the present study, we have utilized a wider range of compounds and developed the model with 5 selected descriptors using 3 latent variables. We can also see from **Table 4.7** that Jain et.al.[375] and Chakraborty et.al.[377] developed QSAR models using a very narrow group of samples (**27** and **30** compounds respectively) and developed MLR and LHM-based models respectively. Hossain et.al.[376] reported 2D-QSAR and 3D-QSAR models along with molecular docking and pharmacophore mapping utilizing **106** compounds. The details of different internal and external validation parameters obtained from our model

and previously reported models are given in **Table 4.7**. The docking and pharmacophore mapping results in this study are also well collaborated with descriptors obtained from the developed QSAR model and justify the significance of the developed model. As the present work deals with diverse classes of compounds, the reported model in the present study may be used for screening purposes for the discovery and development of leads against the BACE1 enzyme.

Table 4.7. Comparison of proposed study with previously published studies against BACE1 enzyme.

Sources	E. L.	LV.	Model	Training set			Test set	
				n	R^2	Q^2	n	R^2pred
Model in this study	**5**	**3**	**PLS**	**76**	**0.826**	**0.795**	**22**	**0.846**
Ambure et al. 2016[374]	5	4	PLS	**52**	0.831	0.764	**22**	0.813
Ambure et al. 2016[374]	5	-	MLR	**51**	0.826	0.764	**22**	0.791
Jain et al. 2013[375]	2	-	MLR	**20**	0.895	0.893	**7**	0.903
Hossain et al. 2013[376]	-	10	CoMFA	**71**	0.998	0.765	**35**	0.772
Hossain et al. 2013[376]	-	10	CoMSIA	**71**	0.992	0.730	**35**	0.713
Hossain et al. 2013[376]	-	7	PLS	**71**	0.941	0.792	**35**	0.713
Chakraborty et al. 2017[377]	4	-	LHM	**20**	0.941	0.913	**10**	0.860

Abbreviations: EL = Equation length, LV= Latent variable, MLR= Multiple linear regression, CoMFA= Comparative Molecular Field Analysis, CoMSIA= Comparative molecular similarity index analysis, and LHM= Linear heuristic method.

4.3. Study 3- Cheminformatic modeling of β-amyloid aggregation inhibitory activity against Alzheimer's disease

4.3.1. 2D QSAR analysis

4.3.1.1. Mechanistic interpretation of modeled descriptors

In this analysis, statistically significant and robust 2D-QSAR models using the PLS regression-based technique were developed; the values of the validation parameters are shown below in **Equations 4.4** and **4.5** (**Box 4.1**). The developed models were validated carefully utilizing the deferent validation metrics (internal and external) for the exploration of robust and statistically significant models. Two PLS models were developed with 12 and 13 descriptors using 6 and 7 latent variables (LVs), respectively. The obtained results suggested that the models were acceptable in terms of fitness, stability, and classical predictivity measures. Descriptors appearing in the developed models demonstrated the structural and functional requirements which can improve the inhibitory activity of molecules against β-amyloid aggregation. Presented scatter plots in **Figure 4.41** show the closeness of the observed and predicted values for the β-amyloid aggregation inhibitors (βAAI). The randomization of models assured that the developed models were not found by any chance correlation. The results obtained from the randomized models (for Model 1: R^2intercepts (int) = 0.00833 and Q^2int = -0.296 and for Model 2: R^2int = 0.00756 and Q^2int = -0.392) suggested that the reported models were not obtained by chance (see S3 **Figure 4.42**).

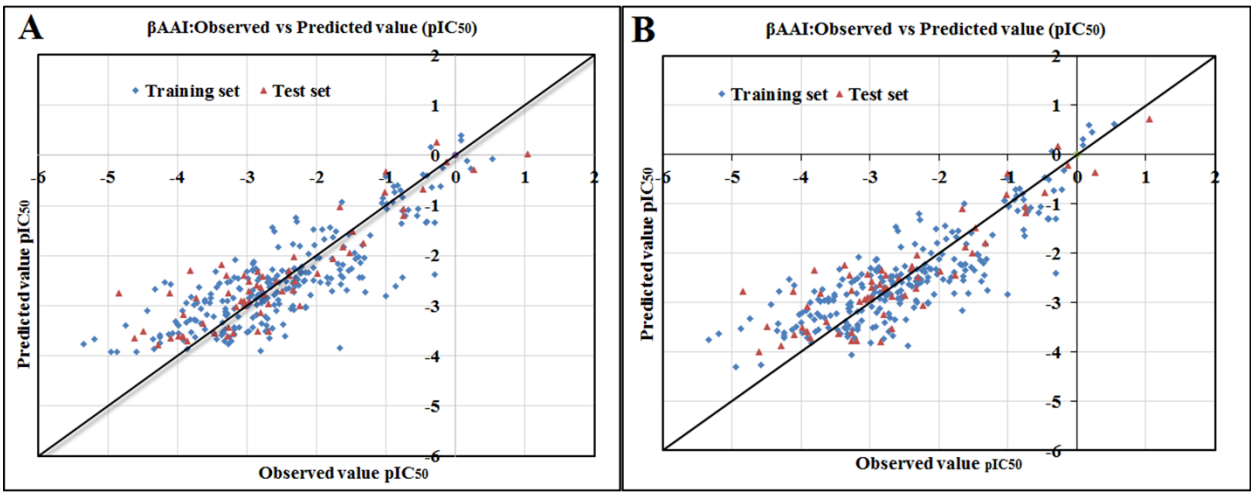


Figure 4.41. The scatter plots of observed and predicted values of final PLS models for β-amyloid aggregation inhibitors (**A**: Model 1, **B**: Model 2).

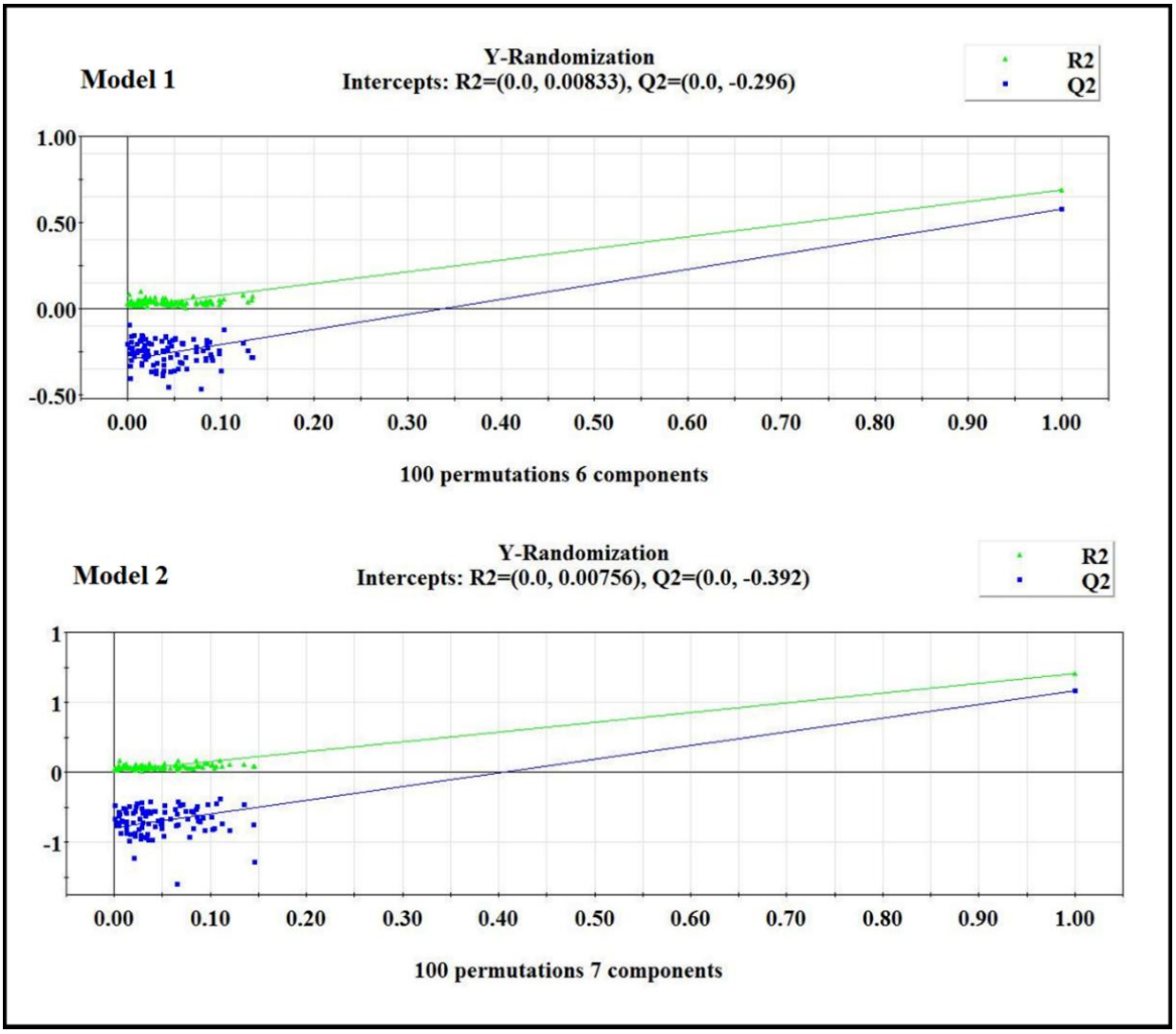


Figure 4.42. Model Randomization plots for final PLS models against β amyloid aggregation.

Box 4.1. PLS-based 2D QSAR models and their statistical validation metrics obtained from the developed models

Model 1: PLS Model against β-amyloid aggregation

$pIC_{50}(nM) = -3.017 + 0.843 \times nArNHR + 0.178 \times SaasC - 0.102 \times F07[C-O] - 1.097 \times NNRS - 0.006 \times D/Dtr12 - 1.185 \times B06[C-N] - 0.730 \times B05[N-N] - 0.655 \times B06[N-O] + 1.610 \times B05[C-N] + 0.079 \times F09[C-C] - 1.051 \times nFuranes - 0.031 \times F06[C-C]$

Equation 4.4

Internal Validation Parameters: $n_{training} = 252$, $r^2 = 0.664$, $Q^2 = 0.621$, Fitting quality = Moderate.

External Validation Parameters: $n_{test} = 62$, $Q^2F_1 = 0.765$, $Q^2F_2 = 0.763$, Avg $R^2m = 0.601$, $\Delta R^2m = 0.199$, CCC = 0.861, MAE = 0.456, SD = 0.394, Fitting quality = Moderate, LV= 6, No. of descriptors = 12

Model 2: PLS Model against β-amyloid aggregation

$pIC_{50}(nM) = -2.842 + 0.754 \times nArNHR + 0.191 \times SaasC - 0.114 \times F07[C-O] - 1.089 \times NNRS - 0.005 \times D/Dtr12 - 0.739 \times B05[N-N] - 1.159 \times B06[C-N] - 0.539 \times B06[N-O] + 1.427 \times B05[C-N] + 0.813 \times F05[O-O] - 1.274 \times nFuranes + 0.076 \times F09[C-C] - 0.032 \times F06[C-C]$

Equation 4.5

Internal Validation Parameters: $n_{training} = 252$, $r^2 = 0.684$, $Q^2 = 0.638$, Fitting quality = Moderate.

External Validation Parameters: $n_{test} = 62$, $Q^2F_1 = 0.771$, $Q^2F_2 = 0.769$, Avg $R^2m = 0.634$, $\Delta R^2m = 0.186$, CCC = 0.867, MAE = 0.462, SD = 0.375, Fitting quality = Good, LV= 7, No. of descriptors= 13

The descriptors in the PLS models are arranged according to their importance and then described separately. The significance level and contribution of the model descriptors towards the β-amyloid aggregation inhibitory activity are determined based on a regression coefficient plot[378] and variable importance plot (VIP)[379] as shown in **Figures 4.43** and **4.44**. The definition and contributions of all the descriptors obtained from the PLS models are mentioned in **Table 4.8**.

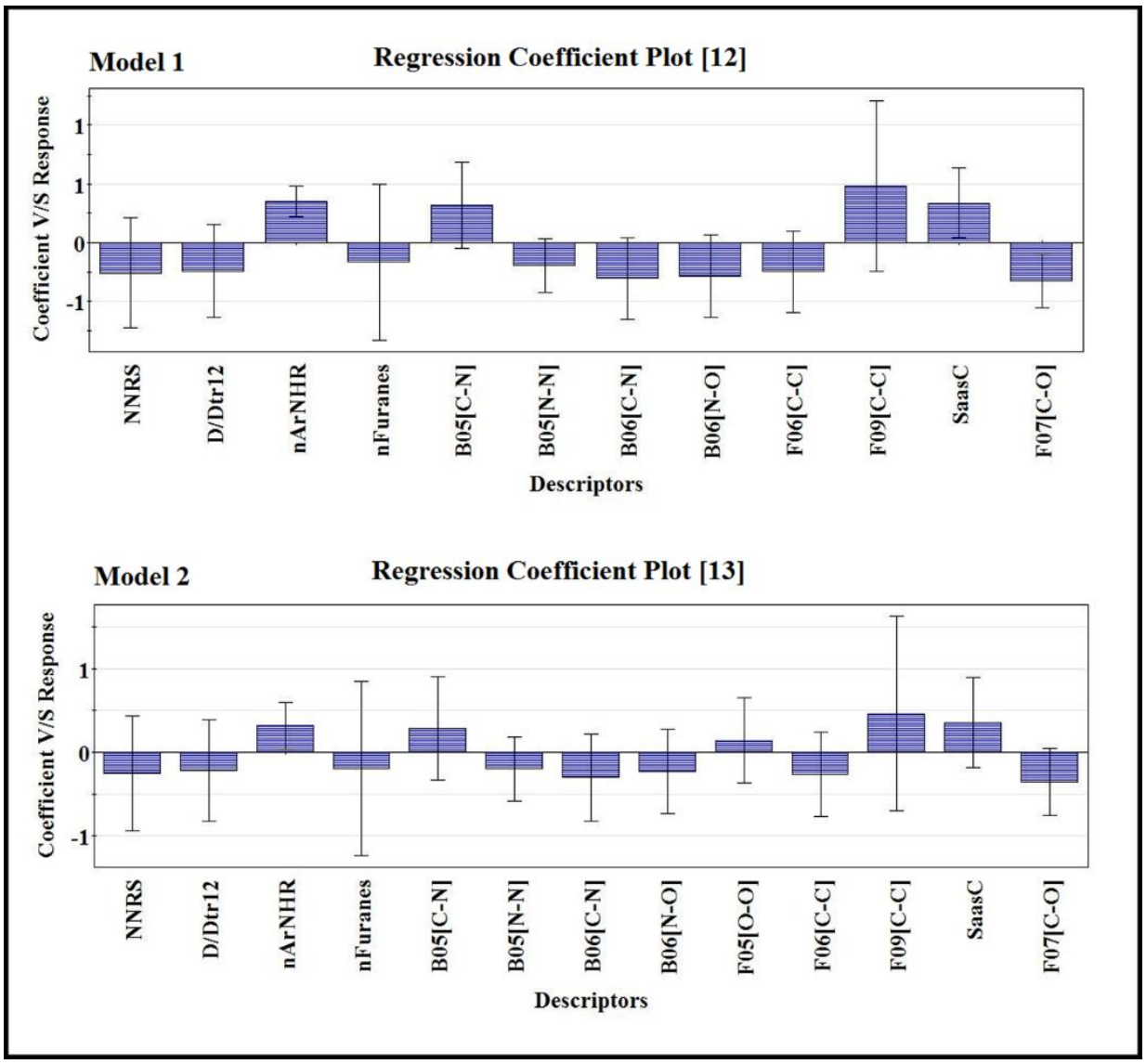


Figure 4.43. Regression coefficient plot of final PLS models against β-amyloid aggregation.

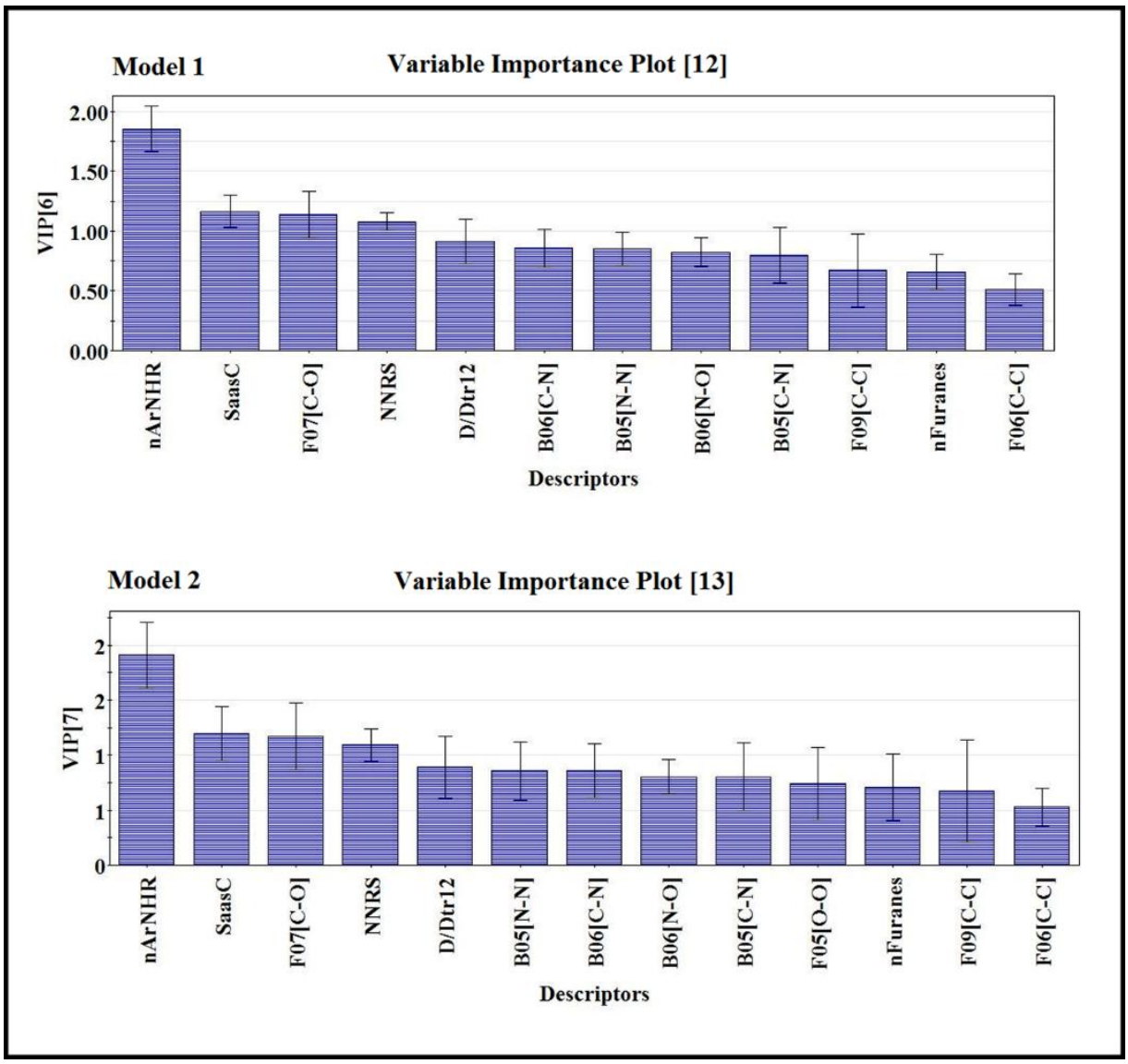
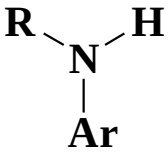


Figure 4.44. Variable importance plot (VIP) of final PLS models against β-amyloid aggregation.

Table 4.8. Definition and contribution of all the descriptors obtained from the PLS models.

S. No.	Name of Descriptors	Contribution	Discussion	Mechanism
1	NNRS	-ve	Normalized number of ring systems	Hydrophobic and π-π interactions
2	D/Dtr12	-ve	Distance/detour ring index of order 12 R	Hydrophobic and π-π interactions
3	nArNHR	+ve	Number of secondary amines (aromatic)	π-π and hydrogen bonding interactions
4	nFuranes	-ve	Number of Furanes	Hydrogen bonding interactions
5	B05[C-N]	+ve	Presence/absence of C - N at topological distance 5	Hydrogen bonding and electrostatics interactions
6	B05[N-N]	-ve	Presence/absence of N - N at topological distance 5	Hydrogen bonding interactions
7	B06[C-N]	-ve	Presence/absence of C - N at topological distance 6	Hydrogen bonding and electrostatics interactions
8	B06[N-O]	-ve	Presence/absence of N - O at topological distance 6	Hydrogen bonding and electrostatics interactions
9	F06[C-C]	-ve	Frequency of C - C at topological distance 6	Hydrophobic interaction
10	F09[C-C]	+ve	Frequency of C - C at topological distance 9	Hydrophobic interaction
11	SaasC	+ve	Sum of aasC E-states	Hydrophobic and π-π interactions
12	F07[C-O]	-ve	Frequency of C - O at topological distance 7	Hydrogen bonding interactions
13	F05[O-O]	+ve	Frequency of O - O at topological distance 5	Hydrogen bonding interactions

The highest contributing descriptor, nArNHR, a functional group count descriptor, denotes the number of secondary aromatic amines present in the compounds.



The lone pair of electrons on nitrogen is delocalized in the aromatic ring thus reducing the electron density of nitrogen[380]. The positive regression coefficient of this descriptor indicated that the activity of inhibitors is directly proportional to the numerical value of nArNHR. Thus, the compounds having a higher number of secondary aromatic amines may enhance the β-amyloid aggregation inhibitory activity as shown in **(Figure 4.45).** Compounds like **88** (pIC_{50}: -0.217), **67** (pIC_{50}: 0.537), and **231** (pIC_{50}: -0.190) and their corresponding descriptor values are 1 each. In contrast, compounds like **294** (pIC_{50}: -5.338), **293** (pIC_{50}: -5.190), and **87** (pIC_{50}: -4.949) have lower inhibitory activity, because these compounds have no such fragment **(Figure 4.45)**. From these observations, we have concluded that a secondary aromatic amine is important for β-amyloid aggregation inhibitory activity.

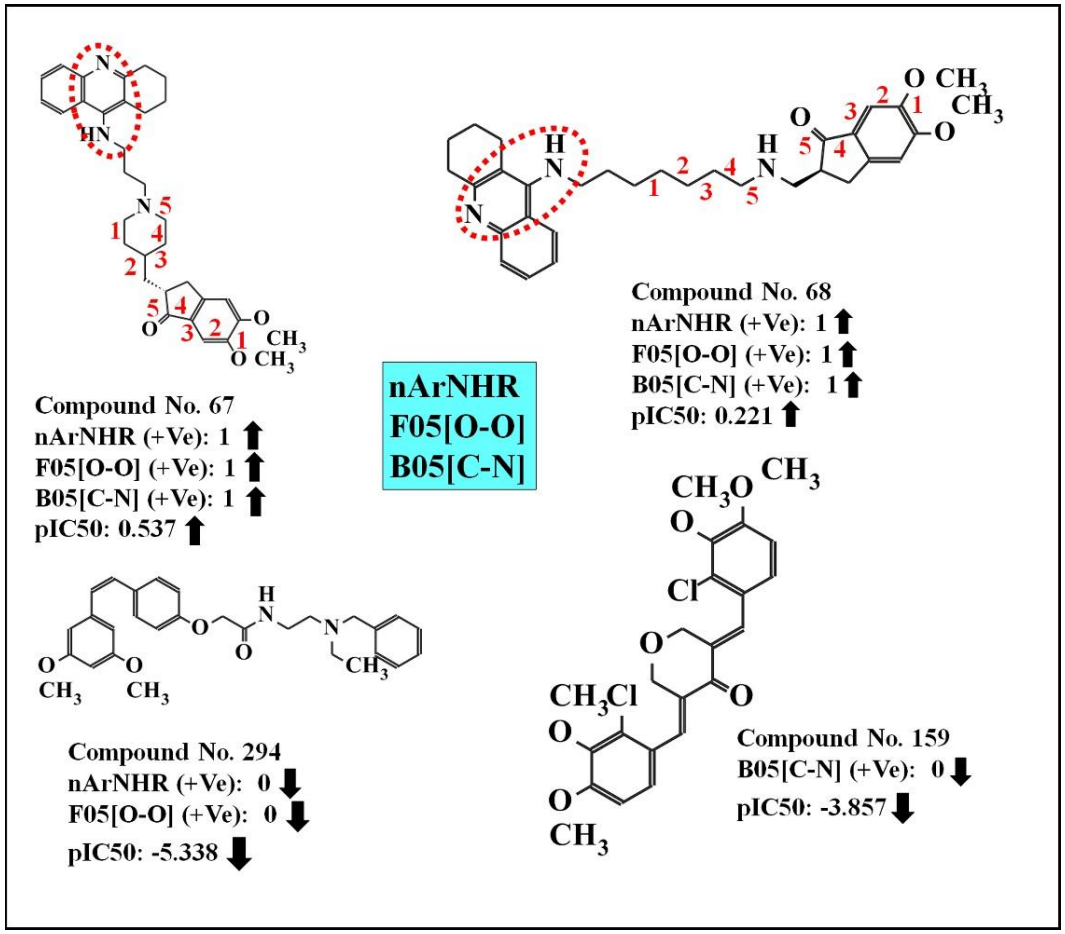


Figure 4.45. Contribution of nArNHR, B05[C-N] and F05[O-O] descriptors on β-amyloid aggregation inhibition.

The next significant descriptor, SaasC, an E-state index descriptor, denotes the sum of the atom-level E-state values for all non-substituted aromatic carbon atoms (Kier and Hall, 1999)[381]. According to Kier and Hall 1999[381], each atom or bond in molecules has its intrinsic state, which is altered by every other atom or bond in the same molecule, encoding information related to electronic distribution and topological aspects. SaasC is related to aromatic carbons with an attached substituent atom[382]. The positive regression coefficient of this descriptor indicates that the presence of aromatic carbons with an attached substituent atom is beneficial for the inhibitory activity as shown in compounds **(Figure 4.46) 71** (pIC_{50}: 0.0861), **70** (pIC_{50}: 0.086), and **74** (pIC_{50}: -0.357), and their corresponding descriptor values are 9.396, 8.807, and 8.779, respectively. On the other hand, the absence of such fragments is detrimental to β-amyloid aggregation inhibitory activity as observed in **(Figure 4.46)** compounds **93** (pIC_{50}: -3.978) and **341** (pIC_{50}: -3.469). The next significant descriptor, F07[C-O] is a 2D atom pair descriptor that accounts for the frequency of C and O atoms at topological distance 7. It contributes negatively toward the endpoint value, which suggests that the numerical values of the descriptor are inversely proportional to the inhibitory activity. Thus, the compounds bearing such fragments show lower values of inhibitory activity as evidenced by **(Figure 4.46)** compounds **153** (pIC_{50}: -3.176), **86** (pIC_{50}: -4.869), and **294** (pIC_{50}: -5.338) (their corresponding descriptor values are 14, 13, and 10, respectively). Whereas, compounds having no such fragments show higher β-amyloid aggregation inhibitory activity as shown in compounds **231** (pIC_{50}: -0.190), **106** (pIC_{50}: -0.301), and **201** (pIC_{50}: -0.301) **(Figure 4.46)**.

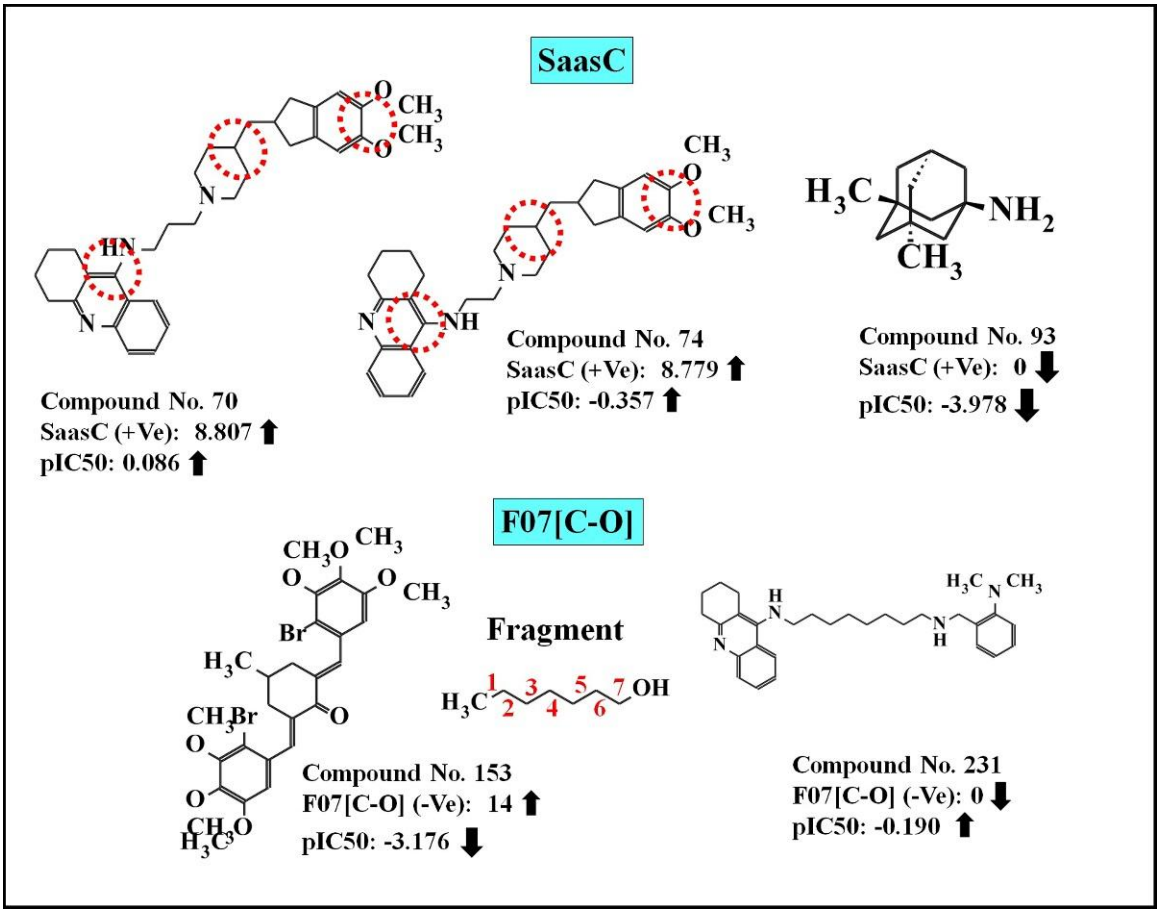


Figure 4.46. Contribution of SaasC and F07[C-O] descriptors on β-amyloid aggregation inhibition.

The ring descriptor, NNRS, indicates a normalized number of ring systems. This can be calculated by the following **equation 4.6**[305].

$$NNRS = NRS/Xmax \qquad \textbf{Equation 4.6}$$

Here, NNRS denotes a normalized number of ring systems, NRS represents the number of ring systems, and an X_{max} means the possibility of a maximum number of ring systems[305]. The negative regression coefficient of this descriptor suggests that NNRS negatively influences the inhibitory activity against β-amyloid plaque. The details of this descriptor were described by Das et al. 2016[305]. From this, it can be suggested that for the development of inhibitors against beta-amyloid aggregation, the normalized number of ring systems should be high, as shown in **(Figure 4.47)** compounds **162** (pIC_{50}: -4.243) (one cyclohexane and two benzene rings), **163** (pIC_{50}: -4.350) (one cyclohexane and two benzene rings), and **293** (pIC_{50}: -5.190) (three benzene rings) and vice versa in the case of compounds **88** (pIC_{50}: -0.217) (two fused rings), **106** (pIC_{50}: -0.301) (one fused ring), and **323** (pIC_{50}: -1.38) (one fused ring, one benzene ring, and one pyridine ring) **(Figure 4.47)**. The ring descriptor, D/Dtr12, denotes the distance/detour ring index of order 12 (size of the ring) in the compounds. The negative impact of this descriptor recommends that a higher numerical value of this descriptor leads to a lower inhibitory activity as evidenced by the compounds **(Figure 4.47) 143** (pIC_{50}: -3.760), **145** (pIC_{50}: -4.746), and **144** (pIC_{50}: -4.068) (their corresponding descriptors values are 98.958, 75.971, and 69.627, respectively). On the contrary, the compounds **(Figure 4.47)** having no such fragments show higher β-amyloid aggregation inhibitory activity as observed in the case of compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221) and **69** (pIC_{50}: 0.173). Another 2D atom pair descriptor, B05[N-N], indicates the presence of two nitrogen atoms at the topological distance of 5. The negative regression coefficient of this descriptor suggests that the presence of an N-N fragment at the topological distance 5 inversely affects the β-amyloid aggregation inhibitory activity. This is observed in compounds **(Figure 4.47) 50** (pIC_{50}: -4.100), **144** (pIC_{50}: -4.068), and **145** (pIC_{50}: -4.74) (all these compounds have a descriptor value of 1). The opposite is observed in the compounds **(Figure 4.47) 67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173).

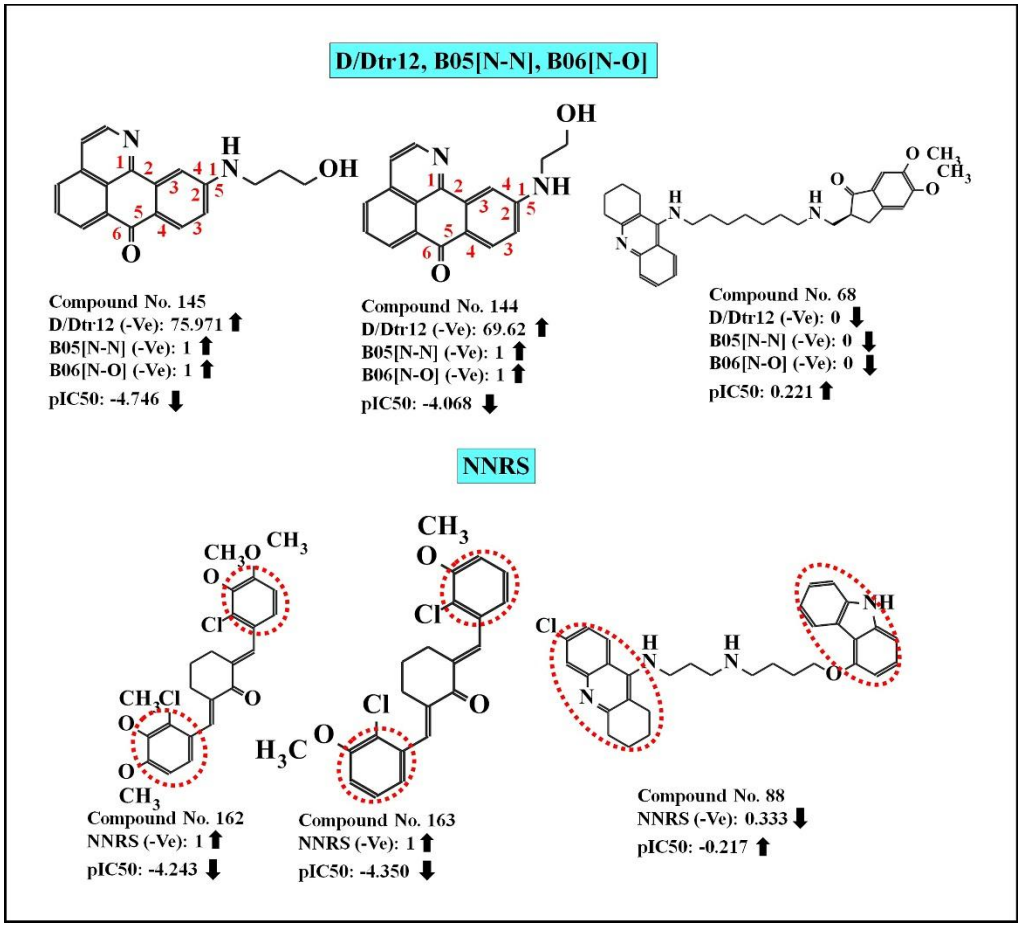


Figure 4.47. Contribution of D/Dtr12, B05[N-N], B06[N-O], and NNRS descriptors on β-amyloid aggregation inhibition.

Another 2D atom pairs descriptor, B06[C-N], indicates the presence/absence of C and N atoms at the topological distance 6. This descriptor contributes negatively towards the β-amyloid aggregation as indicated by the negative regression coefficient. For example, compounds **(Figure 4.48) 50** (pIC_{50}: -4.100), **144** (pIC_{50}: -4.068), and **145** (pIC_{50}: -4.746) (having a descriptor value of 1 each) have lower β-amyloid aggregation inhibitory activity due to the presence of such a fragment at the topological distance 5. On the other hand, the molecules without such a fragment show higher inhibitory activity as shown in compounds **(Figure 4.48) 106** (pIC_{50}: -0.301), **177** (pIC_{50}: -1.531), and **179** (pIC_{50}: -1.892).

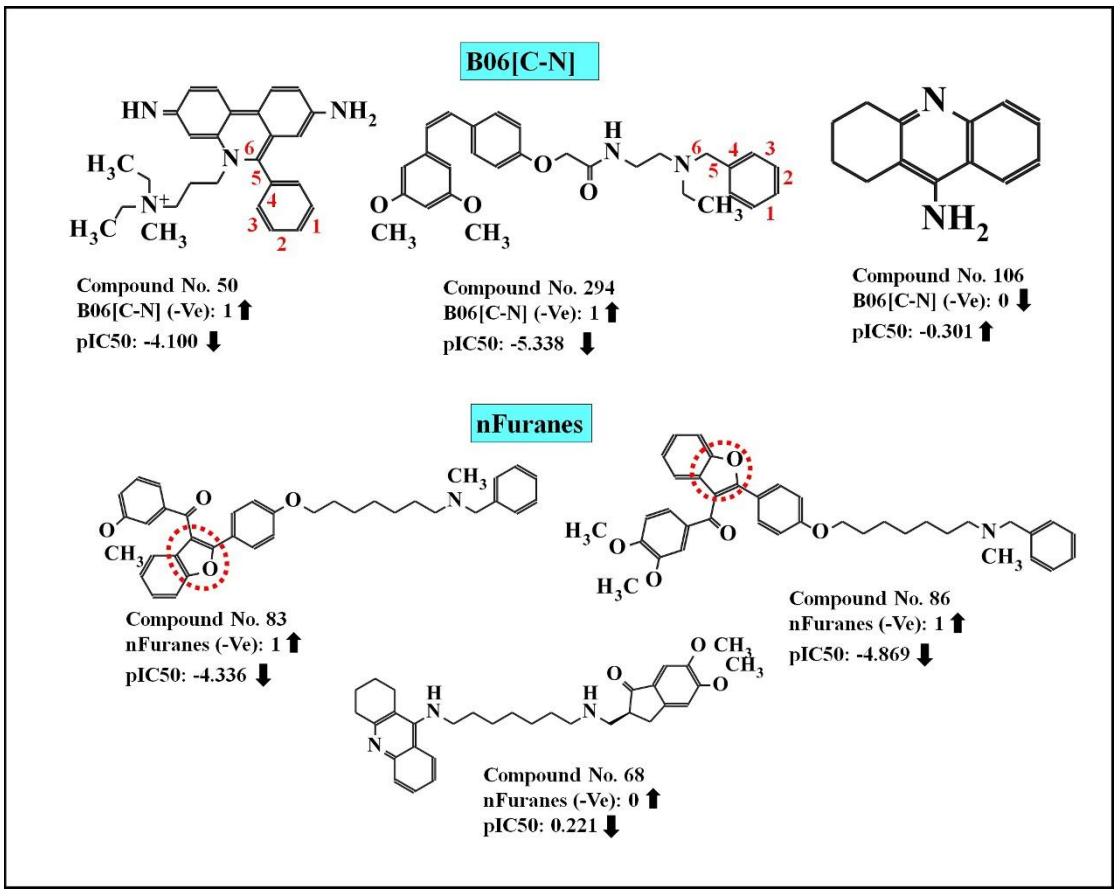


Figure 4.48. Contribution of B06[C-N] and nFuranes descriptors on β-amyloid aggregation inhibition.

The 2D atom pair descriptors, B06[N-O], describe the presence/absence of N-O at topological distance 6. The negative regression coefficient of this descriptor suggests that the absence of such a fragment in the molecules showed good β-amyloid aggregation inhibitory activity as shown in **(Figure 4.47)** compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173). While the presence of a higher number of this fragment shows the lower inhibitory activity as observed in **(Figure 4.47)** compounds **144** (pIC_{50}: -4.068), **145** (pIC_{50}: -4.746), and **293** (pIC_{50}: -5.190) (all of these compounds have a descriptor value of 1).

Another 2D atom pairs descriptor, B05[C-N], denotes the presence/absence of C-N at the topological distance 5. This descriptor positively influences the activity of β-amyloid aggregation inhibitors as suggested by its positive regression coefficient. Thus, the compounds containing a higher number of C-N fragments at topological distance 5 may have high β-amyloid aggregation inhibitory activity as evidenced by **(Figure 4.45)** compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173) (their corresponding descriptors values are 1). On the other hand, the molecules which do not contain such a feature may have lower inhibitory activity as shown in compounds **159** (pIC_{50}: -3.857), **162** (pIC_{50}: -4.243), and **163** (pIC_{50}: -4.350) **(Figure 4.45)**. Another 2D atom pair descriptor, F05[O-O], stands for the frequency of O-O at topological distance 5. For the β-amyloid aggregation inhibitors, this fragment indeed plays an important role in the binding process and may influence the inhibitory activity prominently. The positive impact of this descriptor on the β-amyloid aggregation inhibitors was indicated by its positive regression coefficient. Thus, the information obtained from this descriptor suggested that the molecules containing an O-O the

fragment at the topological distance 5 show higher β-amyloid aggregation inhibitory activity as shown in compounds **(Figure 4.45) 67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173) (containing descriptor value 1 respectively), while compounds **294** (pIC_{50}: -5.338), **293** (pIC_{50}: -5.190), and **87** (pIC_{50}: -4.949) show lower inhibitory activity due to the absence of this fragment **(Figure 4.45)**.

The functional group count descriptor, nFuranes, describes the number of furane rings present in the compounds. The negative regression coefficient of this descriptor suggests that the presence of this ring is inversely proportional to the β-amyloid aggregation inhibitory activity as observed in the case of compounds **(Figure 4.48) 83** (pIC_{50}: -4.33), **84** (pIC_{50}: -4.580), and **86** (pIC_{50}: -4.869) (all of these compounds have descriptor value 1), while the absence of such a ring system in the compounds **(Figure 4.48)** indicated higher inhibitory activity as observed in the compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173). Another 2D atom pair descriptor, F09[C-C], stands for the frequency of C-C at the topological distance 9. The positive regression coefficient of this descriptor suggests that an increase in the frequency of a C-C fragment at topological distance 9 may favor β-amyloid aggregation inhibitory activity. Thus, the molecules bearing such a fragment may enhance the β-amyloid aggregation inhibitory activity as shown in **(Figure 4.49)** compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173) (their corresponding descriptors values are 29, 23, and 27, respectively). The opposite is observed in the case of **(Figure 4.49)** compounds **144** (pIC_{50}: -4.068), **145** (pIC_{50}: -4.746), and **93** (pIC_{50}: -3.978) (containing descriptors values 4, 7, and 0, respectively). Another 2D atom pair descriptor, F06[C-C], indicates the frequency of C-C at topological distance 6. The negative regression coefficient of this descriptor suggests that the presence of higher numbers of this fragment is inversely proportional to the β-amyloid aggregation inhibitory activity as observed in the case of compounds **(Figure 4.49) 86** (pIC_{50}: -4.869), **87** (pIC_{50}: -4.949), and **83** (pIC_{50}: -4.336) (corresponding descriptor values are 49, 48, and 48, respectively). In contrast, a lower numerical value of this descriptor may favor the β-amyloid aggregation inhibitory activity as observed in the case of compound **(Figure 4.49) 106** (pIC_{50}: -0.301) (containing descriptor value 6).

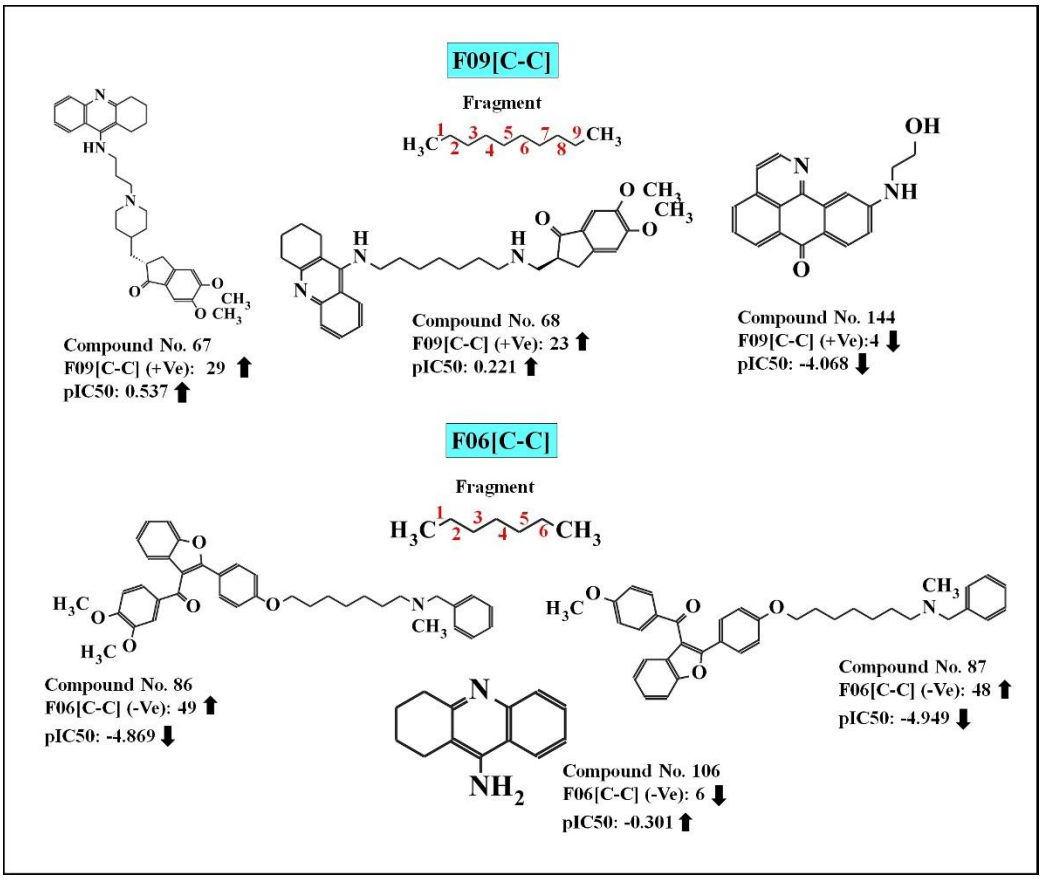


Figure 4.49. Contribution of F09[C-C] and F06[C-C] descriptors on β-amyloid aggregation inhibition.

4.3.1.2. Applicability domain (AD) of PLS models

In this work, the reported PLS models were checked for their applicability domain at a confidence level of 99% according to the DModX (distance to model in the X-space) approach using SIMCA-P 10.0 software (Available from https://umetrics.com/products/simca). In the case of model 1 (see S3 **Figure 4.50**), we found that 5 compounds (i.e., compounds number **26**, **81**, **85**, **262,** and **278**) in the test set are located outside the critical DModX value (D-Crit=1.813). In the case of model 2 (see S3 **Figure 4.50**), we found that 3 compounds (i.e., compounds number **262**, **274,** and **278**) in the test set are located outside the critical DModX value (D-Crit=1.814).

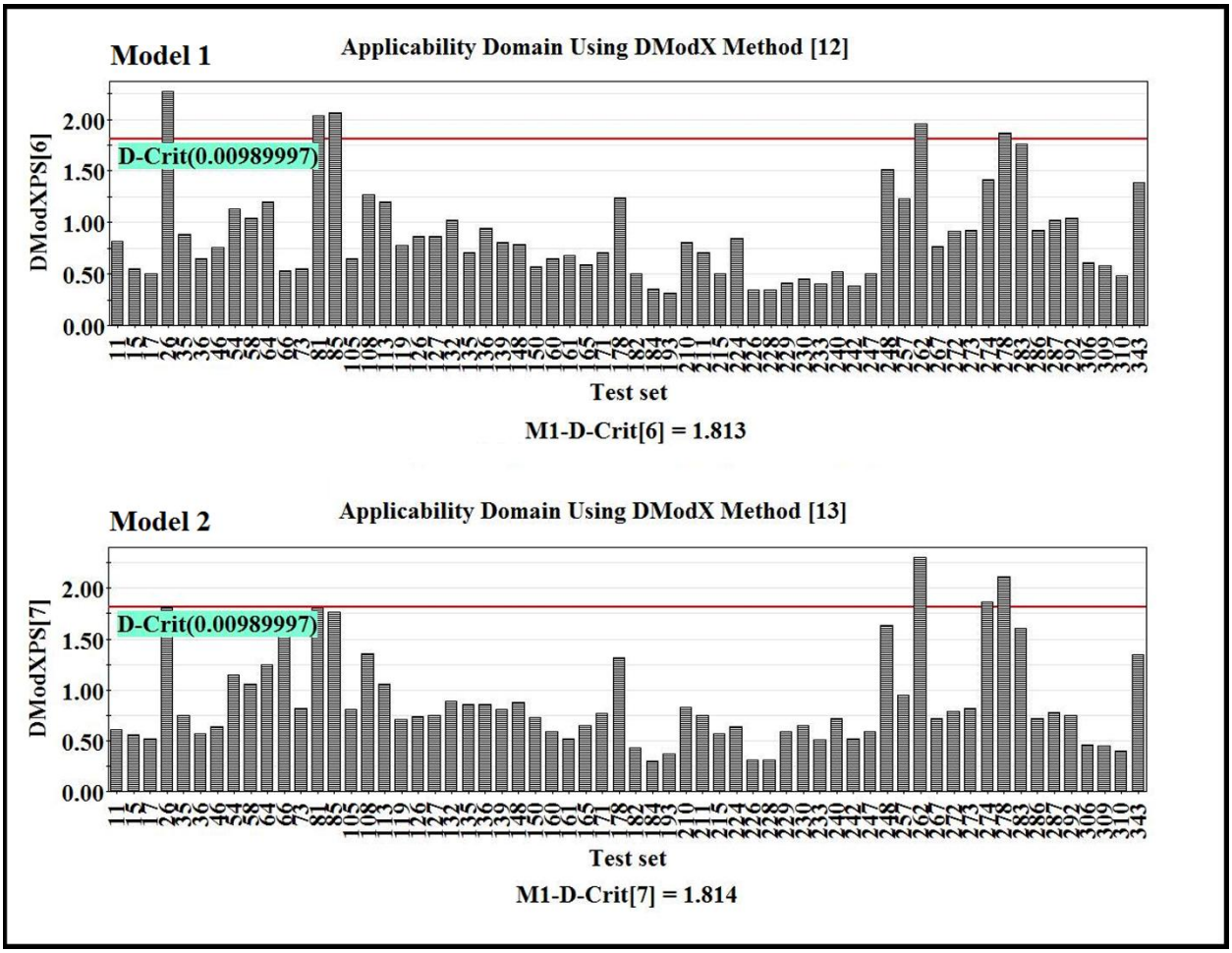


Figure 4.50. Applicability domain DModX values of the test set compounds at 99% confidence level of the developed PLS models against β amyloid aggregation.

4.3.1.3. Loading plot analysis

A loading plot of a PLS model (see **Figure 4.51**) provides information about the relationship between the X-variables and Y-variables. The amount of the loading for each descriptor to the latent variables can be seen from their corresponding loading plot using SIMCA-P 10.0 software (Available from https://umetrics.com/products/simca). In the loading plot, a descriptor that is close to zero is not well associated with the trends contained in the related scores[383]. As we have observed from the loading plot (see **Figure 4.51**), the X-variables nArNHR, SaasC, F09(C-C), B05(C-N), and F05(O-O) are significant for the Y-variable (pIC_{50}), because they are very close to the Y-variable. On the other hand, the variables NNRS, D/Dtr12, nFuranes, B05[N-N], B06[C-N], B06[N-O], F06[C-C] F07[C-O]), which are situated on the opposite side of the plot origin concerning the activity (Y-variable), contribute negatively.

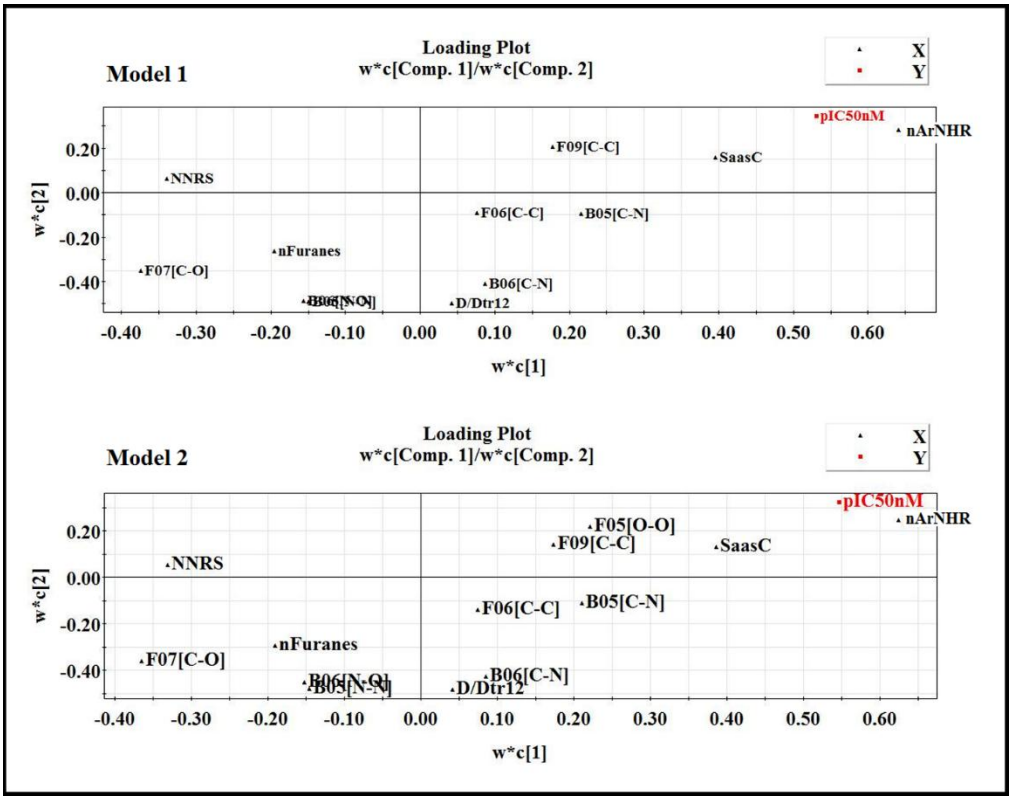


Figure 4.51. Loading plot for final PLS models against β amyloid aggregation.

4.3.2. 3D-Pharmacophore model

In the current work, ten different pharmacophore models were developed from a training set of 62 compounds. For the development of a pharmacophore model, we used Discovery Studio 2016 Client 4 (Available from https://www.3ds.com/products-services/biovia/). In terms of internal validation, the best pharmacophore model (**Hypo 1**) was found in the cost analysis with a higher correlation coefficient (r: 0.724), lower root mean square deviation (rmsd: 2.293), total cost (366.258), maximum fit (6.573), configuration cost (21.364), error (343.071), and weight (1.821). These values indicated that the developed models were inacceptable. The results of ten pharmacophore hypotheses against β-amyloid peptide are given in **Table 4.9**. Based on all reported metrics, Hypo-1 was found to be the best one among the ten hypotheses with one hydrogen bond acceptor (HBA), one hydrophobic (HYD), one hydrophobic aromatic, and one ring aromatic (RA) feature **(Figure 4.52)**. External validation of the model has been carried out by mapping the test set molecules (**Figure 4.52**) on Hypo-1 with the same settings as employed for the pharmacophore generation by the FAST method. After mapping, we found that 240 molecules from the data set of 252 compounds were mapped properly. Only 12 compounds failed to map due to the absence of the features found in the developed pharmacophore model. The results obtained from the qualitative analysis for the training and test sets using Hypo-1 are given in **Table 4.9**. The F-test confirms the non-randomness of the developed pharmacophore (Hypo-1) model. The total cost and correlation values obtained from the original and randomized models of the hypothesis for the F-test are given in the **Figure 4.53**.

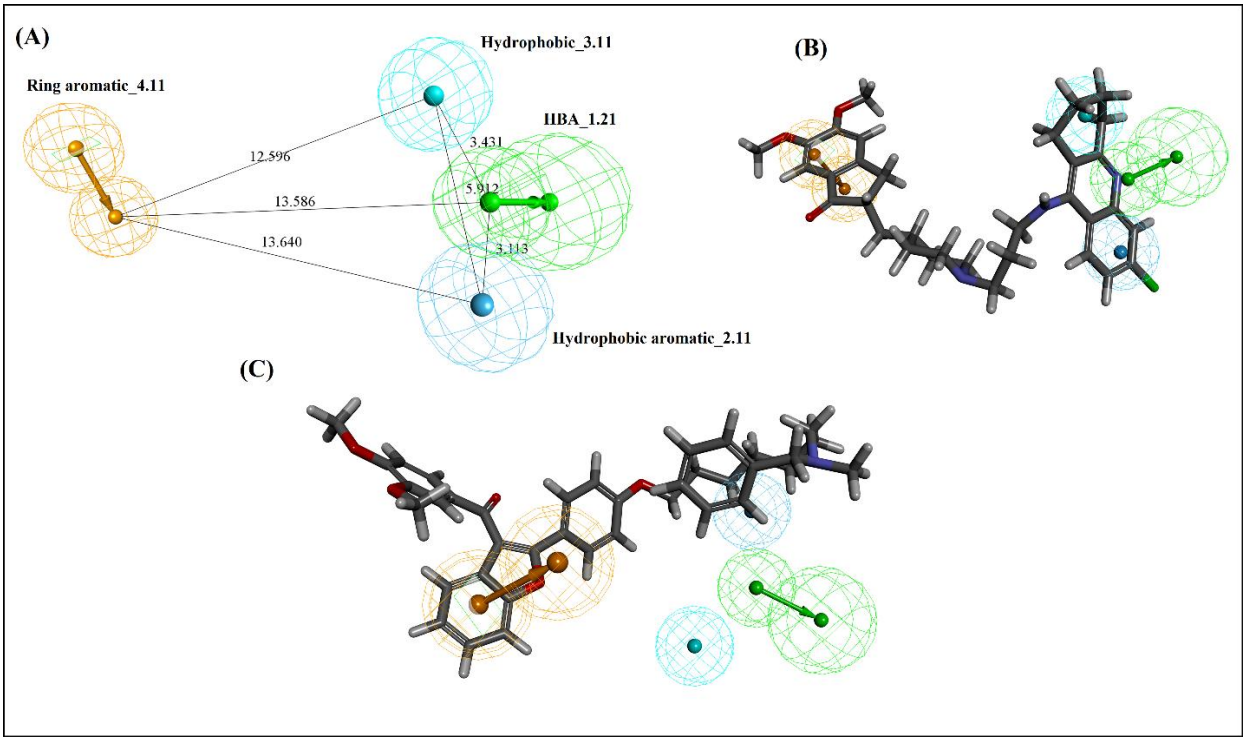


Figure 4.52. The best pharmacophore model (Hypo1) of β-amyloid aggregation inhibitors generated by the HypoGen module: (A) the best pharmacophore model Hypo1 represented with distance constraints (Å), (B) **Hypo-1** mapping with one of the most active compounds **66** of test set compounds and (C) **Hypo-1** mapping with one of the least active compounds **86** of test set compounds. Pharmacophoric features are colored as follows: hydrogen bond acceptor (green), hydrophobic (cyan), hydrophobic aromatic (light blue color), and ring aromatic (orange).

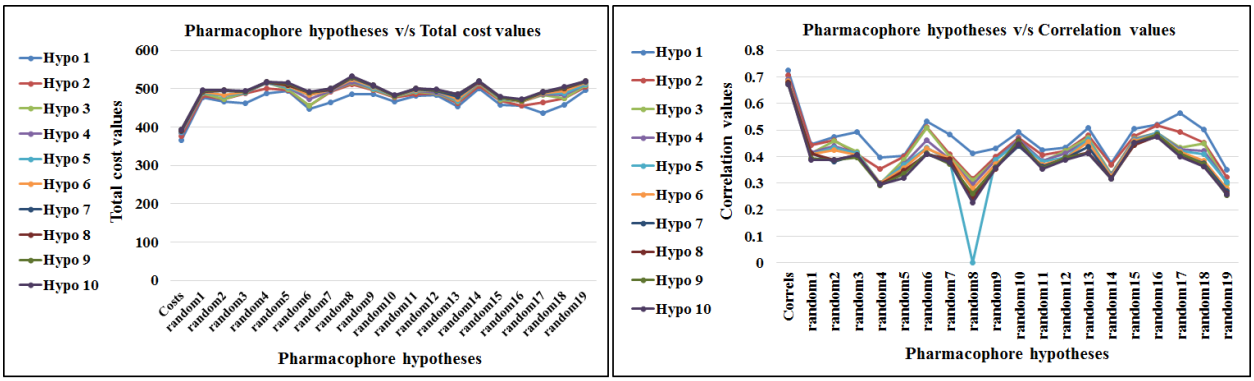


Figure 4.53. The total cost and correlation values were obtained from the original and randomized models of the hypothesis for the F-test.

Table 4.10. Different quantitative and qualitative validation parameters of the **Hypo-1** model were obtained for the training and test sets generated by the HypoGen algorithm.

Quantitative validation parameters						
Hypo.	**Total cost**	**ΔCost[a]**	**ΔCost[b]**	**RMS**	**Correlation**	**Features**
1	366.258	156.90	163.77	2.293	0.724	HBA, HYD_{Aro}, HYD, RA

Dataset	**No. of compounds**	**Qualitative validation parameters**					
		Sensitivity	**Specificity**	**Accuracy**	**Precision**	**F-measure**	**G-Means**
Train	62	88.34	88.23	88.32	95.78	91.91	88.29
Test	252	64.35	89.92	79.16	82.27	72.22	76.07

Cost differencea= Null cost - total cost, Cost differenceb= Total cost - fixed cost, Null cost = 523.154, Fixed Cost = 202.479, Best records in pass: 4, Config. Cost= 21.364, c= Best Hypothesis, Note – RA: Ring aromatic, HYD: Hydrophobic, HYD (aro): Hydrophobic Aliphatic/Aromatic, HBA: Hydrogen bond acceptor, *Compounds with IC_{50} <300 nM: more active (H) and IC_{50}>300 nM: less active (L).

4.3.2.1. Relation of the 3D-pharmacophore model with the 2D-QSAR model

All of the compounds in the dataset have at least one aromatic ring feature. The RA feature is a preliminary requirement for inhibitory activity against β-amyloid aggregation. The RA feature is in accordance with the nArNHR and SaasC descriptors of the 2D-QSAR models. Hydrophobic and hydrophobic aromatic features are in harmony with the F09[C-C] and SaasC descriptors of the 2D-QSAR models. The hydrogen bond acceptor feature is well corroborated with B05[C-N] and F05[O-O] descriptors of the 2D-QSAR models. The most active compound of the training set (**69**, IC_{50}: 0.67 nM) was mapped entirely on Hypo-1 with all of the four features (see **Figure 4.54.**). One benzene ring lies in the RA region, an amino group is in the hydrogen bond acceptor region, and a ring system lies in the hydrophobic region. The least active compound **24** (IC_{50}: 6323 nM) of the training set lacks RA features; thus it does not map completely (see **Figure 4.54.**). The most active compound of the test set (**66**, IC_{50}: 0.09 nM) mapped completely on Hypo-1 with all of the four features **(Figure 4.52)**. The least active compound (**86**, IC_{50}: 74100 nM) of the test set mapped partially with Hypo-1 **(Figure 4.52)**. From the above discussion, we have concluded that the absence of any feature among these four features in a molecule decreases the β-amyloid aggregation inhibitory potency.

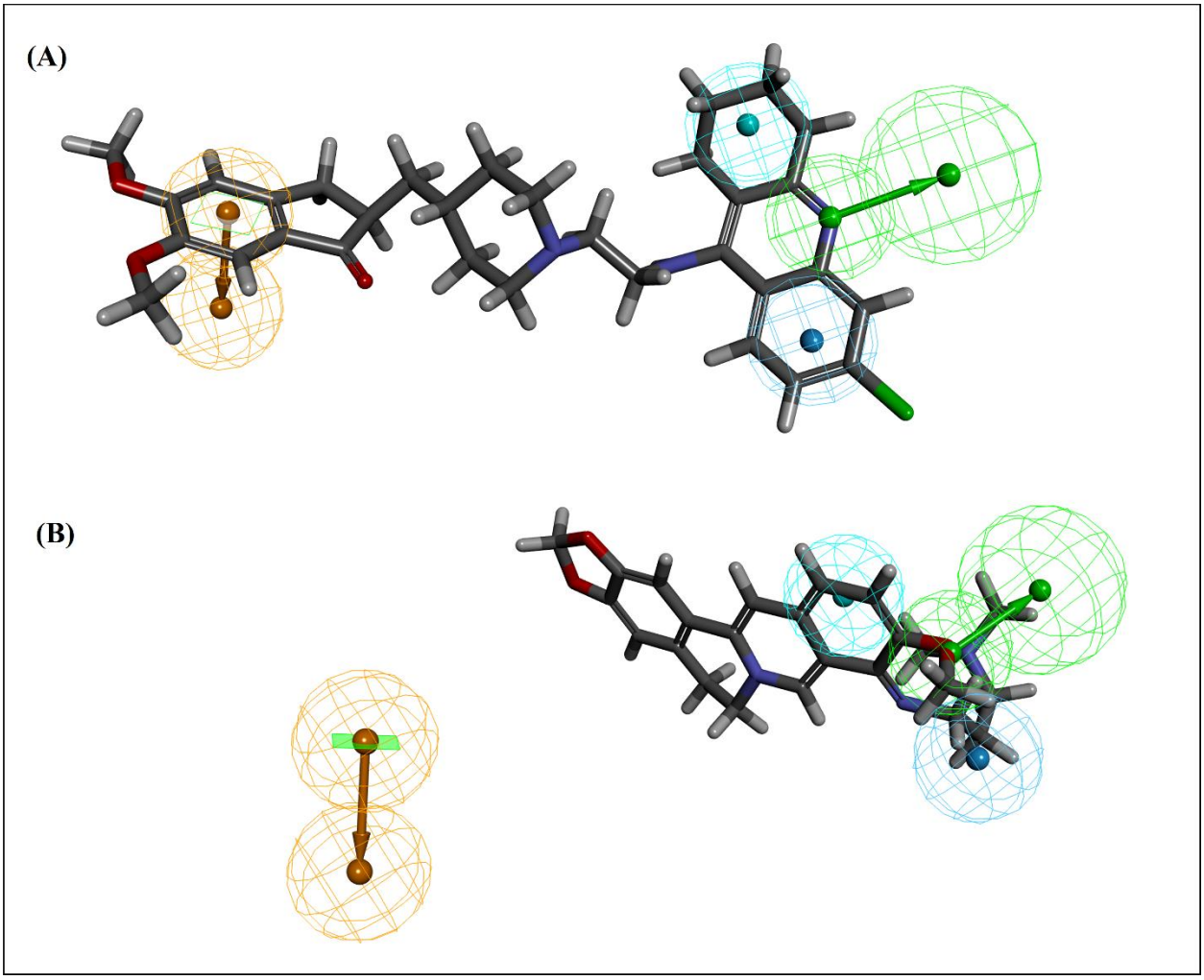


Figure 4.54. Pharmacophore mapping with training set compounds: (A) Hypo1 mapping with one of the most active compounds **69** of training set compounds and (B) Hypo1 mapping with one of the least active compounds **24** of training set compounds.

4.3.3. Molecular docking

In this study, molecular docking analysis was performed using two most active (**67** and **231**), two moderately active (**208** and **276**), and two least active (**87** and **145**) compounds from this dataset. The docking interactions suggest that the molecules interacted with a pocket containing HIS A: 13, (hydrophilic nature), LYS A: 16 (Charged), LEU A: 17, VAL A: 24, PHE A: 20, and ALA A: 21 (Hydrophobic nature) amino acid residues. The details of docking results are tabulated in **Table 4.11**.

4.3.3.1. Molecular Docking of the most active compounds from the dataset

The two most active compounds (**67** and **231)** from the dataset (pIC_{50} = 0.537 and -0.190, respectively) interacted with the active site amino acid residues through different interaction forces like hydrogen bonding interactions (carbon-hydrogen bonds and conventional hydrogen bonds), electrostatic interactions (π-cation), and hydrophobic interactions (π-alkyl bonds, alkyl and π-π-T-shaped). The amino acid residues

involved in interactions with these compounds are HIS A:13, LYS A:16, PHE A:20, LEU A:17, VAL A:24, and ALA A:21 (see **Figure 4.55**).

Figure 4.55 shows that compound **67** (one of the most active compounds in the dataset) interacts with HIS A: 13, LYS A: 16, and LEU A: 17 amino acid residues through a hydrogen bonding interaction, with PHE A: 20 through a π-cation, and VAL A: 24 and ALA A: 21 amino acid residues through alkyl and π-alkyl interactions, respectively. Another of the most active compounds, compound **231,** interacts with the amino acid residue through hydrogen bonding and π-π T shaped interaction **(**HIS A:13**)** and π-alkyl and alkyl (VAL A:24 and ALA A:21) interactions (**Figure 4.55)**.

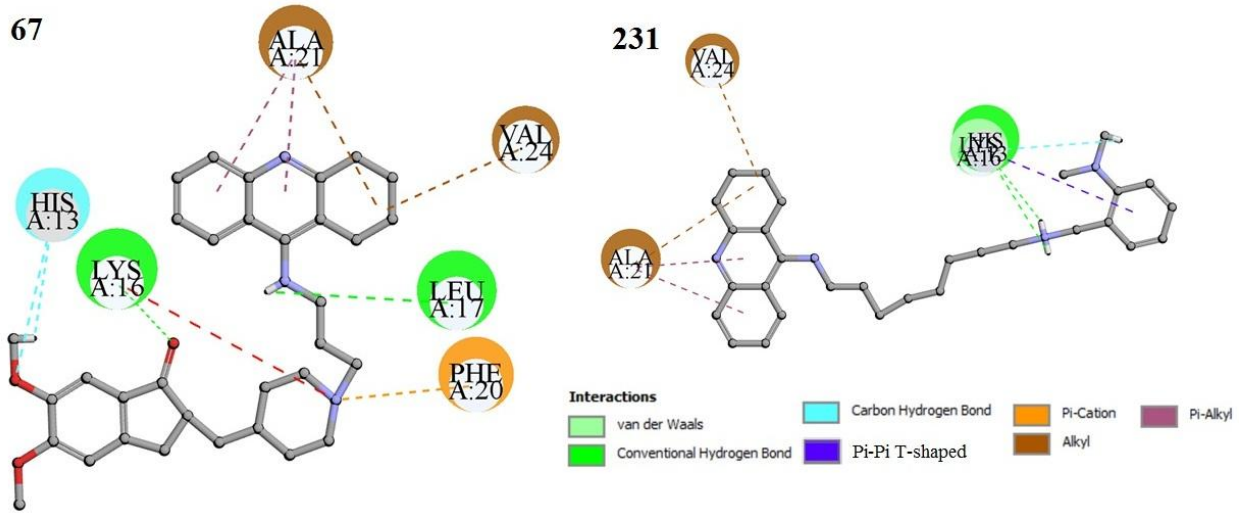


Figure 4.55. Docking interaction in most active compounds **67** and **231** from the dataset.

4.3.3.2. Molecular Docking of the moderately active compounds from the dataset

Two moderately active compounds (**208** and **276)** from the dataset **(**pIC_{50} = -2.908 and -2.989, respectively) interacted with the active site amino acid residues through various interaction forces like hydrogen bonds (carbon-hydrogen bonds and conventional hydrogen bonds) and hydrophobic interactions (π-π-T shaped, π-alkyl and π-sigma). The amino acid residues involved in interactions with these compounds are ALA A:21, LEU A:17, PHE A:20, HIS A:13, VAL A:24, and LYS A:16 **(see Figure 4.56**).

Figure 4.56 shows that compound **208** interacts with LEU A: 17 and LYS A:16 (through hydrogen bonding), PHE A:20 (π-π-T-shaped and pi-sigma), and ALA A:21 (π-alkyl bonds) amino acid residues. Another moderately active compound (compound **276)** from the dataset interacted with the amino acid residues through hydrogen bonding **(**LEU A: 17 and LYS A:16**)** and hydrophobic interactions such as π-π T-shaped and π-alkyl bonding (HIS A:13 ALA A:21 and VAL A:24) (See **Figure 4.56)**.

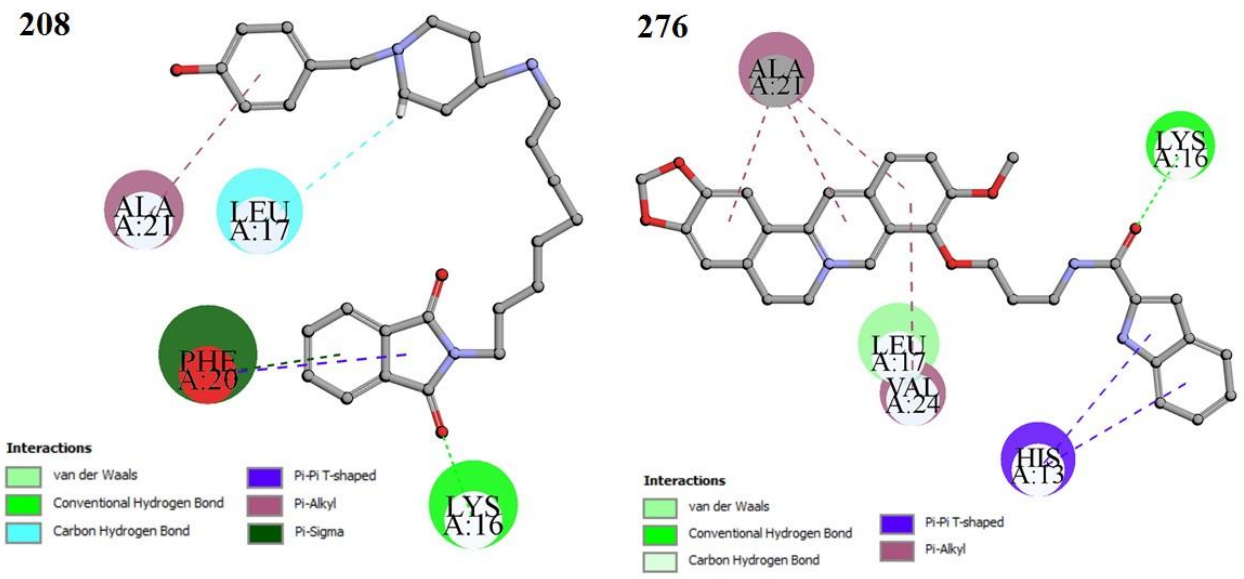


Figure 4.56. Docking interaction in moderately active compounds **276** and **208** from the dataset.

4.3.3.3. Molecular Docking of the least active molecules from the dataset

The two least active compounds from the dataset **(**pIC_{50} = -4.949 and -4.746, respectively) namely **87** and **145** interacted with the active site amino acid residues through different interaction forces like hydrogen bonding (carbon-hydrogen bonds and conventional hydrogen bonds), electrostatic (π-cation) and hydrophobic interactions (π-π-stacking and π-alkyl). The amino acid residues involved in interactions with these compounds are ALA A:21, PHE A:24, LYS A:16, and VAL A:24 **(see Figure 4.57**).

Figure 4.57 shows that compound **87** (one of the least active compounds from the dataset) interacted with ALA A:21 and LYS A:16 amino acid residues through hydrogen bonding interaction, LYS A:16 amino acid residue through π-cation and PHE A:20 and ALA A:21 through π-π -T-shaped and π-alkyl interactions. Another least active compound from the dataset, compound **145,** interacts with the amino acid residues through hydrogen bonding **(**LYS A: 16**)** and hydrophobic interactions such as π-alkyl bonding (ALA A: 21, VAL A: 24**) (See Figure 4.57)**.

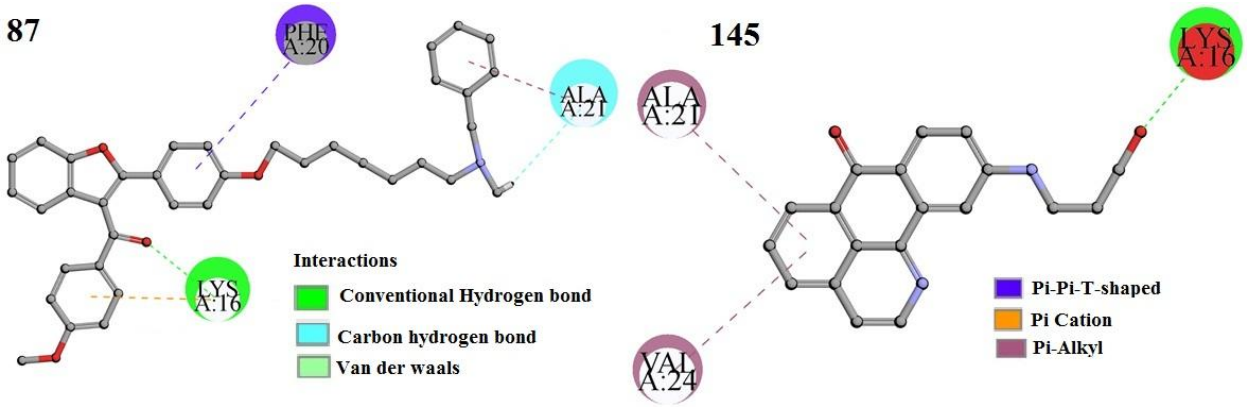


Figure 4.57. Docking interaction in the least active compound **87** and **145** from the dataset.

Finally, molecular docking analysis has demonstrated that the most active compounds from the dataset (**67** and **231)** (**Figure 4.55)** interacted with the maximum number of active amino acid residues with the higher number of interacting forces (non-covalent forces) in comparison with the least active compounds from the data set like **87** and **145** (**Figure 4.57)**. In pharmacophore mapping, the most active compounds (**66** and **67**) from the data set were correctly mapped with all features that appeared in the model, whereas the least active compounds (**86** and **87**) were partially mapped with the model. The least active compound from the data set failed to map in the absence of the features appearing in the developed models, which are most important for inhibitory activity against β-amyloid plaque.

Table 4.11. Docking results and correlation with the 2D-QSAR model in this study.

S. No.	Compound Number	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with the QSAR model
1	**67** (high pIC_{50})	24.939	HIS A:13, LYS A:16, PHE A:20, LEU A:17, VAL A:24 and ALA A:21	Vdw, Hydrogen bonding, π-alkyl, alkyl, unfavorable positive-positive and π-cation	nArNHR, B05[C-N], SaasC and F05[O-O]
2	**231** (high pIC_{50})	23.463	VAL A:24, ALA A:21 and HIS A:13	Vdw, Hydrogen bonding, π-π T-shaped, π-alkyl and alkyl	nArNHR, B05[C-N] and SaasC
3	**208** (Moderate pIC_{50})	25.815	ALA A:21, LEU A:17, PHE A:20, VAL A:24 and LYS A:16	Vdw, hydrogen bonding, π-sigma,π-π-T-shaped and π-alkyl	nArNHR, B05[C-N], SaasC and F05[O-O]
4	**276** (Moderate pIC_{50})	26.173	ALA A:21, LEU A:17, HIS A:13 and LYS A:16	Vdw, hydrogen bonding,π- π-T-shaped and π-alkyl	B05[C-N] and SaasC
5	**145** (low pIC_{50})	22.398	ALA A:21, LYS A:16 and VAL A:24	Vdw, hydrogen bonding and π-alkyl	nArNHR and B05[C-N]
6	**87** (low pIC_{50})	29.676	LYS A:16, PHE A:20 and ALA A:21	Vdw, hydrogen bonding,π- π-T-shaped, π-cation and π-alkyl	B05[C-N] and SaasC

4.3.3.4. Relation with the QSAR models

From the docking studies, it was observed that the formation of a hydrogen bond and alkyl bond between the ligand and receptor plays a vital role in interactions. Hydrogen bonding may correlate with descriptors nArNHR (number of secondary aromatic amines present in the compounds), F05[O-O] (frequency of O-O at topological distance 5), and B05[C-N] (presence/absence of C-N at topological distance 5 in the compounds) in the 2D-QSAR models. Descriptors SaasC (aromatic carbons with an attached substituent atom) and nArNHR are well related with interactions formed via π- interactions (π-alkyl and π-π T-shaped) between the protein and ligands and define the importance of this descriptor as we have observed in compound nos. **67, 231, 208,** and **276** (**Figures 4.55 and 4.56)**. But in contrast, in the case of compounds **87** and **145** (least active), the descriptor NNRS (normalized number of ring systems) contributes negatively to the response and is found to be related to π- π-T-shaped, and π-alkyl bonding interactions with those fragments in the docking experiments (**Figure 4.57)**. Thus, from the above-said information, we can conclude that hydrogen bonding, hydrophobicity, and alkyl (π interaction) features obtained from both the 2D-QSAR model and docking results are essential for inhibitory activity against β-amyloid aggregation.

4.3.4. Comparisons of the performance of the reported models with previously published models

In this investigation, a comparative study was performed of the best models of this study with previously published models (Leal et.al. 2015[384], Zhao et al. 2013[385], Aswathy et al. 2018[386], Hossein et al. 2019[387], Xiangji 2006[388], Yang et al. 2010[389], Najmeh et al. 2014[390] and Sehan et al. 2015[391]) for the prediction of the bioactivity against β-amyloid plaques, as depicted in the **Table 4.12**. The details of different internal and external validation parameters obtained from our models and those obtained from previous models are given in **Table 4.12**. Based on the statistical quality in terms of both internal and external validation criteria, the models reported in this work are statistically significant and robust enough as compared to the previously reported models (**Table 4.12**). Moreover, the models presented in this study are derived from a larger set of molecules than those reported in the previous studies.

Table 4.12. Comparisons of the proposed study with previously published studies against β-amyloid aggregation.

Sources	E. L.	LV	Model	Training set			Test set	
				n	R^2	Q^2	n	R^2_{pred}
Model-1 in this study	**12**	**6**	**PLS**	**252**	**0.664**	**0.621**	**62**	**0.763**
Model-2 in this study	**13**	**7**	**PLS**	**252**	**0.684**	**0.638**	**62**	**0.769**
Leal et.al. 2015[384]	5	-	HQSAR	36	0.937	0.757	10	0.659
Zhao et.al 2013[385]	-	5	CoMFA	32	0.877	0.431	7	0.834
Zhao et.al 2013[385]	-	8	CoMSIA	32	0.836	0.447	7	0.617
Zhao et.al 2013[385]	-	5	CoMFA	34	0.828	0.522	5	0.915
Zhao et.al 2013[385]	-	6	CoMSIA	34	0.800	0.493	5	0.902
Aswathy et.al. 2018[386]	4	-	HQSAR	24	0.931	0.615	6	0.956
Aswathy et.al. 2018[386]	-	5	CoMFA	24	0.787	0.687	6	0.731
Aswathy et.al. 2018[386]	-	3	CoMSIA	24	0.972	0.743	6	0.713
Aswathy et.al. 2018[386]	6	-	MLR	24	0.908	0.747	6	0.807
Hossein et al 2019[387]	4		MLR	28	0.912	0.915	12	0.836
Xiangji 2006[388]	4	-	PLS	22	0.857	-	-	-
Yang et al. 2010[389]	-	6	CoMSIA	21	0.911	0.512	-	-
Najmeh et al. 2014[390]	5	-	PCA	25	0.631	-	-	-
Sehan et al. 2015[391]	-	-	3D-QSAR	63	0.93	0.89	26	0.89

Abbreviations: LV= Latent variables, E. L. = Equation length, MLR= Multiple linear regression, CoMFA=Comparative Molecular Field Analysis, CoMSIA= Comparative molecular similarity index analysis, PCA = Principal component analysis and HQSAR= Hologram QSAR.

4.4. Study 4- Chemometric modeling of structurally diverse carbamates for the inhibition of acetylcholinesterase enzyme (AChE) in Alzheimer's disease

4.4.1. QSAR analysis

The statistically significant 2D-QSAR and GQSAR models derived using the PLS regression-based technique along with the values of their validation parameters are shown below in **equations 4.7** and **4.8.** The obtained results suggested that the models were acceptable in terms of fitness, stability, and classical predictivity measures. The reported 2D-QSAR model was developed by using 6 descriptors with corresponding latent variables of 3, while the GQSAR model was developed by using 7 descriptors with corresponding latent variables of 6. The descriptors appearing in the models define the structural and functional requirements which can improve the inhibitory activity of molecules against the AChE enzyme. The proximity of the observed and predicted values for the AChE enzyme inhibitors in the data set can be further established from the scatter plots as shown in **Figure 4.58**. The quantitative contributions of similar/dissimilar descriptors (similar descriptors are placed in close proximity) and the interrelationships between the X-variables and the Y-response are depicted in the loading plots in **Figure 4.59**. Additionally, we have also performed a Y-Randomization test to check whether the models were obtained by any chance or not. The results obtained from the randomized models in the case of 2D-QSAR (Model 1: R^2rand = -0.0166 and Q^2rand = -0.379) and GQSAR (Model 2: R^2rand = 0.0545 and Q^2rand = -0.606) suggested that the developed models were not obtained by any chance correlation as given in **Figure 4.60**. The list of Carbamate derivatives present in the dataset with their name, SMILES notation of respective compounds and observed and predicted activities (2D-QSAR and GQSAR) against the AChE enzyme are given in **Table 4.13**.

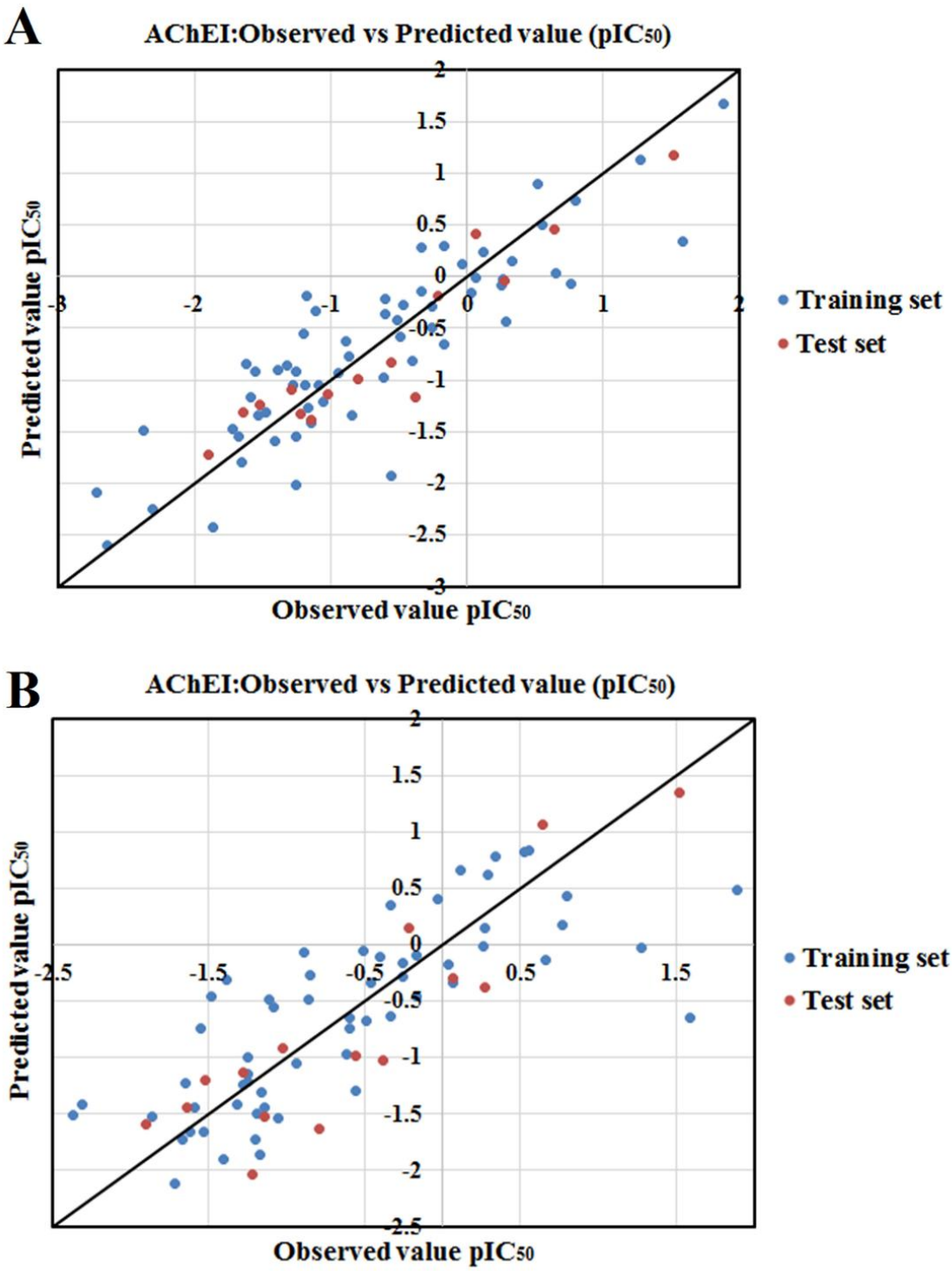


Figure 4.58. Scatter plots of observed and predicted values of final PLS (2D QSAR and GQSAR) models against AChE enzyme (**A**: 2D QSAR, **B**: GQSAR).

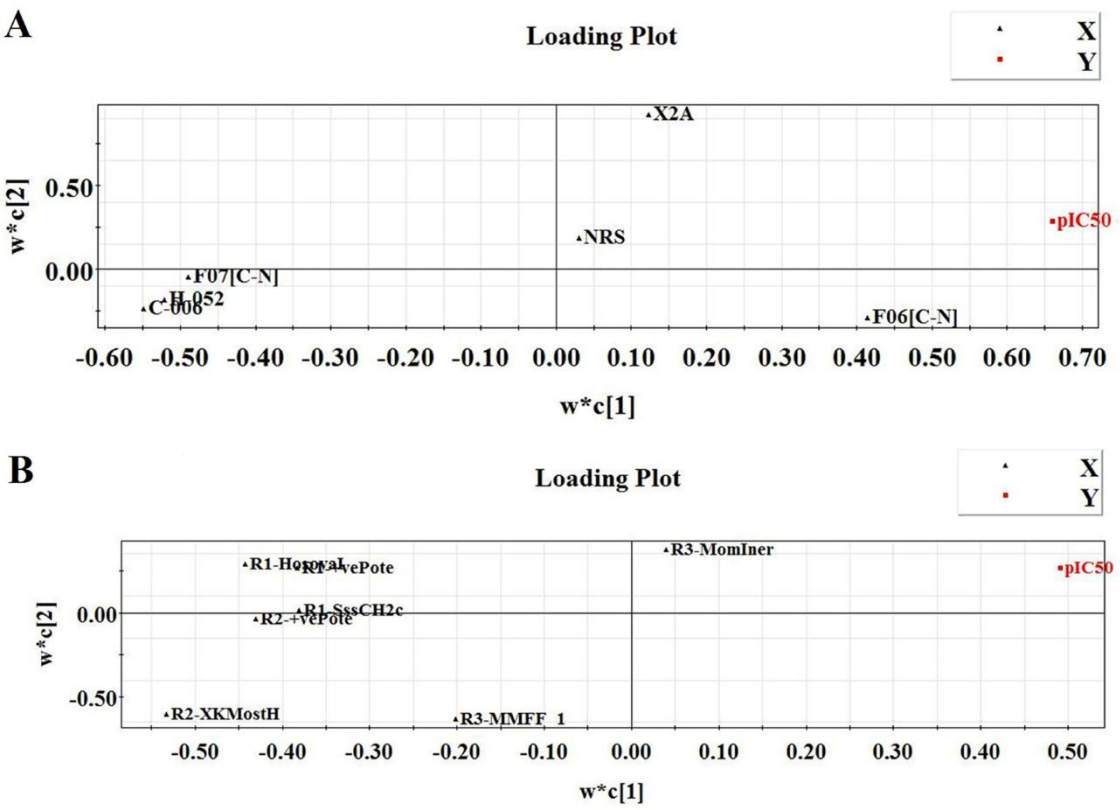


Figure 4.59. Loading plot for final PLS (2D QSAR and GQSAR) models against AChE enzyme (**A**: 2D QSAR, **B**: GQSAR).

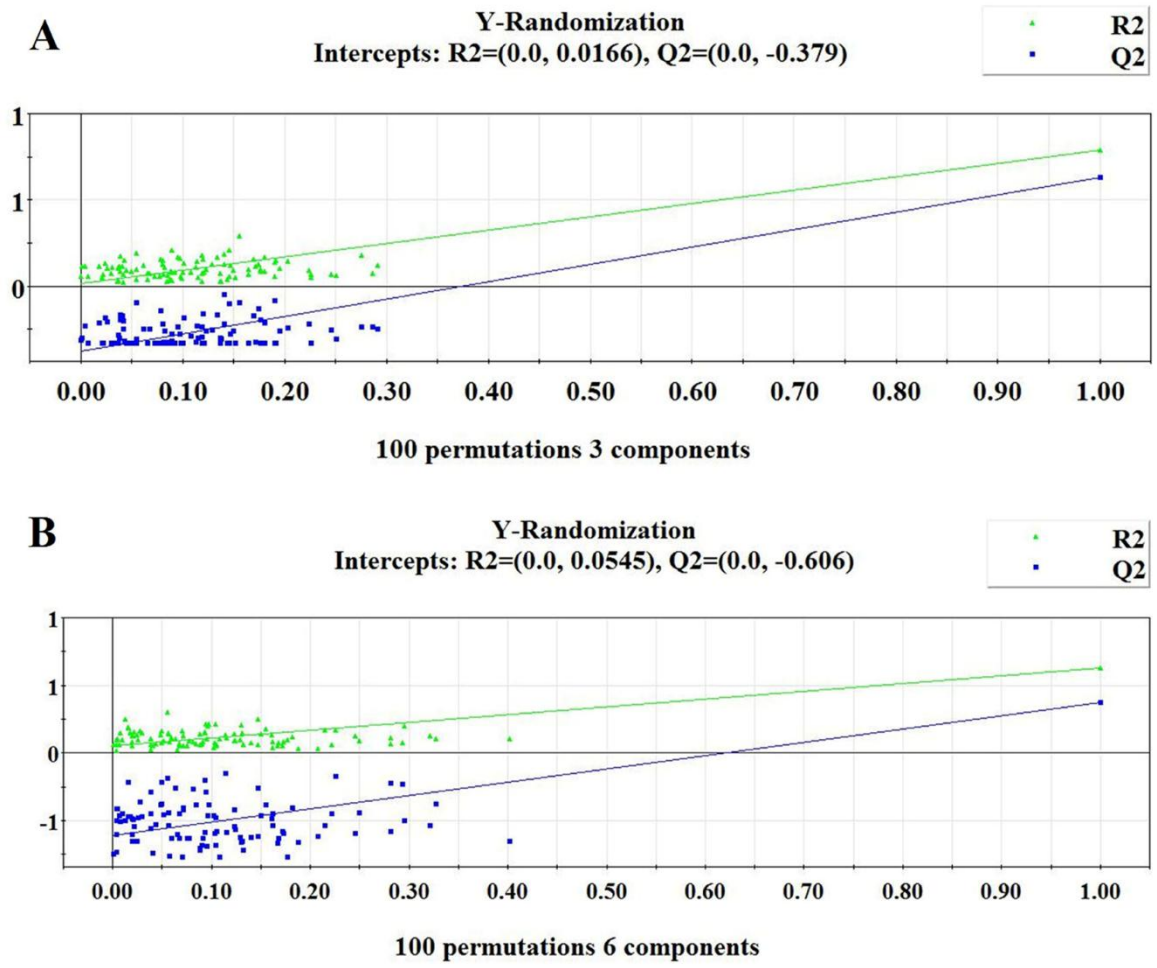


Figure 4.60. Model Randomization plots for final PLS (2D QSAR and GQSAR) models against AChE enzyme (**A**: 2D QSAR, **B**: GQSAR).

4.4.1.1. Mechanistic interpretation of modeled descriptors obtained from the 2D-QSAR model

$pIC_{50}(\mu M) = -9.011 - 0.541 \times C - 006 - 0.466 \times F07[C - N] - 0.190 \times H - 052 + 28.395 \times X2A + 0.554 \times F06[C - N] + 0.481 \times NRS$ **Equation 4.7**

$N_{training}$= 63, R^2= 0.789, Q^2_{LOO}= 0.732, Avgrm^2 $_{(LOO)}$ = 0.634, Δrm^2 = 0.149, LV= 3, EL = 6, Prediction quality= $MODERATE$; N_{test} = 15, Q^2F_1 = 0.883, Q^2F_2 = 0.883, Avg rm^2 = 0.852, Δrm^2 = 0.062, CCC = 0.935, MAE = 0.261, SD = 0.174, Prediction quality = $GOOD$.

The descriptors in the PLS models are arranged accordingly to their importance and then described separately. The significance level and contribution of the model descriptors towards the AChE inhibitory activity are determined based on the variable importance plot (VIP) and regression coefficient plot as shown in **Figure 4.61**[368].

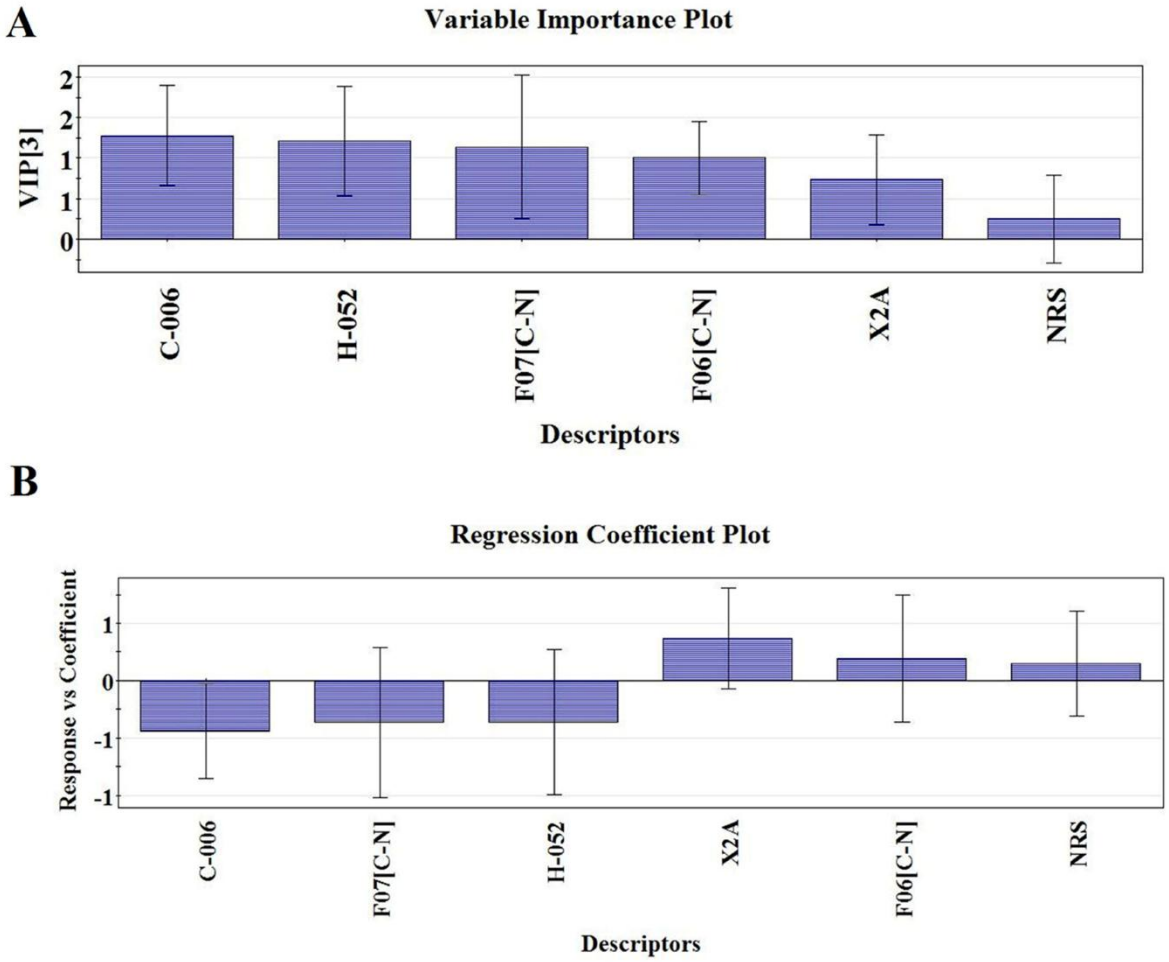


Figure 4.61. Variable importance plot (VIP) and regression coefficient plot of final PLS (2D QSAR) models against AChE enzyme (A: VIP plot and B: Regression coefficient plot).

The descriptor C-006 belongs to the class of atom-centered fragments that encodes information about the topological environment of an atom. This descriptor indicates the number of CH2RX functional groups (X: heteroatom (O, N, S, P, Se or halogens), R: any group linked through carbon) that describes each atom by its atom type and the bond types and atom types of its first neighbors[392]. The neighbors of a carbon atom in this case can be hydrogen (represented as H), carbon (represented as R), or hetero atoms (represented as X). On the other hand, 2D atom pair descriptor, F07[C-N], is simply characterized by the frequency of C-N at topological distance 7. The negative regression coefficients of these descriptors suggest that the absence of such fragments in the compounds may increase the inhibitory activity against the AChE enzyme as observed in (See **Figure 4.62)** case of compounds **18** (pIC_{50}: 2.045) and **55** (pIC_{50}: 1.886), whereas the presence of such fragments correlates to lower inhibitory activity as observed in (See **Figure 4.62)** compounds **52** (pIC_{50}: -2.309) and **73** (pIC_{50}: -2.309).

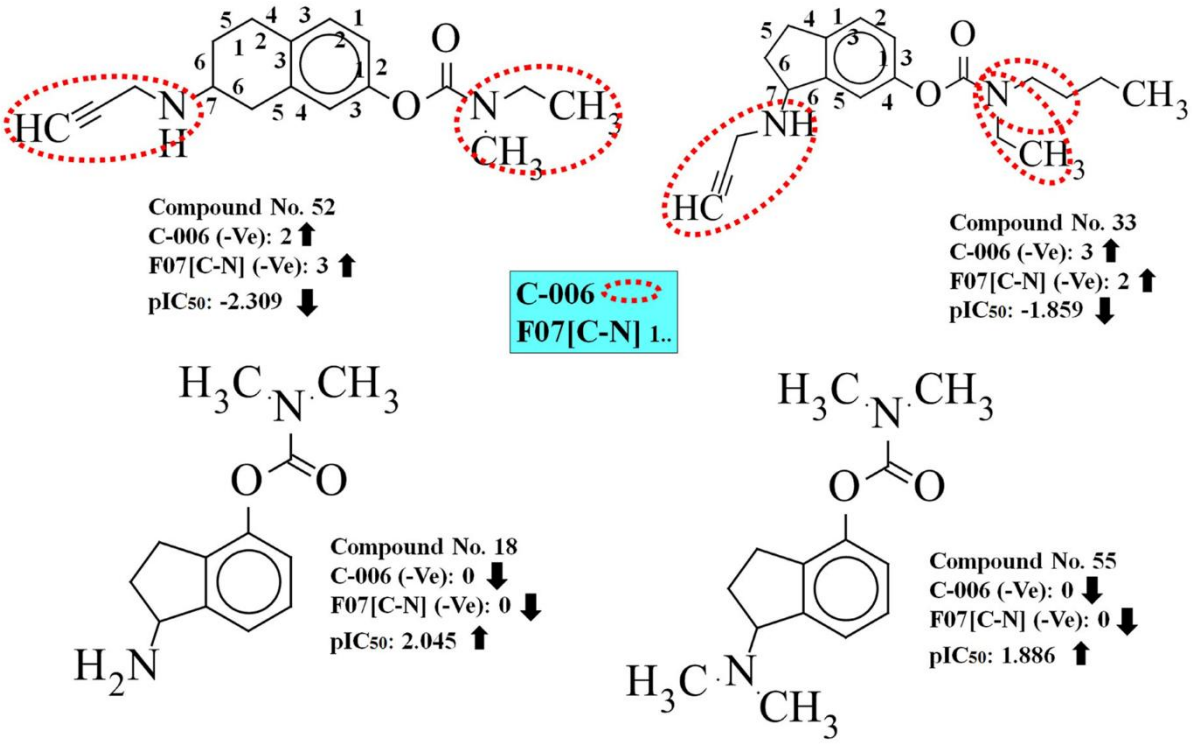


Figure 4.62. Contributions of C-006 and F07[C-N] descriptors on AChE enzyme inhibitory activity.

Another atom-centered fragments descriptor, H-052, describes H (hydrogen) attached to C(sp3) with 1X (heteroatom) attached to the next C[69]. It is a simple molecular descriptor defined as the number of specific atom types in a molecule, and it is calculated by knowing the molecular composition and atom connectivity. The negative sign of H-052 indicates that compounds with the higher number of H atoms attached to C0(sp3) with 1X (X=O) attached to next C would show lower inhibitory activity as observed in (See **Figure 4.63)** compounds **73** (pIC_{50}: -2.369) and **52** (pIC_{50}: -2.309) while lower numerical values of this fragment correlate the higher inhibitory activity as shown in compounds **18** (pIC_{50}: 2.045) and **55** (pIC_{50}: 1.886) (See **Figure 4.63**). Another 2D atom pair descriptor, F06[C-N], indicates the frequency of C-N at the topological distance 6. The positive regression coefficient of this descriptor indicates that the presence of the C-N fragment at the topological distance 6 may favor the inhibitory activity against the AChE enzyme as found in (See **Figure 4.63**) compounds **18** (pIC_{50}: 2.045) and **19** (pIC_{50}: 1.858) (containing descriptor value of 4 for all the cases). On the other hand, compounds with lower numerical values of this descriptor show lower inhibitory activity as observed in (See **Figure 4.63)** compounds **40** (pIC_{50}: -2.720) and **43** (pIC_{50}: -2.642).

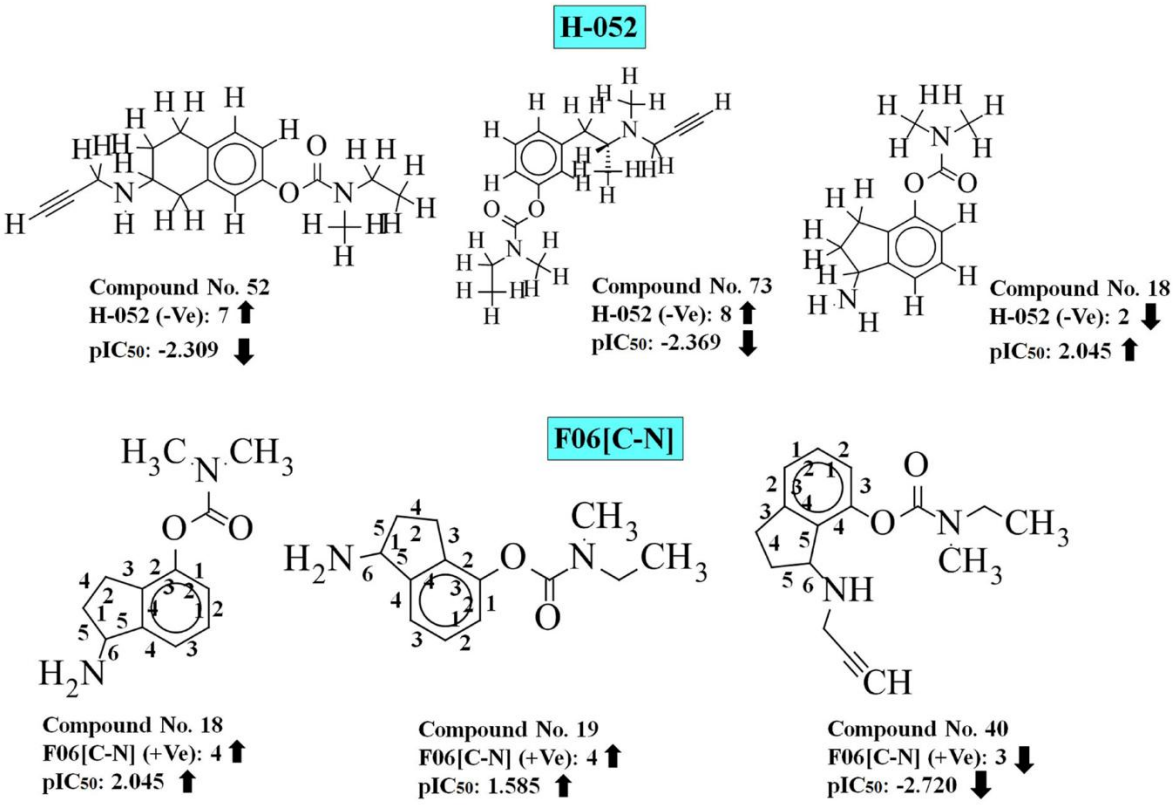


Figure 4.63. Contributions of H-052 and F06[C-N] descriptors on AChE enzyme inhibitory activity.

The connectivity descriptor, X2A, corresponds to the average connectivity index $^{2}\chi$ and represents the steric feature of the molecule. The positive coefficient of X2A indicates that an increase in the values of the descriptor will increase inhibitory activity. This is observed in (See **Figure 4.64)** case of compound **61** (pIC_{50}: 0.795)**, 77** (pIC_{50}: 0.769) (descriptor value 0.333 and 0.32 respectively), and the opposite is seen in compound **40** (pIC_{50}: -2.720) and **43** (pIC_{50}: -2.642) as depicted in **Figure 4.64**. The functional group count descriptor, NRS, indicates the number of ring systems present in the compounds, which contributes positively towards the AChE enzyme inhibitory activity. Hydrophobicity plays an important role in better AChE inhibitory activity as we have observed in compounds (See **Figure 4.64)** such as **8** (pIC_{50}: 0.522) and **35** (pIC_{50}: -0.250) containing descriptor values 4 and 3 respectively showing higher inhibitory activity, while compounds **40** (pIC_{50}: -2.720), 43 (pIC_{50}: -2.642) and **73** (pIC_{50}: -2.369) (containing lower descriptor values 1 in all three cases) (See **Figure 4.64)** show lower inhibitory activity.

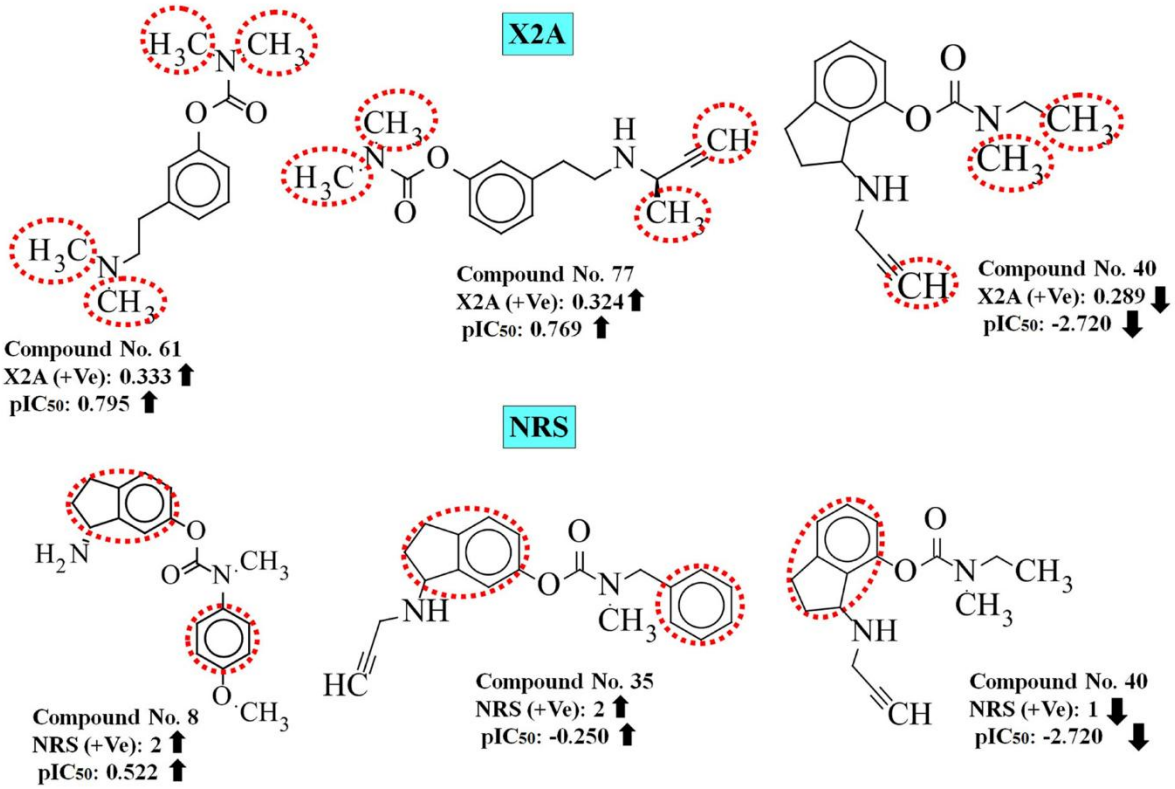


Figure 4.64. Contributions of X2A and NRS descriptor on AChE enzyme inhibitory activity.

4.4.1.2. Mechanistic interpretation of modeled descriptors obtained from GQSAR analysis

$$pIC_{50}(\mu M)=1.681-0.650\times R2-XKMostHydrophobicHydrophilic+0.071\times R2-+vePotentialSurfaceArea-0.0001\times R1-HosoyaIndex-0.483\times R1-SssCH2count-0.015\times R1-+vePotentialSurfaceArea-0.276\times R3-MMFF\_1+0.0007\times R3-MomInertiaX \quad \textbf{Equation 4.8}$$

N_{train} = 63, R^2 = 0.625, Q^2 = 0.538, EL = 7, LV = 6; Ntest = 15, Q^2F_1 = 0.735, Q^2F_2 = 0.734, Avg rm^2 = 0.681, Δrm^2 = 0.151, SD = 0.231, Prediction quality = Moderate.

Equation 4.8 corresponds to the best G-QSAR model that comprises 7 descriptors. The descriptors appearing in the model are arranged accordingly to their importance and then described separately. The VIP and regression coefficient plot[368] defines the importance of each variable obtained from the final PLS models that are responsible to regulate the AChE enzyme inhibitory activity as shown in **Figure 4.65** in the regression coefficient plot.

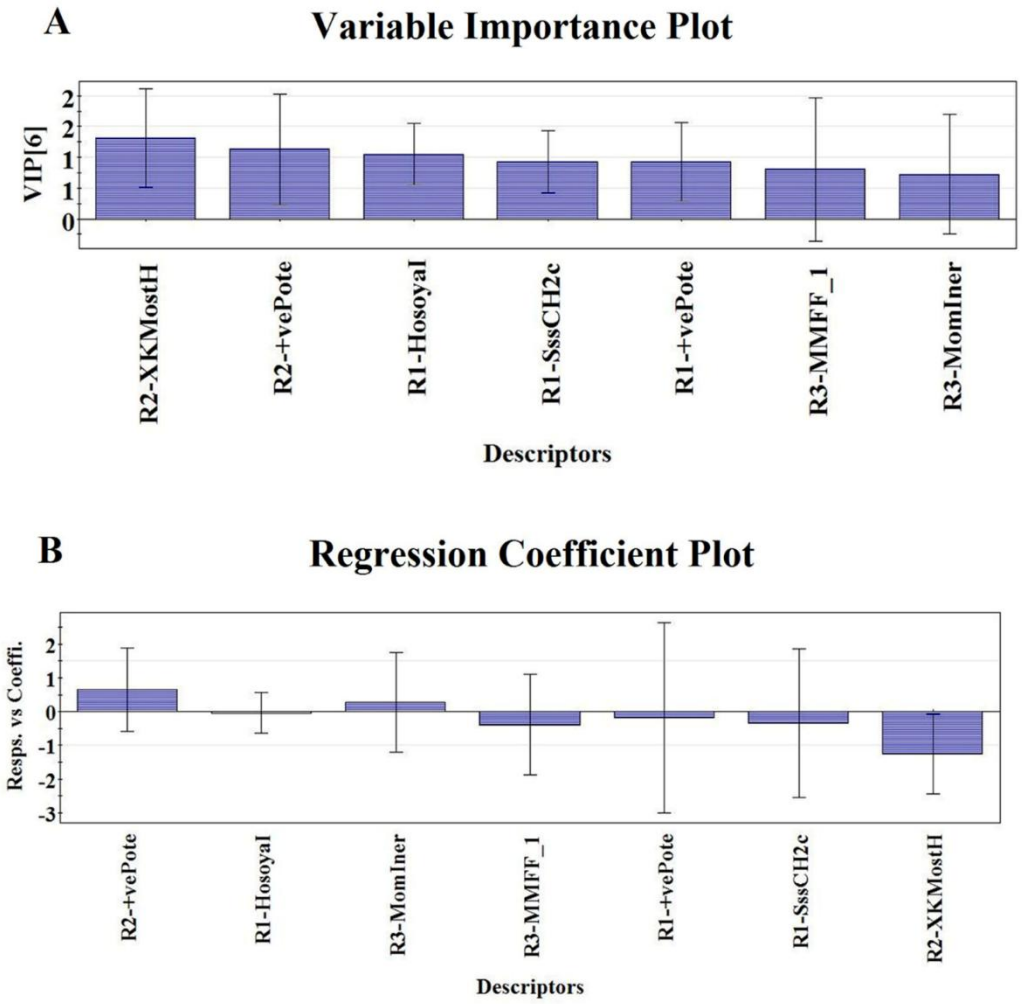


Figure 4.65. Variable importance plot (VIP) and Regression coefficient plot of final PLS (GQSAR) models against AChE enzyme (A: VIP and B: Regression coefficient plot).

The physicochemical descriptor, R2-XKMostHydrophobicHydrophilicDistance, belonging to the subclass Hydrophobicity XlogpK, implies the distance between the most hydrophobic and hydrophilic point on the VdW surface computed using the Kellog XlogP method[392]. For each fragment, the descriptor value is calculated by generating the van der Waals surface of the fragment (R2 position), putting a probe atom at each point on the van der Waals surface, and calculating the distance between most hydrophobic and hydrophilic points. It shows a negative contribution to the inhibitory activity which suggests that relatively less distance between the most hydrophobic and hydrophilic group at the R2 position may favor the inhibitory activity against the AChE enzyme. For instance, most active compounds (See **Figure 4.66**) **76** (pIC_{50}: -0.938) and **42** (pIC_{50}: -0.845) in the dataset contain methyl group at R2 position with minimum distance from nitrogen group in contrast the least active compounds (See **Figure 4.66**) **32** (pIC_{50}: -1.053)

and **33** (pIC_{50}: -1.86) in the dataset containing methyl group at R2 position with maximum distance from nitrogen group.

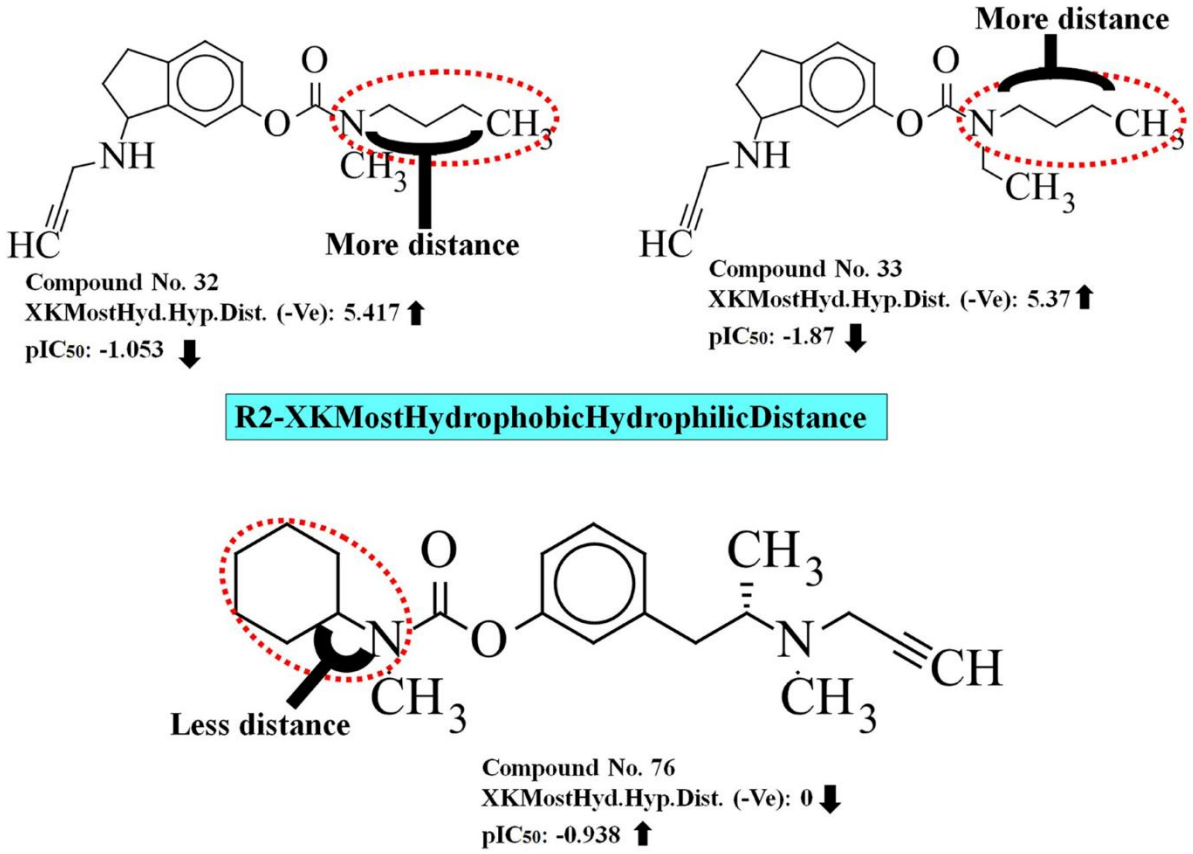


Figure 4.66. Contribution of XKMostHydrophobicHydrophilicDistance descriptor at R2 position on AChE enzyme inhibitory activity.

Another physicochemical descriptor, +vePotentialSurfaceArea, belongs to the subclass Electrostatic descriptors, which signifies the total van der Waals surface area with the positive electrostatic potential of the compounds having electron accepting or positive centers at substitution sites. The most contributing descriptor in position R2 with a positive coefficient value is +vePotentialSurfaceArea which suggests that an increase in the positive electrostatic potential of fragment R2 may lead to an increase in the inhibitory activity against AChE enzyme. We have observed that the most active compounds (See **Figure 4.67**) **75** (pIC_{50}: -0.486) and **58** (pIC_{50}: -0.892) in the dataset showed higher inhibitory activity whereas the least active compounds (See **Figure 4.67**) **52** (pIC_{50}: -2.31) in the dataset with descriptor values 21.624 show lower inhibitory activity. In contrast, +vePotentialSurfaceArea is detrimental for the R1 position in the compounds as we have observed from the least active compounds (See **Figure 4.67) 52** (pIC_{50}: -2.31) and **73** (pIC_{50}: -2.369) in the data set containing higher number electron accepting or positive centers at R1 substitutions site, whereas the most active compounds (See **Figure 4.67**) **18** (pIC_{50}: 2.046) and **19** (pIC_{50}: 1.585) from dataset contain a lower number of the electron-accepting group at the R1 position.

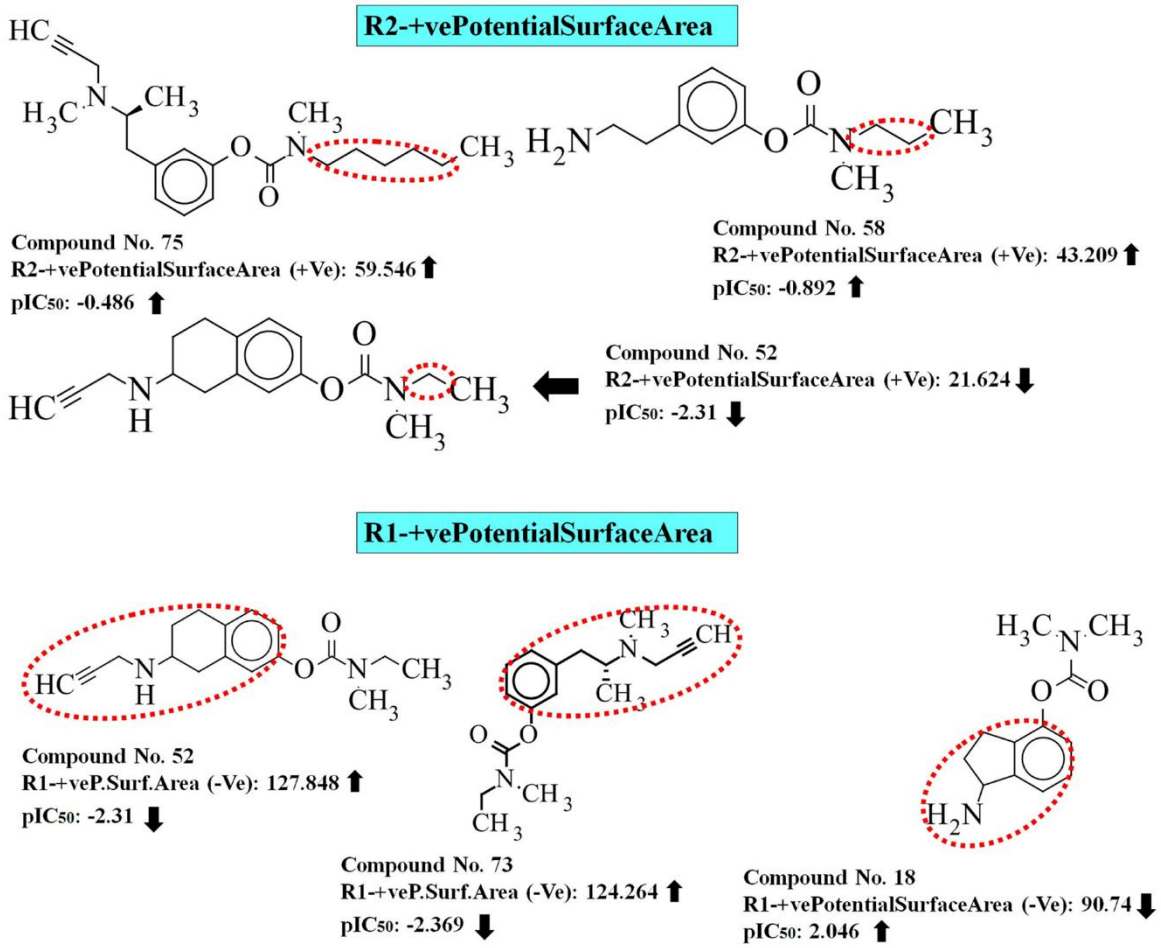


Figure 4.67. Contributions of +vePotentialSurfaceArea descriptor at R2 position and +vePotentialSurfaceArea descriptor at R1 position on AChE enzyme inhibitory activity.

The distance-based topological descriptor, R1-HosoyaIndex, belonging to the family of physicochemical descriptors, signifies the topological index or Z index of a graph which is the total number of matching in it plus 1 ("plus 1" accounts for the number of matchings with 0 edges), and it can be calculated through **equation 4.9**,

$$Z = \sum_{k=0}^{[\frac{n}{2}]} m(G,k) \qquad \textbf{Equation 4.9}$$

where n is the number of the vertices of graph G (order of graph G), [n/2] stands for the integer part of n/2, and m (G, k) is the number of k-matchings of graph G[393]. It shows a negative contribution to the inhibitory activity against the AChE enzyme and suggests that relatively lower numerical values of this fragment may contribute to the inhibitory activity. We have observed that the most active compounds (See **Figure 4.68)** **58** (pIC_{50}: -0.892) and **19** (pIC_{50}: -1.585) in the data set containing a single phenyl ring at R1 position have

lower numerical values for this descriptor, while the least active compounds (See **Figure 4.68) 43** (pIC_{50}: -2.642) and **73** (pIC_{50}: -2.369) in the data set contain indene ring on R1 position (which means a higher number of edge and vertices) show higher numerical values for this descriptor.

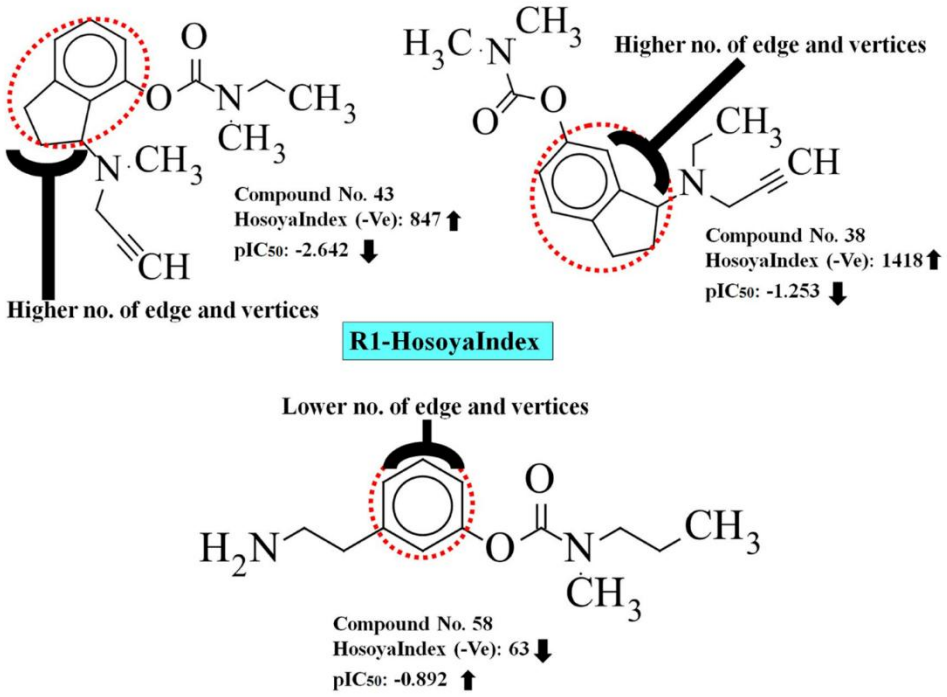


Figure 4.68. Contributions of HosoyaIndex descriptor at R1 position on AChE enzyme inhibitory activity.

The physicochemical descriptor, R1-SssCH2count, belongs to the sub-class E-state numbers. It indicates the total number of –CH2 groups which are connected with the help of two single bonds. It shows a negative contribution at the R1 substitution site of the compounds hinting that a reduction in the number of such groups would be better for the inhibitory activity of the compounds. We have observed that in the most active compounds (See **Figure 4.69**) **2** (pIC_{50}: 0.036) the absence of such fragment at the R1 position shows higher inhibitory activity whereas in the case of the least active compounds (See **Figure 4.69**) **43** (pIC_{50}: -2.642), **52** (pIC_{50}: -2.31) and **40** (pIC_{50}: -2.72), there is the higher number of the such fragment at R1 position which is detrimental for inhibitory activity.

R3-MMFF_1 is an atom type count descriptor and is based on MMFF atom types and their count in each molecule (N-CH3 group at R3 position). It shows a negative contribution to the inhibitory activity and suggests that the absence of this group at the R3 position may favor the inhibitory activity against the AChE enzyme. For instance, the most active compounds (See **Figure 4.69**) **36** (pIC_{50}: 0.26) in the dataset found the absence of methyl group at the R3 position; in contrast, the least active compounds (See **Figure 4.69**) **52** (pIC_{50}: -2.31) and **28** (pIC_{50}: -1.615) in the dataset contain N-CH3 group at the R3 position, which is detrimental for inhibitory activity.

R3-MomInertiaX is a distance-based topological descriptor. It refers to the moment of inertia at the X-axis of the molecules and implies that incorporation of any group at the R3 position that increases the resistance or restricts the internal rotation of the molecule will increase the AChE enzyme inhibitory activity. It shows a positive contribution to the inhibitory activity against the AChE enzyme. We have observed from the most active compounds (See **Figure 4.69) 18** (pIC_{50}: 2.046) and **55** (pIC_{50}: 1.886) in the data set that an increase in the value of this descriptor adds to the activity profiles of the molecules. On the other hand, the least active compounds (See **Figure 4.69**) **73** (pIC_{50}: -2.369) in the data set show lower numerical values of this descriptor.

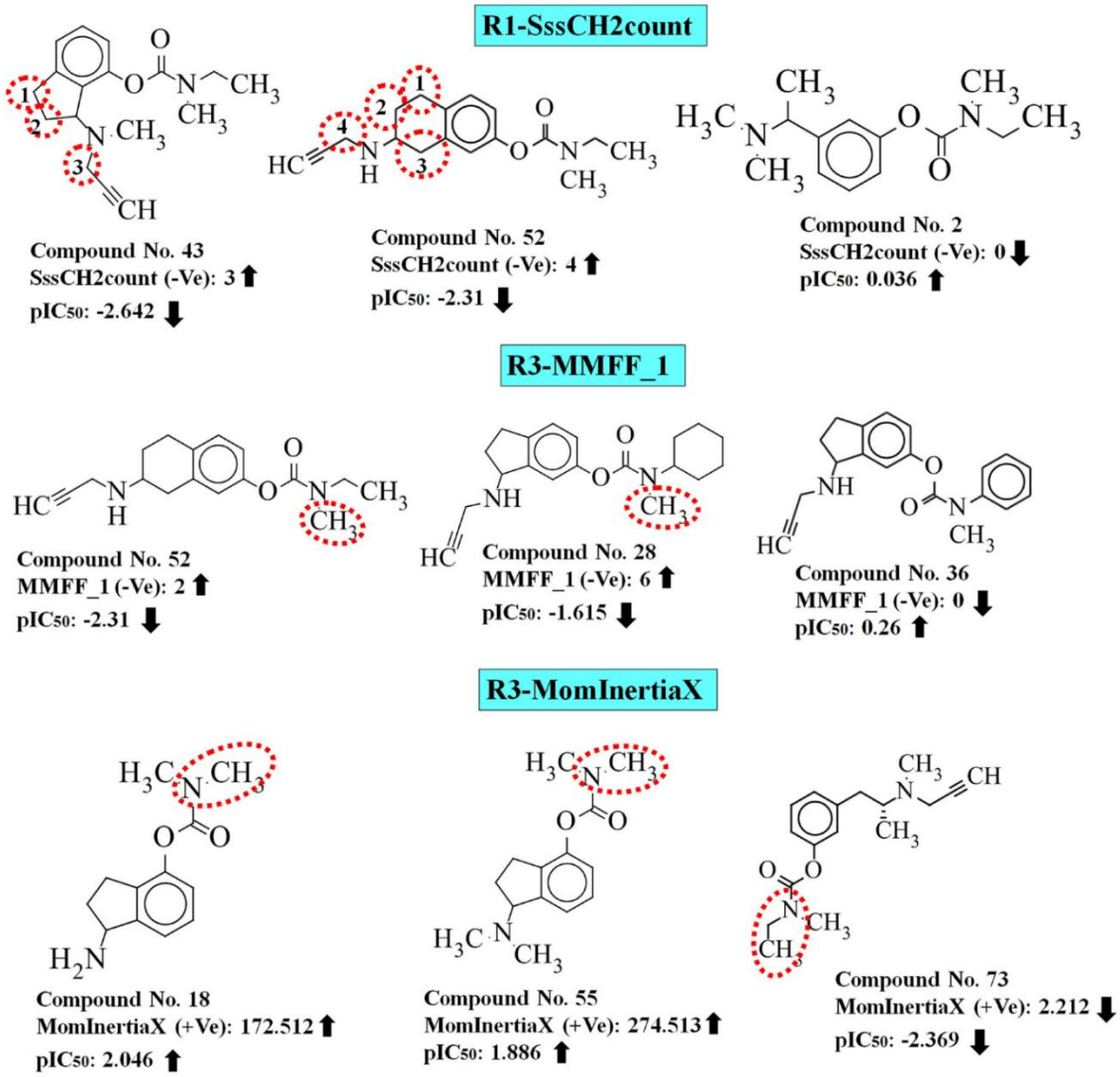


Figure 4.69. Contributions of SssCH2count descriptor at the R1 position, MMFF_1 descriptor at the R3 position, and MomInertiaX descriptor at the R3 position on AChE enzyme inhibitory activity.

Table 4.13. The list of Carbamate derivatives present in the dataset with their name, SMILES notation of respective compounds, observed and predicted activities (2D-QSAR and GQSAR) against the AChE enzyme.

Name	SMILES notation	Observed value (pIC_{50} µM)	Predicted value (pIC_{50} µM)	
			2D QSAR	GQSAR
1*	c1ccc(cc1[C@H](N(C)C)C)OC(=O)N(C)C	1.522	1.171	1
2	c1ccc(cc1[C@H](N(C)C)C)OC(=O)N(CC)C	0.036	-0.153	-0.175
3	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)N	0.119	0.235	0.660
4	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C)N	-1.278	-1.05	-1.237
5	c12CC[C@H](c1cc(cc2)OC(=O)N(CCC)C)N	-0.863	-0.765	-0.481
6*	c12CC[C@H](c1cc(cc2)OC(=O)N(C)CCCCCC)N	0.275	-0.035	-0.375
7	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C1CCCCC1)N	-0.597	-0.364	-0.649
8	c12CC[C@H](c1cc(cc2)OC(=O)N(C)c1ccc(cc1)OC)N	0.522	0.900	0.828
9	c12CC[C@H](c1cc(cc2)OC(=O)NCC)N	-1.247	-0.916	-1.002
10	c12CC[C@H](c1cc(cc2)OC(=O)NCCC)N	-0.170	-0.656	-0.458
11	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)NC	-0.029	0.115	0.409
12	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C)NC	-1.586	-1.162	-1.439
13	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)NCC	-1.389	-1	-0.308
14	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C)NCCC	-0.556	-2	-1.300
15	c1c2c([C@@H](CC2)N)c(cc1)OC(=O)N(C)C	0.337	0.147	0.788
16*	c1c2c(c(cc1)OC(=O)N(CC)C)[C@@H](CC2)N	-1.021	-1	-0.915
17	c1c2c(c(cc1)OC(=O)N(C)C)[C@@H](CC2)NC	0.292	-0.435	0.626
18	c1ccc(c2CC[C@H](c12)N)OC(=O)N(C)C	2.045	1.622	1
19	c1ccc(c2CC[C@H](c12)N)OC(=O)N(CC)C	1.585	0.336	-1
20	c12CCC[C@H](c1cc(cc2)OC(=O)N(C)C)N	-0.170	0.300	-0.091
21*	c12CCC[C@H](c1cc(cc2)OC(=O)N(CC)C)N	-0.793	-1	-2
22	c12CC[C@H](Cc1cc(cc2)OC(=O)N(C)C)N	-0.510	-0.426	-0.056
23*	c12CC[C@H](Cc1cc(cc2)OC(=O)N(CC)C)N	-1.900	-2	-2

24	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)NCC#C	-0.462	-0.268	-0.340
25	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C)NCC#C	-1.672	-2	-2
26	c12CC[C@H](c1cc(cc2)OC(=O)N(CCC)C)NCC#C	-1.164	-1	-1.303
27	c12CC[C@H](c1cc(cc2)OC(=O)N(C)CCCCCC)NCC#C	-1.195	-1	-1.720
28	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C1CCCCC1)NCC#C	-1.614	-1	-1.654
29*	c12CC[C@H](c1cc(cc2)OC(=O)N(C)c1ccc(cc1)OC)NCC#C	0.065	0.414	-0.293
30	c12CC[C@H](c1cc(cc2)OC(=O)NCC)NCC#C	-1.136	-1	-1
31*	c12CC[C@H](c1cc(cc2)OC(=O)NCCC)NCC#C	-0.376	-1	-1.026
32	c12CC[C@H](c1cc(cc2)OC(=O)N(CCCC)C)NCC#C	-1.053	-1	-1.542
33	c12c(cc(cc2)OC(=O)N(CCCC)CC)[C@@H](CC1)NCC#C	-1.859	-2	-2
34	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C1CCCCC1)NCC#C	-1.252	-2	-1.207
35	c12CC[C@H](c1cc(cc2)OC(=O)N(C)Cc1ccccc1)NCC#C	-0.250	-0.492	-0.282
36	c12CC[C@H](c1cc(cc2)OC(=O)N(C)c1ccccc1)NCC#C	0.259	-0.086	-0.282
37	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)N(C)CC#C	-1.110	-0.335	-0.487
38	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)N(CC)CC#C	-1.252	-2	-1
39	c1c2c([C@@H](CC2)NCC#C)c(cc1)OC(=O)N(C)C	-0.397	-1	-0.100
40	c1c2c([C@@H](CC2)NCC#C)c(cc1)OC(=O)N(CC)C	-2.720	-2	-2
41	c1c2c(c(cc1)OC(=O)N(CCC)C)[C@@H](CC2)NCC#C	-1.653	-1.803	-1
42	c1c2c([C@@H](CC2)N(C)CC#C)c(cc1)OC(=O)N(C)C	-0.845	-1.340	-0.272
43	c1c2c([C@@H](CC2)N(C)CC#C)c(cc1)OC(=O)N(CC)C	-2.642	-2.59	-2
44	c1ccc(c2CC[C@H](c12)NCC#C)OC(=O)N(C)C	1.275	1	-0.023
45	c1ccc(c2CC[C@H](c12)NCC#C)OC(=O)N(C)CC	-0.332	-0.139	-1
46	c12CC[C@H](c1cc(c(Cl)c2)OC(=O)N(CC)C)NCC#C	-1.406	-2	-2
47*	c12CC[C@H](c1cc(c(Cl)c2)OC(=O)N(CCC)C)NCC#C	-1.642	-1	-1
48	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)N[C@H](C#C)C	-0.255	-0.287	-0.161
49	c12CCC[C@H](c2cc(cc1)OC(=O)N(C)C)NCC#C	-0.595	-0.212	-1

50	c12CCC[C@H](c2cc(cc1)OC(=O)N(CC)C)NCC#C	-1.719	-1	-2
51	c12CC[C@H](Cc2cc(cc1)OC(=O)N(C)C)NCC#C	-0.612	-1	-1
52	c12ccc(cc2C[C@@H](CC1)NCC#C)OC(=O)N(C)CC	-2.309	-2	-1
53	c1ccc(c2CC[C@H](c12)N(C)CC#C)OC(=O)N(CC)C	-1.173	-0.188	-2
54	c12CC[C@H](c2cc(cc1)OC(=O)N(C)C)N(C)C	-0.332	0.279	0.353
55	c1ccc(c2CC[C@H](c12)N(C)C)OC(=O)N(C)C	1.886	2	0.482
56*	c1ccc(cc1CCN)OC(=O)N(C)C	0.638	0.458	1.071
57	c1ccc(cc1CCN)OC(=O)N(CC)C	-1.550	-1	-1
58	c1ccc(cc1CCN)OC(=O)N(CCC)C	-0.892	-1	-0.067
59	c1ccc(cc1CCNC)OC(=O)N(C)C	0.552	1	0.832
60	c1ccc(cc1CCNC)OC(=O)N(CC)C	-1.315	-1	-1.414
61	c1ccc(cc1CCN(C)C)OC(=O)N(C)C	0.795	1	-0.128
62	c1ccc(cc1CCNCC#C)OC(=O)N(C)C	0.657	0.033	-0.128
63	c1ccc(cc1CCNCC#C)OC(=O)N(C)CC	-1.480	-1.309	-0.450
64	c1ccc(cc1CCNCC#C)OC(=O)N(CCC)C	-1.187	-1.047	-1
65	c1ccc(cc1CCN(CC#C)C)OC(=O)N(C)C	0.070	-0.005	-0.335
66*	c1ccc(cc1CCN(CC#C)C)OC(=O)N(CC)C	-1.220	-1	-2
67	c1ccc(cc1C[C@H](C)NCC#C)OC(=O)N(C)C	0.267	-0.021	0.150
68	c1ccc(cc1C[C@@H](C)NCC#C)OC(=O)N(CC)C	-1.530	-1.345	-2
69*	c1ccc(cc1C[C@@H](C)NCC#C)OC(=O)N(CCC)C	-1.281	-1	-1.126
70*	c1ccc(cc1C[C@@H](NCC#C)C)OC(=O)N(C)C1CCCCC1	-0.556	-1	-1
71	c1ccc(cc1C[C@@H](C)NCC#C)OC(=O)N(C)CCCC	-1.086	-1.054	-1
72*	c1ccc(cc1C[C@H](C)N(CC#C)C)OC(=O)N(C)C	-0.214	-0.193	0.148
73	c1ccc(cc1C[C@@H](C)N(C)CC#C)OC(=O)N(CC)C	-2.369	-1.495	-2
74*	c1ccc(cc1C[C@@H](N(C)CC#C)C)OC(=O)N(CCC)C	-1.519	-1.237	-1.197
75	c1ccc(cc1C[C@@H](N(CC#C)C)C)OC(=O)N(CCCCCC)C	-0.485	-0.581	-1
76	c1ccc(cc1C[C@H](C)N(CC#C)C)OC(=O)N(C)C1CCCCC1	-0.938	-0.935	-1
77	c1c(CCN[C@@H](C#C)C)cc(cc1)OC(=O)N(C)C	0.769	-0.067	0.179
78*	c1ccc(cc1CCN[C@@H](C#C)C)OC(=O)N(CC)C	-1.143	-1.389	-1.521

*Test set compounds

4.4.1.3. Applicability domain of the PLS models

In the current work, applicability domain criteria were checked for test set compounds at a 99% confidence level as described by Roy et al and Khan et al.[352, 394] using the DModX (distance to model in X-space) approach available in SIMCA-P 10.0 software (Available from https://umetrics.com/products/simca). **Figure 4.70** have demonstrated that all the test set compounds in both 2D-QSAR and GQSAR models are within the critical DModX values (Model 1: D-Crit =2.412, Model 2: D-Crit = 3.506).

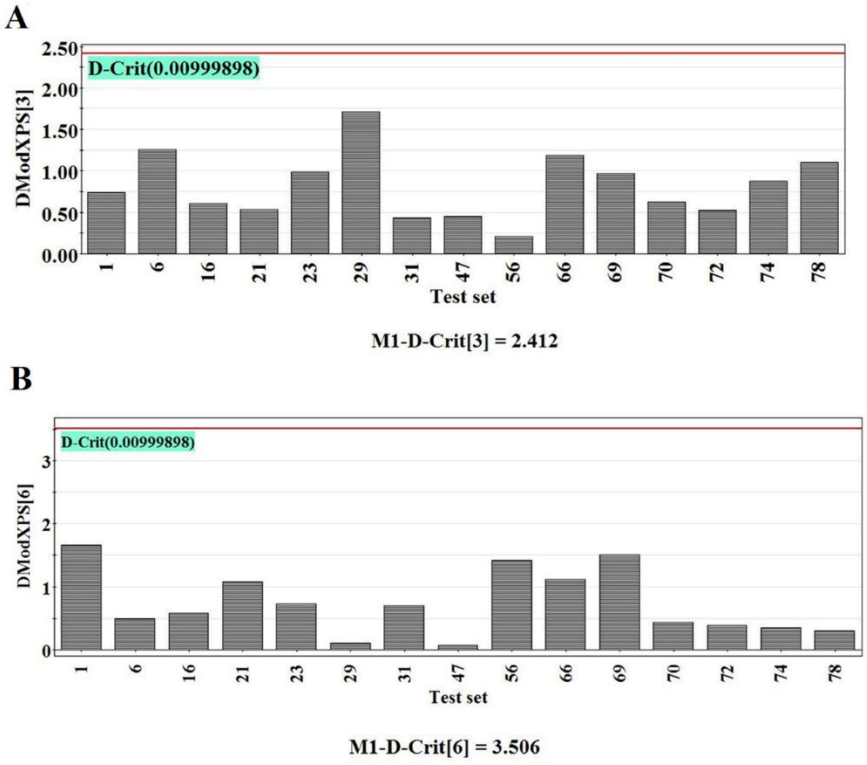


Figure 4.70. Applicability domain DModX values of the test set compounds at 99% confidence level of the developed PLS (2D QSAR and GQSAR) model against AChE enzyme (A: 2D QSAR and B: GQSAR).

4.4.2. 3D-Pharmacophore model analysis

In this study, ten different pharmacophore models were developed using a training set of 23 compounds. From ten different pharmacophore hypotheses, **Hypo-2** with a high correlation coefficient (r: 0.858), lower root mean square deviation (rmsd: 2.061), error 66.77, lower configuration cost (9.909), and weight 1.12 were found to be of acceptable quality. Based on reported metrics (**Table 4.14**), Hypo-2 was found to be the best one among the ten hypotheses with one HBA, one HBD, one RA, and one hydrophobic feature (**Figure 4.71**). External validation of the model has been carried out by mapping the test set molecules (**Figure 4.71**) on Hypo-2 with the same settings as employed for the pharmacophore generation by the BEST method. After mapping, we observed that 50 molecules from the test set of 55 compounds were mapped, and only 5 compounds failed in absence of the features found in the developed model. The values

of different validation parameters for the training, as well as test sets, are given in **Table 4.14** (qualitative validation parameters). The F-test confirms the non-randomness of the developed pharmacophore (Hypo-2). The original and randomized total cost and correlation values of the hypotheses for the F-test are given in **Figure 4.72**. The correlation of the 3D pharmacophore model with QSAR (2D-QSAR and GQSAR) models is depicted in **Table 4.15**.

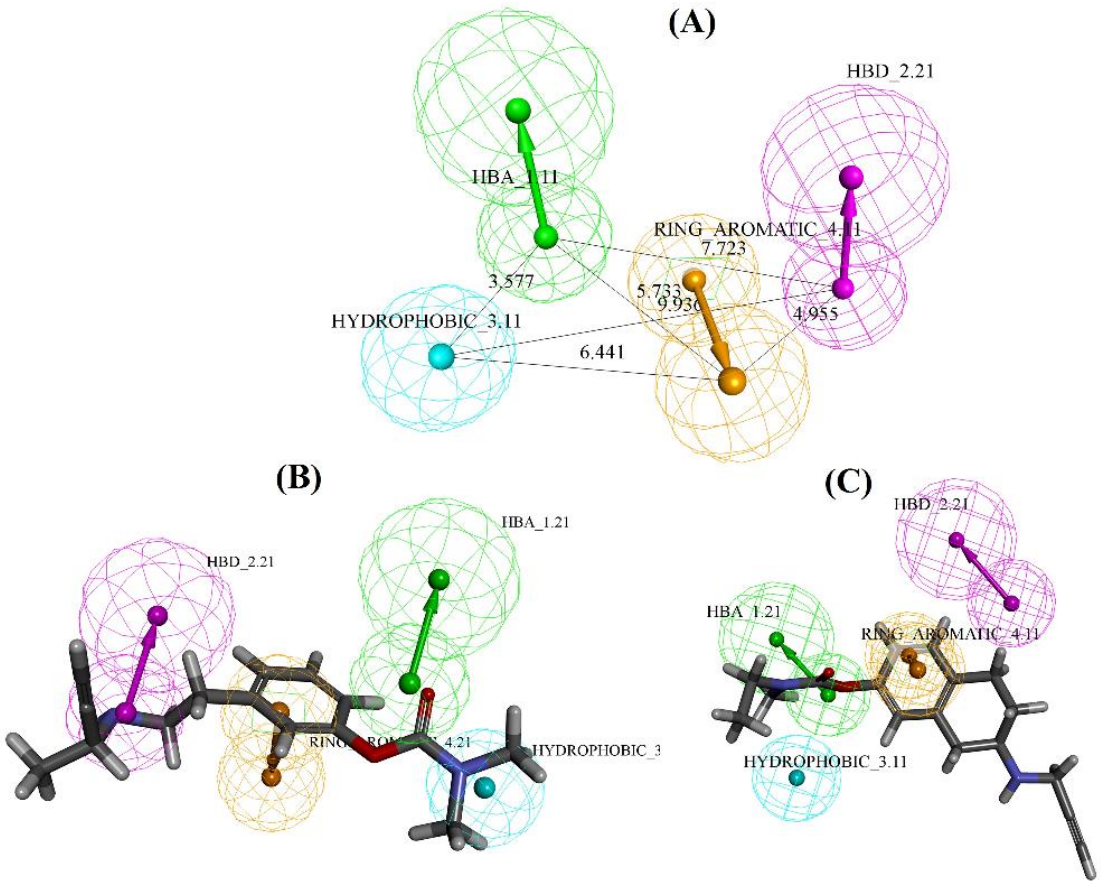


Figure 4.71. (A) Pharmacophore hypothesis (Hypo-2) with one hydrogen bond acceptor (HBA), one hydrophobic (HYD), one hydrogen bond donor, and one ring aromatic (RA) features and interfeature distance (A°); (B) Mapping of the most active compound **77** of the test set (pharmacophore mapping) on the Hypo-2, (C) Mapping of the least active compound **52** of the test set (pharmacophore mapping) on Hypo-2.

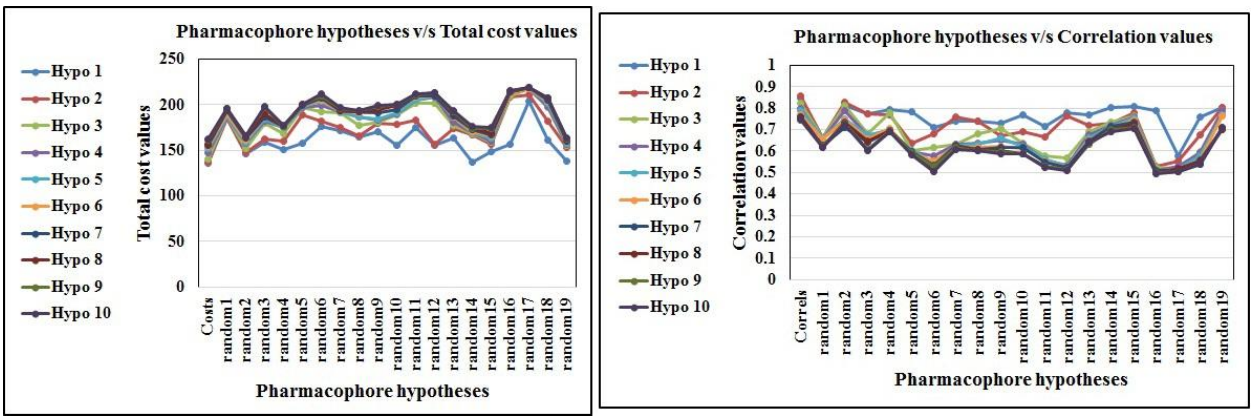


Figure 4.72. The original and randomized total cost and correlation values of the hypotheses for the F-test

Table 4.14. Different quantitative and qualitative validation parameters of the **Hypo-2** model were obtained for the training and test sets generated by the HypoGen algorithm.

Quantitative validation parameters						
Hypo.	**Total cost**	**ΔCost[a]**	**ΔCost[b]**	**RMS**	**Correlation**	**Features**
2	137.453	109.27	59.649	2.061	0.858	HBA, HBD, HYD, RA

Dataset	**No. of compounds**	**Qualitative validation parameters**					
		Sensitivity	**Specificity**	**Accuracy**	**Precision**	**F-measure**	**G-Means**
Train	23	100	95	97.36	94.73	97.29	97.46
Test	55	80	95.91	94.44	66.66	72.72	87.59

Cost difference[a]= Null cost - total cost, Cost difference[b]= Total cost - fixed cost, Null cost = 246.725, Fixed Cost = 77.804, Best records in pass: 3, Config. Cost: 9.909, C= Best Hypothesis, Note: RA: Ring aromatic, HYD: Hydrophobic, HBA: Hydrogen bond acceptor, HBD: Hydrogen bond donor

Table 4.15. Correlation of 3D-QSAR pharmacophore model with QSAR (2D-QSAR and GQSAR) models.

Pharmacophoric features	Correlation with QSAR (2D-QSAR and GQSAR) models
Ring aromatic	The compounds present in the dataset have at least one ring aromatic feature, which is either pyrazole/pyridine/thiazole/phenyl or another heterocycle. The RA features are the preliminary requirements for the inhibitory activity against the AChE enzyme. The RA features are in accordance with the NRS descriptor of the 2D-QSAR model (**Eqs. 7**). This observation we can see from the most active compound of the test set (**77** IC_{50}: 0.17μM) (**Figure 4.71)** one benzene ring lies in the RA region.
Hydrogen bond acceptor	The HBA feature (-CO) of the pharmacophore model is also in accordance with the +vePotentialSurfaceArea and MomInertiaX (moment of inertia at X-axis) descriptors of the GQSAR models at R2 and R3 positions respectively in the compounds (**Eq. 8**). The most contributing descriptor in position R2 is +vePotentialSurfaceArea, which suggests that an increase in the positive electrostatic potential of the fragment at R2 site may lead to an increase in the inhibitory activity against AChE enzyme. As found in the most active compound of the test set (**Figure 4.71**) (**77** IC_{50}: 0.17μM), the presence of these features in the compound shows higher inhibitory activity.
Hydrogen bond donor	Hydrogen bond donor feature from the obtained model in accordance with the F06 [C-N] descriptor of the 2D-QSAR model.
Hydrophobic	The hydrophobic feature (-CH3) from the developed model is in accordance with the X2A descriptor of the 2D-QSAR model (**Eqs**. **7**). We have observed from the most active compound of the test set (**77** IC_{50}: 0.17μM) **Figure 4.71**, hydrophobic feature on **Hypo 2** mapped completely with the molecule.

4.4.3. Molecular docking

A molecular docking study was performed using the most active (**19**, **18, 44, 55, and 61)** (**Figure 4.73**) and least active (**33, 40, 43, 52,** and **73)** (**Figure 4.74**) compounds of the dataset. The details of docking interactions are depicted in **Table 4.16** and their correlations with the final QSAR (2D-QSAR and GQSAR) models are shown in **Table 4.17**.

Table 4.16. Details of docking interactions (Most and Least active compounds from the dataset).

Sr. No.	Name of compound	Docking interactions
1	**19**	The most active compound from the dataset; it interacted with the amino acid residues through hydrogen bonding (SER A: 293, TYR A: 341, and TYR A: 124), alkyl, π-alkyl, π-π stacked and π-cation interactions (TRP A: 286 and TYR A: 341 respectively) and salt bridge formation (LEU A: 289 and ASP A: 74) (Figure 4.73).
2	**18**	Figure 4.73 showed that compound **18** (the most active compound in the dataset) interacted with SER A: 293, TYR A: 341, and TYR A: 124 amino acid residues through hydrogen bonding interaction, TYR A: 341 and TRP A: 286 through π-cation, π-alkyl and π-π stacked) and ASP A: 74 amino acid through salt bridge formation.
3	**44**	Figure 4.73 showed that compound **44** interacted with SER A: 293 and TYR A: 124 (through hydrogen bonding), LEU A: 289, VAL A: 294, TYR A: 124, PHE A: 297, TRP A: 286 and TYR A: 341 (alkyl, π-alkyl and π-π-stacked bond respectively) amino acid residues.
4	**55**	Compound **55** (Figure 4.73) interacted with the amino acid residues through hydrogen bonding (TYR A: 341 and SER A: 293) and π-alkyl and π-π-stacked (TYR A: 337, PHE A: 297, PHE A: 338, TYR A: 341 and TRP A: 286 respectively) amino acid residues.
5	**61**	The most active compound from the dataset, compound **61** (Figure 4.73) interacted with TYR A: 337 and TYR A: 72 (through hydrogen bonding), TYR A: 337, PHE A: 297 TRP A: 286 (through π-alkyl and π-cation respectively), TYR A: 341 and ASP A: 74 (via π-π stacked and pi-cation) amino acid residues.
6	**33**	The least active compound from the dataset, compound **33** (Figure 4.74), interacted with TYR A: 124 and TYR A: 73 (through hydrogen bonding), PHE A: 338, TRY A: 337, PHE A: 297, TYR A: 124, TYR A: 341, TYR A: 286 (through π-alkyl interaction), TRP A: 286 (via π-π stacked interaction) and LEU A: 76 (through alkyl bond) amino acid residues.
7	**40**	Another least active compound from the dataset, compound **40,** interacted with the amino acid residues through hydrogen bonding (SER A: 293, TRP A: 286, TYR A: 341), π-alkyl (TYR A: 72, TRP A: 286 and TYR A: 341) alky (LEU A: 289 and VAL A: 294) and π-π stacked interactions (TRP A: 286 and TYR A: 341) (Figure 4.74).

8	**43**	Figure 4.74 showed that compound **43** (one of the least active compounds from the dataset) interacted with SER A: 293 and TYR A: 73 amino acid residues through hydrogen bonding interaction, TYR A: 72, TRP A: 286, TYR A: 341 amino acid residues through π-alkyl interaction, LEU A: 289 via alkyl bond and TYR A: 341 and TRP A: 286 amino acids via π-π stacked interaction.
9	**52**	Compound **52** (See Figure 4.74 interacted with the amino acid residues through hydrogen bonding (TYR A: 124), π-alkyl (TYR A: 72, TYR A: 337, TRP A: 286 and TYR A: 341), and π-π-stacked interactions (TRP A: 286 and TYR A: 341) amino acid residues.
10	**73**	Figure 4.74 showed that compound **73** interacted with TYR A: 124 (through hydrogen bonding), TYR A: 337, PHE A: 338, TYR A: 72 and TYR A: 341 through π-alkyl, TYR A: 341 and VAL A: 294 via alkyl bond and TRP A: 286 via π-π-stacked bond) amino acid residues.

Table 4.17. Docking results and their correlation with the final QSAR (2D QSAR and GQSAR) models.

S. No.	Compound Number	- CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
1	**18** (high pIC_{50})	40.306	SER A: 293, TYR A: 341, ASP A: 74, TYR A: 124 and TRP A: 286	Hydrogen bonding (Classical and non-classical), electrostatics (Salt bridge, attractive charges and π-cation) and hydrophobic (π-π stacked and π-alkyl)	NRS, R2-+vePotentialSurfaceArea and R3-MomInertiaX
2	**19** (high pIC_{50})	42.26	LEU A: 289, SER A: 293, TRP A: 286, TYR A: 341, ASP A: 74 and TYR A: 124	Hydrogen bonding (Classical, non-classical and Salt bridge), electrostatics (Salt bridge, attractive charges and π-cation) and hydrophobic (π-π stacked, π-alkyl and alkyl)	NRS and R2-+vePotentialSurfaceArea
3	**44** (high pIC_{50})	36.76	LEU A: 289, SER A: 293, TRP A: 286, TYR A: 341, VAL A: 249, PHE A: 297 and TYR A: 124	Hydrogen bonding (Classical, non-classical and Salt bridge) and hydrophobic (π-π stacked, π-alkyl and alkyl)	NRS, X2A, R2-+vePotentialSurfaceArea and R3-MomInertiaX
4	**55** (high pIC_{50})	40.88	TRP A: 286, SER A: 293, TYR A: 341, PHE A: 297,	Hydrogen bonding (Classical and non-classical) and	NRS, R2-+vePotentialSurf

			PHE A: 338 and TYR A: 337	hydrophobic (π-π stacked and pi-alkyl)	aceArea and R3-MomInertiaX
5	**61** (high pIC_{50})	42.625	PHE A: 297, TYR A: 341, TRP A: 286, ASP A: 74, TYR A: 72 and TYR A: 337	Hydrogen bonding (Non-classical), Electrostatics (attractive charges and π-cation), hydrophobic (π-π stacked and π-alkyl)	NRS, R2-+vePotentialSurfaceArea and R3-MomInertiaX
6	**33** (low pIC_{50})	34.295	TYR A: 124, TYR A: 73, PHE A: 338, TRY A: 337, PHE A: 297, TYR A: 341, TRP A: 286 and LEU A: 76	Hydrogen bonding (Classical and non-classical), hydrophobic (π-π stacked, alkyl and π-alkyl)	NRS, X2A, R1-+vePotentialSurfaceArea
7	**40** (low pIC_{50})	31.686	SER A: 293, TRP A: 286, TYR A: 341, TYR A: 72, LEU A: 289, VAL A: 294 and TYR A: 341	Hydrogen bonding (non-classical), hydrophobic (π-π stacked and π-alkyl)	NRS, X2A, R1-+vePotentialSurfaceArea
8	**43** (low pIC_{50})	31.129	SER A: 293, TYR A: 72, TYR A: 341, LEU A: 289 and TRP A: 286	Hydrogen bonding (non-classical), hydrophobic (π-π stacked, alkyl and π-alkyl)	NRS, X2A, R2-+vePotentialSurfaceArea
9	**52** (low pIC_{50})	34.114	TYR A: 124, TYR A: 72, TYR A: 337, TRP A: 286 and TYR A: 341	Hydrogen bonding (Classical and non-classical), hydrophobic (π-π stacked and π-alkyl)	X2A, R2-+vePotentialSurfaceArea
10	**73** (low pIC_{50})	32.807	TYR A: 124, TYR A: 337, PHE A: 338, TYR A: 72, TYR A: 341, VAL A: 294, TRP A: 286	Hydrogen bonding (non-classical), hydrophobic (π-π stacked, alkyl and π-alkyl)	NRS, X2A, R2-+vePotentialSurfaceArea

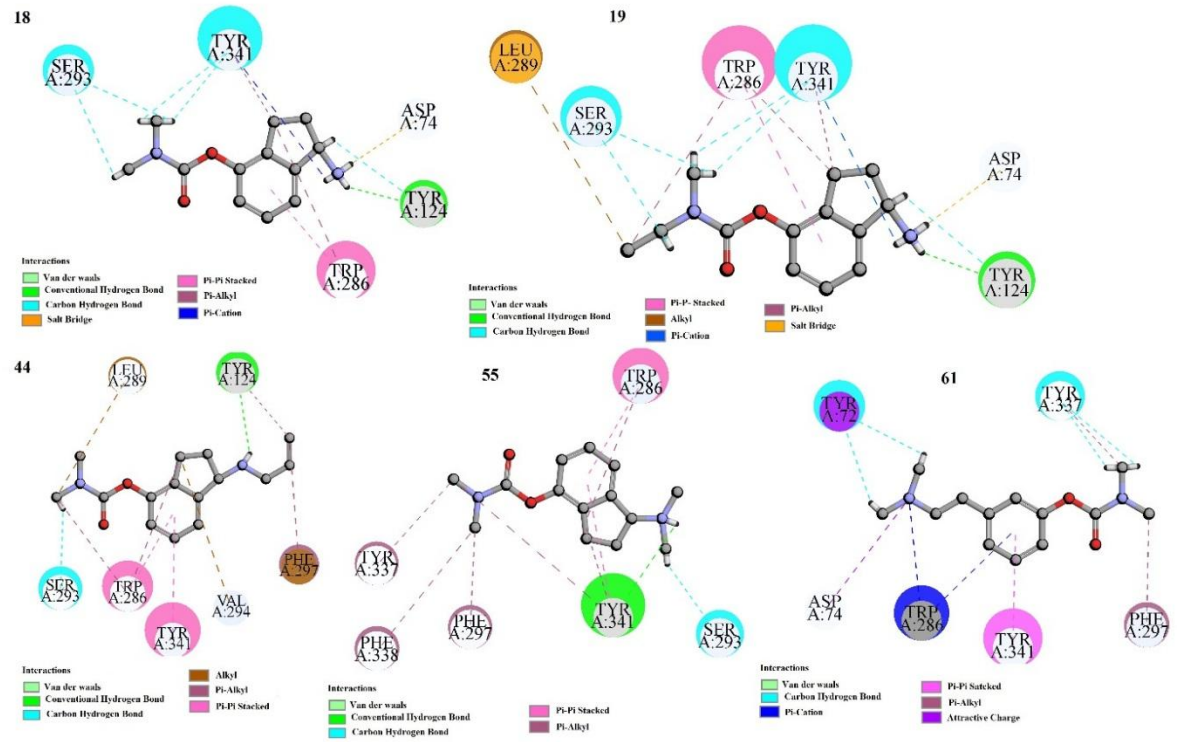


Figure 4.73. Docking interactions of most active compounds of dataset (compound **18**, **19, 44, 55** and **61)**.

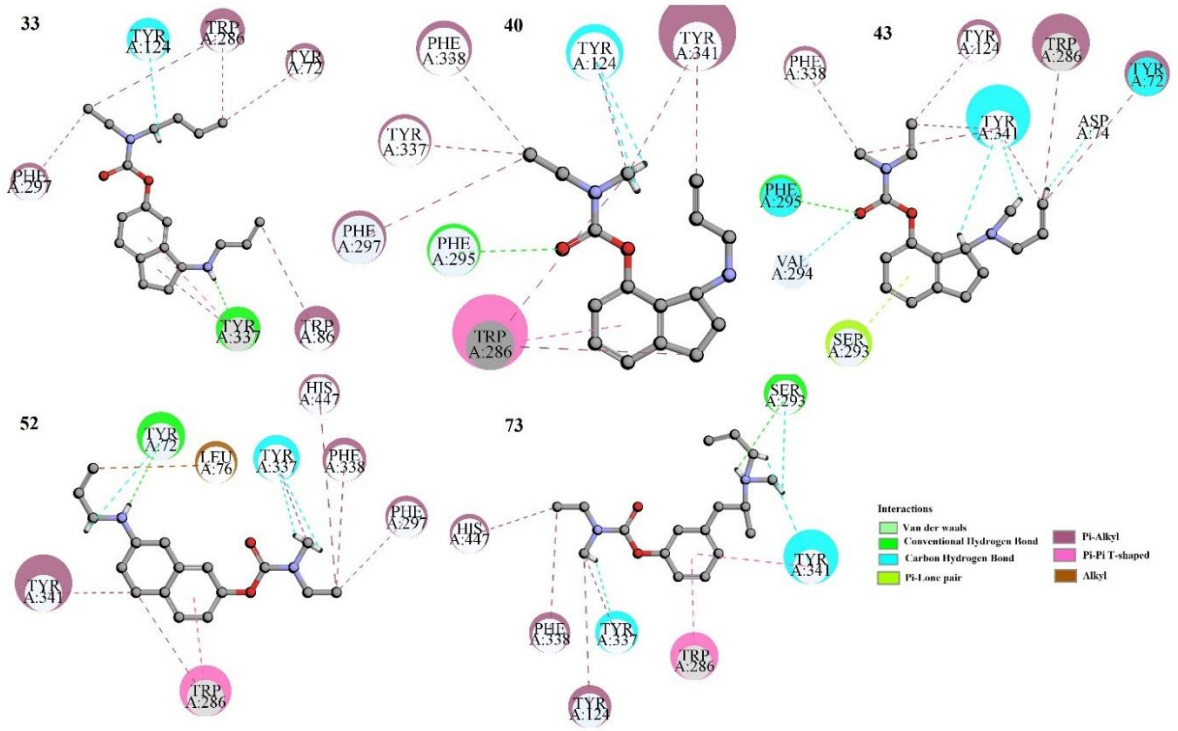


Figure 4.74. Docking interactions of least active compounds of dataset (compound **33, 40, 43, 52** and **73)**.

4.4.3.1. Relationship of the docking results with QSAR (2D QSAR and GQSAR) results

The molecular docking analysis revealed that the formation of hydrogen bonds (classical and non-classical), hydrophobic bonding (π-π Stacked, π-π-T-Shaped, alkyl, and π-alkyl) and some other electrostatic interactions such as π-cation and attractive charges between the ligand and the protein play a vital role in the binding process. Hydrogen and hydrophobic bonding can be correlated with X2A, R2-+vePotentialSurfaceArea, R3-MomInertiaX, and NRS descriptors in the QSAR models. X2A, R2-+vePotentialSurfaceArea, and R3-MomInertiaX are related to hydrogen bonding and electrostatic interactions between protein and ligand. The descriptor, NRS, gives evidence of hydrophobic interaction. The descriptor +vePotentialSurfaceArea contributes positively to the substitution site R2; but in contrast, it contributes negatively in the case of the R1 substitution site as we have observed in the least active compounds from dataset like **33** and **40** (**Figure 4.74**). Furthermore, the R3-MomInertiaX (moment of inertia at X-axis) descriptor supports the evidence of hydrophobic interactions along with hydrogen bonding interactions as we have observed in (**Figure 4.73**) most active compounds from the dataset (**19, 18**, **44**, **55,** and **61**), Thus, from above-stated information, we can conclude that hydrogen bonding, hydrophobicity, electrostatic interactions and unsaturation (π-π interaction) features as obtained from both QSAR and docking study are essential for the inhibitory activity against the AChE enzyme.

4.4.4. Comparison of the performance of the present QSAR models with previously published models

In the present work, a comparison of the best models of this study with previously published models[395, 101, 104, 396, 102, 397, 103, 398] was performed for the prediction of the bioactivity against the AChE enzyme, as depicted in **Table 4.18**. The details of different internal and external validation parameters obtained from our models and those obtained from previous models are given in **Table 4.18.** It is important to note that the models developed in this study show better quality models of low equation length and less number of variables of LVs and consider more diverse compounds as compared to previously reported models.

Table 4.18. Comparisons of the proposed study with previously published studies against AChE enzyme.

Sources	EL	LV	Model	Training set			Test set	
				n	R^2	Q^2	n	R^2_{pred}
Model 2 (2D QSAR)	**6**	**3**	**PLS**	**63**	**0.789**	**0.732**	**15**	**0.88**
Model 2 (GQSAR)	**7**	**6**	**PLS**	**63**	**0.625**	**0.538**	**15**	**0.735**
Brahmachari et al. 2015[395]	8	5	PLS	325	0.647	0.625	105	0.675
Shen et al. 2007[101]	-	5	CoMFA	36	0.974	0.784	9	0.968
Shen et al. 2007[101]	-	4	CoMSIA	36	0.947	0.736	9	0.927
de Souza et al. 2012[104]	-	8	HQSAR	29	0.965	0.787	7	-
Goyal M et al. 2014[396]	4	4	PLS	19	0.822	0.683	5	0.789
Solomon KA et al. 2009[102]	5	-	GFA	62	0.857	0.803	26	0.882
Karmakar A et al. 2019[397]	3	-	GFA	28	0:683	0:589	-	0:641
Gupta et al 2011[103]	6	-	GFA	31	0.88	0.838	11	0.75
Gupta et al 2011[103]	6	-	GPLS	31	0.889	0.739	11	0.706
Gupta et al 2011[103]	-	-	SVM	31	0.798	-	11	0.762
Gupta et al 2011[103]	-	-	ANN	31	0.753	-	11	0.694
Bernd., et al 2003[398]	-	-	CoMFA	28	0.974	0.671	4	-

Abbreviations: LV= Latent variables, MLR= Multiple linear regression, CoMFA= Comparative Molecular Field Analysis, CoMSIA= Comparative molecular similarity index analysis, ANN = Artificial neural network, SVM = Support vector machine, PLS= Partial least square, GFA= genetic function approximation and HQSAR= Hologram QSAR.

4.5. Study 5- In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease

4.5.1. 2D QSAR analysis

In this research, PLS regression-based 2D-QSAR models were developed for AChE and BuChE enzyme inhibitory activities. In the datasets, some compounds have both AChE and BuChE inhibitory values. A PLS regression-based 2D-QSAR model has been developed based on their selectivity (AChEI-BuChEI). The details of the different validation metrics values for the models are summarized in **equations 4.10, 4.11,** and **4.12**. The statistical results obtained from the models suggested that the developed models are acceptable in terms of stability, predictive, and fitness criteria. The nearness of the observed and predicted values for the AChE and BuChE enzyme inhibitors and also the selectivity values can be further established from the scatter plots as shown in **Figures 4.75 and 4.76**. The quantitative contributions of similar/dissimilar descriptors and the interrelationships between the X-variables and the Y-response are depicted in the loading plots of **Figure 4.77**. Moreover, we have also employed the Y-Randomization test to cross-verify whether the models were obtained by any chance or not using Simca-P 10.0 software (Available from https://umetrics.com/products/simca). The results obtained from the randomized models, in the case of AChE inhibitors (Model 1: R^2int (intercept values) = -0.0011 and Q^2int = -0.157), whereas in the case of BuChE inhibitors (Model 2: R^2int = 0.0017 and Q^2int = -0.184) and in case of selectivity based model (Model 3: R^2int = -0.0013 and Q^2int = -0.26), suggested that the developed models were not obtained by any chance correlation as depicted in **Figure 4.78**.

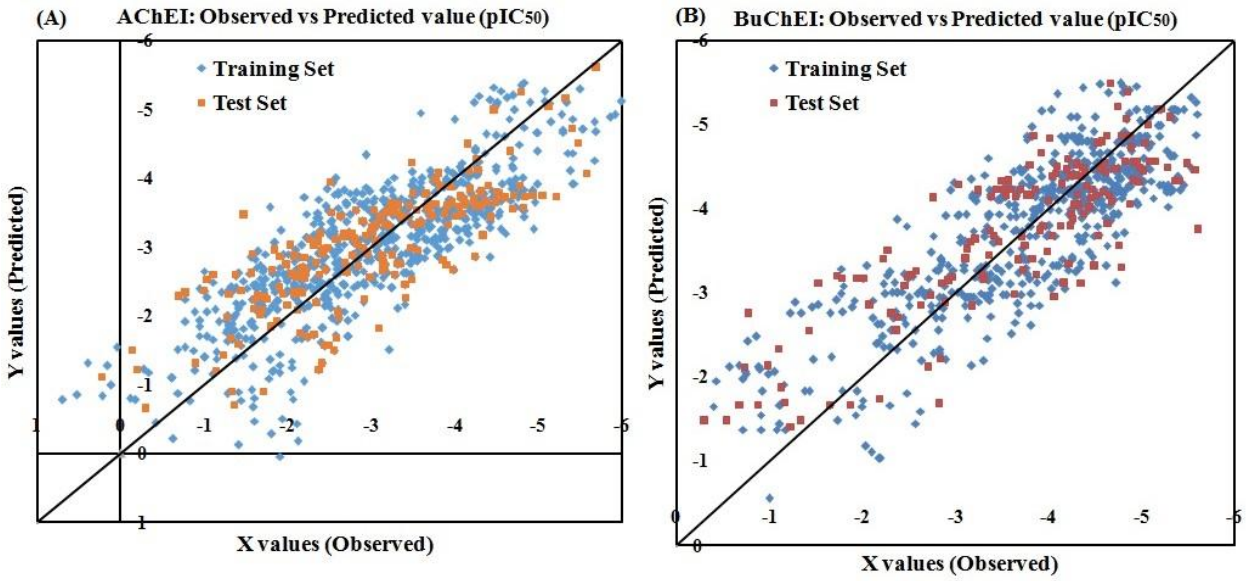


Figure 4.75. Scatter plots of developed PLS models (observed and predicted values) against (A) AChE and (B) BuChE enzymes.

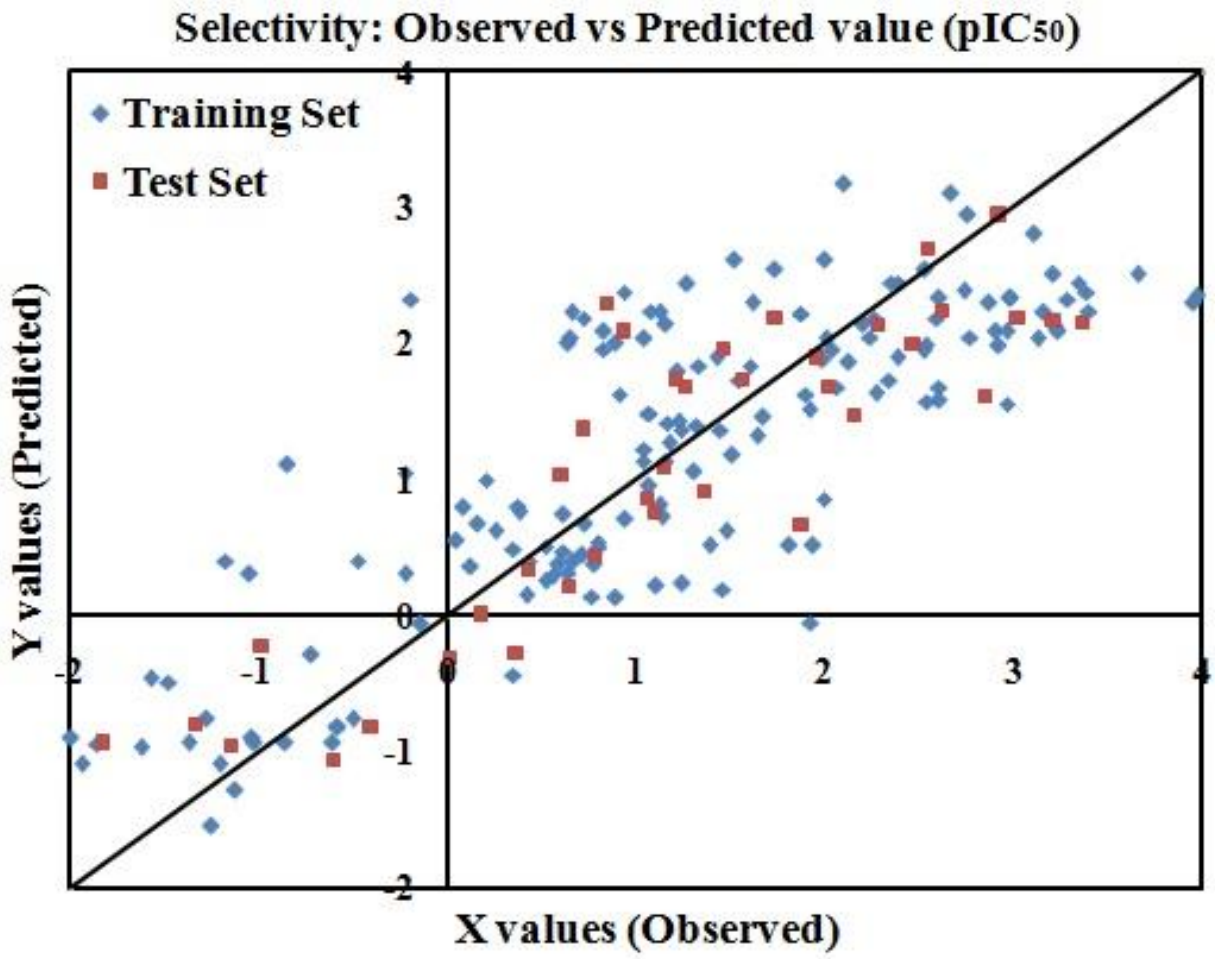


Figure 4.76. Scatter plot of the model based on selectivity (AChE-BuChE enzyme) (observed and predicted values).

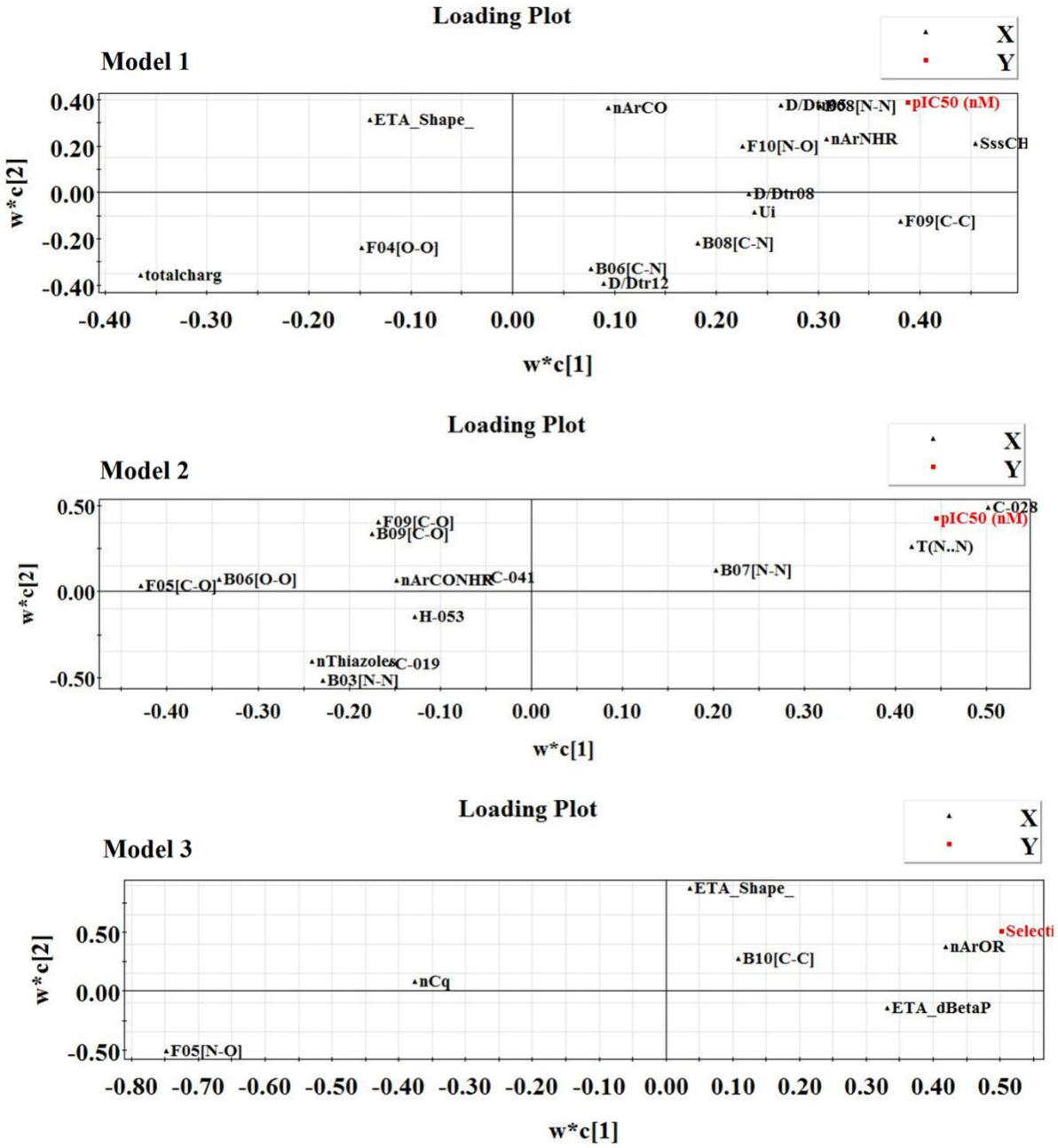


Figure 4.77. Loading plot for the final PLS models: Model 1 = AChE, Model 2 = BuChE enzyme and Model 3 = based on selectivity.

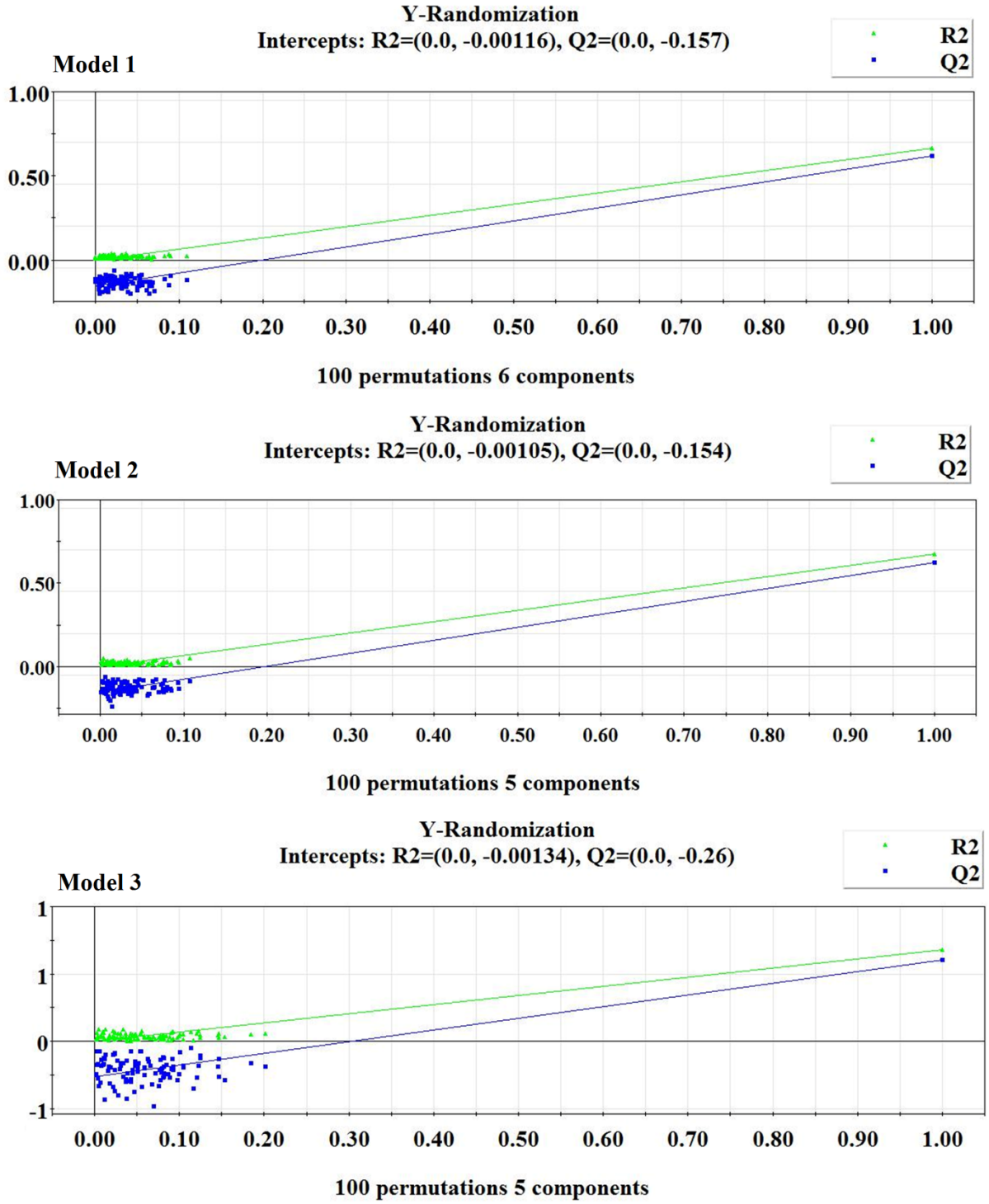


Figure 4.78. Model Randomization plots for the final PLS models: Model 1 = AChEI, Model 2 = BuChEI and Model 3 = based on selectivity.

4.5.1.1. Mechanistic interpretation of modeled descriptors

4.5.1.1.1. 2D-QSAR analysis against AChE enzyme: Model 1

Box 4.2. 2D-QSAR model and statistical validation parameters obtained from the developed model against the AChE enzyme.

$$pIC_{50} = -2.806 + 0.148 \times SssCH2 - 0.905 \times totalcharge - 0.027 \times F09[C-C] + 0.504 \times nArNHR + 0.004 \times D/Dtr05 + 0.588 \times B08[N-N] + 6.32076 \times ETA\_Shape\_P - 0.007 \times D/Dtr12 + 0.009 \times D/Dtr08 + 1.125 \times Ui - 0.284 \times B06[C-N] - 0.337 \times B08[C-N] + 0.361 \times F10[N-O] + 0.669 \times nArCO - 0.209 \times F04[O-O] \quad \textbf{Equation 10}$$

Internal Validation Parameters: n_{train}= 798, r^2= 0.662, Q^2= 0.645, EL = 15, LV = 6, Prediction quality = Good.

External Validation Parameters: n_{test}= 199, Q^2F_1= 661, Q^2F_2= 660, Prediction quality = Good.

The descriptors appearing in the model have been ranked accordingly to their significance, and then described individually. The importance and contribution of the obtained descriptors in the models towards the AChE inhibitory activity are identified based on the variable importance plot (VIP) and regression coefficient plot as shown in **Figure 4.79**[111]. The importance of each descriptor against the AChE enzyme has been analyzed with their appropriate examples.

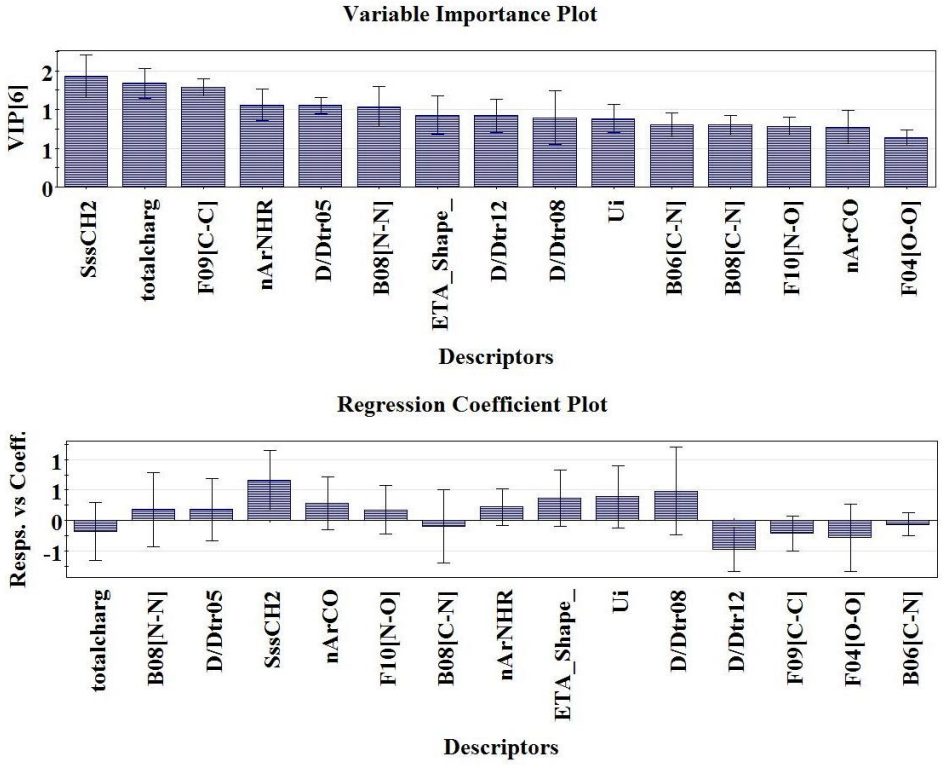


Figure 4.79. Variable importance plot (VIP) and regression coefficient plot of final PLS model against AChE enzyme.

According to the VIP plot, SssCH2 is the most important descriptor contributing positively to the response. It is an atom-type E-state index encoding information about the sum of ssCH2 count in the compounds. Sharma *et al.*[399] suggested that the descriptor represents the electro-topological state for the number of –CH_2- groups. The positive contribution suggests that the AChE enzyme inhibitory activity may be increased by an increase in the number of such -CH_2- groups in the molecules. A higher number of –CH2– groups (not more than the descriptor SssCH2 values 40) in compounds leads to better inhibitory activity against the AChE enzyme as longer chain compounds with a higher number of -CH2 groups would be more lipophilic resulting in improved brain permeability. Compounds like **597** and **841** (**Figure 4.80**) have a higher number of -CH_2- groups in their structure, showing higher descriptor values (32.05 and 24.37), leading to their higher range of AChE enzyme inhibitory activity (-1.418 and -0.365 respectively). On the other hand, compounds 884 and **896** showed AChE enzyme inhibitory activity in very a lower range (-5.505 and -5.361 respectively), due to the absence of such a group in the compounds. The extended topochemical atom (ETA) descriptor, ETA_Shape_P, represents the effect of branching in the cationic structure (one central atom is attached to three other non-hydrogen atoms) on the inhibitory activity against the AChE enzyme. This descriptor particularly denotes the branching where one central atom is attached to three other non-hydrogen atoms, making a Y-shaped structural fragment[305]. The positive contribution of this descriptor indicates that the inhibitory activity of compounds is directly proportional to the numerical value of ETA_Shape_P. Therefore, the compounds with the higher numerical value of this descriptor may enhance the AChE enzyme inhibitory activity as shown in (**Figure 4.80**) compounds like **735** (pIC_{50}: -1.929) and **740** (pIC_{50}: -1.875) and their corresponding descriptor values are 0.367 and 0.314, respectively. In contrast, compounds like **379** (pIC_{50}: -5.414) and **723** (pIC_{50}: -5) have no such fragment that shows lower inhibitory activity (**Figure 4.80**).

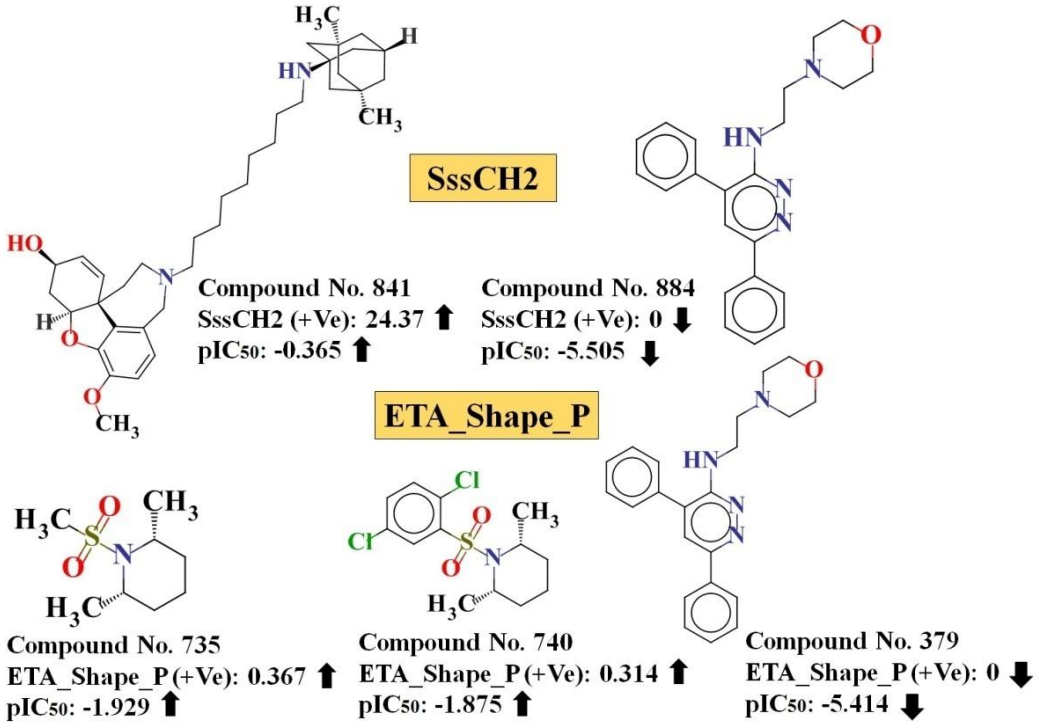
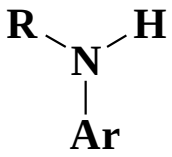


Figure 4.80. Impact of SssCH2 and ETA_Shape_P descriptors on AChE enzyme inhibitory activity.

The functional group count descriptor, nArNHR**,** simply represents the higher number of secondary aromatic amines (not more than descriptor values 4) in the compounds. The fragment contributes significantly to increasing the intermolecular interactions by strong H-bonds (discussed later in the docking part). Here we have represented the general structure of secondary aromatic amine (**1**) to allow the beginners a proper understanding of the important structural fragment responsible for the inhibition.



(1)

As per the regression coefficient plot, this descriptor contributed positively toward AChE enzyme inhibitory activity. Thus, the compounds with a higher number of this fragment (secondary aromatic amines) may enhance the AChE inhibitory activity as shown in compounds **3** (pIC_{50}: 0.698) and **18** (pIC_{50}: 1.154) (containing descriptor values 2) (**Figure 4.81**). On the other hand, the compounds containing no such fragments have lower inhibitory activity as observed in compounds **675** (pIC_{50}: -7.683) and **681** (pIC_{50}: -7.152).

The descriptor Ui belongs to the class of molecular properties of an unsaturation index. The positive contribution (as per the regression coefficient plot) of this descriptor implies that it has a positive impact on the AChE enzyme inhibitory activity. Therefore, the compounds bearing higher unsaturation index may have enhanced inhibitory activity as presented in compounds (**Figure 4.81**) **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698) (containing descriptor value 3.584), while the compounds with a lower unsaturation index may have lower AChE enzyme inhibitory activity as displayed in compounds (**Figure 4.81**) **848** (pIC_{50}: -5.414).

The 2D atom pair descriptor, B08[N-N], signifies the presence/absence of N-N at the topological distance 8. As per the regression coefficient plot, this descriptor positively influences the inhibitory activity against the AChE enzyme. Therefore, the compounds containing a higher number of N-N fragments at the topological distance 8 may have higher AChE enzyme inhibitory activity as shown by compounds **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698) (their corresponding descriptors values are 1) (**Figure 4.81**). Again, the compounds with no such fragments show lower inhibitory activity as shown in compounds **675** (pIC_{50}: -7.683), and **681** (pIC_{50}: -7.152).

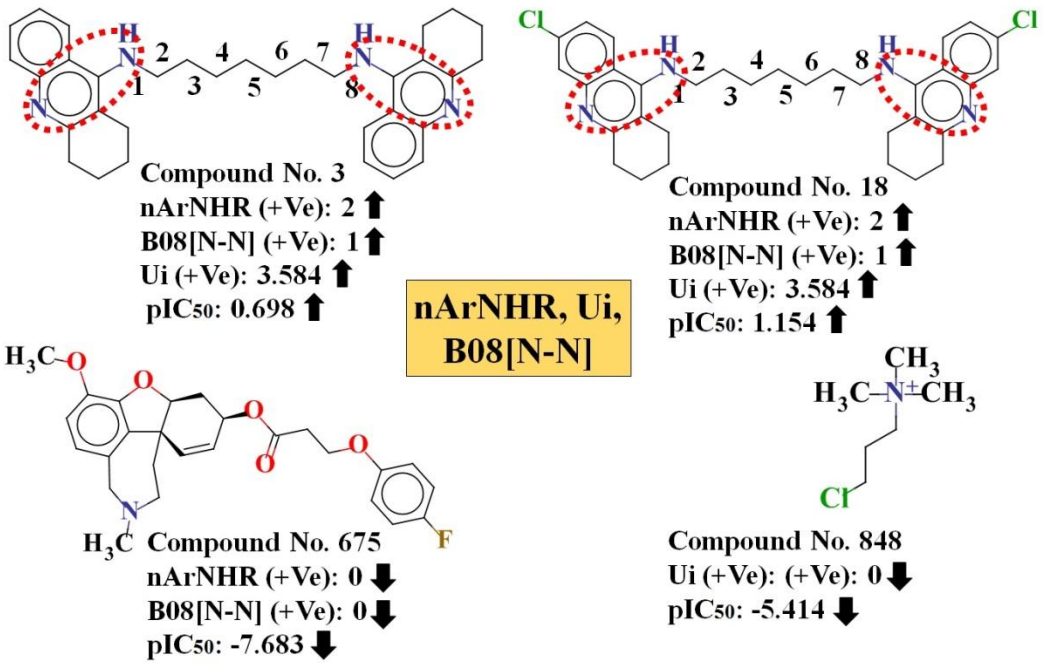


Figure 4.81. Impact of nArNHR, Ui and B08[N-N] descriptors on AChE enzyme inhibitory activity.

2D atom pair descriptor, F09[C-C]**,** represents the frequency of C - C at the topological distance 9. This descriptor negatively influences the inhibitory activity against the AChE enzyme as per the negative regression coefficient value. So, the higher number of this fragment correlates to lower AChE enzyme inhibitory activity as observed in compounds numbers **675** (pIC_{50}:-7.683), and **681** (pIC_{50}: -7.152) (**Figure 4.82**), while the absence of this feature correlates to the higher potency of AChE enzyme inhibitory activity as observed in compounds **747** (pIC_{50}: -1.785) and **740** (pIC_{50}: -1.875).

The descriptor, B06[C-N] belong to the family of 2D atom pair descriptor and defines the presence/absence of C-N at the topological distance 6. The negative contribution (as per the regression coefficient plot) of this descriptor suggests that the descriptor is inversely related to the AChE inhibitory activity. The same has been observed in compounds (**Figure 4.82**) **675** (pIC_{50}: -7.683) and **681** (pIC_{50}: -7.152) (lower enzyme inhibitory activity as their corresponding numerical descriptor values are in the higher range), whereas the inverse phenomena have been observed in compounds (**Figure 4.82**) **747** (pIC_{50}: -1.785), **740** (pIC_{50}: -1.875) (increases in enzyme inhibitory activity as their corresponding numerical descriptor value is in the lower range).

Another 2D atom pair descriptor, B08[C-N], signifies the presence/absence of C-N at the topological distance 8. This descriptor negatively affects the activity of AChE enzyme inhibitors as indicated by its negative regression coefficient. So, the compounds with the higher number of C-N fragments at the topological distance 8 may have lower AChE enzyme inhibitory activity as evidenced by compounds **675** (pIC_{50}: -7.683), **681** (pIC_{50}: -7.152) (their corresponding descriptors values are 1). Again, the compounds with no such fragments show improved inhibitory activity as supported by compounds **747** (pIC_{50}: -1.785), and **740** (pIC_{50}: -1.875) (**Figure 4.82**).

The ring descriptors, D/Dtr05, D/Dtr08, and D/Dtr12, denote the distance/detour ring index of order 5, 8, and 12 respectively (size of the ring system) in the compounds. Among these descriptors, D/Dtr05 and D/Dtr08 contribute positively to the inhibitory activity, whereas the descriptor D/Dtr12 influences the inhibitory activity negatively against the AChE enzyme. As per the above information, it can be concluded that a lower size of the ring (size of the ring not more than ring index of order 8) may be more favorable for the inhibitory activity against the AChE enzyme instead of a larger size of the ring. The positive regression coefficients of the descriptors (D/Dtr05 and D/Dtr08) suggest that a higher numerical value of these descriptors leads to improved inhibitory activity as verified by the compounds (**Figure 4.82**) **947** (pIC_{50}: -1.633) and **717** (pIC_{50}: -1.278) in case of D/Dtr05 and **841** (pIC_{50}: -0.365) and **837** (pIC_{50}: -1.907) in case of D/Dtr08. Again, the compounds with no such fragments show lower AChE enzyme inhibitory activity as found in the compounds, such as **123** (pIC_{50}: -6.143) and **871** (pIC_{50}: -6) (D/Dtr05 and D/Dtr08). The descriptor D/Dtr12 negatively influences the inhibitory activity as suggested by the regression coefficient plot. The negative contribution indicates the higher number of this fragment in the compounds shows the lower inhibitory activity as supported by compounds (**Figure 4.83**) **675** (pIC_{50}: -7.683) and **681** (pIC_{50}: -7.152). On the other hand, the absence of this feature in the compounds shows higher AChE enzyme inhibitory activity as observed in the case of compounds (**Figure 4.83**) **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698).

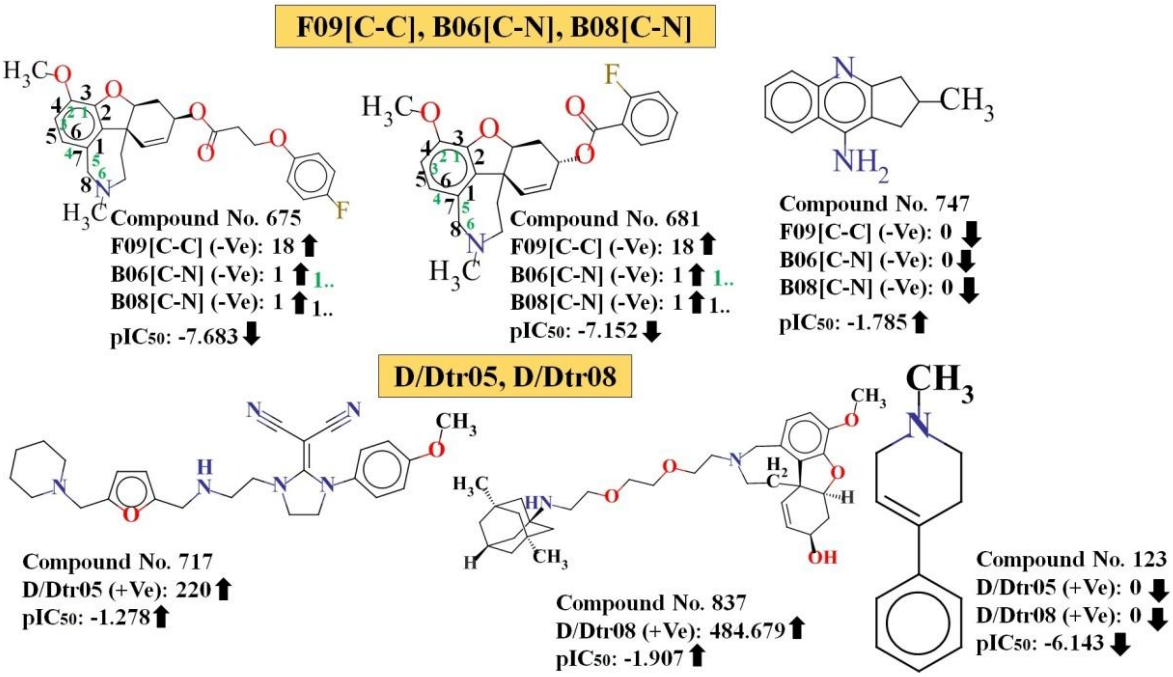


Figure 4.82. Impact of F09[C-C], B06[C-N], B08[C-N], D/Dtr05 and D/Dtr08 descriptors on AChE enzyme inhibitory activity.

Totalcharge is a constitutional index that negatively influences the inhibitory activity against the AChE enzyme. In the dataset, there are only 44 compounds with non-zero values for this descriptor, with a descriptor value of 2 in the case of compound number **872,** and 1 in the rest of all 43 compounds (compound numbers **685-687, 847-849, 851-858, 860, 862-866, 869-871, 873-877, 879, 880, 882-884, 888, 890, 892, 893, 897, 898, 900, 934, 937** and **938**). The descriptor encoded the information about the total charges

present in the compounds. The descriptor negatively influences the inhibitory activity and suggested that the presence of this feature in the compounds correlates to low inhibitory activity as shown in compounds **872** (pIC_{50}: -5.079) and **871** (pIC_{50}: -6) (containing descriptor values 2 and 1 respectively) (**Figure 4.83**). On the other hand, the uncharged compounds have higher inhibitory activity as shown in compounds **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698).

The last descriptor in this model, F04[O-O] belongs to the family of 2D atom pair descriptors that defines the frequency of two oxygen atoms at the topological distance 4. As per the regression coefficient plot, the descriptor contributes negatively to the inhibitory activity against the AChE enzyme. So, the compounds with such fragments express the lower inhibitory activity as proved by the (**Figure 4.83**) compounds **902** (pIC_{50}: -4.799) and **133** (pIC_{50}: -5.423) (their corresponding descriptor values are 6 and 5 respectively), whereas compounds with no such fragments show higher AChE enzyme inhibitory activity as shown in compounds **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698) (**Figure 4.83**).

2D atom pair descriptor, F10[N-O], defines the frequency of N-O at the topological distance 10. As per the regression coefficient plot, the descriptor contributed positively toward the inhibitory activity against the AChE enzyme. Thus, the evidence obtained from this descriptor suggests that the molecules containing the N - O fragment at the topological distance 10 show higher AChE enzyme inhibitory activity as shown in compounds (**Figure 4.83**) **107** (pIC_{50}: -0.692) and **719** (pIC_{50}: -0.342) (containing descriptor value 2 respectively), while compounds **675** (pIC_{50}: -7.683), **681** (pIC_{50}: -7.152) show lower inhibitory activity due to the absence of this fragment (**Figure 4.83**).

The functional group count descriptor, nArCO, denotes the number of aromatic ketone groups present in molecules. The positive regression coefficient of this descriptor suggests that the ketone group attached with an aromatic ring (not less than 2 fragments) is favorable for AChE enzyme inhibitory activity, as found in the case of compounds (**Figure 4.83**) **653** (pIC_{50}: -0.619) and **1131** (pIC_{50}: -0.619) and vice versa founds in case of compounds **675** (pIC_{50}: -7.683) and **681** (pIC_{50}: -7.152).

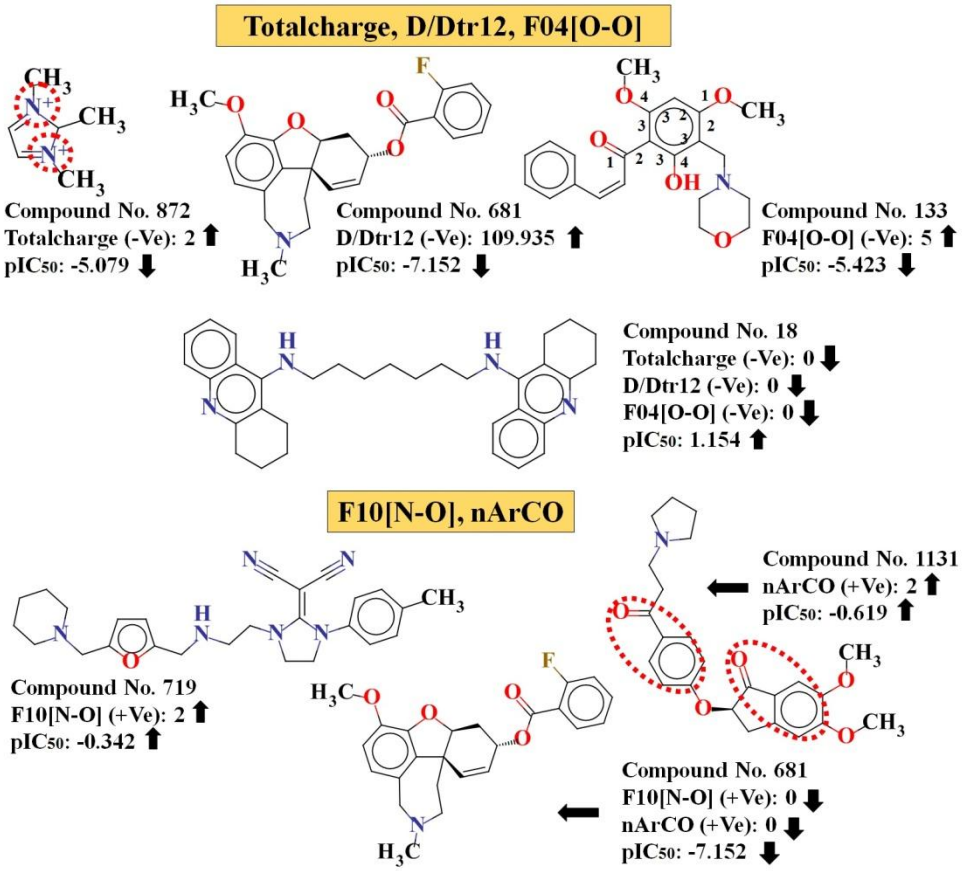


Figure 4.83. Impact of Totalcharge, D/Dtr12, F04[O-O], F10[N-O], and nArCO descriptors on AChE enzyme inhibitory activity.

4.5.1.1.2. 2D-QSAR analysis against BuChE enzyme: Model 2

Box 4.3. 2D QSAR model and statistical validation parameters obtained from the developed model against BuChE enzyme.

$$pIC_{50} = -4.13393 + 0.617 \times C-028 - 0.059 \times F05[C-O] + 0.018 \times T(N..N) - 0.219 \times B06[O-O] - 0.931 \times B03[N-N] + 0.041 \times F09[C-O] - 2.192 \times nThiazoles + 0.276 \times B09[C-O] + 0.727 \times C-041 - 0.981 \times C-019 + 0.374 \times nArCONHR + 0.194 \times B07[N-N] - 0.097 \times H-053$$

Equation 4.11

Internal Validation Parameters: n_{train}= 603, r^2= 0.674, Q^2= 0.656, Average rm^2= 0.54, EL = 13, LV = 5.

External Validation Parameters: n_{test}= 158, Q^2F_1= 0.663, Q^2F_2 = 0.660, Average rm^2= 0.499,

Equation 4.11 corresponds to the best PLS regression-based 2D-QSAR model that comprises 13 descriptors with corresponding latent variables 5 against the BuChE enzyme. The descriptors appearing in the model are organized accordingly to their significance and then defined individually. The VIP and regression coefficient plot defines the significance level of each variable found from the final PLS model that is responsible to regulate the BuChE enzyme inhibitory activity as presented in **Figure 4.84** in the regression coefficient plot.

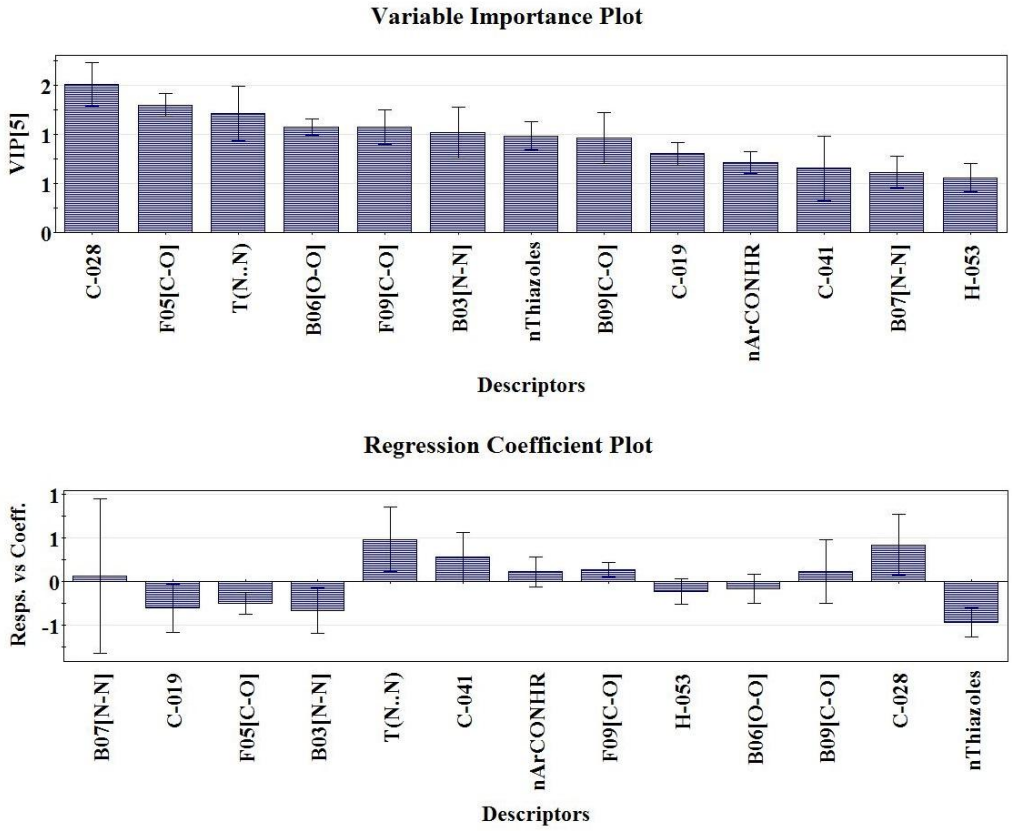


Figure 4.84. Variable importance plot (VIP) and regression coefficient plot of final PLS model against BuChE enzyme.

The next 2D atom pair descriptor, F09[C-O], describes the frequency of C-O at the topological distance 9. The positive contribution (as per regression coefficient plot) of this descriptor points out that the number of the C-O group at the topological distance 9 may favor the inhibitory activity against BuChE enzyme as found in compounds (**Figure 4.85**) **513** (pIC_{50}: -0.863) and **536** (pIC_{50}: -1.778) (containing descriptor values 6 and 7, respectively) and opposite found in case of compounds (**Figure 4.85**) **624** (pIC_{50}: -5.605) and **224** (pIC_{50}: -5.430) (absence of such fragment).

The next atom-centered fragment descriptor**,** C-041, exemplifies the fragment (X-C(=X)-X) indicating the number of fragments containing C(sp2) atoms that are attached with two electronegative atoms (O, N, S, Se, and halogens), i.e., one by a single bond and another by a double bond [10]. The positive regression coefficient suggests the influential effect of the feature containing C(sp2) atoms directly attached to two electronegative atoms toward the BuChE inhibitory activity. This is witnessed by the compounds (**Figure**

4.85) **748** (pIC_{50}: -1.462) and **750** (pIC_{50}: -1.531) (descriptor value 1 in both cases), and the opposite is seen in compounds **624** (pIC_{50}: -5.605) and **772** (pIC_{50}: -5.594) as depicted in **Figure 4.85**.

The functional group count descriptor, nArCONHR, stands for the number of secondary amides (aromatic) in the compounds. The fragment contributes significantly to increasing the intermolecular interactions by strong H-bonds (discussed later in the docking part). The descriptor contributes positively to the BuChE inhibitory activity as specified by the positive regression coefficient. Thus, the molecules bearing such fragments may have enhanced BuChE inhibitory activity as presented in compounds (**Figure 4.85**) **391** (pIC_{50}: -0.578) and **388** (pIC_{50}: -0.892) (containing descriptor values 1). Although the compounds containing no such fragments have lower inhibitory activity as shown in compounds (**Figure 4.85**) **624** (pIC_{50}: -5.605) and **772** (pIC_{50}: -5.594).

The 2D atom pair descriptor, B06[O-O] represents the presence/absence of two oxygen atoms at the topological distance 6. The negative contribution of this descriptor indicates that the presence of two oxygen atoms at the topological distance 6 may be detrimental to BuChE enzyme inhibitory activity. This is evidenced by compounds (**Figure 4.85**) such as **772** (pIC_{50}: -5.594) and **621** (pIC_{50}: -5.593) (with descriptor value 1 in each case), while an absence of this fragment in the compounds leads to a higher inhibitory activity as observed in compounds (**Figure 4.85**) **13** (pIC_{50}: -0.397) and **547** (pIC_{50}: -0.428).

The functional group count descriptor, nThiazoles, designates the number of thiazole ring present in the molecules. As per the regression coefficient plot, this descriptor contributed negatively to the BuChE enzyme inhibitory activity. The information obtained from the regression coefficient plot suggests that the presence of such fragments in the compounds is inversely proportional to the BuChE enzyme inhibitory activity as witnessed by the compounds (**Figure 4.85**) **135** (pIC_{50}: -5.202) and **137** (pIC_{50}: -5.073) (both compounds have descriptor value 1), even though the absence of such ring system in the compounds (**Figure 4.85**) leads to an improved inhibitory activity as detected in the compounds **13** (pIC_{50}: -0.397) and **547** (pIC_{50}:-0.428).

The next atom-centered fragment descriptor, C-019, simply refers to CRX, where R represents any group connected through carbon atom; X signifies any heteroatom (O, N, S, P, Se, and halogens) [10]. This descriptor negatively influences the inhibitory activity of BuChE enzyme inhibitors as specified by its negative regression coefficient, which indicates that this feature does not enhance the BuChE enzyme inhibitory activity of molecules as found in compounds (**Figure 4.85**) **66** (pIC_{50}: -4.100) and **793** (pIC_{50}: -5.096) (containing descriptor values 2 in both cases). But, the compounds with no such fragment have higher inhibitory activity as shown in compounds (**Figure 4.85**) **13** (pIC_{50}: -0.397) and **547** (pIC_{50}: -0.428).

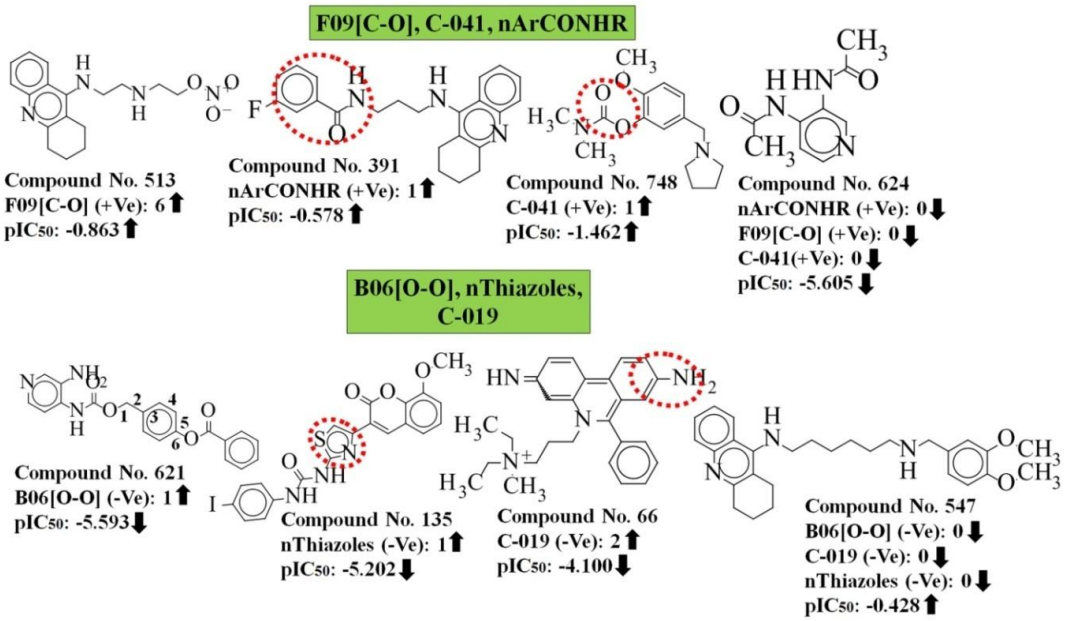


Figure 4.85. Contribution of F09[C-O], C-041, nArCONHR, B06[O-O], nThiazoles, and C-019 descriptors for the inhibition of the BuChE enzyme.

The most contributing descriptor, C-028, an atom-centered fragment descriptor, describes each atom by its atom type and the bond types and atom types of its first neighbors (R--CR-X)[305]. In this situation, R--CR-X can be defined as a central carbon atom (C) on an aromatic ring that has one carbon neighbor (R) and one heteroatom neighbor (X) on the same aromatic ring and the third neighbor outside this ring is a carbon (R)[305]. The positive contribution of this descriptor to the BuChE inhibitors indicates that by increasing the number of heteroatoms (with R--CR-X format) in compounds, the value of this descriptor increases, increasing its activity values. This has been noticed in compounds (**Figure 4.86**) **13** (pIC_{50}:-0.397) and **547** (pIC_{50}: -0.428) having corresponding descriptor value 2 in each case showing higher inhibitory activity, whereas, in the case of compounds (**Figure 4.86**) like **624** (pIC_{50}: -5.594) and **772** (pIC_{50}: -5.605), the absence of such fragments in the compounds shows lower inhibitory activity. The 2D atom pair descriptor, B09[C-O], stands for the presence/absence of C-O at the topological distance 9. According to the regression coefficient plot, this feature positively affects the activity of the BuChE enzyme inhibitors. So, the compounds with the higher number of C-O fragments at the topological distance 9 may have improved BuChE enzyme inhibitory activity as verified by (**Figure 4.86**) compounds **13** (pIC_{50}: -0.397) and **547** (pIC_{50}: -0.428) (their corresponding descriptors values are 1). But, the compounds with no such fragment show lower inhibitory activity as revealed in compounds **624** (pIC_{50}: -5.605) and **224** (pIC_{50}: -5.430). The 2D atom pair descriptor in this model, B07[N-N], describes the presence/absence of two nitrogen atoms at the topological distance 7. As per the regression coefficient plot, the descriptor positively influences the inhibitory activity against the BuChE enzyme. This phenomenon is well observed in compounds (**Figure 4.86**) **547** (pIC_{50}: -0.428) and **555** (pIC_{50}: -0.715), and the reverse is seen in the case of compounds (**Figure 4.86**) **624** (pIC_{50}: -5.605) and **772** (pIC_{50}: -5.594) (no such fragment at topological distance 7). The descriptor, B03[N-N] belongs to the family of 2D atom pairs and describes the presence/absence of two nitrogen atoms at the topological distance 3. As per the regression coefficient plot, the descriptor contributed negatively to the inhibitory activity against BuChE enzyme, suggesting that compounds

containing a lower number or absence of such fragments have good inhibitory activity against BuChE enzyme as presented in compounds (**Figure 4.86**) **13** (pIC_{50}: -0.397) and **547** (pIC_{50}: -0.428) (absence of such fragment), while a higher number of this fragment shows the lower inhibitory activity as detected in compounds (**Figure 4.86**) **624** (pIC_{50}: -5.605) and **772** (pIC_{50}: -5.594).

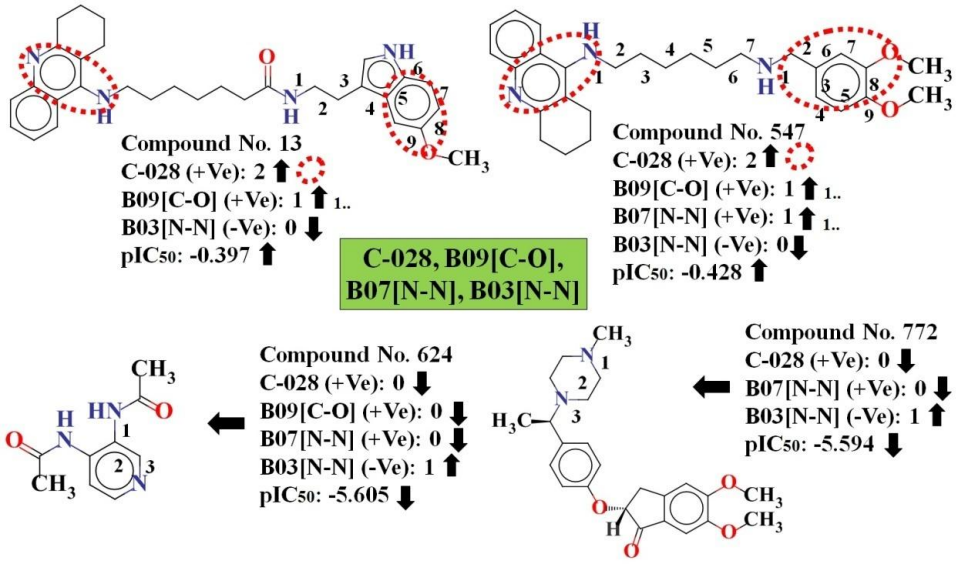


Figure 4.86. Contribution of C-028, B09[C-O], B07[N-N], and B03[N-N] descriptors for the inhibition of the BuChE enzyme.

The next important descriptor, F05[C-O], belongs to the family of 2D atom pair descriptor that defines the frequency of C and O atoms at the topological distance 5. As per the regression coefficient plot, the descriptor negatively influenced the inhibitory activity against the BuChE enzyme. Therefore, the compounds having such fragment show lower values of inhibitory activity as proved by (**Figure 4.87**) compounds **74** (pIC_{50}: -4.212) and **75** (pIC_{50}: -4.394) (their corresponding descriptor values are 21 and 19 respectively), whereas the absence of such fragment in the compounds show higher BuChE enzyme inhibitory activity as presented in compounds **466** (pIC_{50}: -0.929) and **2** (pIC_{50}: -1) (**Figure 4.87**). The descriptor, T(N..N), stands for the 2D atom pair descriptor, and simply characterizes the sum of topological distances between two nitrogen atoms. This descriptor contributes positively to the BuChE inhibitory activity as suggested by the positive regression coefficient. Thus, the molecules bearing higher topological distance between two nitrogen atoms may have higher BuChE inhibitory activity as presented in (**Figure 4.87**) compounds **8** (pIC_{50}: -0.908) and **466** (pIC_{50}: -0.929) (containing descriptor values 59 and 124 respectively), whereas in contrary, compounds **770** (pIC_{50}: -5.539) and **779** (pIC_{50}: -5.459) which do not contain any such fragment showed less BuChE inhibitory activity. From this observation, it can be concluded that the topological distances between two nitrogen atoms should be higher for improved inhibitory activity against the BuChE enzyme. The last atom-centered fragment descriptor in this model, H-053, simply refers to H atoms attached to $C^0(sp^3)$ with 2X connected to the next C, where X signifies any heteroatom (O, N, S, P, Se, and halogens) and the superscript characterizes the formal oxidation number[69]. The formal oxidation number of a carbon atom equals the sum of the conventional bond orders with electronegative atoms[69]. This descriptor is defined as the number of definite atom types in a compound and

can be calculated by knowing only molecular composition and atom connectivity[69]. The negative regression coefficient of this descriptor suggested that compounds containing a larger number of such hydrogen atoms have lower inhibitory activity against BuChE enzyme as shown in compounds (**Figure 4.87**) **280** (pIC_{50}: -4.989) and **296** (pIC_{50}: -4.935), and their corresponding descriptor values are 6 in both cases. On the other hand, compounds **13** (pIC_{50}: -0.397) and **597** (pIC_{50}: -0.428) show higher inhibitory activity because of the absence of such an H atom.

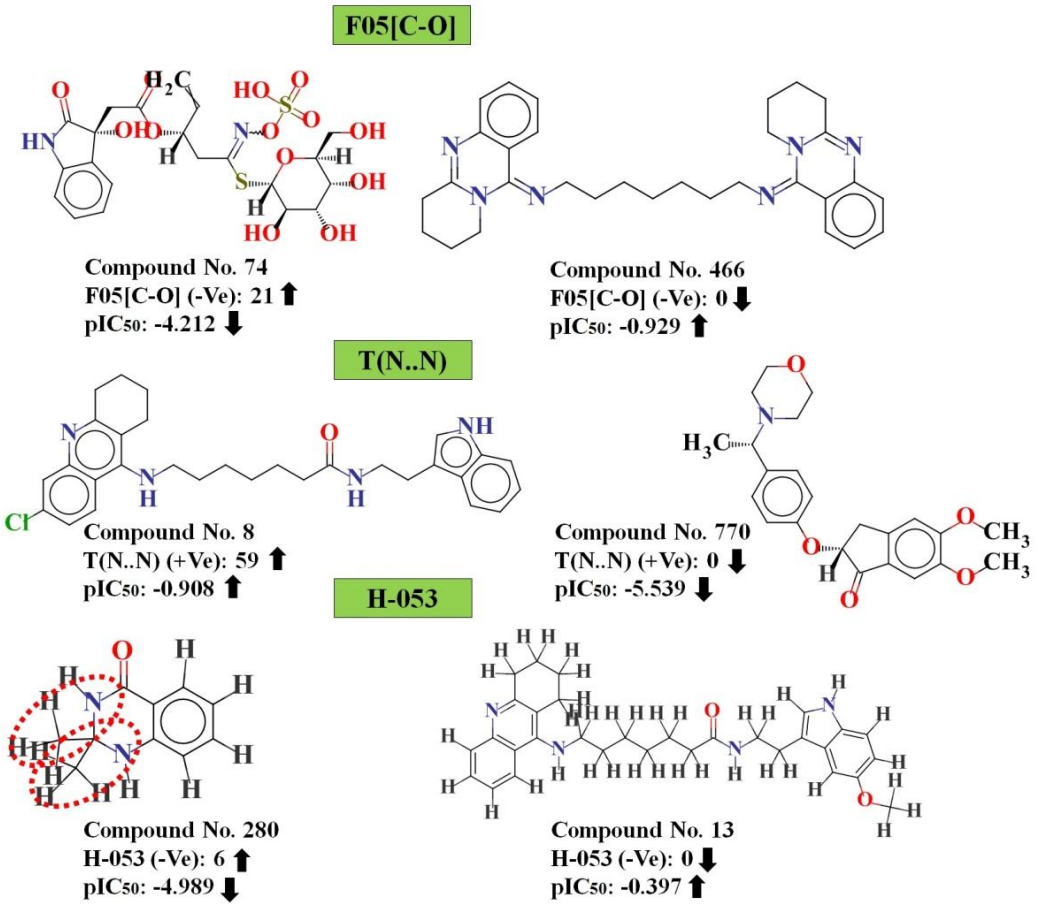


Figure 4.87. Contribution of F05[C-O], T(N..N) and H-053 descriptors for the inhibition of BuChE enzyme.

4.5.1.1.3. 2D-QSAR analysis based on the selectivity of AChE and BuChE enzyme inhibitors: Model 3

Box 4.4. 2D QSAR model and statistical validation parameters obtained from the developed model based on selectivity **(AChEI-BuChEI)**.

Selectivity[pIC_{50}]= $-0.891 - 0.720 \times F05[N-O] + 79.839 \times ETA\_Shape\_X - 2.607 \times nCq + 0.505 \times nArOR + 5.599 \times ETA\_dBetaP + 2.962 \times B10[C-C]$ **Equation 4.12**

Internal Validation Parameters: n_{train}= 159, r^2= 0.679, Q^2= 0.650, Average rm^2= 0.524, EL = 6, LV = 5, **External Validation Parameters:** n_{test}= 39, Q^2F_1= 0.787, Q^2F_2= 0.785, Average rm^2= 0.686, Δrm^2= 0.176,

In the brain, two varieties of cholinesterase enzymes (AChE and BuChE) are capable of hydrolyzing acetylcholine neurotransmitters[400]. Throughout the development of AD, BuChE activity increases by 40–90% in the temporal cortex and hippocampus, while at the same time AChE activity decreases[400]. Therefore, the concurrent inhibition of both AChE and BuChE should provide extra benefits in the treatment of AD. In this research, PLS regression-based 2D-QSAR model was developed based on selectivity for both AChE and BuChE (difference of pIC_{50} values between them). The descriptors found in the developed model are organized as per the VIP plot (**see Figure 4.88**) and then defined individually. The regression coefficient plot[111] describes the contribution of each descriptor in the model for the inhibition of both cholinesterase enzymes (**see Figure 4.88**).

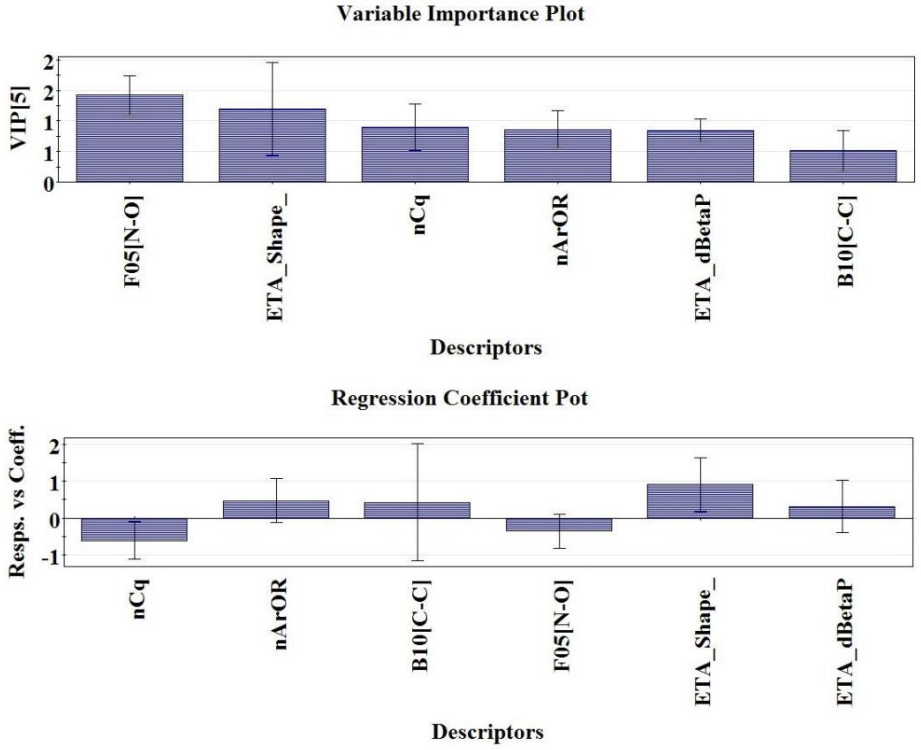


Figure 4.88. Variable importance plot (VIP) and regression coefficient plot of final PLS selectivity-based model.

The next most significant descriptor, ETA_Shape_X, simply refers to the shape index X[369]. We can find out the effect of molecular shape on the inhibitory activity against cholinesterase enzymes with the help of this descriptor. The positive regression coefficient of the descriptor suggests that the above feature of the compounds is more specific for AChE enzyme inhibitory activity as compared to the BuChE enzyme inhibitory activity. The higher numerical value of this descriptor correlates to higher AChE enzyme inhibitory activity as observed in the case of compounds (**Figure 4.89**) **31** (pIC_{50}: 2.986) and **104** (pIC_{50}: 2.98) and opposite in the case of compounds (**Figure 4.89**) **57** (pIC_{50}: -2.238) and **58** (pIC_{50}: -1.995). The extended topochemical atom (ETA) descriptor, ETA_dBetaP, demonstrates the measure of the unsaturation content relative to molecular size[369]. The positive regression coefficient of this descriptor proposes that the highest unsaturation content related to their molecular size in the compounds is more specific for AChE enzyme inhibitory activity. It has been found that with an increase in the numerical value of this descriptor, the inhibitory activity of the compound also increases against AChE enzyme as presented in the case of compounds (**Figure 4.89**) **203** (pIC_{50}: 3.212) and **11** (pIC_{50}: 4.325), while the lower numerical value indicated the less specificity against the AChE enzyme as found in case of compounds (**Figure 4.89**) **160** (pIC_{50}: -1.122) and **158** (pIC_{50}: -1.477). Thus, we can conclude from the above information that unsaturation content related to their molecular size is more specific for AChE enzyme inhibitory activity compared to the BuChE enzyme inhibitory activity.

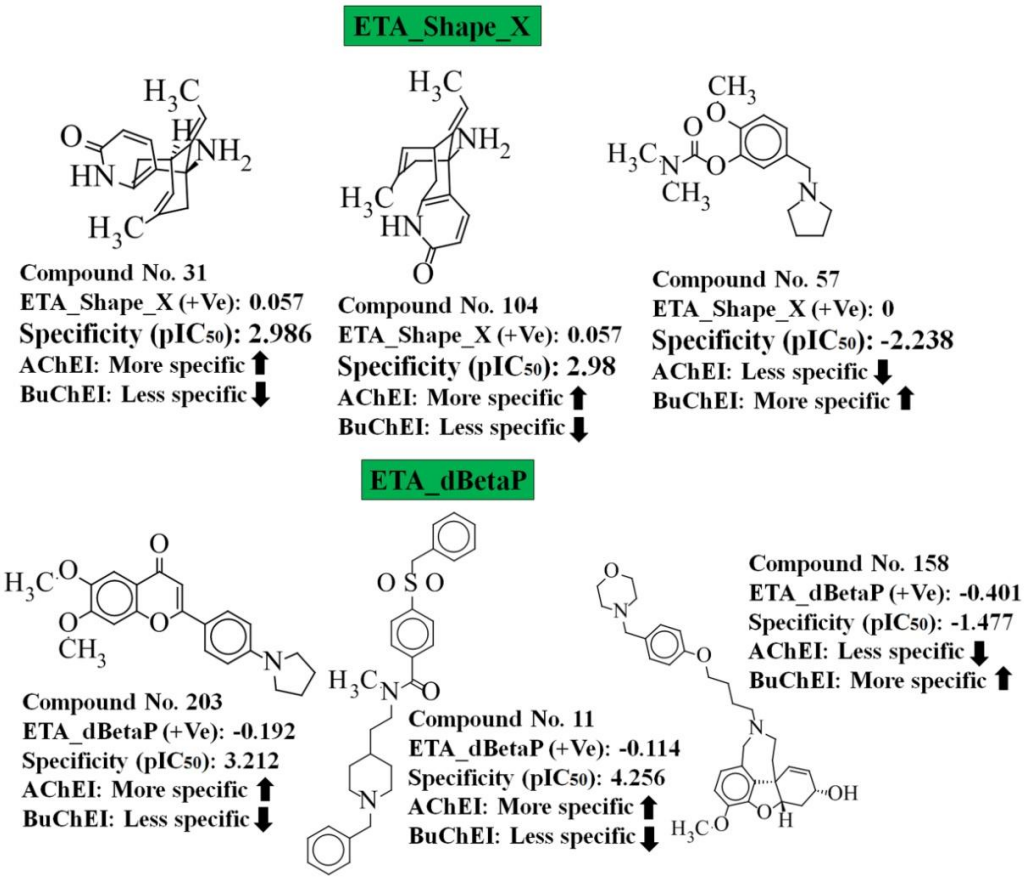


Figure 4.89. Contribution of ETA_Shape_X and ETA_dBetaP descriptors for the dual inhibition of AChE and BuChE enzyme.

The descriptor, F05 [N-O], belonging to the class of 2D atom pairs, indicates the frequency of N-O at the topological distance 5. The negative contribution (as per the regression coefficient plot) of this descriptor suggested that the frequency of N-O at the topological distance 5 is more specific to BuChE enzyme inhibitory activity than AChE enzyme inhibitory activity. Thus, the higher number of this fragment correlates with lower AChE enzyme inhibitory activity as noticed in the case of compounds (**Figure 4.90**) **160** (pIC_{50}: -1.122) and **57** (pIC_{50}: -2.238). On the other hand, compounds having no such fragments show better AChE enzyme inhibitory activity values as observed in the compounds (**Figure 4.90**) **11** (pIC_{50}: 4.256) and **44** (pIC_{50}: 4.041).

The functional group count descriptor, nCq, denotes the number of total quaternary carbons (sp3) present in the molecules. As per the regression coefficient plot, the descriptor negatively influences the inhibitory activity and suggests that the presence of the number of quaternary carbon in the compounds is more specific to the BuChE enzyme inhibitory activity compared to the AChE enzyme inhibitory activity. From the descriptor contribution, it can be suggested that molecules containing this fragment may not be favorable for the AChE enzyme inhibitory activity as presented in compound numbers (**Figure 4.90**) **164** (pIC_{50}: -1.852) and **158** (pIC_{50}: -1.477). The reverse is seen in compounds (**Figure 4.90**) **11** (pIC_{50}: 4.256) and **44** (pIC_{50}: 4.041).

The descriptor, nArOR, belongs to the family of functional group count descriptors and designates the number of aromatic ether groups present in compounds. The positive regression coefficient of this descriptor advises that the ether group attached with an aromatic ring is more specific and favorable for the AChE enzyme inhibitory activity, as found in the case of compound numbers (**Figure 4.90**) **97** (pIC_{50}: 3.974) and **88** (pIC_{50}: 3.960), whereas, the lack of such fragment in the compounds leads to a decrease in the AChE enzyme inhibitory activity as shown in compound number (**Figure 4.90**) **2** (pIC_{50}: -1.249) and **204** (pIC_{50}: -1.175). Thus, it can be concluded that the fragment is more specific for AChE enzyme inhibitory activity compared to the BuChE enzyme inhibitory activity.

The last 2D atom pair descriptor in this model, B10[C-C], describes the presence/absence of two carbon atoms at the topological distance 10. The positive contribution (as per the regression coefficient plot) of this descriptor specifies that the presence of two carbon fragments at the topological distance 10 is more specific for AChE enzyme inhibitory activity than the BuChE enzyme inhibitory activity. This phenomenon is well noticed in compounds (**Figure 4.90**) **11** (pIC_{50}: 4.256) and **44** (pIC_{50}: 4.041) and the reverse is seen in the case of compounds (**Figure 4.90**) **195** (pIC_{50}: -1.570) and **2** (pIC_{50}: -1.249) (no such fragment at the topological distance 10).

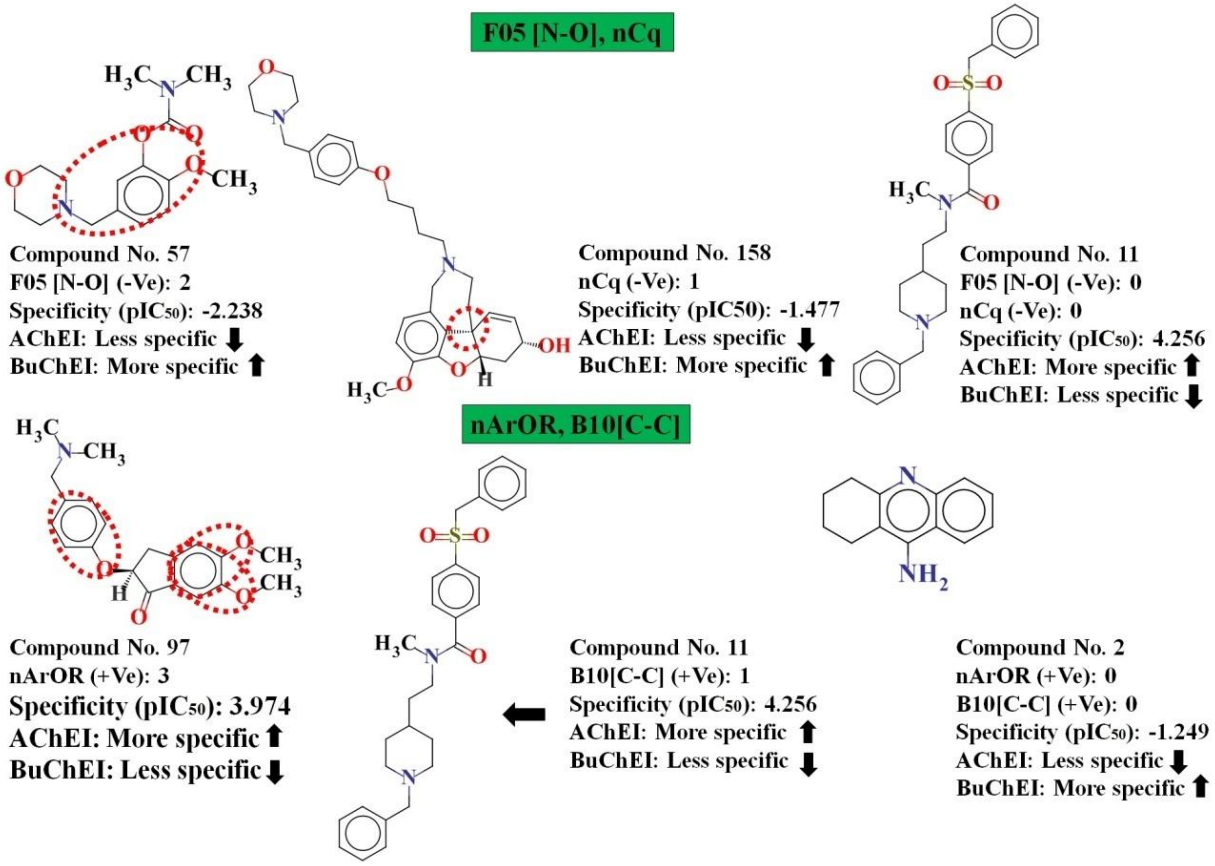


Figure 4.90. Contribution of F05 [N-O], nCq, nArOR, and B10[C-C] descriptors for the inhibition of both AChE and BuChE enzymes.

4.5.1.2. Applicability domain criteria

The applicability domain for the test set compounds was checked at a 99% confidence level by applying the DModX (distance to model in X-space) approach available in SIMCA-P 10.0 software (available from https://umetrics.com/products/simca) as described by the Roy *et al.* and Khan *et al.*[394, 352]. **Figure 4.91** revealed that there is only one compound (compound **666**) in the test set found outside the DModX value (D-Crit=1.623) in the case of **Model 1** (AChEI model). In the case of the reported **Model 2** (see **Figure 4.91**), there was only one compound (compound **138**) in the test set traced outside the critical DModX value (D-Crit=1.67), whereas in case of **Model 3** (selectivity-based model) (see **Figure 4.91**), there was also only one compound (Compound **30**) in the test set is found outside the DModX value (D-Crit=3.066).

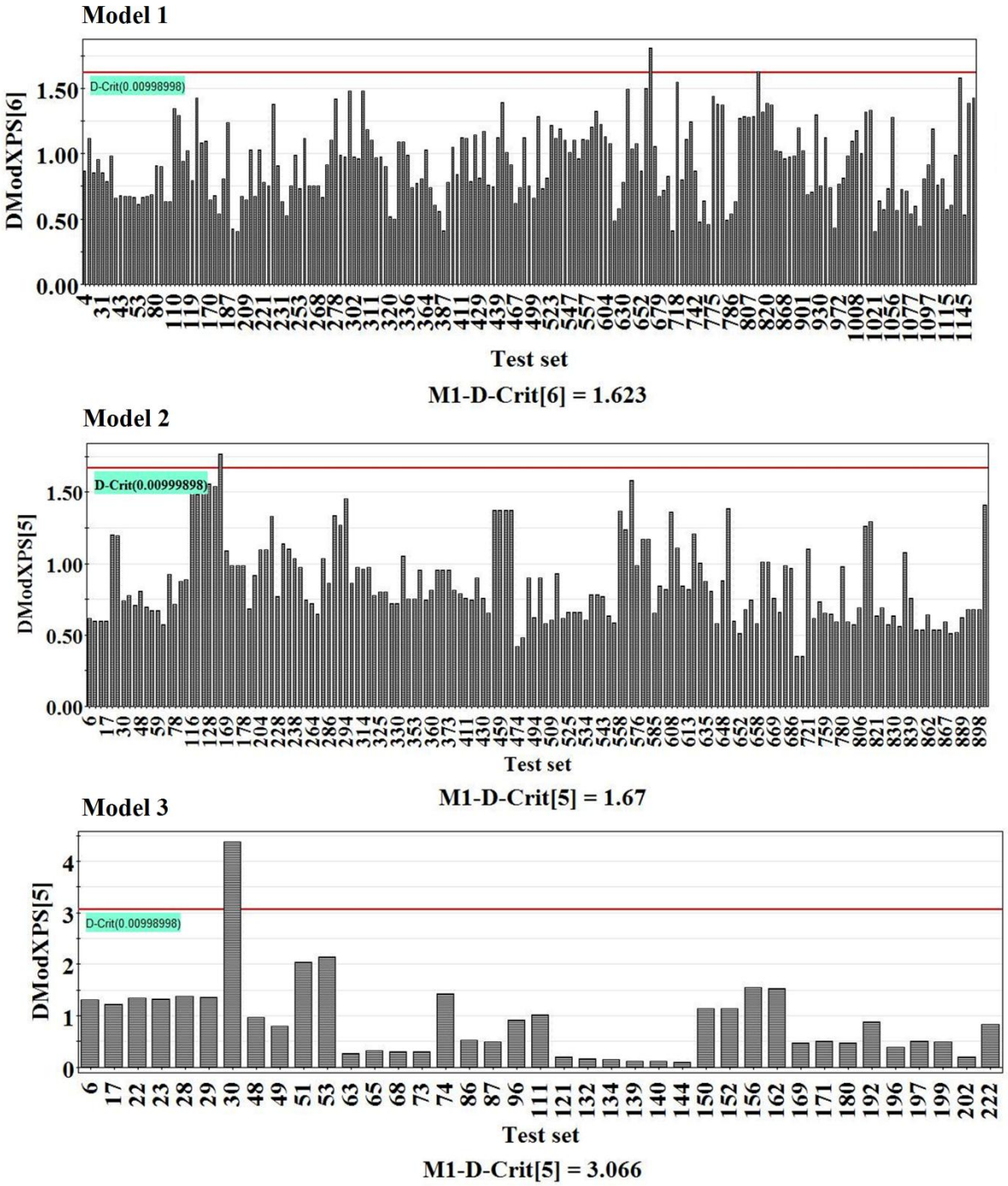


Figure 4.91. Applicability domain DModX values of the test set compounds at 99% confidence level of the developed PLS models: Model 1 = AChEI, Model 2 = BuChEI, and Model 3 = based on selectivity.

4.5.2. Molecular docking

In the present work, molecular docking studies were performed using the most and least active compounds from the two datasets, i.e., AChE and BuChE enzyme inhibitors. In the case of AChE enzyme inhibitors, we have selected the two most active compounds, i.e., **15** and **19,** and the two least active compounds **123** and **674**, from the dataset; in the case of BuChE inhibitors dataset, we have selected two most active compounds, such as **13** and **547** and least active compounds **621** and **624** to understand the docking interactions with the active site of enzymes. The details of docking interactions and their relation with 2D-QSAR results are depicted in **Table 4.19**. Here, we have discussed the details of docking interaction as well as their analysis below.

4.5.*2.1. Molecular docking analysis of the selected compounds from the AChE enzyme inhibitors dataset*

In this analysis, the two most active compounds (**15** and **19**) from the dataset (pIC_{50} =0.221 and0.522 respectively) linked with the active site amino acid residues (LEU A:289, TRP A:286, SER A: 293, TYR A: 124, TRP A: 86, HIS A:447, GLU A:292, GLN A: 291, ASN A: 87, TYR A: 72, TYR A:337 and TYR A: 341) through interacting forces, such as hydrogen bonding (conventional and carbon-hydrogen bonds), π-bonding (π-donor hydrogen bond, π-π stacked, π-alkyl, π-cation, π-π T-shaped), alkyl, and halogen. One of the most active compounds form the dataset, compound **15** (see **Figure 4.92.**), interacts with the active site cavity through Hydrogen bonding (TYR A:337, ASN A:87, SER A:293, GLU A:292, and GLN A:291), π-donor hydrogen bonding (TYR A:124), Halogen (fluorine) (TRP A: 86 and TRP A: 87), π-π stacked (TRP A:286, TRP A:86), π-π T-shaped (TYR A:124), Alkyl (LEU A:289) and π-Alkyl (TRP A: 86, HIS A:447, TYR A:72, TRP A:286, and LEU A:289). Another most active compound **19** of the dataset interacted with amino acid residues, such as TYR A: 341, SER A: 293 and TYR A: 337 (hydrogen bond), TYR A: 124 (π-donor hydrogen bond), TRP A: 286 (π-cation), TRP A: 286 and TRP A: 86 (π- π stacked), PRO A: 88 and LEU A: 289 (alkyl bond), TRP A: 86 and HISA: 447 (π-alkyl bonding) and ASN A: 87 (halogen bond) (see **Figure 4.92.**).

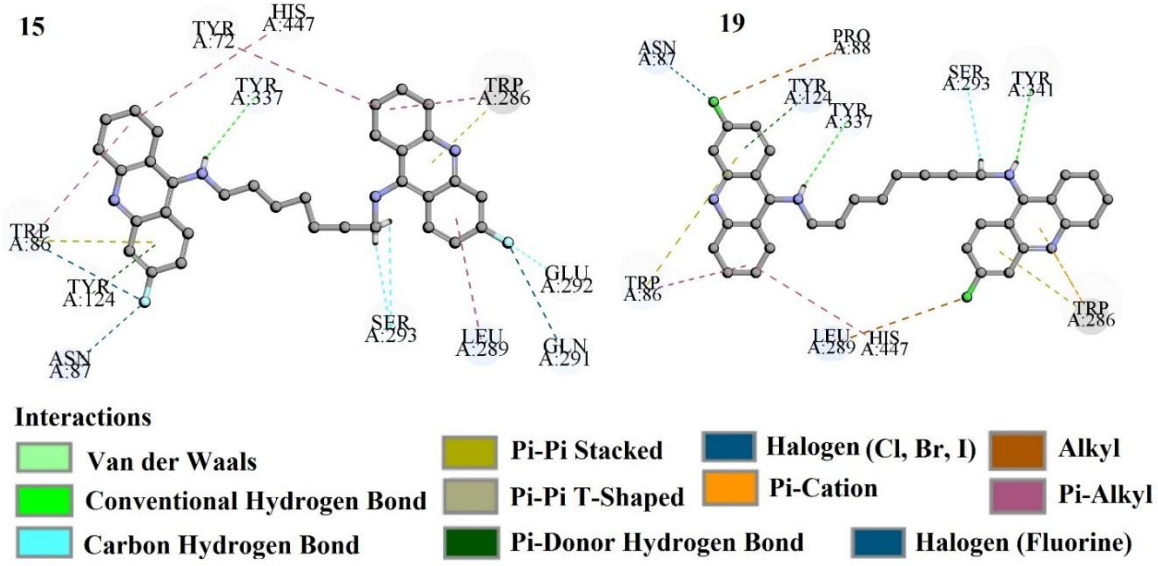


Figure 4.92. Docking interactions of most active compounds (**15 and 19**) from the dataset of AChE enzyme inhibitors.

In this investigation, two least active compounds (**123** and **674**) from the dataset (pIC_{50} = -6.143 and -6.495 respectively) linked with the active site amino acid residues (ARG A:296, TYR A:341, SER A:293, TRP A: 286, HIS A:287, PHE A: 338, TYR A: 337,) through interacting forces, such as Hydrogen bonding (conventional and carbon-hydrogen bonds) and π-bonding (π-π stacked, π-alkyl). One of the least active compounds of the dataset, compound **123** (**Figure 4.93**) interacts with amino acid residues like ARG A: 296 and SER A: 293 through hydrogen bonding and TYR A: 341 via π- π stacked bonding. **Figure 4.93** displays that compound **674** interacts with amino acid residues, such as HIS A: 287, SER A: 293 (through hydrogen bonding) and TRP A: 286, TYR A: 337 (via π-alkyl bond).

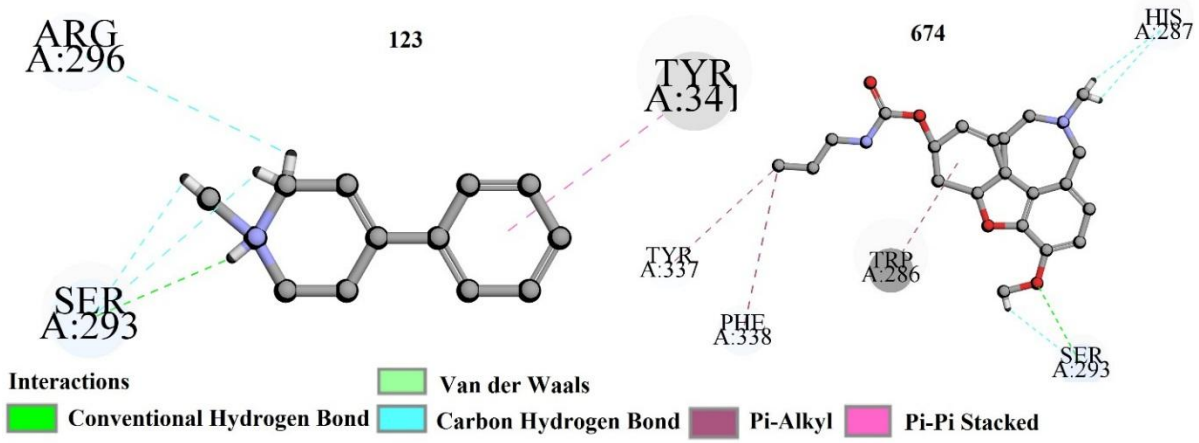


Figure 4.93. Docking interactions of least active compounds (**123 and 674**) from dataset of AChE enzyme inhibitors.

4.5.*2.2. Molecular docking analysis of the selected compounds from BuChE enzyme inhibitors dataset*

In this work, the two most active compounds (**13** and **547**) from the dataset (pIC_{50} = -0.397 and -0.428 respectively) are linked with the active site amino acid residues, like THR A: 120, GLU A: 197, TRP A: 82, HIS A: 438, TYR A: 440, GLY A: 116, PHE A: 329, SER A: 287, SER A: 79, ASP A: 70, TYR A: 332, SER A: 198 and ALA A: 328, through interacting forces such as hydrogen bonding (conventional and carbon-hydrogen bonds), π-bonding (π-π stacked, π-alkyl, amide π-stacked, π-cation, π-π T-shaped), alkyl, and attractive charges. **Figure 4.94** shows that compound **547** interacts with amino acid residues, such as SER A: 79, ASP A: 70, TYR A: 332, HIS A: 438, SER A: 198 (hydrogen bonding), PHE A: 329, TRP A; 82 (π-π T-shaped), ASP A: 70, GLU A: 197 (attractive charges), TYR A: 332 (π-π stacked, π-cation), TRP A; 82, ALA A: 328 (π-alkyl), GLY A: 116 (Amide π-stacked). One of the most active compounds from the dataset, compound **13** (**Figure 4.94**), interacts with amino acid residues, THR A: 120, GLU A: 197 TRP A: 82, HIS A: 438, TYR A: 440, SER A: 287 via hydrogen bonds, PHE A: 329 through π-π T-shaped, TRP A: 82 via π-π stacked, GLY A: 116 Amide π-stacked.

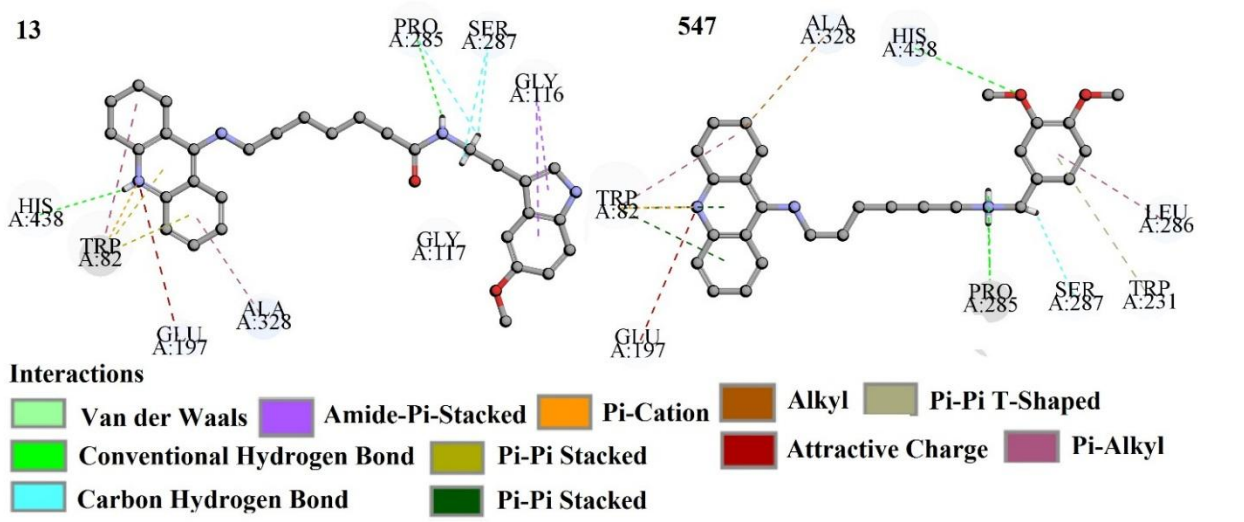


Figure 4.94. Docking interactions of most active compounds (**13 and 547**) from the dataset of BuChE enzyme inhibitors.

In this analysis, we have selected the two least active compounds (**621** and **624**) from the dataset (pIC_{50} = -5.593 and -5.605 respectively) which interacted with the active site amino acid residues, such as SER A: 79, ASP A: 70, GLN A: 71, TYR A: 332, GLU A: 197 and TRP A: 82, through interacting forces, such as Hydrogen bonding (conventional, carbon-hydrogen bonds and π-donor hydrogen bond), π-bonding (π-π stacked, π-cation) and salt bridge. One of the least active compounds from the dataset, compound **624** (**Figure 4.95**), interacts with amino acid residues such as SER A: 79, ASP A: 70, and TRP A: 82 through hydrogen bonding and TRP A: 82 via the π-π stacked bond. Another least active compound of the dataset, compound **621** (**Figure 4.95)**, interacts with amino acid residues such as GLU A: 197, GLN A: 71 (via hydrogen bonding), TYR A: 332, TRP A: 82 (π-π stacked), TRP A: 82 (through π-cation) and GLU A: 197 (via salt bridge).

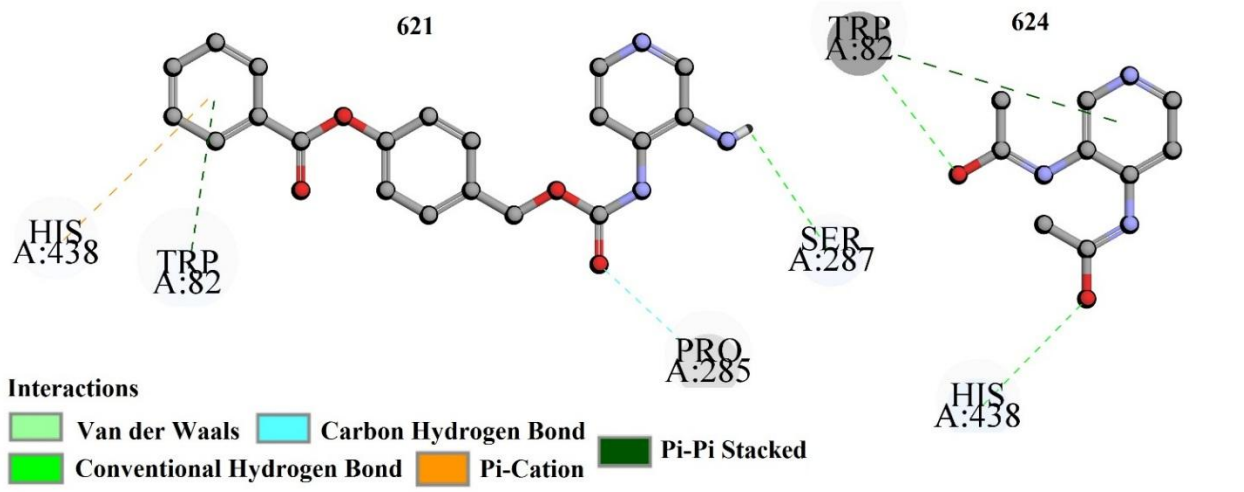


Figure 4.95. Docking interactions of least active compounds (**621 and 624**) from the dataset of BuChE enzyme inhibitors.

Table 4.19. Docking results and correlation with 2D-QSAR models in this study.

S. No.	Compound Number	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
AChE enzyme inhibitors					
1	**15** (high pIC_{50})	70.674	TYR A:337, ASN A:87, SER A:293, GLU A:292, GLN A:291, TYR A:124, TRP A: 87, TRP A:286, TRP A:86, LEU A:289, HIS A:447, TYR A:72 and LEU A:289	Hydrogen bonding, π-donor hydrogen bonding, Halogen (fluorine), π-π stacked, π-π T-shaped, Alkyl and π-Alkyl	SssCH2, nArNHR, ETA_Shape_P and Ui,
2	**19** (high pIC_{50})	80	TYR A: 341, SER A: 293, TYR A: 337, TYR A: 124, TRP A: 286, TRP A: 86, PRO A: 88, LEU A: 289, HISA: 447 and ASN A: 87	Hydrogen bond, π-donor hydrogen bond, π-cation, π- π stacked, alkyl, π-alkyl and Halogen (fluorine)	SssCH2, nArNHR, ETA_Shape_P and Ui
3	**123** (low pIC_{50})	31.26	ARG A: 296, SER A: 293 and TYR A: 341	Hydrogen bonding and π- π stacked	B06[C-N]
4	**674** (low pIC_{50})	43	HIS A: 287, SER A: 293, TRP A: 286 and TYR A: 337	Hydrogen bonding and π-alkyl	D/Dtr12, F09[C-C], F04[O-O] and B06[C-N]
BuChE enzyme inhibitors					
1	**13** (high pIC_{50})	63.977	THR A: 120, GLU A: 197, TRP A: 82, HIS A: 438, TYR A: 440, GLY A: 116, PHE A: 329, SER A: 287	Hydrogen bonding, π-π stacked, Amide π-stacked and π-π T-shaped	F09[C-O], B09[C-O] and C-028
2	**547** (high pIC_{50})	72.494	SER A: 79, ASP A: 70, TYR A: 332, HIS A: 438, TRP A: 82, SER A: 198, GLY A: 116, GLU A: 197, PHE A: 329, ALA A: 328	Hydrogen bonding, attractive charges, π-cation, π-π stacked, π-π T shaped, Amide π-stacked, alkyl and π-alkyl	B07[N-N], T(N..N), F09[C-O], B09[C-O] and C-028

3	**624** (low pIC_{50})	20.491	SER A: 79, ASP A: 70 and TRP A: 82	Hydrogen bonding, π-π stacked and π-donor hydrogen bond	F05[C-O], B03[N-N]
4	**621** (low pIC_{50})	38.333	GLN A: 71, TYR A: 332, GLU A: 197 and TRP A: 82	Hydrogen bonding, π-cation, π-donor hydrogen bond and π-π stacked bond	F05[C-O], B03[N-N], B06[O-O]

4.5.*2.3. Correlation with the developed 2D-QSAR models*

From the above investigation, we have concluded that hydrogen bonding and π-interaction among the ligand and receptor play important roles in the interactions. Hydrogen bonding may associate through the descriptor nArNHR (against AChE enzyme inhibitor model), C-028, F09[C-O], B09[C-O], B07[N-N] and T(N..N) (against BuChE enzyme inhibitors model) of the developed 2D-QSAR models. Descriptors nArNHR, ETA_Shape_P, and Ui (against AChE enzyme), C-028, B07[N-N], and T(N..N) (against BuChE enzyme) are well corroborated with interactions made via π- interactions (π-π stacked, π-cation, π-alkyl, and π-π T-shaped) between the protein and ligand. The above-mentioned features are observed in compounds in the case of AChE enzyme inhibitors such as **15** and **19** (most active) (**Figure 4.92**) and the case of BuChE enzyme inhibitors, such as **547** and **13** (most active) (**Figure 4.94**). But in contrast, the descriptors B06[C-N], D/Dtr12, F09[C-C], F04[O-O] and B06[C-N] (against AChE enzyme), F05[C-O], B03[N-N], B06[O-O] (against BuChE enzyme) contributed negatively in the 2D-QSAR model and this has been observed in case of AChE enzyme inhibitors such as **123**, **674**, **621** and **624** (least active) respectively (see **Figure 4.93, and Figure 4.95**). Thus, from the above investigation, we can conclude that features obtained from molecular docking studies and 2D-QSAR models are in agreement and essential for the inhibitory activity against both AChE and BuChE enzymes.

4.5.*3.* Comparisons of the performance of the reported models with previously published models

In this investigation, a comparison of the best models of this study with previously published models (Shrivastava et al. 2019[401], Bukhari et al 2014[402], De Souza et al. 2012[104] and Pang et al 2017 et al.[183]) for the prediction of the bioactivity against AChE and BuChE enzymes was performed, as depicted in the **Table 4.20**. The details of different internal and external validation parameters obtained from our models and those obtained from previous models are given in **Table 4.20**. Based on the statistical quality in terms of both internal and external validation criteria, the models reported in this work are statistically significant and robust enough as compared to the previously reported models (**Table 4.20**). Moreover, the models presented in this study are derived from a larger set of molecules than those reported in the previous studies.

Table 4.20. Comparisons of the proposed study with previously published studies against AChE and BuChE enzymes.

Sources	E. L.	LV	Model	Training set			Test set	
				n	R^2	Q^2	n	Q^2F_1
AChE (Present work)	**15**	**6**	**PLS**	**798**	**0.662**	**0.645**	**199**	**0.661**
BuChE (Present work)	**13**	**5**	**PLS**	**603**	**0.674**	**0.656**	**158**	**0.663**
Selectivity (Present work)	**6**	**5**	**PLS**	**159**	**0.679**	**0.650**	**39**	**0.787**
Shrivastava et al. 2019[401]	-	-	PLS	26	0.792	0.713	6	0.542
Bukhari et al 2014[402]	2	-	GA-MLR	14	0.855	0.792	3	0.771
De Souza et al. 2012[104]	2	-	HQSAR	29	0.965	0.787	7	-
De Souza et al. 2012[104]	2	-	HQSAR	29	0.952	0.904	7	-
Pang et al 2017 et al.[183]	-	-	3D-QSAR	35	0.925	-	10	0.850
Pang et al 2017 et al.[183]	-	-	3D-QSAR	35	0.883	-	10	0.881

Abbreviations: LV= Latent variables, E.L. = Equation length, PLS = Partial least square, GA-MLR= Genetic algorithm multiple linear regression, 3D-QSAR = Three dimensional quantitative structure activity relationship and HQSAR= Hologram QSAR.

4.6. Study 6- Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease

The current research aimed at developing statistically significant 2D-QSAR models against 12 major targets with easily interpretable descriptors and using them to check the applicability domain of four chemical drug-like databases (ZINC[12], Asinex, NCI, and InterBioscreen databases) and providing a prioritized set of compounds for experimental detection of their performance as anti-Alzheimer's drugs. The interpretations from the 2D-QSAR models were further confirmed by molecular docking strategies. The current work comprises five phases: (1) development of a well-validated 2D-QSAR model (individual, selectivity-based, and QSAAR) against 12 major targets; (2) chemical Read-Across analysis; (3) prediction of inhibitory activity of four chemical compound databases against 12 major targets, using the developed 2D-QSAR models; (5) molecular docking of the most and least active compounds of the modeled datasets.

4.6.1. 2D QSAR analysis

In this analysis, PLS-based 2D-QSAR models against 12 major targets (AChE enzyme, BuChE enzyme, BACE1 enzyme, β-amyloid aggregation, 5-HT6, CDK-5, Gamma-secretase enzyme, Glutaminyl Cyclase enzyme, GSK-3β enzyme, MAO-B enzyme, NMDA receptor and phosphodiester enzyme (PDE 10A)) were developed to search for novel anti-Alzheimer's agents and identify important structural features responsible for inhibiting the enzymes involved in AD. Additionally, 10 Selectivity-based models (6-PLS-based and 4-MLR-based models) and 17 QSAAR-based models (15-PLS-based and 2-MLR-based models) were also developed for the identification of features with dual inhibitory activity. The details about the developed models and their various validation metrics are given in **Tables 4.21, 22,** and **23**. The statistical results indicated that all the models were acceptable in terms of stability, predictive ability, and fitness. The obtained features define the structural and functional requirements for compounds to improve their inhibitory activity against the respective enzymes. The scatter plot (**Figs. 4.96-102**) describes the closeness of the observed and predicted values for the modeled enzyme inhibitors. Moreover, we have also performed the Y-randomization test using the SIMCA-P 10.0 software and MLRPlusValidation1.3 tool by randomly reshuffling (100 permutations) the dependent variable to ensure that the developed models were not obtained by any chance. The R^2 and Q^2 values for the random models (Y-axis) are plotted against the correlation coefficient between the original Y values and the permuted Y values (X-axis). The statistical results obtained from randomized models suggested that the developed models were not found by any chance. The variable importance plot (VIP) and regression coefficient plot, were used to determine the significance and contribution of the obtained features in the models toward the inhibitory activity.

Table 4.21. Individual 2D QSAR models and their statistical validation matrics were obtained from the developed models.

Target	Equation	LV	Training Set				Test Set			
			Train	R^2	$Q^2_{(LOO)}$	PQ	Test	Q^2F_1	Q^2F_2	PQ
5-HT6 antagonist	pIC_{50} (nM) $= -1.671 + 0.449 \times minssNH + 0.155 \times F04[C-O] + 0.465 \times nCp + 1.710 \times nArNH2 - 0.226 \times MaxTD + 0.393 \times nBnz$	5	60	0.800	0.742	B	20	0.786	0.784	M
AChE inhibitors	pIC_{50} (nM) $= -5.907 + 0.192 \times X2v + 0.30397 \times DBI + 0.56431 \times B07[N-N] + 0.00419 \times D/Dtr05 - 0.01895 \times NdsCH - 0.15889 \times C-016 - 0.28519 \times F04[O-O] - 0.25533 \times NsssN + 0.89374 \times F04[N-Cl] - 0.01257 \times T(N..Cl) + 1.44516 \times B05[C-N] - 1.08240 \times B01[N-N] - 3.08774 \times nROCON - 0.56172 \times B09[N-O] + 0.04808 \times minssO - 1.06153 \times totalcharge - 0.38974 \times CATS2D\_02\_AP - 0.57236 \times B02[N-N] + 1.05258 \times B04[O-S] - 0.07341 \times CATS2D\_08\_AL + 1.42829 \times nR\#CH/X - 0.41181 \times C-008 + 0.26843 \times CATS2D\_08\_AA - 1.01794 \times B06[O-S] + 1.95405 \times CATS2D\_01\_DD + 0.47598 \times CATS2D\_06\_AP - 1.58252 \times B09[O-F] + 0.36591 \times nArCO$	7	1325	0.635	0.621	-	408	0.678	0.678	M
BACE1 inhibitors	pIC_{50} (nM) $= -11.3326 + 0.2442 \times SAscore + 0.2420 \times F08[N-N] + 2.0777 \times Ui + 0.4070 \times X5v - 0.1594 \times F04[O-O] + 0.0720 \times mindO + 0.3921 \times nR\#CH/X - 0.4098 \times B02[O-O] - 0.0706 \times SaaCH - 0.3112 \times F04[F-F] - 0.9326 \times minaaaC - 0.5385 \times O-058 - 0.0108 \times D/Dtr04 + 0.5356 \times nR04 + 0.1804 \times B06[O-O] + 0.1890 \times F09[C-Cl] + 0.9600 \times nN = C-N < -0.4325 \times B05[N-Cl]$	9	680	0.669	0.650	-	225	0.675	0.674	M
β-amyloid inhibitors	pIC_{50} (nM) $= -9.3579 + 0.1369 \times SssCH2 - 0.2513 \times F02[N-N] - 0.1060 \times F05[C-O] + 20.6914 \times PW3 - 0.3449 \times MAXDN + 0.3534 \times F04[N-O]$	4	197	0.729	0.705	M	65	0.844	0.844	G
BuChE inhibitors	pIC_{50} (nM) $= -0.64834 + 0.10456 \times C-002 + 0.87920 \times N-070 - 0.02751 \times nCs - 0.18997 \times nArNHR + 0.55922 \times MaxaaaC + 1.07712 \times nArOCON - 0.60012 \times Psi\_e\_A - 0.08769 \times H-051 + 0.00608 \times MDEC-22 + 0.07105 \times F04[C-N] - 0.10931 \times MAXDP + 0.44817 \times B07[N-N] + 1.01229 \times N-077 + 1.95654 \times nArC = N-$	8	1882	0.689	0.668	M	625	0.702	0.702	M

	$2.11509 \times C-036-0.55208 \times B03[N-N]+1.64662 \times C-035+3.47045 \times nN-N+1.34851 \times BLI-0.32650 \times F02[N-N]+0.08134 \times SdsN+0.02974 \times minsF+0.76341 \times C-009+2.59201 \times nROCON-1.20613 \times C-037+0.24057 \times B04[O-O]-0.23948 \times nR09-0.19739 \times nCsp$									
CDK-5 inhibitors	pIC_{50} (nM) $= -2.82109+0.52376 \times NaaaC-0.33618 \times PBF-0.01011 \times T(S..S)+0.21105 \times SaasC-0.02307 \times T(N..F)-1.65453 \times S-106-0.37066 \times F07[N-S]$	4	169	0.675	0.652	-	56	0.790	0.790	G
Gamma-secretase inhibitors	pIC_{50} (nM) $= -15.4364+0.0155 \times P\_VSA\_MR\_7-1.1252 \times nOHp-0.6246 \times B05[N-N]-0.9356 \times nRCOOH-0.5702 \times CATS2D\_07\_NL-0.5050 \times F05[N-O]+2.3424 \times ATS6s-0.5725 \times B09[C-S]+0.4953 \times F08[O-S]+0.1089 \times VE3sign\_D+0.4441 \times B10[O-F]+0.1313 \times VE3sign\_D/Dt-1.4220 \times B07[O-Cl]-0.1858 \times F06[F-F]-1.9204 \times nS(=O)2+0.7022 \times C-029$	7	172	0.773	0.720	-	45	0.734	0.734	M
Glutaminyl Cyclase inhibitors	pIC_{50} (nM) $= -1.24512+1.87780 \times C-034-6.43315 \times ETA\_Shape\_Y-0.21722 \times F05[C-S]-0.43309 \times PBF+0.46170 \times B09[C-O]-0.24517 \times F02[N-N]+0.02157 \times T(N..S)$	5	99	0.944	0.934	G	33	0.956	0.956	G
GSK-3β inhibitors	pIC_{50} (nM) $= -5.65151+0.98713 \times nThiazoles+0.09410 \times SaasC+2.20185 \times PDI+0.08050 \times SaaaC-0.14490 \times mindsCH+0.57896 \times B03[O-Br]+0.21992 \times F06[N-O]-0.31342 \times B03[C-O]-0.00636 \times nPyrroles+0.09118 \times B05[N-O]$	5	118	0.703	0.648	-	41	0.763	0.763	M
MAO-B inhibitors	pIC_{50} (nM) $= -7.74797-0.48129 \times mindsCH+1.46202 \times F02[O-O]+0.09291 \times C\%+0.63918 \times nRCN-0.54786 \times B06[O-Cl]-0.10809 \times ALOGP-0.34911 \times F05[C-N]-1.36772 \times B01[C-O]+0.71091 \times B03[N-N]-0.39179 \times minssCH2+1.04117 \times SAscore-1.99700 \times ETA\_Eta\_B\_RC-0.94480 \times nArNHR$	6	125	0.722	0.649	-	45	0.639	0.639	-
NMDA receptor antagonist	pIC_{50} (nM) $= 0.81523-0.04386 \times SaaN-0.87822 \times MDEN-23-0.00678 \times TPSA(NO)-1.66542 \times C-043-0.04382 \times ATSC2m+0.00139 \times ATSC1p-0.81419 \times MATS4i+0.10333 \times C-028+0.00021 \times TPSA(Tot)-0.12093 \times MDEC-24-0.92317 \times GATS2m-0.39657 \times GATS8v-0.11764 \times F05[C-O]+0.19022 \times B10[C-O]$	7	267	0.740	0.708	G	89	0.640	0.639	G

Phosphodiester enzyme	pIC_{50} (nM) $= -7.87353 + 1.09932 \times N - 070 + 0.00916 \times AMR + 0.99679 \times minsCH3 - 1.22178 \times MDEN - 23 - 0.06459 \times MaxDD - 0.34705 \times nFuranes + 7.19158 \times Eta\_epsi\_2 - 1.74389 \times B03[N - F] - 1.36496 \times F09[N - N] - 0.57580 \times nArNR2 - 0.82639 \times nRCOOR + 1.07168 \times B09[N - N] - 1.49240 \times NNRS - 0.52332 \times mindssC - 1.39784 \times B06[O - F]$	9	222	0.722	0.677	M	67	0.739	0.730	M

Table 4.22. Selectivity-based models and their statistical validation matrics were obtained from the developed models.

Target (Selectivity)	Equation	LV	Training Set				Test Set			
			Train	R^2	$Q^2_{(LOO)}$	PQ	Test	Q^2F_1	Q^2F_2	PQ
AChEI - BACE1	Selectivity (AChEI-BACE1) $= -0.42092 + 1.61441 \times F08[N - O] - 1.82470 \times B05[N - O] + 0.45021 \times X2v - 0.56114 \times F10[O - O] + 0.90094 \times SsssCH - 0.10198 \times F05[C - C]$	5	34	0.827	0.744	G	9	0.627	0.624	G
AChEI -β-amyloid	Selectivity (AChEI-β-Amyloid) $= -2.06170 + 0.70243 \times B07[C - N] - 1.34266 \times nArNH2 + 1.01867 \times F06[C - Cl] + 0.30819 \times SAscore - 1.12566 \times F10[C - Cl] + 1.04712 \times ETA\_Shape\_Y - 0.10865 \times PJI2$	6	62	0.787	0.679	G	21	0.901	0.898	G
AChEI -BuChEI	Selectivity (AChEI-BuChEI) $= 3.15944 + 0.81697 \times F06[N - O] - 3.26082 \times B03[N - N] - 0.71013 \times B03[N - O] - 1.80289 \times nROR + 1.63240 \times nR12 - 0.22007 \times N\% - 8.65969 \times GD - 1.24833 \times B02[N - N] + 0.20227 \times nDB - 0.21524 \times ICR - 0.04797 \times PHI$	6	90	0.785	0.717	G	23	0.732	0.709	G
AChEI - MAO-B	Selectivity (AChEI-MAO-B) $= 5.01821 - 1.81468 \times SsssCH - 0.53685 \times O\% - 0.23837 \times F10[C - C] - 0.47260 \times F10[C - N] + 0.17398 \times F09[C - C] + 1.97941 \times MaxsCH3 - 2.36195 \times minsCH3$	6	42	0.778	0.651	-	10	0.725	0.723	M
BACE1 - GSK-3β	Selectivity (BACE1-GSK3B) $= 0.401 + 1.85 \times LipinskiFailures - 0.299 \times nRotB - 2.83 \times AATSC1s$	-	15	0.889	0.766	G	5	0.876	0.870	G
BuChEI - BACE1	Selectivity (BuChEI-BACE1) $= 0.260 + 0.877 \times nArCO - 3.90 \times nR = Ct + 0.176 \times NaasC + 2.51 \times minaaaC - 0.0159 \times ALOGP2 + 0.000332 \times ZM1Kup$	-	39	0.903	0.863	G	12	0.866	0.828	G

BuChEI - β-amyloid	Selectivity (BuChEI-β-amyloid) $= 3.69262 + 0.83031 \times B04[N-N] + 0.00558 \times D/Dtr10 - 0.46510 \times B10[N-O] - 3.56920 \times B10[C-N] \times 3.24186 \times B08[C-N] + 0.48410 \times B10[C-O] - 0.62086 \times B08[O-O] + 2.72960 \times ETA\_dEpsilon\_B - 0.30788 \times LOGP99 - 9.01720 \times ETA\_Shape\_Y - 2.20437 \times PJI2$	7	93	0.821	0.763	M	30	0.696	0.656	M
BuChEI - MAO-B	Selectivity (BuChEI-MAO-B) $= 2.92274 - 2.35759 \times SsssCH - 0.97195 \times B08[N-O] - 0.26699 \times NdssC + 0.72489 \times B06[N-O] - 1.99605 \times minsCH3 - 0.01919 \times SAacc + 0.94967 \times SsssN$	3	38	0.756	0.637	-	10	0.826	0.801	M
*AChEI – GSK-3β	Selectivity (AChEI-GSK3B) $= -17.8 - 113 \times ETA\_BetaP\_ns\_d + 10.7 \times Psi\_i\_A - 0.143 \times F02[C-O]$	-	18	0.766	0.664	M	3*	0.784	0.765	G
*BuChEI - GSK-3β	Selectivity (BuChEI-GSK3B) $= -6.54 - 2.39 \times ETA\_Beta\_ns\_d + 4.43 \times Psi\_i\_A + 0.867 \times nCL$	-	18	0.735	0.647	M	3*	0.770	0.752	G

* Models were developed by SmallDataModeler_betaVersion

Table 4.23. QSAAR models and their statistical validation matrics were obtained from the developed models.

Variable		Equation	LV	Training Set				Test Set			
Y	X			Train	R^2	$Q^2_{(LOO)}$	PQ	Test	Q^2F_1	Q^2F_2	PQ
AChEI	BACE1	pIC_{50} (nM)_AChEI $= -9.09333 + 0.41826 \times nCrs \times 1.26840 \times MaxaasC + 0.00358 \times log\_BACE1 + 9.75454 \times PW3$	3	33	0.896	0.843	G	10	0.746	0.744	G
BACE1	AChEI	pIC_{50} (nM)_ BACE1 $= -3.37330 + 0.27402 \times X5v - 0.23660 \times C - 001 - 0.20158 \times nCconj - 0.55996 \times nArOH + 0.97466 \times B08[O - O] + 0.02858 \times log\_AChEI$	4	33	0.766	0.637	-	10	0.772	0.749	M
AChEI	β-amyloid	pIC_{50} (nM)_AChEI $= -3.32884 + 0.85551 \times log\_Beta\ Amyloid + 0.34000 \times nArCO + 2.12019 \times B10[N - Cl] + 6.27610 \times PW3 + 0.70658 \times B07[C - N]$	3	63	0.962	0.949	G	20	0.968	0.968	G
β-amyloid	AChEI	pIC_{50} (nM)_ β-amyloid $= 3.16964 + 1.00643 \times log\_AChEI - 1.78556 \times ETA\_EtaP\_F\_L - 1.99421 \times B10[N - Cl] - 0.65971 \times B07[C - N] - 0.55678 \times SAscore$	4	63	0.947	0.924	G	20	0.972	0.972	G
AChEI	BuChEI	pIC_{50} (nM)_AChEI $= -5.82692 + 0.71516 \times NssNH + 0.40986 \times nR10 + 1.23241 \times ICR + 0.94089 \times B04[N - O] - 0.24740 \times F08[C - N] + 0.09780 \times log\_BuChEI + 0.61060 \times F06[N - O] - 1.26313 \times nROR + 1.04963 \times B04[N - Cl] + 0.51418 \times nCrt$	5	89	0.867	0.808	G	29	0.749	0.736	G
BuChEI	AChEI	pIC_{50} (nM)_ BuChEI $= -0.767384 + 0.126111 \times NaaCH - 0.753111 \times F06[N - O] - 0.883081 \times MaxsCH3 + 0.419744 \times DECC + 0.302945 \times log\_AChEI + 0.756970 \times B10[N - O] - 0.293615 \times ETA\_EtaP\_F - 0.124508 \times nCrs - 0.126071 \times Ram - 0.064782 \times nCb -$	6	89	0.789	0.724	-	29	0.767	0.756	M
AChEI	GSK-3β	pIC_{50} (nM)_ AChEI $= -8.9587 - 0.6585 \times C - 026 + 0.6051 \times log\_GSK3B + 15.8263 \times ETA\_Epsilon\_1$	2	14	0.847	0.761	-	4	0.837	0.824	G
GSK-3β	AChEI	pIC_{50} (nM)_ GSK-3β $= -4.25506 + 0.31626 \times log\_AChEI + 2.20735 \times MaxaasC$	1	14	0.891	0.858	-	4	0.943	0.939	G

AChEI	MAO-B	pIC_{50} (nM)_ AChEI $= 37.6420 + 0.4495 \times F01[C-N] - 0.7457 \times N\% + 1.0504\ MaxdssC - 27.2159 \times ETA\_Epsilon\_5 - 4.0017 \times MaxaaCH + 0.3758 \times B05[N-Cl] - 2.2041 \times MAXDP + 0.1345 \times log\_MAO - B$	6	42	0.793	0.677	-	10	0.755	0.638	M
BACE1	GSK-3β	pIC_{50} (nM)_ BACE1 $= -3.10715 + 0.11713 \times max\_conj\_path - 0.12209 \times minsOH + 0.62254 \times C - 025 + 0.81474 \times log\_GSK - 3\beta$	3	15	0.862	0.662	-	5	0.808	0.779	G
GSK-3β	BACE1	pIC_{50} (nM)_ GSK-3β $= -2.50148 - 2.03802 \times MATS2e + 0.14716 \times log\_BACE1 - 0.02185 \times ZMIC4 + 0.10199 \times MAXDN - 0.04842 \times AATS3s$	3	15	0.808	0.569	-	5	0.668	0.666	-
BACE1	BuChEI	pIC_{50} (nM)_ BACE1 $= -3.5140 + 0.58485 \times MaxsCH3 - 0.13063 \times F01[C-O] + 0.94806 \times nFuranes + 0.06244 \times log\_BuChEI + 1.37864 \times B05[O-O] - 0.07915 \times nCIC$	5	39	0.943	0.882	G	12	0.942	0.940	G
BuChEI	BACE1	pIC_{50} (nM)_ BuChEI $= -3.12300 - 1.57604 \times nFuranes + 9.30594 \times RFD - 0.00055 \times F08[C-C] + 1.97203 \times nR09 + 0.09518 \times log\_BACE1 - 2.62602 \times B04[O-O]$	-	39	0.956	0.857	G	12	0.942	0.941	G
β-amyloid	BuChEI	pIC_{50} (nM)_ β-amyloid $= -2.12671 - 4.83670 \times ETA\_EtaP\_F + 0.13649 \times log\_BuChEI + 1.32042 \times minssCH2 + 0.16273 \times SaasC + 0.09384 \times F08[C-C] - 0.11038 \times F06[C-C] + 1.01759 \times Uc$	4	93	0.805	0.771	M	30	0.776	0.776	M
BuChEI	β-amyloid	pIC_{50} (nM)_BuChEI $= -2.17 + 0.00370 \times D/Dtr10 - 0.0654 \times nCp + 1.45 \times nCrt + 1.73 \times nRNHR + 0.403 \times log\_Beta\ Amyloid - 0.0696 \times nAB$	-	93	0.879	0.837	G	30	0.910	0.910	G
GSK-3β	BuChEI	pIC_{50} (nM)_ GSK-3β $= -3.38300 + 0.22068 \times N\% + 0.14374 \times log\_BuChEI + 0.30867 \times minssCH2$	2	14	0.946	0.858	M	4	0.922	0.922	G
MAO-B	BuChEI	pIC_{50} (nM)_ MAO-B $= -7.47908 + 1.04311 \times nCt + 2.00343 \times minsCH3 + 0.61755 \times B08[N-O] + 0.24486 \times F06[C-O] + 0.06530 \times log\_BuChEI + 1.33039 \times nROH + 2.42118 \times nRNHR - 0.67456 \times ETA\_Beta\_ns\_d$	6	39	0.758	0.638	-	12	0.771	0.771	-

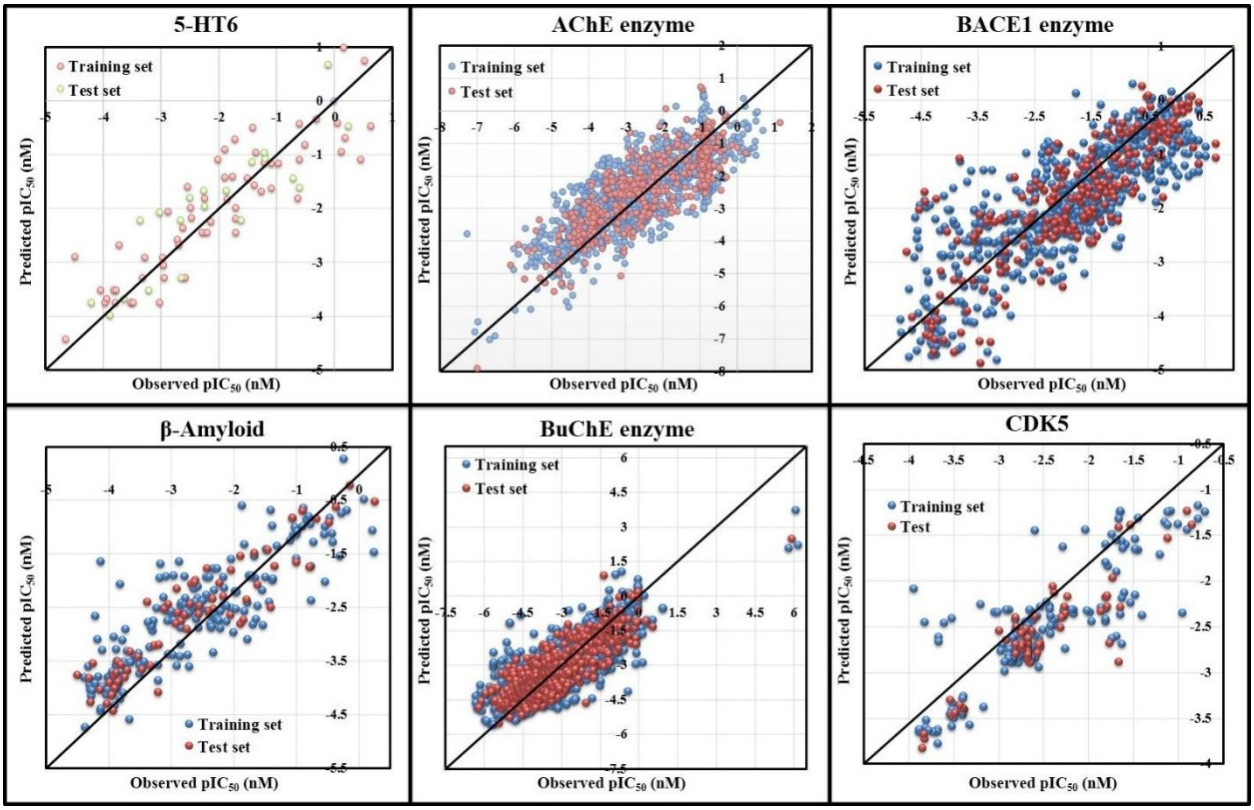


Figure 4.96. Scatter plots of observed v/s predicted values for 5-HT6, AChE, BACE1, β-amyloid, BuChE, and CDK-5 models.

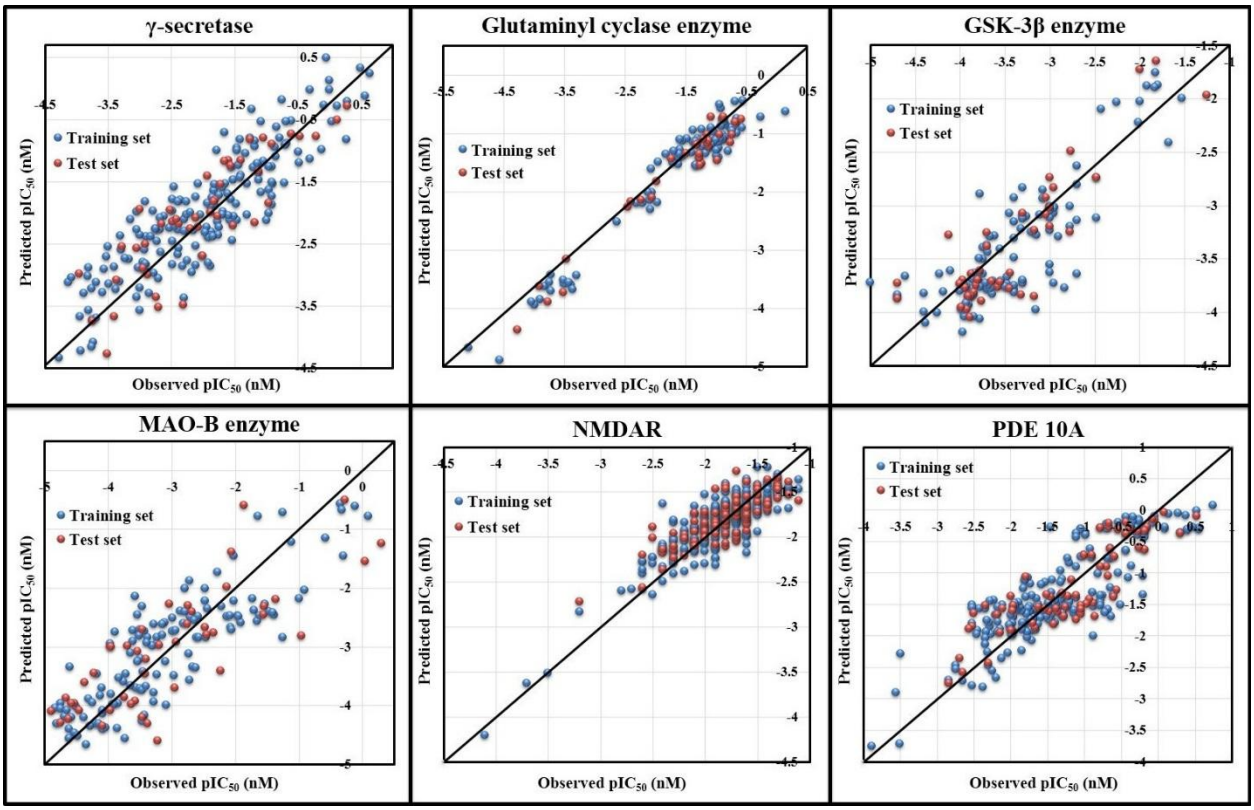


Figure 4.97. Scatter plots of observed v/s predicted values for γ-secretase, Glutaminyl Cyclase, GSK-3β, MAO-B, NMDAR, and PDE 10A models.

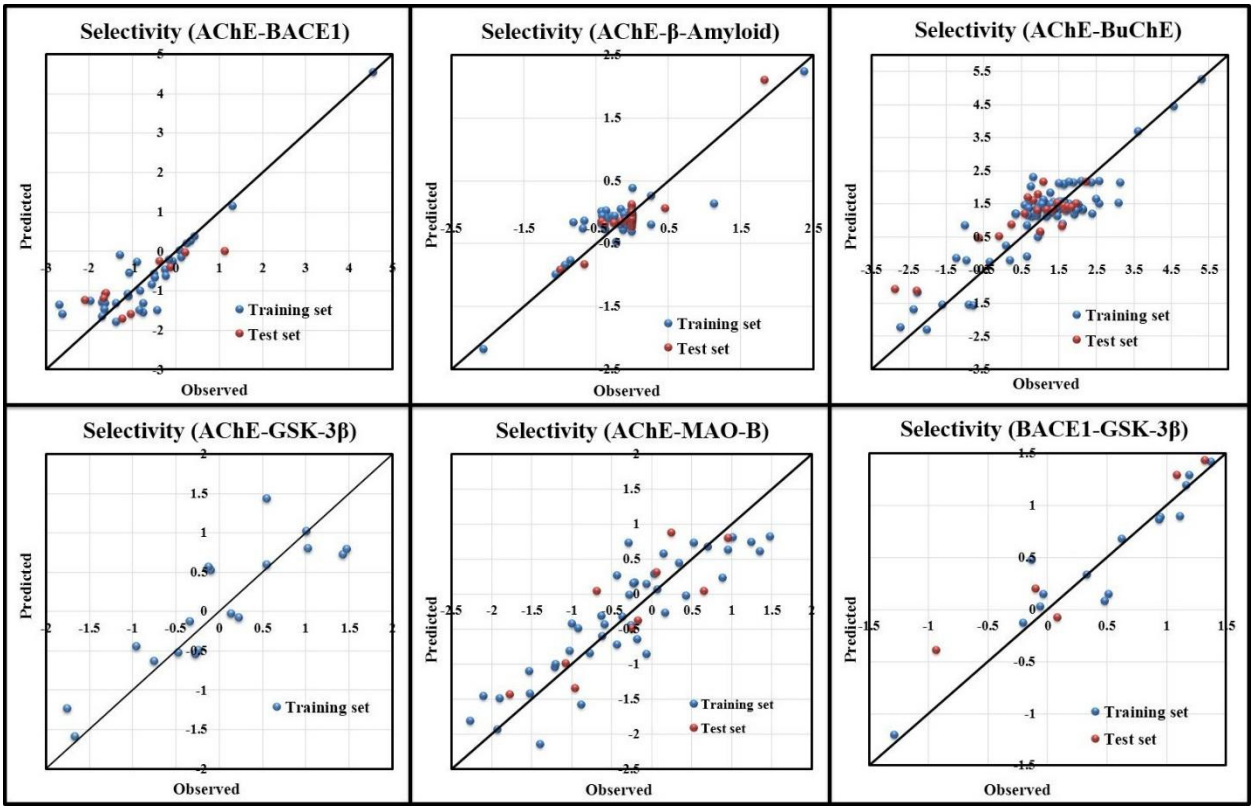


Figure 4.98. Scatter plots of observed v/s predicted values for selectivity based (AChE-BACE1, AChE-β-Amyloid, AChE-BuChE, AChE-GSK-3β, AChE-MAO-B, and BACE1-GSK-3β) models.

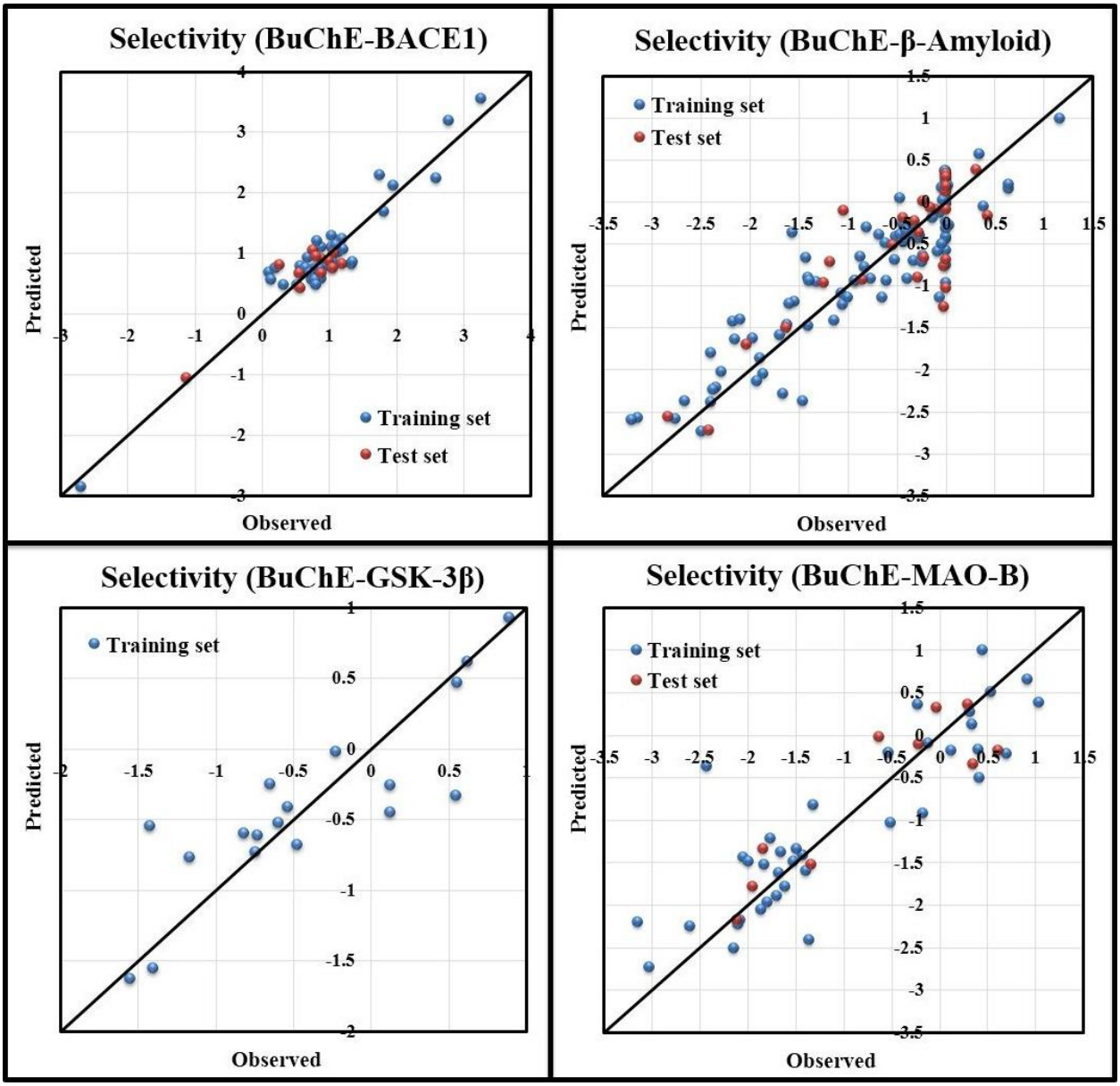


Figure 4.99. Scatter plots of observed v/s predicted values for selectivity-based (BuChE-BACE1, BuChE-β-Amyloid, BuChE-GSK-3β, and BuChE-MAO-B) models.

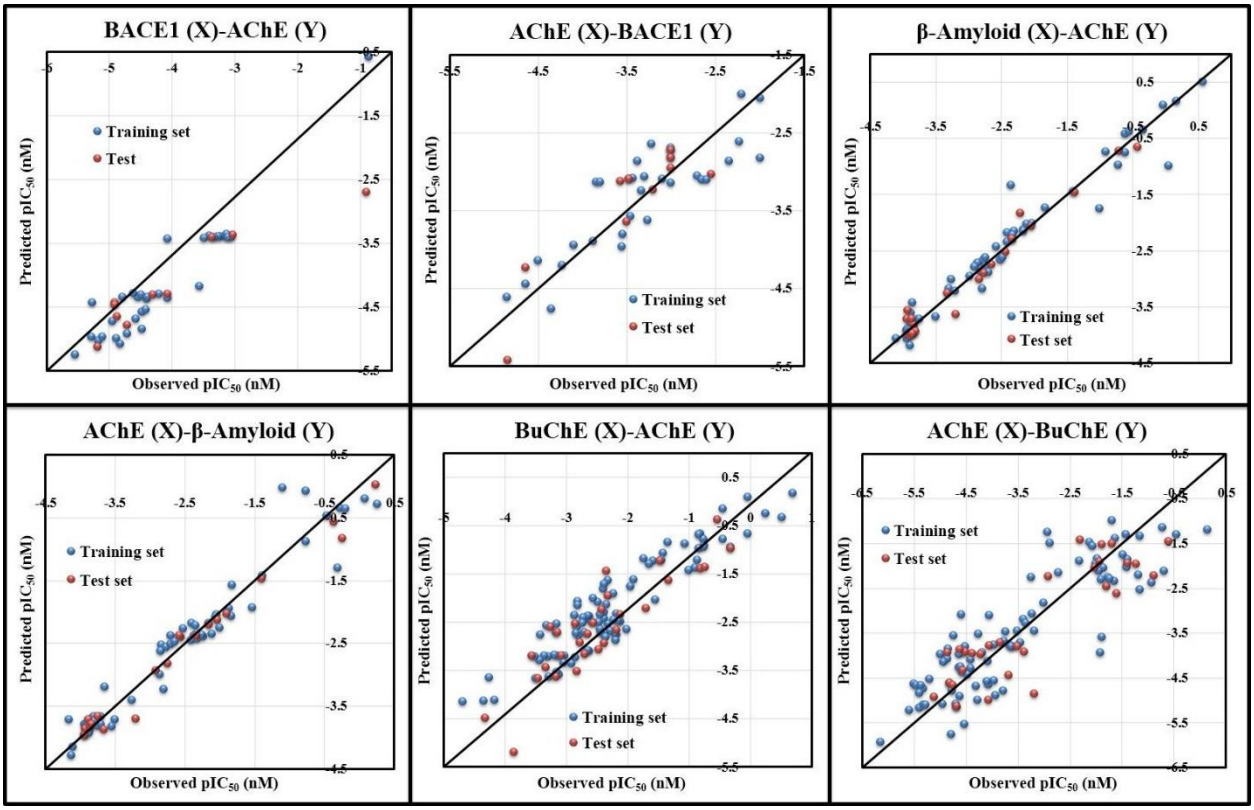


Figure 4.100. Scatter plots of observed v/s predicted values for QSAAR (BACE1 (X)-AChE (Y), AChE (X)-BACE1 (Y), β-Amyloid (X)-AChE (Y), AChE (X)-β-Amyloid (Y), BuChE (X)-AChE (Y) and AChE (X)-BuChE (Y)) models.

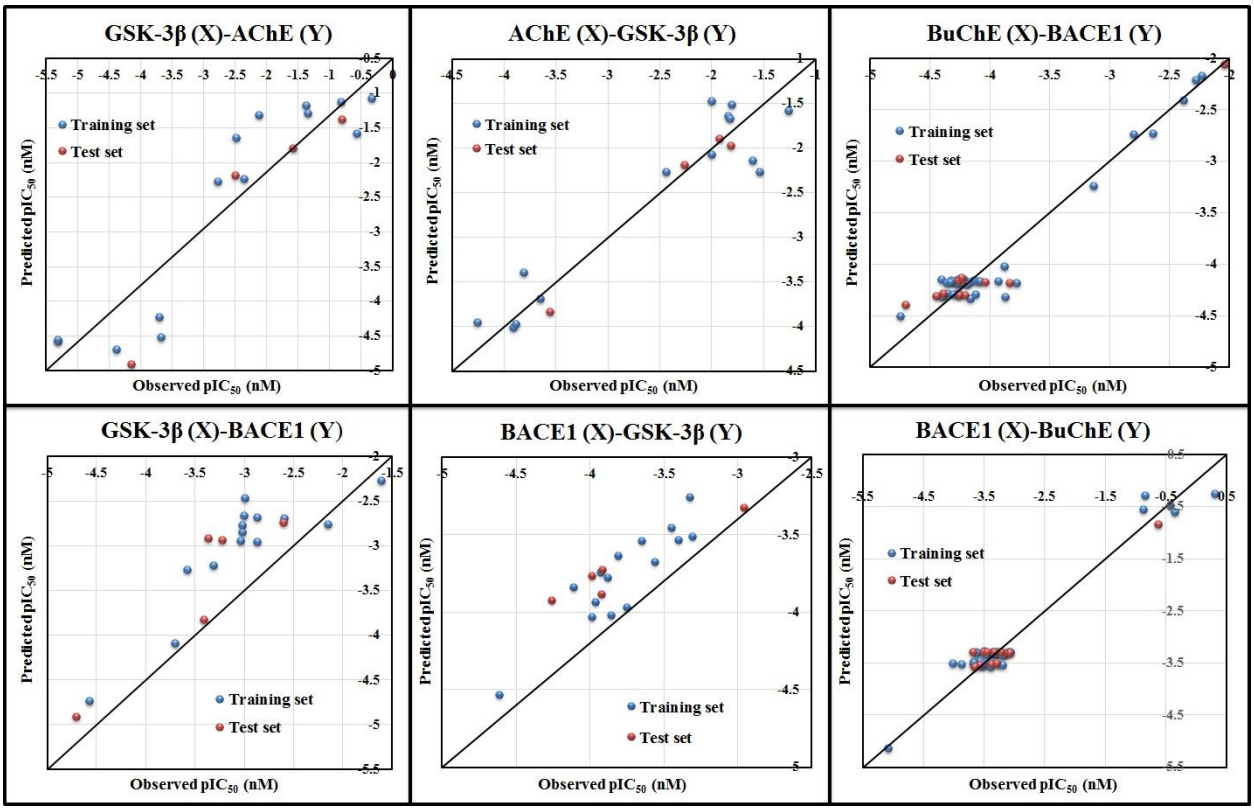


Figure 4.101. Scatter plots of observed v/s predicted values for QSAAR (GSK-3β (X)-AChE (Y), AChE (X)-GSK-3β (Y), BuChE (X)-BACE1 (Y), GSK-3β (X)-BACE1 (Y), BACE1 (X)-GSK-3β (Y) and BACE1 (X)-BuChE (Y)) models.

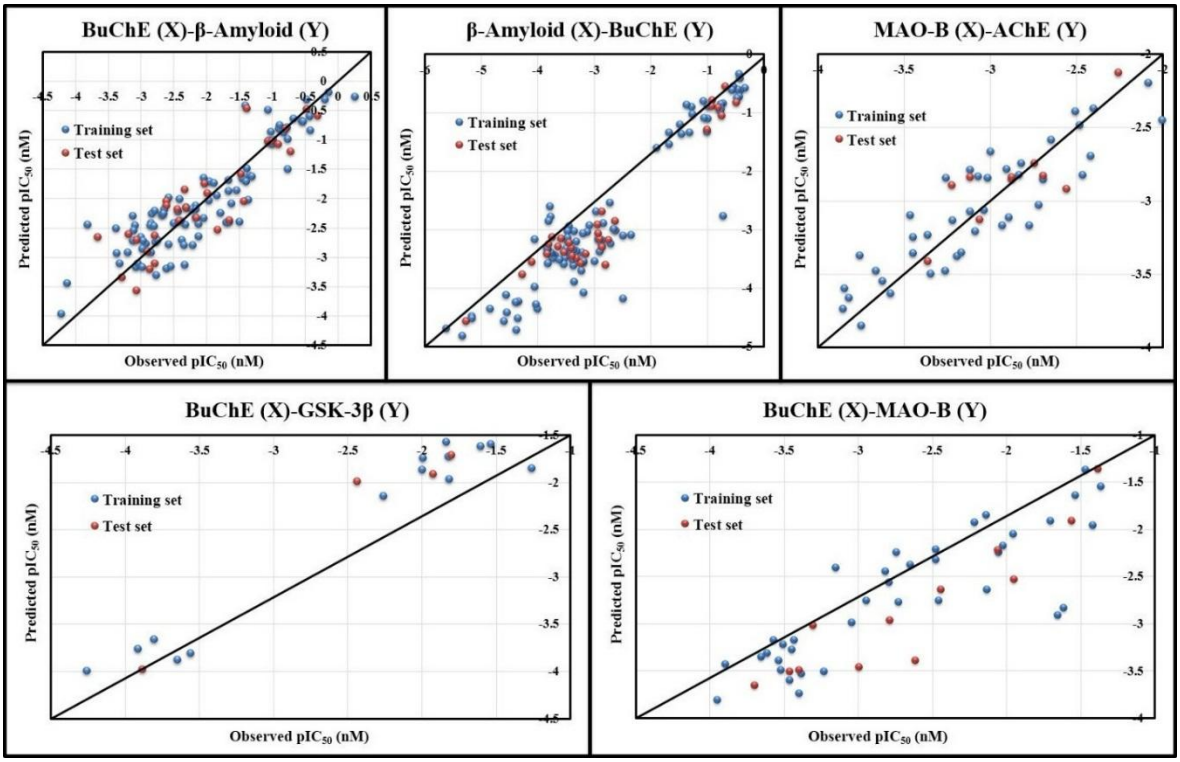


Figure 4.102. Scatter plots of observed v/s predicted values for QSAAR (BuChE (X)-β-Amyloid (Y), β-Amyloid (X)-BuChE (Y), MAO-B (X)-AChE (Y), BuChE (X)-GSK-3β (Y) and BuChE (X)-MAO-B (Y)) models.

4.6.1.1. Mechanistic interpretation of the Descriptors Involved in the Developed Individual QSAR Models

4.6.1.1.1. 5-hydroxytryptamine receptor 6 (5-HT6)

The first most important descriptor in this model is minssNH, the value of minssNH is 2.844 for the most active compound **1** and 0 for the least active compound **74**. The positive regression coefficient of this descriptor indicates that the higher value of the descriptor leads to an improved inhibitory activity against the enzyme. The second most significant descriptor in this series is F04[C-O], the value of F04[C-O] descriptor is 6 for the most active compound **3** and 0 for the least active compound **74**. The positive sign of the coefficient of this descriptor confirms that the larger the value of the descriptor higher the inhibitory activity towards the enzyme. According to the discussion made above, the minssNH, and F04[C-O] of the most active compounds must be higher than those of the less active compounds (see **Figure 4.104**).

4.6.1.1.2. Acetylcholinesterase (AChE) enzyme

In this model, the first most significant descriptor is X2v which belongs to the class of connectivity index and is related to molecular branching and shape information. The positive regression coefficient sign confirms that the AChE enzyme inhibitory activity may be increased by an increase in molecular branching and shape (bond angles) in the compounds. For instance, we can see the most active compound **1754** has a higher value leading to better inhibitory activity, while the least active compound **698** has the lowest value for this descriptor. Another positively correlated descriptor, DBI, represents the branching nature of the compound. With an increase in the branching index, the inhibitory activity will increase as observed in

compound **3** (most active). The next positively correlated descriptor, B07[N-N], leads to better inhibitory activity against the enzyme as the value of the descriptor is 1 for the most active **compound 842** and zero for the compound **725**. The ring descriptor, D/Dtr05, having a positive regression coefficient suggests that a higher numerical value of this descriptor leads to an improved inhibitory activity as verified by the compound **841** (most active), and the compounds with no such fragment show lower AChE enzyme inhibitory activity as found in the compound **710** (least active) (see **Figure 4.103**).

4.6.1.1.3. Butyrylcholinesterase (BuChE) enzyme

Among the essential features enhancing BuChE enzyme inhibitory activity are nArOCON, N-070, MaxaaaC, C-002, F04[C-N], and MDEC-22. As lipophilicity is an important parameter for AD drugs, it can be traced from variables nArOCON, C-002, MDEC-22, and MaxaaaC which contributed positively towards the inhibitory activity as evidenced by the compounds **593**, **416,** and **415** (most active) respectively. The enhanced concentration of electronegative atom count in a molecule has a direct impact on the improved inhibitory activity. This hypothesis can be confirmed by the presence of variables N-070 and F04[C-N] which contributed positively as evidenced by the compounds **415** and **416** (most active) respectively. Again, variables nCs, nArNHR, H-051, and Psi_e_A contributed negatively which means that the presence of these features in the molecules leads to lower inhibitory activity as observed in the compounds **2330**, **402**, **399**, and **411** (least active) respectively(see **Figure 4.103**).

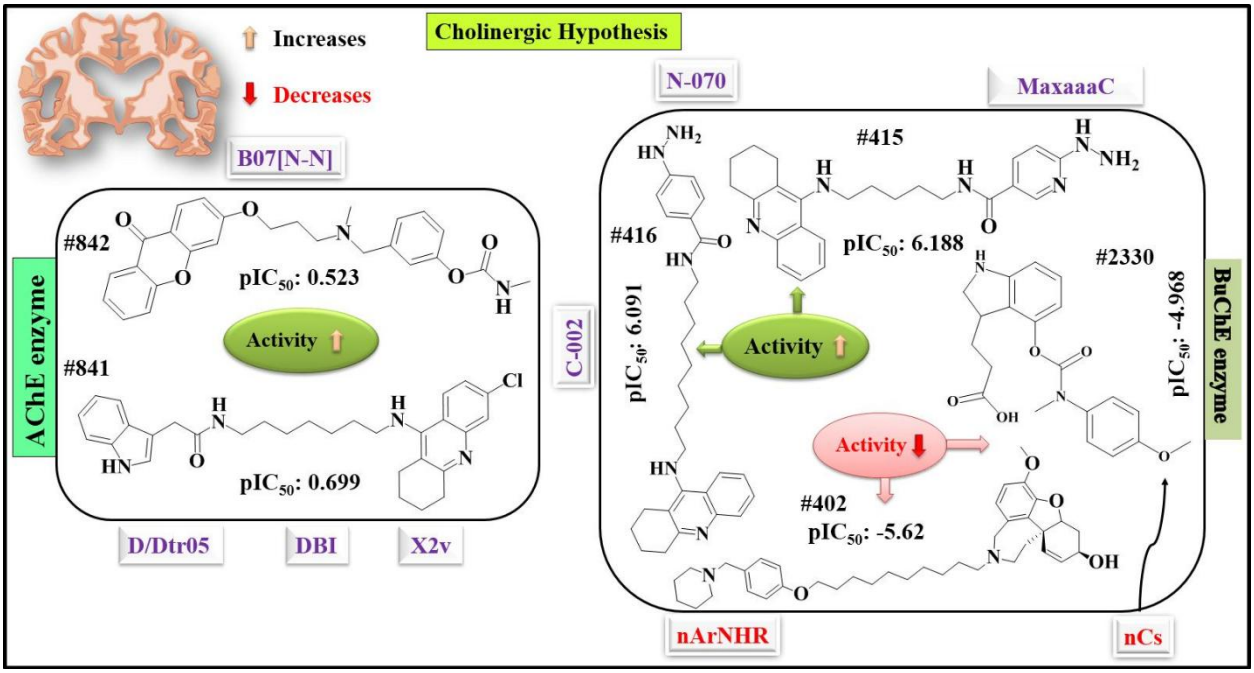


Figure 4.103. Possible mechanistic interpretation of the most significant descriptors obtained from the models against AChE and BuChE enzymes.

4.6.1.1.4. Beta-secretase 1 (BACE1) enzyme

The final equation comprises 28 descriptors, among which five descriptors (SA score, F08[N-N], Ui, X5v, and F04[O-O]) contributed most against the enzyme as shown by the VIP statistic (VIP score > 1). Among them, four variables showed positive contribution towards an inhibitory activity which include SA score, F08[N-N], Ui, and X5v, which means that the presence of these features in the molecules leads to better inhibitory activity as found in the compounds **455** and **463** (in case of X5v). The negative contribution of

F04[O-O] denotes the frequency of two oxygen atoms at topological distance 4, suggesting that a higher number of electronegative atom count in a molecule has a direct impact on the inhibitory activity as observed in the compound **819** (least active) (see **Figure 4.104**).

4.6.1.1.5. β-amyloid aggregation

The major group of positive contributing descriptors involved in the developed equation by PLS is subgroups like SssCH2, PW3, and F04[N-O]. The direct relationship of the descriptor SssCH2 suggested that the presence of the number of such -CH2- groups in the molecules leads to better β-amyloid aggregation inhibitory activity as longer chain compounds with a higher number of -CH2- groups would be more lipophilic resulting in improved brain permeability. This assumption can be confirmed by compound **57** having a higher number of -CH2- groups in their structure, showing higher descriptor values leading to their higher range of inhibitory activity. The descriptor, PW3 can be considered as a shape descriptor whose value increases with increased branching in the vertices. The descriptor F04[N-O] denotes the frequency of nitrogen and oxygen atoms at the topological distance 4. These two contribute positively towards inhibitory activity as evident by the compound **59**. Again, descriptors F05[C-O], F02[N-N], and MAXDN contributed negatively which means that the presence of these features in the molecules shows lower inhibitory activity as observed in the compounds **149**, **262** and **175** respectively (see **Figure 4.104**).

4.6.1.1.6. Cyclin Dependent Kinase 5 (CDK-5) protein

In this model, NaaaC is the most contributing descriptor toward inhibitory activity. It denotes the number of atoms of type aaaC (π), aromatic fused carbons. The fragment has a positive contribution toward inhibitory activity against protease. A compound like **195** shows higher inhibitory activity due to the presence of a high number of aaaC fragments. The next most important descriptor PBF denotes the Plane of best fit in the compounds having a negative coefficient towards the inhibitory activity. Therefore, for improved activity, their values must be augmented as low as possible. This statement can be confirmed by compound **146** having higher descriptors values, showing their lower range of inhibitory activity (see **Figure 4.105**).

4.6.1.1.7. Gamma-secretase enzyme

As per the VIP plot, P_VSA_MR_7 is the most significant with a positive coefficient feature towards the inhibitory activity in this model. P_VSA_MR_7 belongs to P_VSA-like descriptors that are calculated from the amount of van der Waals surface area (VSA), thus being related to the lipophilic feature. Compound **209** with a high value of P_VSA_MR_7 showed strong inhibitory activity (see **Figure 4.104**). The variables nOHp, B05[N-N], and nRCOOH, contributed negatively suggesting that the presence of these features in the compounds will result in a lower inhibitory activity as observed in the compounds **41**, **207**, and **66** (least active) respectively (see **Figure 4.104**).

4.6.1.1.8. Glutaminyl Cyclase (QCs) enzyme

The developed QSAR model showed that the features C-034, ETA_Sh_Y, F05[C-S], and PBF are important for defining the inhibitory activity against the QCs enzyme. Lipophilicity is a significant parameter for AD drugs as can be noticed from the variable C-034 which contributed positively towards the inhibitory activity as evidenced by compound **34** (most active). C-034 enhancing lipophilicity represents the feature R-CR..X, where R is any group linked through carbon; X is an electronegative atom (O, N, S, P, Se, halogens); '-'is an aromatic bond as in benzene or delocalized bonds such as the N, O bond in a nitro group; .. denotes aromatic single bond). In this series, variables ETA_Sh_Y, F05[C-S], and PBF contributed negatively

suggesting that the presence of these features in the compounds will result in a lower inhibitory activity as evidenced by the molecules **33**, **31,** and **72** (least active) respectively (see **Figure 4.104**).

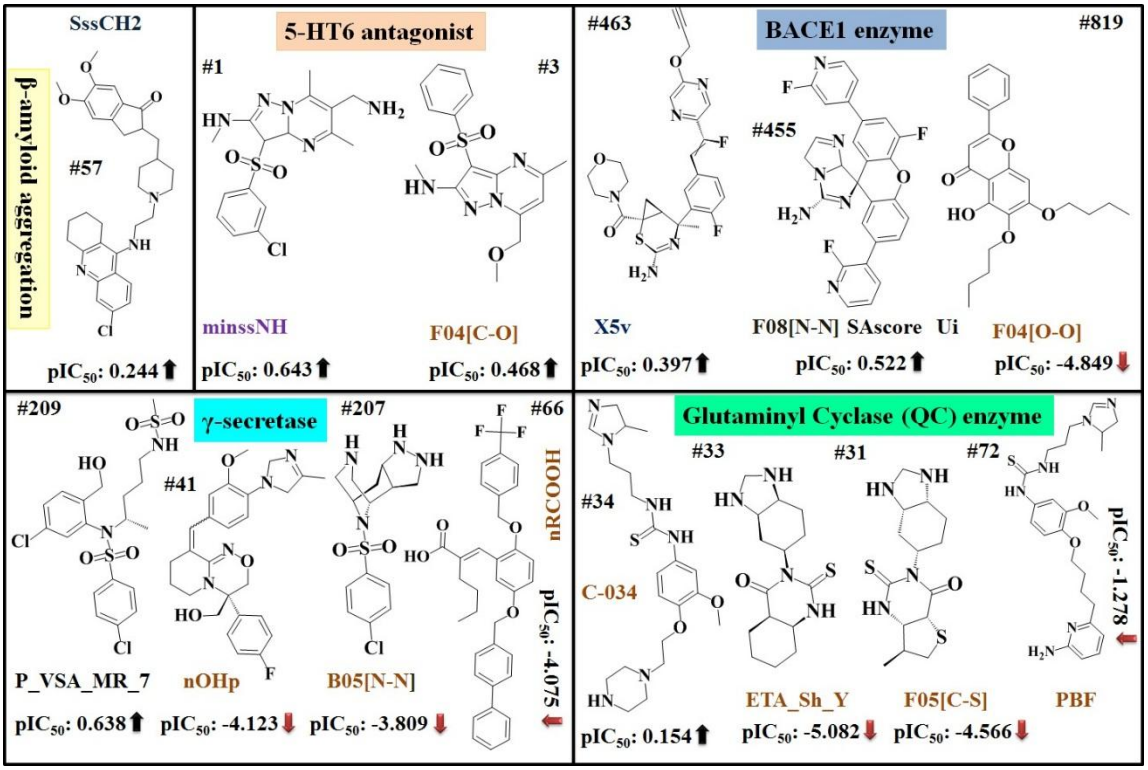


Figure 4.104. Mechanistic interpretation of the most significant descriptors obtained from the models against β-amyloid aggregation, 5-HT6, BACE1, γ-Secretase, and Glutaminyl Cyclase.

4.6.1.1.9. Glycogen synthase kinase-3β (GSK-3β) enzyme

The first and most crucial feature important for the GSK-3β enzyme inhibitory activity was nThiazoles, which designates the number of thiazole rings present in the molecules. The lipophilicity accompanying molecular bulk was the next most significant feature responsible for enzyme inhibitory activity which appeared as SaaaC, SaasC, and PDI. SaaaC is an atom-type E-state index for carbons with three aromatic connections and SaasC is for aromatic carbons with an attached substituent atom. The positive coefficients of SaaaC and SaasC in the model were considered to be a consequence of the importance of aromatic rings in controlling the inhibitory activity of the compounds, both in determining the compound's hydrophobicity and its π-π interactions with the target as witnessed in the molecular docking study. The packing density index (PDI) is a molecular property descriptor. PDI is designated as the ratio between the McGowan volume and the total surface area[403]. The positive sign of the coefficient of these descriptors confirms that the larger the value of descriptors higher is the inhibitory activity towards the enzyme as witnessed by compound **153** (most active). Again, the next significant descriptor mindsCH designates the minimum atom-type E-State: =CH- contributed negatively means the presence of this feature in the compounds shows the lower inhibitory activity as observed in compound **161** (least active) (see **Figure 4.105**).

4.6.1.1.10. Monoamine oxidase B (MAO-B) enzyme

As per the VIP score, mindsCH is the most significant descriptor that appeared in this equation which contributed negatively towards the inhibitory activity as observed in compound **25** (least active). This feature also appeared in the previous model (GSK-3β) with a negative coefficient toward inhibitory activity. In this series next most significant descriptors are F02[O-O], C%, and nRCN. The positive regression coefficient sign confirms that the inhibitory activity increases with the presence of the above fragments which can be verified by the most active compounds **86** (in the case of F02[O-O] and C%) and **72** (in the case of nRCN) in their structure(see **Figure 4.105**). The other significant features were polar oxygen groups (B06[O-Cl], B01[C-O]), molecular branching (Eta_B), secondary aromatic amine (nArNHR), minssCH2, logP, and F05[C-N] contributed negatively suggesting that the presence of these features in the compounds will result in lower inhibitory activity.

4.6.1.1.11. N-methyl-D-aspartate (NMDA) receptor

The higher concentration of electronegative atom count in a compound has a direct impact on the lowering of antagonistic activity against the target protein. This hypothesis can be confirmed by the presence of variables SaaN, TPSA(NO), C-043, ATSC2m, and MDEN-23 contributed negatively as evidenced by the compounds **483** and **446** (least active). According to the VIP scores, SaaN is the most contributing descriptor in the model which encodes information about both the topological environment of the particular atom and the electronic interactions due to all other atoms in the molecule. SaaN is the sum of E-State values of all nitrogen atoms with two aromatic bonds found in the molecule (see **Figure 4.105**).

4.6.1.1.12. Phosphodiester 10A (PDE 10A) enzyme

The leading group of positive contributing descriptors involved in the developed model is subgroups like N-070, and AMR. The direct association of the descriptor N-070 suggested that the presence of the number of N atoms in the Ar-NH-Al group with Al representing aliphatic groups in the compounds leads to better enzyme inhibitory activity. The hydrophobicity associated with molecular bulk was the next most significant feature responsible for the inhibitory activity which appeared as AMR variables, as the higher the features in the compounds higher would be the lipophilicity resulting in improved brain permeability. This concept can be confirmed by compound **226** (most active) (see **Figure 4.105**).

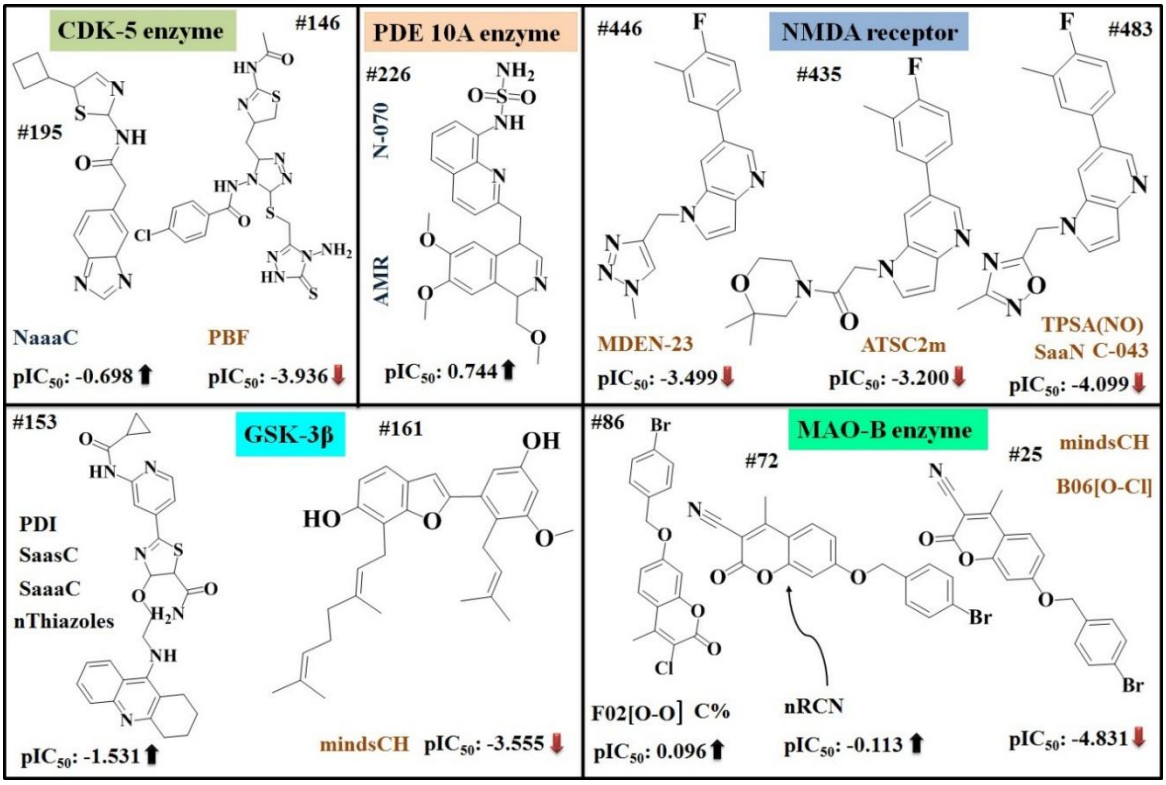


Figure 4.105. Contributions of the most significant descriptors obtained from the models against CDK-5, PDE 10A, NMDAR, GSK-3β, and MAO-B enzymes.

4.6.1.2. Mechanistic interpretation of the descriptors involved in the development of selectivity-based models

4.6.1.2.1. Selectivity of AChE and BACE1 enzyme inhibitors

According to the VIP score, F08 [N–O] is the most significant positively contributing descriptor in the model for more specific inhibition to AChE enzyme than BACE1 enzyme inhibitory activity. Thus, the higher number of this fragment leads to better AChE enzyme inhibitory activity as noticed in the case of compound **2**. Again, the variable B05[N-O] contributed negatively and is more specific to BACE1 enzyme inhibitory activity than AChE enzyme inhibitory activity. So, the higher number of this fragment correlates with lower AChE enzyme inhibitory activity as observed in the case of compound **28** (see **Figure 4.106**).

4.6.1.2.2. Selectivity of AChE enzyme and β-amyloid aggregation inhibitors

The important group of the positive contributing variables more specific to the AChE enzyme inhibitory activity involved in the developed model is subgroups like B07[C-N], F06[C-Cl], SAscore, and ETA_Shape_Y. In these features, ETA_Shape_Y signifies the measure of molecular shape. The straight connotation of these descriptors suggested that the presence of the above groups in the compounds leads to better AChE enzyme inhibitory activity. This concept can be confirmed from compound **60** (most active). The next significant features of negative contributing more specific to β-amyloid aggregation inhibitory activity than AChE enzyme inhibitory activity appeared in the developed equation are subgroups like nArNH2, F10[C-Cl], and PJI2. In this series, variable nArNH2 signifies the presence of the number of primary aromatic amines in the compounds. The above specification can be confirmed by compounds **3** (in the case of nArNH2) and **10** respectively (see **Figure 4.106**).

4.6.1.2.3. Selectivity of AChE and BuChE enzyme inhibitors

Among the crucial features improving AChE enzyme inhibitory activity are F06[N-O] and nR12, as lipophilicity is an important parameter for AD drugs and can be outlined from variable nR12. Thus, the higher number of these fragments leads to better AChE enzyme inhibitory activity as noticed in the case of compound **81** (most active). Again, variables constituting electronegative atoms (B03[N-N], B03[N-O], nROR, N%, and GD) contributed negatively suggesting that the presence of these features in the compounds more specific to BuChE inhibitory activity as observed in the compounds **93**, **92** and **94** respectively (see **Figure 4.106**).

4.6.1.2.4. Selectivity of AChE and MOA-B enzyme inhibitors

The most significant features enhancing lipophilicity (SsssCH, and F10[C-C]), and constituting electronegative and polar atoms (O%, F10[C-N], and minsCH3) contributed negatively suggesting that the presence of these features in the compounds are more favorable for MOA-B enzyme inhibitory activity as evidenced by the compounds **52, 44, 25** and **35**. In this series, the next most significant descriptors that contributed positively are F09[C-C], and MaxsCH3, which are more specific to the AChE enzyme inhibitory activity than MOA-B enzyme inhibitory activity. Thus, the higher number of this fragment leads to better AChE enzyme inhibitory activity as observed in compounds **27** and **18** respectively(see **Figure 4.106**).

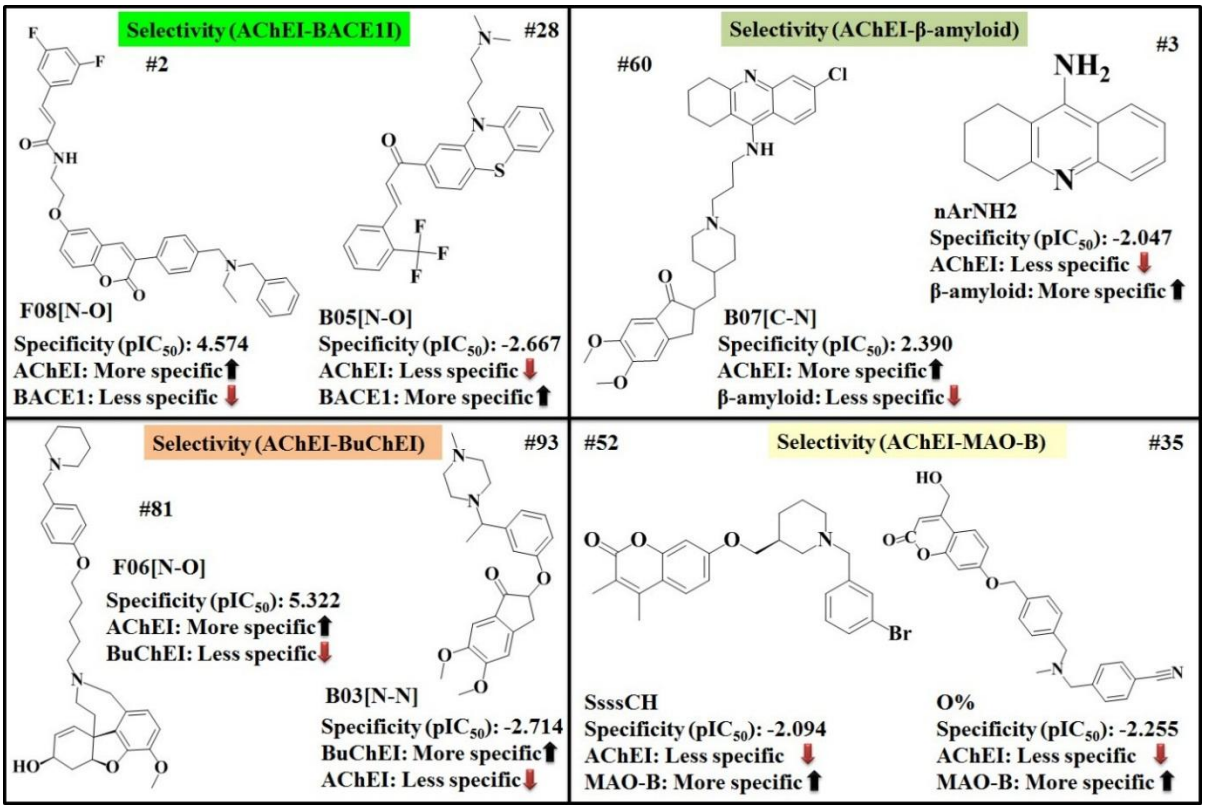


Figure 4.106. Mechanistic interpretation of the most significant descriptors obtained from the selectivity-based (AChE-BACE1, AChE-β-Amyloid, AChE-BuChE, and AChE-MAO-B) models.

4.6.1.2.5. Selectivity of BACE1 and GSK-3β enzyme inhibitors

The significant descriptors in this series are nRotB and AATSC1swhichcontributed negatively toward the inhibitory activity against the BACE1 enzyme; accordingly, these features are more specific to GSK-3βenzyme inhibitory activity. Thus, the higher number of this fragment shows lower BACE1 enzyme inhibitory activity as seen in compounds **5** and **12** respectively (see **Figure 4.107**).

4.6.1.2.6. Selectivity of BuChE and BACE1 enzyme inhibitors

In the MLR equation, variables NaasC, minaaaC, nArCO, and ZM1Kup contributing positively were the most significant descriptors toward the inhibitory activity against the BuChE enzyme, which means the presence of these features in the compounds are most specific to the BuChE enzyme inhibitory activity than BACE1 as observed in the molecules **7, 2,** and **3** respectively. While variables constituting lipophilicity (ALOGP2, and nR=Ct) contributed negatively suggesting that the presence of these features in the molecules is more specific to BACE1inhibitory activity as observed in compound **8** (see **Figure 4.107**).

4.6.1.2.7. Selectivity of BuChE enzyme and β-amyloid aggregation inhibitors

The higher concentration of electronegative atom count and aromaticity in a compound has a direct impact on the better inhibitory activity against the BuChE enzyme. This premise can be confirmed from the presence of variables D/Dtr10 (aromatic ring count) and B04[N-N], B10[C-O], B08[C-N], ETA_dEpsilon_B (electronegative atom counts) contributed positively as evidenced by the compounds **35, 1** and **4** (most active) respectively. The other features B10[C-N], LOGP99, B10[N-O], PJI2, B08[O-O], and ETA_Shape_Y contributed negatively to the model are more specific to the β-amyloid aggregation inhibitory activity than BuChE enzyme inhibitory activity as observed in the compounds **124, 118, 120, and 14** respectively (see **Figure 4.107**).

4.6.1.2.8. Selectivity of BuChE and MOA-B enzyme inhibitors

The molecules constituting variables containing tertiary nitrogen atoms (SsssN) and nitrogen and oxygen atoms at the topological distance 6 (B06[N-O])are more specific to the BuChE enzyme inhibitory activity than the MOA-B enzyme inhibitory activity as seen in the compound **23**. The variables imparting lipophilicity (minsCH3, SsssCH, NdssC, and SAacc) and electronegativity (B08[N-O]) of the compound are more specific to the MOA-B enzyme inhibitory activity than the BuChE enzyme inhibitory activity as observed in the compounds **4, 47**, and **51** (see **Figure 4.107**).

4.6.1.2.9. Selectivity of AChE and GSK-3β enzyme inhibitors

The variable Psi_i_A contributed positively suggesting that the presence of such fragment in the molecule is more specific to the AChE enzyme inhibitory activity than GSK-3β enzyme inhibitory activity as seen in compound **10**. The negatively contributing descriptors ETA_BetaP_ns_d, and F02[C-O] are more specific to the GSK-3β enzyme inhibitory activity than the AChE enzyme inhibitory activity as observed in the compound **14** (see **Figure 4.107**).

4.6.1.2.10. Selectivity of BuChE and GSK-3β enzyme inhibitors

As we can see in both selectivity-based models, we found similar features contributed accordingly to the inhibitory activity. The variables Psi_i_A and nCL suggested that the presence of such fragments in the compounds is more specific to the BuChE enzyme inhibitory activity as seen in compounds **13** and **7** respectively. As the previous model descriptor, ETA_Beta_ns_d contributed negatively, the same effect exerted in this equation also suggests that the presence of this feature in the molecule is more specific to

the GSK-3β enzyme inhibitory activity than BuChE enzyme inhibitory activity as observed in the compound **12** (see **Figure 4.107**).

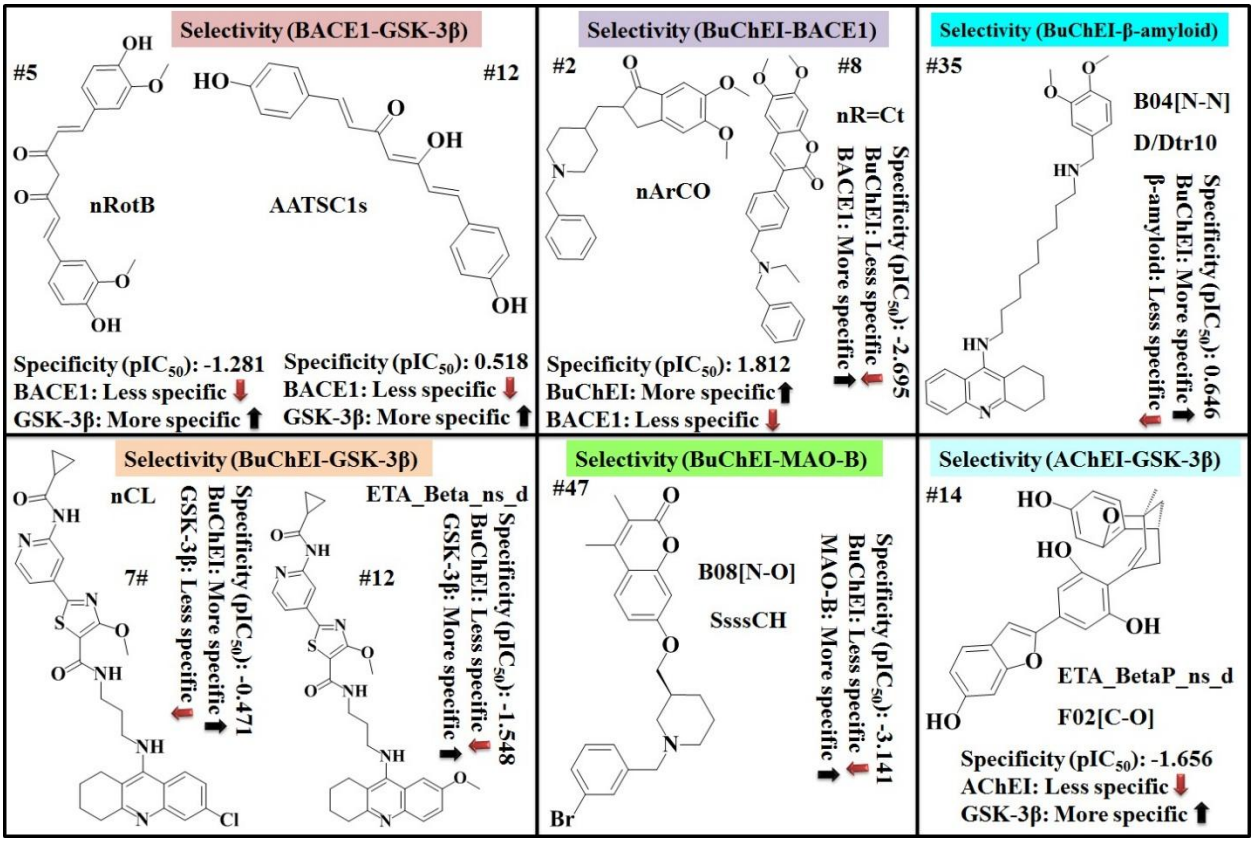


Figure 4.107. Probable mechanistic interpretation of the most significant descriptors obtained from the selectivity-based (BACE1-GSK-3β, BuChE-BACE1, BuChE-β-Amyloid, BuChE-GSK-3β, BuChE-MAO-B, and AChE-GSK-3β) models.

4.6.1.3. Mechanistic interpretation of the descriptors involved in the development of QSAAR-based models

The QSAAR models are the *in silico* analysis of activity-activity correlations that can be used to evaluate the dual inhibitory activity of a molecule. The QSAAR analysis usually comprises the use of activity values of one of the endpoints as the response variable (i.e., Y-variable), whereas the activity to the other endpoint is used as one of the predictor variables (i.e., X-variable). If the activity values of the identified molecules for one endpoint correspond strongly to the values for another endpoint, the chemicals are likely to have a similar mode of action for both and vice versa. This section provides a concise and more relevant description of the mechanistic interpretation of various QSAAR models.

4.6.1.3.1. BACE1 enzyme inhibitory activity as a predictor (X) and AChE enzyme inhibitory activity as the response (Y) and vice versa

The activity of both targets is shown to be the most significant predictor variable, positively contributing to each other. Other important descriptors in the BACE1 (X)-AChE (Y) and AChE (X)-BACE1 (Y) models include nCrs, MaxaasC, PW3, X5v, B08[O-O] (having positive regression coefficients), C-001, nCconj and nArOH (having negative regression coefficients) respectively, as established by the VIP plots. The next

most important inhibitory activity-enhancing features identified were hydrophobicity, molecular bulk, and electronegativity. The descriptors proving this hypothesis were hydrophobicity variables (nCrs, and MaxaasC), molecular bulk (PW3), electronegativity (B08[O-O]), and size index (X5v). Thus, the presence of these features in the compounds leads to better inhibitory activity (see **Figure 4.108**).

4.6.1.3.2. AChE enzyme inhibitory activity as a predictor (X) and β-amyloid inhibitory activity as a response (Y) and vice versa

The activity terms (both contributing positively) are the highest contributing descriptors by the corresponding VIP and loading plots. In the β-amyloid (X)-AChE (Y) model, all the identified variables (nArCO, B10[N-Cl], PW3, and B07[C-N]) contributed positively toward inhibitory activity. The other recognized significant features were B10[N-Cl] and B07[C-N] having positive regression coefficients in the β-amyloid (X)-AChE (Y) model while negative regression coefficients in AChE (X)-β-amyloid (Y) model, as per the VIP plot. The remaining two descriptors of the AChE (X)-β-amyloid (Y) models are SAscore and ETA_EtaP_F_L inversely correlated to the response as shown by their negative regression coefficients (see **Figure 4.108**).

4.6.1.3.3. BuChE enzyme inhibitory activity as a predictor (X) and AChE enzyme inhibitory activity as the response (Y) and vice versa

The activities against both targets are directly related (both having positive regression coefficients) and the most relevant predictor variables for the respective QSAAR models. In the BuChE (X)-AChE (Y) model, the essential features enhancing inhibitory activity were lipophilicity (nR10, and nCr), electronegative atoms (NssNH, B04[N-O], F06[N-O] and B04[N-Cl]) and other feature like ICR (radial centric information index). As per the regression coefficient plot, the presence of these features in the compounds leads to better inhibitory activity. The features contributing negatively were F08[C-N], and nROR which indicate that the decrease in the values of these variables affects an increase in the inhibitory activity. In the AChE (X)-BuChE (Y) model, the variable constituting the lipophilicity (NaaCH), shape (DECC) and electronegative atoms (B10[N-O]) are significant for enhancing inhibitory activity. Apart from the above descriptors, the other identified variables were F06[N-O], MaxsCH3, ETA_EtaP_F, nCrs, Ram, and nCb- inversely correlated to the response as suggested by their negative regression coefficients (see **Figure 4.108**).

4.6.1.3.4. GSK-3β enzyme inhibitory as a predictor (X) and AChE enzyme inhibitory activity as the response (Y) and vice versa

The responses against both targets are shown to be the most significant predictor variables, contributing positively to each other. Other significant variables in the GSK-3β (X)-AChE(Y) and AChE (X)-GSK-3β (Y) models comprise C-026 (having negative regression coefficients), ETA_Epsilon_1 and MaxaasC (having positive regression coefficients) respectively, as established by the VIP plots. C-026 inversely correlated to the response. Furthermore, variables constituting fragments imparting lipophilicity (MaxaasC) and electronegative atoms (ETA_Epsilon_1) in the molecules lead to improved inhibitory activity (see **Figure 4.108**).

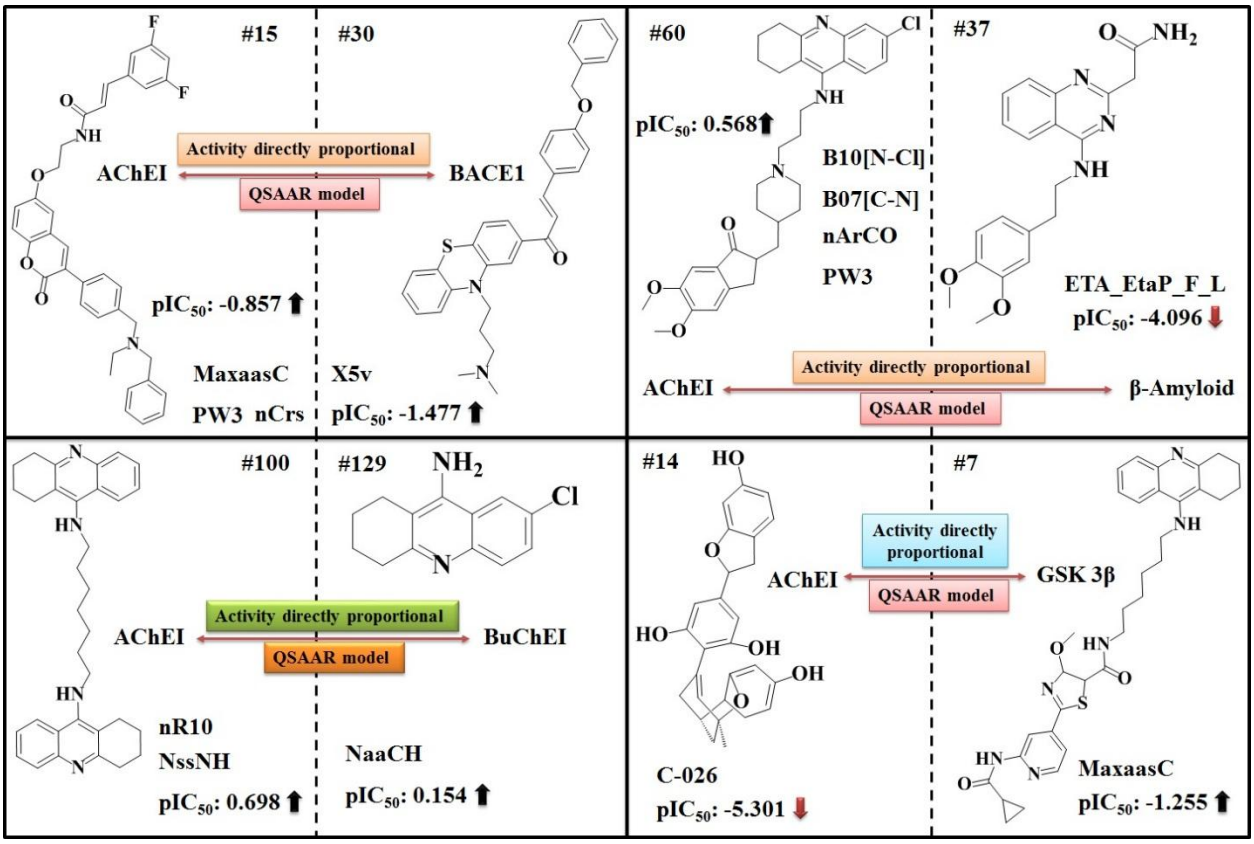


Figure 4.108. Probable mechanistic interpretation of the most significant descriptors obtained from the QSAAR (AChE (X)-BACE1 (Y), BACE1 (X)-AChE (Y), β-Amyloid (X)-AChE (Y), AChE (X)-β-Amyloid (Y), BuChE (X)-AChE (Y) and AChE (X)-BuChE (Y), GSK-3β (X)-AChE (Y), and AChE (X)-GSK-3β (Y)) models.

4.6.1.3.5. MAO-B enzyme inhibitory as a predictor (X) and AChE enzyme inhibitory activity as the response (Y)

The positive regression coefficient of the activity variable specifies that the MAO-B enzyme inhibitory value is proportionally related to the AChE enzyme inhibitory activity (response). In this series, the most significant variables (having positive regression coefficients) were F01[C-N], MaxdssC, and B05[N-Cl], enhancing the inhibitory activity. The other recognized significant features were N%, MAXDP, ETA_Epsilon_5, and MaxaaCH inversely associated with the response as suggested by their negative regression coefficients (see **Figure 4.109**).

4.6.1.3.6. GSK-3β enzyme inhibitory as a predictor (X) and BACE1 enzyme inhibitory activity as the response (Y) and vice versa

The activity end points of both targets (with positive regression coefficients) contributed the most to the development of the respective models. The QSAAR model for GSK-3β (X)-BACE1 (Y) includes max_conj_path and C-025, which contributed positively towards inhibitory activity. The last variable in this model was minsOH having a negative regression coefficient towards the inhibitory activity. In the BACE1 (X)-GSK-3β (Y) model, the identified important features were MAXDN (having a positive regression coefficient) signifying the maximal electrotopological negative variation in the compounds. The

other recognized important variables were MATS2e, ZMIC4, and AATS3s inversely associated with inhibitory activity (see **Figure 4.109**).

4.6.1.3.7. BuChE enzyme inhibitory as a predictor (X) and BACE1 enzyme inhibitory activity as the response (Y) and vice versa

Similar to the aforementioned QSAAR models delineated above, the activity end points of both targets (with positive regression coefficients) contributed the most to the development of the respective models. The significant variables in the BuChE(X)-BACE1(Y) model includes MaxsCH3, nFuranes, and B05[O-O], having positive regression coefficients, thus, the presence of these features in the compounds leads to better inhibitory activity. Again, the variables F01[C-O], and nCIC have negative regression coefficients and inversely affect the inhibitory activity. In the BACE1(X)-BuChE (Y) model, the second most significant inhibitory activity-improving feature was identified as hydrophobicity. The variable evidencing this premise was hydrophobicity variables (RFD, F08[C-C], and nR09), thus, the presence of these fragments in the molecules leads to enhanced inhibitory activity (see **Figure 4.109**).

4.6.1.3.8. BuChE enzyme inhibitory as a predictor (X) and β-amyloid inhibitory activity as the response (Y) and Vice Versa

The inhibitory activities against both targets are shown to be the most important predictor variables, contributing positively to each other. Amongst the essential features enhancing inhibitory were hydrophobic moieties (SaasC, F08[C-C], Uc, and nCrt), size (D/Dtr10), constituting secondary aliphatic amines (nRNHR), and hybrid group (minssCH2), means the presence of these fragments in the molecules leads to improved inhibitory activity. In contrast, fragments with inhibitory activity lowering potential against the enzyme were ETA_EtaP_F, F06[C-C], nCp, and nAB (see **Figure 4.109**).

4.6.1.3.9. BuChE enzyme inhibitory as Predictor (X) and GSK-3β enzyme inhibitory activity as the response (Y)

The positive regression coefficient of the activity variable indicates that the BuChE enzyme inhibitory value is proportionally related to the GSK-3 enzyme inhibitory activity (the endpoint). The other important variables in this model were N% and minssCH2, having positive regression coefficients towards the inhibitory activity (see **Figure 4.110**).

4.6.1.3.10. BuChE enzyme inhibitory as Predictor (X) and MAO-B enzyme inhibitory activity as the response (Y)

The positive regression coefficient of the activity variable suggests that the BuChE enzyme inhibitory value is consistently connected to the MAO-B enzyme inhibitory activity (endpoint). The essential features improving inhibitory activity were hydrophobic moiety (nCt), polar moieties (nROH, F06[C-O], and B08[N-O]), constituting secondary aliphatic amines (nRNHR) and hybrid group (minsCH3), which suggests that the presence of these features in the compounds leads to better inhibitory activity. On the other hand, fragments with inhibitory activity lowering potential against the enzyme are ETA_Beta_ns_d (see **Figure 4.110**).

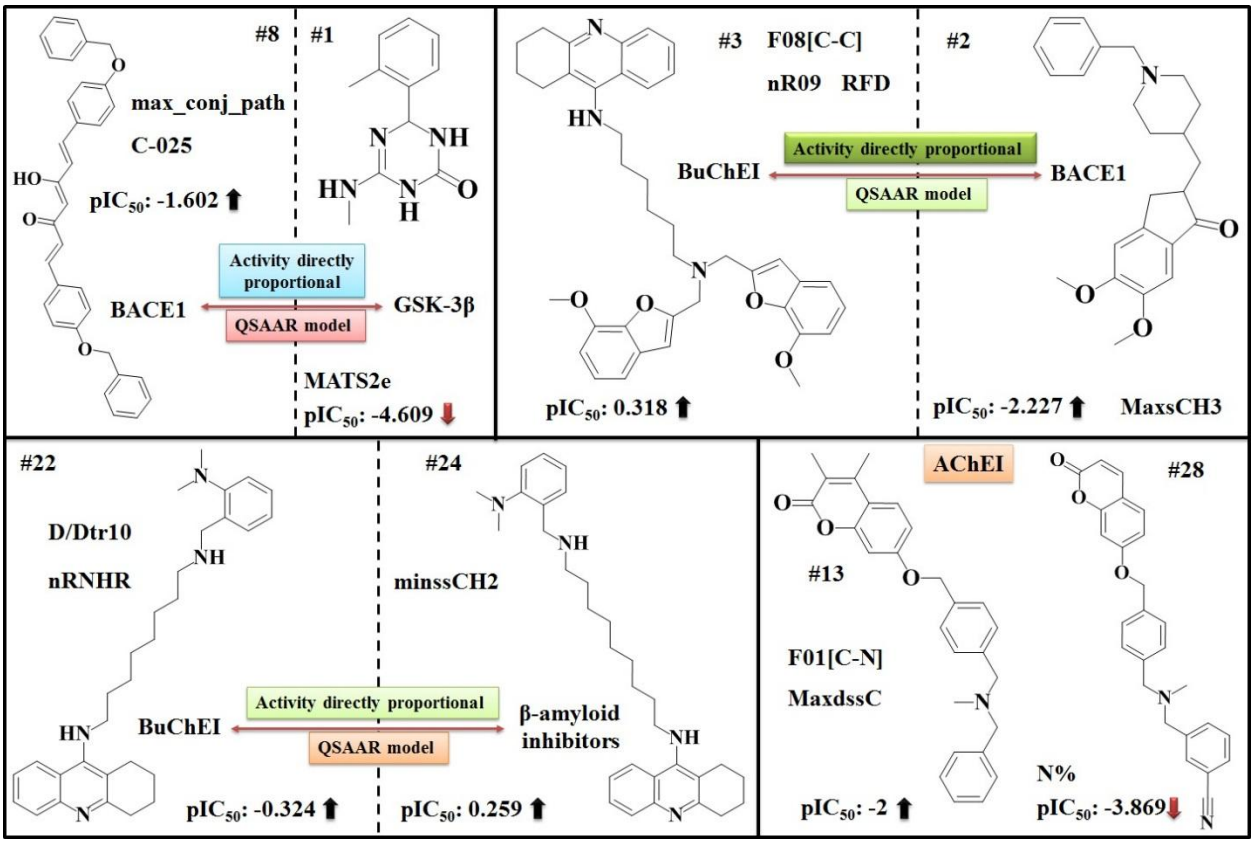


Figure 4.109. Probable mechanistic interpretation of the most significant descriptors obtained from the QSAAR (GSK-3β (X)-BACE1 (Y), BACE1 (X)-GSK-3β (Y), BuChE (X)-BACE1 (Y), BACE1 (X)-BuChE (Y), BuChE (X)-β-Amyloid (Y), β-Amyloid (X)-BuChE (Y) and MAO-B (X)-AChE (Y)) models.

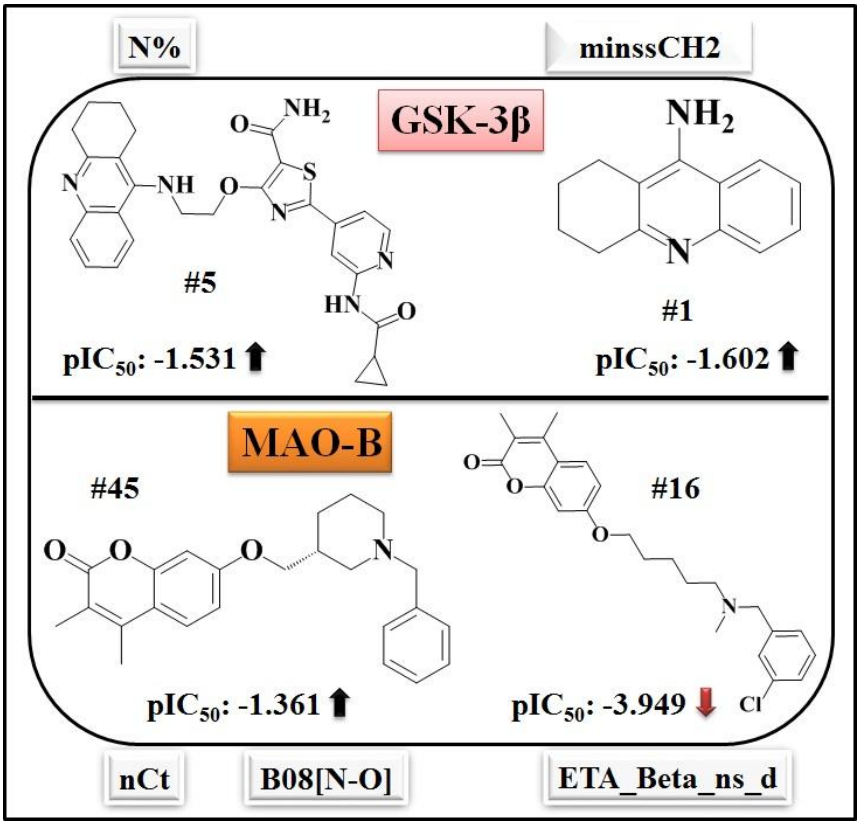


Figure 4.110. Possible mechanistic interpretation of the most significant descriptors obtained from the QSAAR (BuChE (X)-GSK-3β (Y) and BuChE (X)-MAO-B (Y)) models.

4.6.1.4. Applicability domain (AD) study

In the current investigation, the DModX approach[404-405] was applied at the 99% confidence level using SIMCA-P version 10.0 software to perform the AD evaluation of established PLS models, whereas the AD of the developed MLR models was estimated using the standardization approach[300] with the help of tool developed in our laboratory (http://teqip.jdvu.ac.in/QSAR Tools/).

4.6.1.4.1. Individual QSAR models

The DModX outcomes of the PLS models revealed that most of the developed models (against 5-HT6, β-amyloid aggregation, GSK-3β enzyme, and MAO-B enzyme) had no compound in the test set outside the AD. However, in the case of the AChE enzyme model, 48 compounds (i.e. compounds **241, 300, 364, 412, 458, 520, 655, 720, 764, 765, 776, 814, 839, 1077, 1115, 1135, 1138, 1141, 1212, 1216, 1271, 1275, 1299, 1300, 1338, 1352, 1357, 1364, 1366, 1368, 1371, 1379, 1398, 1399, 1411, 1463, 1651, 1652, 1656, 1673, 1679, 1680, 1680, 1688, 1689, 1690, 1693,** and **1695**) in the test set are located outside the AD. In the case of the BuChE enzyme model, 57 compounds (i.e. compounds **186, 210, 255, 257, 281, 294, 419, 438, 451, 478, 580, 635, 637, 641, 645, 651, 807, 879, 928, 933, 955, 966, 970, 972, 981, 1003, 1006, 1012, 1015, 1016, 1405, 1580, 1582, 1617, 1733, 1793, 1797, 1799, 1802, 1817, 1819, 1914, 1928, 2052, 2104, 2156, 2216, 2224, 2237, 2242, 2249, 2256, 2257, 2354, 2374, 2386, and 2387**) in the test set are placed outside the AD. In the case of the BACE1 enzyme model, we found 16 compounds (i.e. compounds **7, 23, 33, 147, 336, 409, 464, 574, 684, 695, 788, 791, 820, 824, 830, and 857**) in the test set are positioned outside the AD. In the case of the CDK-5 protein model, only 3 compounds (i.e. compounds **32, 184,** and **186**) in the

test set are situated outside the AD. In the case of the Gamma-secretase enzyme model, only 2 compounds (i.e. compounds **76** and **80**) in the test set are positioned outside the AD. In the case of the Glutaminyl Cyclase (QCs) enzyme model, only 1 compound (i.e. compound **3**) in the test set is located outside the AD. In the case of the NMDA receptor model, only 3 compounds (i.e. compounds **107**, **338,** and **434**) in the test set are placed outside the AD. In the case of the PDE 10A enzyme model, only 5 compounds (i.e. compounds **58, 79, 182, 204,** and **206**) in the test set are located outside the AD.

4.6.1.4.2. Selectivity-based QSAR models

The AD results of the models revealed that the majority of the developed models (BACE1-GSK-3β enzyme inhibitors, AChE-GSK-3β enzyme inhibitors, BuChE-GSK-3β enzyme inhibitors, AChE-BACE1 enzyme inhibitors, AChE enzyme-β-amyloid aggregation inhibitors, and AChE-MOA-B enzyme inhibitors) had no molecules in the test set outside the AD. However, in the case of the AChE-BuChE enzyme inhibitors model, only 2 compounds (i.e. compounds **29** and **137**) in the test set are positioned outside the AD. In the case of the BuChE-BACE1 enzyme inhibitors model, only 1 compound (i.e. compound **1**) in the test set is located outside the AD. In the case of the BuChE enzyme-β-amyloid aggregation inhibitors model, only 2 compounds (i.e. compounds **62** and **76**) in the test set were placed outside the AD. In the case of the BuChE -MOA-B enzyme inhibitors model, only 1 compound (i.e. compound **9**) in the test set is situated outside the AD.

4.6.1.4.3. QSAAR-based models

AD plots of the developed QSAAR models revealed that except for the β-amyloid (X)-BuChE (Y) model which contained one test set molecule (molecule **14**) and the AChE (X)-β-amyloid (Y) model which contain two test set compounds (i.e. compound **26** and **80**) outside the AD, all of the developed QSAAR models exhibited 100% domain of applicability for all potential combinations.

4.6.1.5. Activity prediction using the developed 2D-QSAR models

In the current work, well-validated QSAR models were individually used to predict the inhibitory activity of four chemical drug-like databases, which had no reported quantitative experimental response values in their source files against the respective targets. Primarily, the established models were utilized to compute the predicted values of database compounds; the validated models were capable of precisely predicting the inhibitory activity of the majority of the molecules, as suggested by the 'alvaRunner version 2.0.4' tool (https://www.alvascience.com/alvarunner/). After prediction, we arranged the compounds based on predicted values (highest to least active). In a detailed analysis of the predicted chemical databases, we have identified the top 56 lead compounds with multitarget inhibitory activity, which are stated in **Table 4.24**.

Table 4.24. List of identified top lead compounds from the databases with multitarget inhibitory activity using developed models.

Compound ID	Smiles	Targets and their predicted inhibitory activity (IC_{50}nm)			
STOCK7S-62250	CC(C1CCC(CC1)(C)c1cc2Cc3cc(c(cc3O)Cc3cc(O)c(cc3C3(C)CCC(CC3)C(C)C)Cc3cc(c(cc3O)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2O)c(c6)C1(C)CCC(CC1)C(C)C)O)c(O)c5)C1(C)CCC(CC1)C(C)C)c(c4)C1(C)CCC(CC1)C(C)C)O)cc3C1(C)CCC(CC1)C(C)C)O)C1(C)CCC(CC1)C(C)C)C1(C)CCC(CC1)C(C)C)C	**AChE**	**BACE1**	**CDK-5**	**GSI**
		0.000474	2.663E-08	0.00243	0.0282
118199	O=CN(C)C(=S)S[Sb](SC(=S)N(C=O)C)SC(=S)Nc1ccc(cc1)S(=O)(=O)Nc1ccc(/N=N/c2ccccc2)cc1	**AChE**	**BACE1**	**GSI**	**BuChE**
		0.00157	2.794E-05	0.0097	7.870
118200	S(C(=S)NC(=O)OCC)[Sb](SC(=S)NC(=O)OCC)SC(=S)Nc1ccc(cc1)S(=O)(=O)Nc1ccc(/N=N/c2ccccc2)cc1	0.00108	0.00061	7.97E-05	1.629
BAS 01060168	c12c3c(C(c4c5c6c(cc4)c(ccc6c(n[nH]5)C)OC)c4ccccc4)ccc1c(ccc2c(n[nH]3)C)OC	**5-HT6**	**BuChE**		**CDK-5**
		0.0119	1.875		0.059
BAS 01060169	C(c1c2c3c(cc1)c(ccc3c(n[nH]2)C)OC)(c1c2c3c(cc1)c(ccc3c(n[nH]2)C)OC)c1ccc(cc1)OC	0.0087	1.954		0.046
STOCK7S-67258	CCCC[Sn](OC(=O)CCCC(N1C(=O)C(=C(C1=O)Cl)Cl)C(=O)O[Sn](CCCC)(CCCC)CCCC)(CCCC)CCCC	**AChE**	**BACE1**		**GSI**
		6.2906E-06	6.96627E-07		0.0948
STOCK7S-70271	CCCCCCC1c2cc3c4cc2OP(=S)(Oc2c1cc1c(c2)OP(=S)(Oc2c(C1CCCCCC)cc1C(c5cc(C3CCCCCC)c(OP(=S)(O4)N(C)C)cc5OP(=S)(Oc1c2)N(C)C)CCCCCC)N(C)C)N(C)C	0.00986	0.000221		2.486E-05
127500	S(C(=S)NC(=O)OCC)[Sb](SC(=S)NC(=O)OCC)Sc1nc2c(s1)cc(/N=N/c1ccc(cc1)NS(=O)(=O)c1ccc(cc1)NC(=O)C(N1CC1)(N1CC1)N1CC1)cc2	8.182E-05	5.273E-06		0.001002
381279	[Si](N([Si](C)(C)C)[Sn](C(Cl)(Cl)Cl)(Cl)N([Si](C)(C)C)[Si](C)(C)C)(C)(C)C	1.5761E-06	1.0843E-25		0.062
633331	P1([Fe+2](P(c2ccccc2)(c2ccccc2)C=C1)([I-])([C-]#[O+])([C-](C(F)(F)F)(C(F)(F)F)F)[C-]#[O+])(c1ccccc1)c1ccccc1	2.878E-07	1.298E-23		0.1194
633993	P1([Fe+2](P(CC1)(c1ccccc1)c1ccccc1)([C-]#[O+])([C-]#[O+])([I-])[C-](C(F)(F)F)(F)C(F)(F)F)(c1ccccc1)c1ccccc1	1.637E-07	2.149E-28		0.119

635009	[Mn+]1([O-]C(=O)C(F)(F)F)([C-]#[O+])([C-]#[O+])([C-]#[O+])P(C=CP1(c1ccccc1)c1ccccc1)(c1ccccc1)c1ccccc1	0.00635	1.429E-19	0.1122
643862	[Sn](CCC)(O[Sn](CCC)(OC(=O)c1c(nccc1)SC)CCC)(OC(=O)c1cccnc1SC)CCC	0.00505	6.882E-13	0.0073
677704	c1c2c(ccc1[N+](=O)[O-])[O-]1[Cu+2]3([N](=C2)[N-]C(=[S]3)NCC)[O-]2c3c(cc(cc3)[N+](=O)[O-])C=[N]3[N-]C(=[S][Cu+2]123)NCC	0.00131	3.599E-09	0.0320
677707	c1c2c(ccc1[N+](=O)[O-])[O-]1[Cu+2]3([N](=C2)[N-]C(=[S]3)N(CC)CC)[O-]2c3c(cc(cc3)[N+](=O)[O-])C=[N]3[N-]C(=[S][Cu+2]123)N(CC)CC	0.00021	1.3122E-09	0.0182
28796-39-6	O=CN(C)C(=S)S[Sb](SC(=S)N(C=O)C)Sc1nc2c(s1)cc(/N=N/c1ccc(cc1)NS(=O)(=O)c1ccc(cc1)NC(=O)C(N1CC1)(N1CC1)N1CC1)cc2	0.00036	2.548E-07	0.112
29878-72-6	[Sn](c1ccccc1)(c1ccccc1)([Sn](c1ccccc1)(c1ccccc1)OC(=O)CCl)OC(=O)CCl	4.6586E-07	3.2537E-27	0.089
69272-27-1	c1c(c(c(nc1C)S[Sn](Cl)(Cl)Sc1c(C(=O)OCC)c(cc(n1)C)C)C(=O)OCC)C	8.9729E-08	0.00046	0.1039
82475-53-4	ClCCN(P(=O)(O[Si](c1ccccc1)(C(C)(C)C)c1ccccc1)O[Si](c1ccccc1)(C(C)(C)C)c1ccccc1)CCCl	0.0034	1.1551E-06	0.026
STOCK7S-64595	CCCCC1(CCCC1)c1cc2Cc3cc(c(cc3O)Cc3cc(O)c(cc3C3(CCCC)CCCC3)Cc3cc(c(cc3O)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2O)c(c6)C1(CCCC)CCCC1)O)c(O)c5)C1(CCCC)CCCC1)c(c4)C1(CCCC)CCCC1)O)cc3C1(CCCC)CCCC1)O)C1(CCCC)CCCC1)C1(CCCC)CCCC1	**BACE1** 1.7698E-07	**GSI** 0.024	**CDK-5** 0.00105
STOCK7S-65432	Oc1cc2Cc3cc(O)c(cc3C3(C)CCCC3)Cc3cc(c(cc3O)Cc3cc(O)c(cc3C3(C)CCCC3)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2C1(C)CCCC1)c(O)c6)C1(C)CCCC1)c(c5)C1(C)CCCC1)O)c(O)c4)C1(C)CCCC1)cc3O)C1(C)CCCC1)C1(C)CCCC1	6.2808E-05	0.069	0.0028
STOCK7S-67132	NC(=O)COc1cc2Cc3cc(OCC(=O)N)c(cc3C3(C)CCCCC3)Cc3cc(c(cc3OCC(=O)N)Cc3cc(OCC(=O)N)c(cc3C3(C)CCCCC3)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2C1(C)CCCCC1)c(OCC(=O)N)c6)C1(C)CCCCC1)c(c5)C1(C)CCCCC1)OCC(=O)N)c(OCC(=O)N)c4)C1(C)CCCC	0.00012	0.0248	0.00244

	C1)cc3OCC(=O)N)C1(C)CCCCC1)C1(C)CCCCC1			
STOCK7S-67182	Oc1cc2Cc3cc(O)c(cc3C3(C)CCCCC3)Cc3cc(cc(c3O)Cc3cc(c(Cc4c(cc(Cc5c(c(Cc1cc2C1(C)CCCCC1)cc(c5)C1(C)CCCCC1)O)c(O)c4)C1(C)CCCCC1)cc3O)C1(C)CCCCC1)C1(C)CCCCC1	0.000313	0.1742	0.0732
STOCK7S-67536	Oc1cc2Cc3cc(O)c(cc3C3(C)CCCCC3)Cc3cc(c(cc3O)Cc3cc(O)c(cc3C3(C)CCCCC3)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2C1(C)CCCCC1)c(O)c6)C1(C)CCCCC1)c(c5)C1(C)CCCCC1)O)c(O)c4)C1(C)CCCCC1)cc3O)C1(C)CCCCC1)C1(C)CCCCC1	9.8853E-07	0.0513	0.00241
STOCK7S-70545	CCCCCCCC1c2cc3c4cc2OP(Oc2c1cc1c(c2)OP(OCCCCCCCC)Oc2c(C1CCCCCCC)cc1C(c5cc(C3CCCCCCC)c(OP(O4)OCCCCCCCC)cc5OP(Oc1c2)OCCCCCCCCC)CCCCCCC)OCCCCCCCC	0.00174	0.1682	0.00034
619178	n12ccc(c3c1c1n(ccc(c1cc3)c1ccccc1)[Co+3]132(n2ccc(c4c2c2n1ccc(c2cc4)c1ccccc1)c1ccccc1)n1ccc(c2c1c1n3ccc(c1cc2)c1ccccc1)c1ccccc1)c1ccccc1.O=C([O-])[C@@H]([O-])[C@@H]([O-])C(=O)[O-]	6.6985E-05	0.0053	0.043
STOCK7S-65614	CCCCOc1ccc(cc1)/C/1=C/2\C=CC(=N2)/C(=c\2/cc/c(=C(/C3=N/C(=C(\c4[nH]c1cc4)/c1ccc(cc1)OCCCC)/C=C3)\c1ccc(cc1)OCCCC)/[nH]2)/c1ccc(cc1)OCCCC	**GSI**	**CDK-5**	**BuChE**
		0.161	0.0044	1.819
332889	O(C(=O)Nc1nc(c(c(c1)Cl)N=C(c1ccccc1)c1ccccc1)N)CC.O(C(=O)Nc1nc(c(c(c1)Cl)N=C(c1ccccc1)c1ccccc1)N)CCC	**5-HT6**	**AChE**	**GSI**
		0.00045	0.0062	2.848E-05
618825	P(c1ccccc1)(c1ccccc1)(c1ccccc1)[Cu+](P(c1ccccc1)(c1ccccc1)c1ccccc1)(/[S]=c\1/cccc[nH]1)[Cl-]	**5-HT6**	**AChE**	**BACE1**
		0.0079	0.000171	7.793E-17
1262-78-8	[Sn](CC(c1ccccc1)(C)C)(CC(c1ccccc1)(C)C)(CC(c1ccccc1)(C)C)CC(c1ccccc1)(C)C	0.0082	0.0049	0.00045
60042-87-7	c1c(cc2c(c1C(C)(C)C)O[Sb](N2)(c1ccccc1)(c1ccccc1)c1ccccc1)C(C)(C)C	9.0369E-06	0.0081	0.0164
60042-88-8	C(c1cc(c2c(c1)N[Sb](O2)(c1ccccc1)(c1ccccc1)c1ccccc1)C(C)(C)C)(c1ccccc1)(c1ccccc1)c1ccccc1	6.0399E-06	0.00498	0.028
81928-48-5	c1cc(c(cc1)[Sn](c1ccccc1C)(Cl)O[Sn](c1ccccc1C)(Cl)c1ccccc1C)C	0.0039	0.00059	3.674E-14
5424-36-2		**5-HT6**	**BACE1**	**CDK-5**

	[Sn](c1cccc2ccccc12)(c1c2c(cccc2)ccc1)(c1cccc2ccccc12)c1c2ccccc2ccc1	0.0112	0.0056	0.00155
26246	C12=C3C4=C5[C-]1[Fe+2]16782345C2=C7[C-]8C6=C12.II	**AChE**	**BACE1**	**CDK-5**
		0.00134	0.000212	0.0681
176220	[C-]12C3=C4C5=C1[Fe+2]16782345[C-]2C1=C6C7=C82.II	0.00134	0.000212	0.0681
118016	n1c(c2[nH]cnc2nc1)S[Sn](c1ccccc1)(c1ccccc1)c1ccccc1.n1c(c2ncn(c2nc1)[Sn](c1ccccc1)(c1ccccc1)c1ccccc1)S[Sn](c1ccccc1)(c1ccccc1)c1ccccc1	**AChE**	**CDK-5**	**GSI**
		3.0972E-06	0.0058	0.00041
118017	n1c(c2[nH]cnc2nc1)S[Pb](c1ccccc1)(c1ccccc1)c1ccccc1.n1c(c2ncn(c2nc1)[Pb](c1ccccc1)(c1ccccc1)c1ccccc1)S[Pb](c1ccccc1)(c1ccccc1)c1ccccc1	6.6093E-09	0.0055	0.000409
403635	c1ccc(cc1CCCCCCCCCCCCCCC)OC(=O)Nc1cccc2c1cccc2.c1ccc(cc1CCCCCCC/C=C/CCCCCC)OC(=O)Nc1cccc2c1cccc2.c1ccc(cc1CCCCCCC/C=C/C/C=C/CCC)OC(=O)Nc1cccc2c1cccc2.c1ccc(cc1CCCCCCC/C=C/C/C=C/CC=C)OC(=O)Nc1cccc2c1cccc2	0.00381	0.000211	0.00644
700533	c12c(c(c3c(n1)c1c(CC3)ccc(c1)c1c3c(nc4c1cccc4)cccc3)c1ccc(cc1)OCc1cc(cc(c1)OCc1cc(cc(c1)C(C)(C)C)C(C)(C)C)OCc1cc(cc(c1)C(C)(C)C)C(C)(C)C)CCc1c2cc(cc1)c1c2c(nc3c1cccc3)cccc2	0.00815	5.435E-08	0.085
682371	[Pt+2]([Cl-])([Cl-])(n1ccc(cc1)/C/1=c/2\cc/c(=C(\c3ccc(cc3)C)/C3=N/C(=C(\c4ccc(/C(=C\5/C=CC1=N5)/c1ccc(cc1)C)[nH]4)/c1ccc(cc1)C)/C=C3)/[nH]2)[O]=S(C)C	**BACE1**	**CDK-5**	**QC**
		0.00696	0.00593	2.973E-05
1803-10-7	N(c1ccccc1)(c1ccccc1)C(=S)S[Sn](SC(=S)N(c1ccccc1)c1ccccc1)(c1ccccc1)c1ccccc1	**AChE**	**BACE1**	**BuChE**
		0.00646	1.9209E-07	4.130
118195	S(C(=S)NC(=O)OCC)[Sb](SC(=S)NC(=O)OCC)Sc1nc2c(s1)cc(cc2)/N=N/c1ccc(cc1)NS(=O)(=O)c1ccc(cc1)NC(=O)C(Cl)Cl	**BACE1**	**GSI**	**BuChE**
		1.4771E-05	0.00182	4.130
636592	c12c(c(c3c4c1c(=O)oc1c4c(c(=O)o3)c(c(c1O)O)c1c(c(O)c(O)cc1C(=O)OC[C@H]1[C@@H](OC(=O)c3c2c(c(O)c(O)c3)O)[C@H]2OC(=O)c3cc(O)c(O)c(O)c3c3c(C(=O)O[C@H]2[C@@H](O1)O)cc(O)c(O)c3O)O)O)O	**AChE**	**GSI**	**MAO-B**
		0.00087	0.0297	3.672E-07
676818	c12c3c(c(c(c1O)O)O)c1c(cc(c(c1O)O)O)C(=O)O[C@H]1[C@H]([C@@H]([C@H]4[C@@H](C5=C([C@H]2[C@](C5=	0.00124	0.0230	5.799E-06

	O)(C(=O)OCC)O)C(=O)O4)O)OC3=O)OC(=O)c2c(c(c(c(c2)O)O)O)c2c(c(c(cc2C(=O)OC1)O)O)O			
676822	c12c3c(c(c(c1O)O)O)c1c(cc(c(c1O)O)O)C(=O)O[C@H]1[C@H]([C@@H]([C@@H]4[C@H]5c6c7c(c(cc6O[C@]65[C@H](C2=C(C6=O)O)C(=O)O4)O)C[C@@H]([C@H](O7)c2cc(c(cc2)O)O)O)OC3=O)OC(=O)c2c(c(c(c(c2)O)O)O)c2c(c(c(cc2C(=O)OC1)O)O)O	0.000485	0.0199	0.00028
676825	[C@H]1([C@H](OC(=O)c2c(c3c(c(c(cc3C(=O)O1)O)O)O)c(c(c(c2)O)O)O)[C@H]1OC(=O)c2c(c(c(c(c2)O)O)O)c2c3c4c5c(c(=O)oc4c(c2O)O)c(c(c(c5oc3=O)O)O)c2c(c(c(cc2C(=O)OC[C@@H]1O)O)O)O)C=O	0.00378	0.01570	4.694E-06
24312-00-3	c1(c(c2c3c(c1O)[C@@H]([C@H]([C@H]1OC(=O)c4c2c(c(c(c4c2c(c(c(cc2C(=O)O[C@H]2[C@H]1OC(=O)c1cc(c(c(c1c1c(c(c(cc1C(=O)OC2)O)O)O)O)O)O)O)O)O)O)O)O)OC3=O)O)O)O	0.000905	0.0338	0.000144
36001-47-5	c1c2c(c3c(cc(c(c3O)O)O)C(=O)O[C@H]3[C@H](COC2=O)OC(=O)c2cc(c(c(c2c2c4c(c5c6c(c(c(c5O)O)O)O)[C@@H]([C@H]([C@@H]3OC4=O)OC6=O)O)c(c(c2O)O)O)O)O)O)c(c(c1O)O)O	0.000905	0.0338	0.000144
173121	[CH-]12[CH]3=[CH]4[CH]5=[CH]1[Co+]162345[CH]2=[CH]1[S]6C2	**AChE**	**BACE1**	**β-amyloid**
		0.00106	3.5773E-12	0.143
633340	[C]123=C[Rh+]456783([C]1(=C8)CC(=C)CC[C]5(=C7)[C]4(=C6)CC2)[Cl-]	7.4994E-06	7.015E-16	0.528
81741-72-2	[Ge]([Ge](N(CC)CC)(CC)CC)(CC)(CC)CC	7.948E-06	1.1578E-07	0.353
81741-73-3	N(CC)(CC)[Ge](CC)(CC)[Ge](CC)(CC)N(CC)CC	1.0373E-05	3.3865E-11	0.457
993-62-4	[Ge](CC)(CC)(CC)[Ge](CC)(CC)CC	2.839E-06	0.00352	0.218
993-63-5	[Sn]([Sn](CC)(CC)CC)(CC)(CC)CC	5.587E-11	5.527E-12	0.236
847446-05-3	N(c1c(C)cc(C=CC#N)cc1C)c1nc(Nc2ccc(C#N)cc2)ncc1.C([C@H](OCP(=O)(O)O)C)n1c2c(nc1)c(N)ncn2.N(c1c2c(n(cn2)[C@@H]2C[C@H](CO)C=C2)nc(N)n1)C1CC1	**CDK-5**	**5-HT6**	**BACE1**
		0.1225	0.00424	0.177

4.6.2. Chemical Read-Across analysis

In the present investigation, the similarity-based quantitative read-across prediction was performed using the same training and test set combinations as used in 2D-QSAR modeling. The current approach employs three distinct similarity-based measures: Laplacian kernel similarity-based (LK) predictions, Gaussian kernel similarity-based (GK), and Euclidean distance-based (ED) estimations, and after hyperparameter optimization, we found that the external validation results obtained from a quantitative read-across algorithm using Gaussian Kernel Similarity-based functions (in case of 5-HT-6, CDK-5, PDE 10A models) and Laplacian kernel similarity-based function (in case of AChE and BACE1 enzyme models) were better compared to the results obtained from the individual regression-based 2D-QSAR models. The details about the external validation results are given in **Table 4.25**.

Table 4.25. External validation results of the **Read-Across analysis.**

Target	Yeuc(Test)		Ygk(Test)		Ylk(Test)		σ-value	γ-value	No. of Similar train comp.
	Q^2F_1	Q^2F_2	Q^2F_1	Q^2F_2	Q^2F_1	Q^2F_2			
5-HT6	0.745	0.743	**0.776**	**0.775**	0.738	0.737	1	1.75	8
AChE	0.729	0.728	0.755	0.753	**0.783**	**0.782**	1.25	1.25	6
BACE1	0.624	0.623	0.666	0.666	**0.675**	**0.675**	0.75	0.75	10
CDK-5	0.903	0.903	**0.917**	**0.917**	0.881	0.880	0.5	0.5	6
PDE 10A	0.678	0.667	**0.784**	**0.777**	0.737	0.729	0.25	0.25	6

4.6.3. Molecular docking analysis

In this exploration, molecular docking was performed using the most and least active compounds from the initial datasets to explore the molecular interactions at the active pocket of the respective targets. The evidence of docking interactions, CDocker interaction energy, and their correlation with the features obtained from the developed 2D-QSAR models are demonstrated in **Tables 4.26** and **4.27**. Moreover, we have also performed the molecular docking analysis using the top predicted compounds from the databases at the active pocket of the respective targets. The details of the docking analysis are demonstrated in **Table 4.28.**

Tables 4.26. Molecular docking results (5-HT6 antagonist, AChE, BACE1 enzyme, β-amyloid, BuChE enzyme, CDK-5 inhibitors) and correlation with 2D-QSAR models in this study.

Compound	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
5-hydroxytryptamine 6 (5-HT6) antagonist				
1 (most active)	35.318	ALA A:266, LYS A:262, HIS A:259, SER A:61, TYR A:134, ASP A:123, LUE A:127	Hydrogen Bond (conventional), Attractive charge, π-sulfur, Alkyl, π-Alkyl	F04[C-O], minssNH, nCp, nArNH2, nBnz
5 (most active)	32.544	SER A:61, ASP A:123, TYR A:134, HIS A:259, LYS A:262, LUE A:127	Hydrogen Bond (conventional and Carbon), Attractive charge, π-sulfur, π-Alkyl, π-π T-shaped	F04[C-O], minssNH, nCp, nArNH2, nBnz
70 (least active)	19.878	ASN A:62, ASN A:59, LYS A:262, HIS A:259	Hydrogen Bond (conventional and Carbon)	F04[C-O]
71 (least active)	17.953	LYS A:262, LEU A:127	Hydrogen Bond (Carbon), π-Alkyl	nBnz
Acetylcholinesterase (AChE) enzyme inhibitor				
841 (most active)	67.407	TYR A:72, TRP A:286, SER A:293, TYR A:341, TRP A:86, TYR A:337, GLU A:202, HIS A:447	Hydrogen Bond (conventional, Carbon and π-donor), π-π T-shaped, π-π stacked, π-Lone pair, Halogen (Cl, Br, I), π-Alkyl, π-Cation	D/Dtr05, DBI, X2v, F04[N-Cl], B05[C-N]
842 (most active)	51.491	THR A:75, TRP A:286, SER A:293, TRP A:86, TYR A:337, TYR A:124, TYR A:341,	Hydrogen Bond (conventional and Carbon), π-π T-shaped, π-π stacked	X2v, DBI, B07[N-N], B05[C-N], minssO, nArCO
714 (least active)	38.559	PHE A:297, TRP A:286, TYR A:72	Hydrogen Bond (conventional), π-Alkyl	minssO

721 (least active)	39.518	TYR A:341, TYR A:72	Hydrogen Bond (conventional), π-Alkyl	minssO
β-secretase 1 (BACE1) enzyme inhibitors				
458 (most active)	56.095	ALA A:335, TYR A:14, THR A:232, GLY A:230, ASP A:228, GLN A:73, TYR A:198, GLY A:34, SER A:35, TYR A:71	Hydrogen Bond (conventional, Carbon and π-donor), π-π T-shaped, π-Anion, π-Lone pair, Halogen (Fluorine), Alkyl, π-Alkyl	F08[N-N], Ui,
463 (most active)	51.796	LEU A:30, SER A:229, THR A:231, TYR A:14, LYS A:9, VAL A:170, GLU A:339, GLY A:13, GLY A:11, ALA A:335, THR A:232, GLY A:230, TYR A:71, THR A:72, ASP A:228, ASP A:32	Hydrogen Bond (conventional, Carbon), Salt bridge, Attractive charge, Sulfur-X, Halogen (Fluorine), Alkyl, π-Alkyl	F08[N-N], Ui, X5v, mindO, nR#CH/X
816 (least active)	31.749	ARG A:128, TYR A:198, ASP A:228	Hydrogen Bond (conventional), Attractive charge, π-Anion	B06[O-O]
831 (least active)	27.353	LYS A:107, GLY A:230, LEU A:30	Hydrogen Bond (conventional and carbon), π-Alkyl	mindO
β-amyloid inhibitors				
208 (most active)	30.398	LEU A:17, LYS A:16, HIS A:13, ALA A:21, VAL A:18	Hydrogen Bond (conventional and carbon), π-Alkyl, Alkyl, π-π T-shaped	SssCH2, PW3, F04[N-O]
124 (least active)	19.144	ALA A:21, LYS A:16	π-Alkyl, π-Cation	PW3
Butyrylcholinesterase (BuChE) enzyme inhibitors				
415 (most active)	62.027	TRP A:82, ALA A:328, PRO A:285, HIS A:438, TRP A:231, PHE A:329, LEU A:286, SER A:287	Hydrogen Bond (conventional, Carbon and π-donor), π-π T-shaped, π-π stacked, π-Cation, π-Alkyl	C-002, N-070, MaxaaaC, MDEC-22, F04[C-N], B07[N-N], nN-N, BLI,

420 (most active)	64.064	ASP A:70, TYR A:332, GLU A:197, HIS A:438, TRP A:231, LEU A:286, ALA A:328, TRP A:430, TRP A:82	Hydrogen Bond (conventional, Carbon), π-π T-shaped, Attractive charge, Alkyl, π-Alkyl	C-002, N-070, MaxaaaC, MDEC-22, F04[C-N], B07[N-N], nN-N, BLI
417 (least active)	31.984	HIS A:438, ASP A:70	Hydrogen Bond (conventional, Carbon)	C-002
2353 (least active)	38.996	GLU A:197, GLY A:115, ILE A:69	Hydrogen Bond (Carbon)	MDEC-22
Cyclin-dependent kinase 5 (CDK-5) inhibitors				
1 (most active)	54.534	LYS A:89, ASP A:86, LEU A:133, ALA A:31, VAL A:18, GLY A:11, GLU A:12, ILE A:10	Hydrogen Bond (conventional, Carbon), Salt bridge, π-Sigma, Alkyl, π-Cation, π-Alkyl	NaaaC, SaasC
2 (most active)	58.694	LYS A:89, ASP A:86, ALA A:31, LEU A:133, PHE A:80, VAL A:18, VAL A:64, CYS A:83, ILE A:10, LYS A:9	Hydrogen Bond (conventional, Carbon), π-π T-shaped, Attractive charge, Alkyl, π-Alkyl	NaaaC, SaasC
146 (least active)	47.848	LYS A:33, LEU A:133, VAL A:18	Hydrogen Bond (Carbon), π-Alkyl	SaasC
194 (least active)	30.142	LEU A:133, PHE A:82, ILE A:10, LYS A:89	Hydrogen Bond (Conventional), π-Alkyl, π-Sulfur	SaasC

Table 4.27. Molecular docking results (γ-secretase, GSK-3β, MAO-B inhibitors, NMDA receptor antagonist, PDE 10A enzyme, Glutaminyl Cyclase (QC) inhibitors) and correlation with 2D-QSAR models in this study.

Compound	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
γ-secretase inhibitors				
180 (most active)	126.087	PRO A:424, PRO A:141, TYR A:453, ARG A:281, GLU A:333, ASN A:142, TYR A:337	Hydrogen Bond (Conventional and Carbon), Alkyl, π-Cation	P_VSA_MR_7, ATS6s, VE3sign_D, VE3sign_D/Dt
208 (least active)	52.566	HIS A:444, PRO A:424, ARG A:281, PRO A:141, TYR A:453, SER A:425	Hydrogen Bond (Carbon), π-Alkyl, π-Cation, Alkyl	P_VSA_MR_7, ATS6s
GSK-3β inhibitors				
41 (most active)	38.688	VAL A:70, ILE A:62, VAL A:135, ALA A:83, TYR A:134, LEU A:188, ASP A:133	Hydrogen Bond (Conventional, Carbon), π-π stacked, π-Alkyl	SaasC, PDI, SaaaC, F06[N-O], nPyrroles, B05[N-O]
112 (least active)	18.996	LEU A:188, ALA A:83	π-Alkyl	SaaaC
MAO-B inhibitors				
72 (most active)	47.999	LYS A:296, TRP A:388, VAL A:294, GLY A:57, CYS A:397, TYR A:435, ILE A:14, SER A:15, GLY A:13, THR A:426, ARG A:42, ALA A:439	Hydrogen Bond (Conventional, Carbon), Amide-π stacked, π-Alkyl, Alkyl	F02[O-O], C%
87 (most active)	44.213	ILE A:14, SER A:15, GLY A:13, THR A:426, ALA A:439, TYR A:435, ARG A:42, CYS A:397, GLY A:57, GLY A:434	Hydrogen Bond (Conventional, Carbon), Amide-π stacked, π-Alkyl, Alkyl, Halogen (Fluorine)	F02[O-O], C%

14 (least active)	37.933	TYR A:398, CYS A:397, ARG A:42, ILE A:14, ALA A:439, THR A:426	Hydrogen Bond (Carbon), π-Alkyl, Alkyl	C%
25 (least active)	32.254	GLY A:58, GLY A:40, ARG A:42, MET A:436, TYR A:435, TYR A:398, GLN A:206	Hydrogen Bond (Carbon), π-π T-shaped, π-Sigma, π-Alkyl, Alkyl, Halogen (Cl, Br, I)	C%
NMDA receptor antagonist				
485 (most active)	32.035	TYR A:184, GLN A:144, LEU A:146, SER A:180, THR A:126, HOH A:1036, HOH A:1111, GLN A:13, TRP A:223, PRO A:15, VAL A:227, ALA A:226, PHE A:16, ASP A:224	Hydrogen Bond (Conventional, Carbon and Water), π-π T-shaped, π-Lone pair, π-Sigma, π-Alkyl, Alkyl, Halogen (Fluorine)	ATSC1p, C-028, TPSA(Tot), B10[C-O]
486 (most active)	32.578	PHE A:92, LEU A:125, THR A:126, PRO A:124, HOH A:1078, HOH A:1036, ASP A:224	Hydrogen Bond (Conventional, Carbon and Water), π-Anion, π-Alkyl, Alkyl,	ATSC1p, C-028, TPSA(Tot)
398 (least active)	27.011	GLN A:144, ASP A:224, PRO A:124, PHE A:92, THR A:126	Hydrogen Bond (Carbon), π-Lone pair, π-Alkyl, Halogen (Fluorine)	C-028
455 (least active)	21.838	PHE A:92, PHE A:246, LEU A:146, TYR A:184	Hydrogen Bond (Carbon), π-π T-shaped, π-π stacked, π-Alkyl	TPSA(Tot)
Phosphodiester 10A (PDE 10A) enzyme inhibitors				
226 (most active)	51.832	TYR A:524, ASP A:674, SER A:677, ILE A:692, LEU A:675, GLN A:726, PHE A:729, PHE A:696, LEU A:635, ALA A:636, SER A:571, ASN A:572, GLU A:592	Hydrogen Bond (Conventional and Carbon), π-π T-shaped, π-π stacked, π-Alkyl	N-070, AMR, minsCH3, Eta_epsi_2

228 (most active)	57.044	LEU A:675, SER A:677, PHE A:729, GLN A:726, ILE A:692, MET A:713, PHE A:696, PHE A:570, HIS A:525, GLU A:592, THR A:633, LEU A:635, ASP A:674	Hydrogen Bond (Conventional and Carbon), Salt bridge, π-π T-shaped, π-π stacked, π-Sigma, Alkyl, π-Alkyl, π-Cation, Attractive charge	N-070, AMR, minsCH3, Eta_epsi_2
217 (least active)	36.980	LEU A:635, HIS A:567, PHE A:696, PHE A:729, ILE A:692, GLN A:726	Hydrogen Bond (Conventional and π-Donor), π-π T-shaped, π-π stacked, π-Alkyl	Eta_epsi_2, AMR
221 (least active)	35.608	LEU A:675, ILE A:692, PHE A:729, TYR A:524, HIS A:525, ASP A:634, LEU A:635	Hydrogen Bond (Carbon), π-π T-shaped, π-π stacked, π-Alkyl, π-Cation	Eta_epsi_2, AMR
Glutaminyl Cyclase (QC) inhibitors				
87 (most active)	131.348	TRP A:329, LYS A:144, HIS A:330, ASP A:159, GLU A:202, GLU A:201, SER A:160, MET A:167, PRO A:163, CYS A:139, LEU A:137, LEU A:246, ALA A:138, LEU A:247, ASP A:248, ASP A:305	Hydrogen Bond (Conventional, Carbon and π-Donor), Attractive charge, Alkyl, π-π T-shaped, π-Anion, π-Cation, π-Alkyl	C-034, B09[C-O], T(N..S)
88 (most active)	128.118	TRP A:329, TRP A:207, HIS A:330, ASP A:159, GLU A:201, SER A:160, MET A:167, PRO A:163, CYS A:139, LEU A:137, VAL A:245, LEU A:247, ASP A:248, ASP A:305	Hydrogen Bond (Conventional and Carbon), Attractive charge, Alkyl, π-Sigma, π-Anion, π-Alkyl	C-034, B09[C-O], T(N..S)
31 (least active)	15.879	TRP A:329, TRP A:207, ILE A:303	Alkyl, π-Alkyl	T(N..S)
33 (least active)	17.698	GLN A:304, TRP A:207	Hydrogen Bond (Conventional), π-Cation	T(N..S)

Table 4.28. Molecular docking analysis of top predicted compounds from the databases.

Compound	Smiles notation	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	IC_{50}nM (Predicted)
5-hydroxytryptamine 6 (5-HT6) antagonist					
434-91-3	Fc1c(c2c(N)c(c(N)c(c2F)C)C)c(N)c(c(N)c1C)C	24.504	LEU A: 57, ASN A: 62, LEU A: 65, SER A: 61, ARG A: 124, HIS A: 259, LEU A: 127, TYR A: 134	Conventional Hydrogen Bond, Carbon hydrogen bond, Halogen (Fluorine), Alkyl, Pi-Alkyl	5.62341E-06
2475-45-8	O=C1c2c(C(=O)c3c1c(N)ccc3N)c(N)ccc2N	20.305	MET A: 324, ARG A: 325, ASP A: 326, LYS A: 262, HIS A: 259	Conventional Hydrogen Bond, Carbon hydrogen bond, Pi-Alkyl, Pi-sulfur, Pi-Cation	5.04661E-06
Acetylcholinesterase (AChE) enzyme inhibitor					
ZINC00628557	O=C1Nc2ccc(S(=O)(=O)N3CCCC3)c3ccc(S(=O)(=O)N4CCCC4)c1c23	47.905	LEU A: 76, TYR A: 72, TYR A: 337, TRP A: 286, TYR A: 124, PHE A: 338, VAL A: 294, TYR A: 341, PHE A: 295	Conventional hydrogen bond, Carbon hydrogen bond, Pi-Pi stacked, Pi-Sulfur, Pi-Alkyl, Alkyl	0.00818
ZINC67805358	CCOCCN1CCN(S(=O)(=O)c2sc3c(c2C(=O)O)CCNC3)CC1	47.629	TYR A: 341, SER A: 293, TYR A: 124, TYR A: 72, TRP A: 86, GLY A: 121, PHE A: 338	Conventional hydrogen bond, Carbon hydrogen bond, Pi-Pi T-Shaped, Pi-Sulfur, Pi-lone pair	0.00853
β-secretase 1 (BACE1) enzyme inhibitors					
ZINC00952595	C[C@H]1CCc2c(sc3ncn(CCn4cnc5sc6c(c5c4=O)CC[C@H](C)C6)c(=O)c23)C1	46.620	GLN A: 73, TYR A: 71, ILE A: 110, ASP A: 32, GLY A: 230, THR A: 231, ASP A: 228, TYR A: 198	Conventional hydrogen bond, Carbon hydrogen bond, Pi-donor hydrogen bond, Pi-Pi T-shaped, Pi-Anion Pi-Alkyl, Alkyl	0.147

ZINC72147875	CCn1cc(-c2c(C#N)c(N)nc3c2CN(c2ncccn2)CC3)cc1C#N	39.711	THR A: 232, GLY A: 11, GLY A: 230, ASP A: 228, ASP A: 32, TYR A: 71, GLN A: 73, LYS A: 107	Conventional hydrogen bond, Carbon hydrogen bond, Pi-donor hydrogen bond, Pi-Pi T-shaped, Pi-Anion Pi-Alkyl	0.163
β-amyloid inhibitors					
STOCK1S-18299	c12c(C(=NO)c3c(C1=NO)[n+](on3)[O-])[n+](on2)[O-]	21	LYS A: 16, GLN A: 15, VAL A: 12	Conventional Hydrogen Bond, Carbon hydrogen bond, Pi-Alkyl, Pi-sigma	0.1213
ZINC43762221	C1C[C@@]23CO[C@@]45CC[C@@]61CC[C@@] (CO2)(OC4)[C@@]	32	PHE A: 20, LEU A: 17, HIS A: 13, LYS A: 16	Conventional Hydrogen Bond, Carbon hydrogen bond, Pi-Alkyl, Pi-sigma, Alkyl	0.0952
Butyrylcholinesterase (BuChE) enzyme inhibitors					
6304-69-4	O(C(=O)N(NC(=O)OCC)c1c(cc(N(NC(=O)OCC)C (=O)OCC)c(c1)C)C)CC	55.125	PRO A: 84, PHE A: 329, GLY A: 117, LEU A: 286, TRP A: 231, PHE A: 398, HIS A: 438, TRP A: 82, GLU A: 197, GLY A: 116	Conventional Hydrogen Bond, Carbon hydrogen bond, Pi-Alkyl, Alkyl	0.0460
231883	C(=O)(OC[C@H](OC(=O)Nc1ccccc1)[C@H](OC(=O)Nc1ccccc1) COC(=O)Nc1ccccc1)Nc1ccccc1	70.527	GLY A: 115, GLY A: 117, ALA A: 199, HIS A: 438, GLY A: 116, LEU A: 286, TRP A: 82, ALA A: 328, ASP A: 70, ILE A: 69	Conventional Hydrogen Bond, Carbon hydrogen bond, Alkyl, Pi-Alkyl, Pi-Pi T-shaped, Pi-donor hydrogen bond, Pi-Cation, Amide-Pi stacked, Pi-anion	0.01425
Cyclin-dependent kinase 5 (CDK-5) inhibitors					
ZINC00352612	c1ccc2c(c1)nc1n2c2nc3ccccc3n2c2nc3ccccc3n12	44.799	ALA A: 31, VAL A: 18, LEU A: 133, CYS A: 83, ILE A: 10, LYS A: 89, ASP A: 86, GLU A: 12, ASN A: 144	Conventional Hydrogen Bond, Pi-Alkyl, Pi-sigma, Salt Bridge	0.0073

ZINC22712768	Cc1c(c2cc(c3nc4c5sc6nc(C)cc(C)c6c5ncn4n3)c3ccccc3n2)cnn1C	45.249	ASN A: 144, ALA : 31, LEU A: 133, ILE A: 10, GLU A: 12, ASP A: 86, VAL A: 18	Conventional Hydrogen Bond, Carbon Hydrogen bond, Pi-Anion, Alkyl, Attractive charge, Pi-Alkyl, Pi-sigma, Pi-donor hydrogen bond	0.00427
γ-secretase inhibitors					
ZINC25228619	CC(C)c1ccc(N2C(=O)Cc3c2nc(N)c2c(N)nc(N4CCOCC4)c(C#N)c32)cc1	81.769	SER A: 426, PRO A: 424, GLU A: 364, SER A: 137, THR A: 334, PRO A: 141, TYR A: 453, ARG A: 281, ALA A: 298	Conventional hydrogen bond, Carbon hydrogen bond, Pi-donor hydrogen bond, Pi-Cation, Alkyl, Attractive charge, Pi-Alkyl	0.00279
ZINC25748138	CCN(CC)c1nc(N)c2c(N)nc3c(cc(O)n3c3cc(Cl)cc(Cl)c3)c2c1C	45.159	PRO A: 141, TYR A: 453, ARG A: 281, PRO A: 423, VAL A: 440, SER A: 425, PRO A: 424, GLU A: 364, GLU A: 333, ASN A: 142, ARG A: 285, SER A: 137, TYR A: 173, THR A: 334	Conventional Hydrogen Bond, Carbon hydrogen bond, Alkyl, Pi-Alkyl, Pi-Cation, Pi-lone pair	2.46037E-05
GSK-3β inhibitors					
1349658-81-6	[C@H](Cc1ccccc1)([C@@H]([C@@H]([C@H](Cc1cncs1)NC(=O)OCc1cncs1)O)O)NC(=O)[C@@H](NC(=O)N(Cc1nc(C(C)C)sc1)C)C	42.81	LEU A: 81, ARG A: 141, ILE A: 62, VAL A: 70, CYS A: 199, VAL A: 110, LEU A: 188, LEU A: 132, VAL A: 135, ALA A: 83, ASP A: 133, GLN A: 185, PRO A: 136, GLN A: 185, TYR A: 134	Conventional hydrogen bond, Carbon hydrogen bond, Pi-lone pair, Pi-sigma, Alkyl, Pi-Alkyl	0.132

1236407-24-1	[C@H](Cc1ccccc1)([C@@H]([C@@H]([C@H](Cc1cncs1)NC(=O)OCc1cncs1)O)O)NC(=O)[C@@H](NC(=O)N(Cc1nc(C(C)C)sc1)C)C	43.537	CYS A: 199, PHE A: 67, ASN A: 186, VAL A: 70, TYR A: 134, GLN A: 72, ILE A: 62, ARG A: 141, ASN A: 64, LYS A: 183	Conventional hydrogen bond, Carbon hydrogen bond, Pi-Sulfur, Pi-sigma, Alkyl, Pi-Alkyl	0.297
MAO-B inhibitors					
STOCK1S-91910	OS(=O)(=O)c1cccc(c1)c1csc2n1c1cc(ccc1n2)S(=O)(=O)O	34.116	TRP A: 388, GLY A: 57, CYS A: 397, MET A: 436, CYS A: 172, TYR A: 435, TYR A: 398	Conventional hydrogen bond, Carbon hydrogen bond, Pi-Sulfur, Sulfur-X, Alkyl, Pi-Alkyl, Pi-Pi stacked, Pi-Anion	1.7338E-06
ZINC40061664	COC(=O)c1ccc(C)c(S(=O)(=O)Oc2ccc3ccc(=O)oc3c2)c1	50.764	TYR A: 60, GLY A: 57, TYR A: 398, GLY A: 58, MET A: 436, CYS A: 397, GLY A: 425, THR A: 426, ALA A: 439, ARG A: 42, TYR A: 435, GLN A: 206	Conventional hydrogen bond, Carbon hydrogen bond, Alkyl, Pi-Alkyl, Amide-Pi stacked	0.000176
NMDA receptor antagonist					
STOCK1S-09681	n1(c(cc(=N)c2c1cccc2)C)C	25.842	ALA A: 226, ASP A: 224, VAL A: 227, PRO A: 124, THR A: 126, PHE A: 92	Conventional Hydrogen Bond, Alkyl, Pi-Pi stacked, Pi-Anion, Pi-Cation	0.1815
STOCK2S-07694	C(=O)(N[O-])COc1c(cc(cc1)Cl)Cl	31.112	TRP A: 223, GLN A: 144, HOH A: 1036, HOH A: 1078, PHE A: 250, PHE A: 92, PRO A: 124, ALA A: 226, ASP : 224	Conventional Hydrogen Bond, Water hydrogen bond, Alkyl, Pi-Alkyl, Pi-Pi stacked, Pi-Anion	0.1945
Phosphodiester 10A (PDE 10A) enzyme inhibitors					

1091628-86-2	N(C12CC3CC(C2)CC(C1)C3)c1nc(NCC(=O)O)nc(NCCCCC)n1	46.270	ALA : 689, VAL A: 678, ILE A: 692, PHE A: 729, LEU A: 635, ASP A: 674, TYR A: 524, HIS A: 529, HIS A: 563, HIS A: 525, HIS A: 567	Conventional hydrogen bond, Carbon hydrogen bond, Alkyl, Pi-Alkyl, Pi-Pi stacked, Attractive charge	0.000192
1422258-88-5	C(Nc1nc(NCCNC(=O)CCl)nc(NCC#C)n1)(C)C12CC3CC(C2)CC(C1)C3	42.86	SER A: 571, HIS A: 567, THR A: 633, ASP A: 674, VAL A: 678, HIS A: 525, LEU A: 675, ILE A: 692, SER A: 677, PHE A: 696, PHE A: 729, MET A: 713, LEU A: 635	Conventional hydrogen bond, Carbon hydrogen bond, Alkyl, Pi-Alkyl, Pi-Cation	0.000609
Glutaminyl Cyclase (QC) inhibitors					
691803	s1c(ccc1c1sc(cc1)c1sc(cc1)c1sc(cc1)CN)CN	44.8781	GLU A: 201, ASP A: 248, TRP A: 329, TRP A: 207, HIS A: 330	Conventional Hydrogen Bond, Carbon hydrogen bond, Pi-Pi stacked, Attractive charge, Pi-Sulfur, Salt Bridge	6.45654E-12
ZINC23130127	Cc1cc(c2n[nH]c3c2CN(C(=O)c2cc4sccc4[nH]2)CC3)c(C)o1	27.4382	LYS A: 144, HIS A: 330, TRP A: 207, GLN A: 304, ILE A: 303, GLU A: 201, TRP A: 329, HIS A: 140	Conventional Hydrogen Bond, Carbon hydrogen bond, Pi-Pi stacked, Pi-donor hydrogen bond, Pi-Sulfur, Pi-Sigma, Pi-Cation, Pi-Pi T-shaped, Pi-anion	1.65959E-08

4.6.4. Comparisons of proposed QSAR models with our previous published studies

In the present exploration, a comparison of the best models of this study with our previously published models[60, 61, 49] was performed. The details of different internal and external validation parameters obtained from this study and those obtained from previous models are given in **Table 4.29**.

Table 4.29. Comparisons of proposed QSAR models with our previous published studies.

	Sources	E. L.	LV	Model	Training set			Test set		
					n	R^2	Q^2	n	Q^2F_1	Q^2F_2
Models in this study	AChE model	28	7	PLS	1325	0.635	0.621	408	0.678	0.678
	BuChE model	28	8	PLS	1882	0.689	0.668	625	0.702	0.702
	BACE1 model	18	9	PLS	680	0.669	0.650	225	0.675	0.674
	β-Amyloid model	6	4	PLS	197	0.729	0.705	65	0.844	0.844
Published models	AChE model[60]	15	6	PLS	798	0.662	0.645	199	0.661	0.660
	BuChE model[60]	13	5	PLS	603	0.674	0.656	158	0.663	0.660
	BACE1 model[111]	5	3	PLS	76	0.826	0.795	22	0.846	0.846
	β-Amyloid model[112]	12	6	PLS	252	0.664	0.621	62	0.765	0.763

CHAPTER - 5

CONCLUSION

Chapter 5: Conclusion

In the current thesis work, multiple *in silico* approaches were used to investigate prospective treatments for Alzheimer's disease, which is currently incurable. The main objective was to employ appropriate computational tools for finding and developing possible anti-Alzheimer agents against a number of key targets involved in AD. In this work, we have focused on twelve main biological targets, namely, **AChE**, **BuChE**, **BACE1**, **5-HT6**, **CDK-5**, **Gamma (γ)-secretase**, **Glutaminyl Cyclase (QCs)**, **GSK-3β**, **NMDA receptor**, **amyloid-beta plaques (Aβ)**, **PDE 10A** and **MAO-B** enzyme are targeted to find appropriate treatment in AD. We have emphasized both on single-target drug designing and multi-target drug designing strategies using various *in silico* methodologies, particularly QSAR. Designing multi-target directed ligands (MTDLs) is essential since AD is multifactorial and linked to the concurrent dysfunction of numerous enzymes in the affected brain. In **study 1**, using 2D QSAR and molecular docking techniques we have managed to understand the structural features that are involved in improving the inhibitory activity against the BuChE enzyme. In **study 2**, we performed 2D QSAR analyses, pharmacophore mapping, and molecular docking study by employing **98** heterocyclic compounds as human BACE1 inhibitors to identify the vital structural features that are responsible for their activity. In **study 3**, we have implemented 2D QSAR analyses, pharmacophore mapping, and molecular docking study by using **314** heterocyclic compounds as β-amyloid aggregation inhibitors to identify the vital structural features that are responsible for their inhibition. In **study 4**, a congeneric series of 78 carbamate derivatives with inhibitory activity information against the AChE enzyme was analyzed using several *in silico* techniques (2D QSAR, GQSAR, pharmacophore mapping, and molecular docking study) to recognize the vital structural features and also understand the chemical features required for interaction with the AChE enzyme and the key active site residues involved in the intermolecular interactions. In **study 5**, we implemented 2D-QSAR modeling using two different datasets, namely, AChE and BuChE enzyme inhibitors. A third dataset has been derived based on their selectivity against both these enzymes to identify which structural fragments/properties are essential specifically to inhibit a specific type of enzyme. Additionally, molecular docking analysis was performed using the most and least active compounds from the datasets and tried to rationalize the influences of different descriptors/features as apparent from the 2D-QSAR models. The final *in silico* study (**study 6)** was specifically designed to identify potential compounds that are capable of simultaneously hitting multiple targets (*all twelve above-mentioned targets were studied*) against AD. These *in silico* studies provide us with the necessary knowledge for future research on changing the scaffolds for improved therapeutic efficacy against the devastating Alzheimer's disease, in addition to the crucial information gathered about achieving the activity against various biological targets explored.

Additionally, all the computational models (including the 2D QSAR and pharmacophore) developed in this investigation were properly validated using strict internal and external validation methodologies. To avoid chance correlation, these models were also subjected to randomization testing. As a result, these models can be extensively employed to screen query molecules or databases, predict the biological activity of recently designed compounds, classify compounds into active or inactive categories against studied targets, and last but not least, for designing novel analogs with enhanced activity. Finally, the precise information revealed from all of the studies that were implemented in this thesis work is described individually as follows:

5.1. Study 1- A Multi-layered Variable Selection Strategy for QSAR Modeling of Butyrylcholinesterase Inhibitors

Alzheimer's disease (AD) is an age-related neurodegenerative disorder, which is the most common cause of dementia in elderly individuals. The current treatment strategy for AD patients is the use of AChE enzyme inhibitors, which give only symptomatic relief. However, recent studies indicated a long-lasting effect in a certain percentage of patients[2, 5]. There is accumulating evidence that AChE and BuChE have secondary non-cholinergic functions including the processing and deposition of β-amyloid (Aβ)[2, 5]. BuChE and AChE could play a role in the Aβ metabolism and during an early step in the development of the senile plaque, as revealed by the finding that AChE and BuChE accelerate Aβ deposition[2, 5]. In the present study, a PLS-regression-based 2D-QSAR model was developed using a multi-layered feature selection strategy for the prediction of inhibitory activity against the BuChE enzyme using a large dataset containing 1130 diverse molecules. The PLS model was developed by following strict OECD guidelines (a defined endpoint, unambiguous algorithm, acceptable quantitative metrics, applicability domain analysis, and mechanistic interpretation). The QSAR model was analyzed, and the structural features (hydrophobic, ring aromatic, and hydrogen bond acceptor/donor) responsible for the enhancement of the activity were identified. The developed model further suggests that the presence of hydrophobic features like a long carbon chain would increase the BuChE inhibitory activity and the presence of amino group and hydrazine fragment promoting the hydrogen bond interactions would be important for increasing the inhibitory activity against the BuChE enzyme. Furthermore, molecular docking studies have been carried out to understand the molecular interactions between the ligand and receptor, and the results are then correlated with the structural features obtained from the QSAR models. The information obtained from the QSAR models is well corroborated with the results of the docking study.

5.2. Study 2- Exploring 2D-QSAR for prediction of Beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease

The pathogenesis of Alzheimer's disease (AD) is highly complex. The beta-secretase 1 (BACE1) enzyme plays an important role in the abnormal production of β-amyloid plaques (Aβ), which is a major hallmark of the pathophysiology of the disease. In the present work, a PLS-regression-based 2D-QSAR model was developed from **98** diverse classes of compounds having defined BACE1 enzyme inhibitory activity to investigate the structural requirements or molecular properties essential for the enzyme inhibitory activity. The 2D QSAR model was developed with simple, meaningful, and easily interpretable descriptors. Before the development of the final model, a multilayered variable selection strategy using stepwise regression followed by best subset selection was applied for the selection of significant descriptors. The statistical results of the developed model show good predictivity based on both internal and external validation parameters. The PLS model was developed by following strict OECD guidelines. The information obtained from the PLS model, the authors have concluded that:

i. The presence of hydrogen bond donor groups like –OH, $–NH_2$, –SH, etc. may enhance the inhibitory activity against the BACE1 enzyme;
ii. The higher number of heteroatoms with a lone pair of electrons capable of resonance with an aromatic nucleus is essential to increase the inhibitory activity against the BACE1 enzyme;
iii. The R--CH..X fragment is favorable to enhance the inhibitory activity of β-secretase enzyme inhibitors;
iv. A higher number of H atoms attached to C^1(sp3)/C^0(sp2) and the presence of N–S fragments at topological distance 3 in the molecules are detrimental to the enzyme inhibitory activity.

Additionally, pharmacophore mapping revealed that the ring aromatic (RA), hydrophobic, and hydrogen bond acceptor (HBA) features are an initial necessity for the inhibitory activity against the BACE1 enzyme, and these features are well corroborated with the descriptors obtained from the 2D-QSAR model. Furthermore, the results obtained from the molecular docking study are well supported by the QSAR and pharmacophore analysis. The developed 2D-QSAR model thus may be helpful for the prediction of the activity of new analogs even before their synthesis and evaluation.

5.3. Study 3- Cheminformatic modeling of β-amyloid aggregation inhibitory activity against Alzheimer's disease

In this research, chemoinformatic tools were applied to examine a set of 314 heterocyclic compounds with defined β-amyloid aggregation inhibitory activity to identify the structural requirements essential for the inhibitory activity. The QSAR models were developed with simple, meaningful, and easily interpretable descriptors. Before the development of the final models, a multilayered variable selection strategy using stepwise regression followed by best subset selection was applied to investigate the meaningful descriptors, and the final models were built by using the PLS regression-based methodology following strict OECD guidelines. The obtained results suggest that the developed model showed good predictivity based on both internal and external validation parameters. The information obtained from the PLS models suggested the following: (i) The presence of a secondary aromatic amine in compounds, (ii) The presence of C-N fragments at the topological distance 5, and (iii) Aromatic carbons with an attached substituent atom are favorable for enhancing β-amyloid aggregation inhibitory activity; whereas, (iv) A higher number of the normalized number of ring systems and a furane ring in the molecules is detrimental to the inhibitory activity. Furthermore, pharmacophore mapping revealed that the RA, hydrophobic, hydrophobic aromatic and HBA features are an initial necessity for the inhibitory activity and these features are well corroborated with the nArNHR, SaasC, F09[C-C], SaasC, B05[C-N] and F05[O-O] descriptors of the 2D-QSAR model. Moreover, the results obtained from molecular docking analysis are well supported by the QSAR and pharmacophore analysis. The developed 2D-QSAR models may be useful for predicting the activity of new analogs even before their synthesis and evaluation. Features obtained from the developed models (2D-QSAR and 3D-pharmacophore) and molecular docking can be helpful for the design of novel inhibitors against β-amyloid aggregation, and the overall approach established can be adopted for ligand-based drug-design campaigns.

5.4. Study 4- Chemometric modeling of structurally diverse carbamates for the inhibition of acetylcholinesterase enzyme (AChE) in Alzheimer's disease

In the current study, 2D-QSAR and GQSAR models were developed for the prediction of inhibitory activity against the AChE enzyme using a dataset containing 78 carbamate derivatives. The statistical results of the developed models show good predictivity based on both internal and external validation parameters. From the evidence obtained from developed models (2D QSAR and GQSAR) authors have concluded that:

i. The higher number of steric features in the molecule may enhance the inhibitory activity against the AChE enzyme,
ii. The higher number of ring systems present in the molecules is essential to increase the inhibitory activity against the AChE enzyme as also corroborated with the developed pharmacophore model,
iii. The CH2RX fragment and H attached to C0(sp3) with 1X attached to the next "C" are detrimental to the AChE enzyme inhibitory activity,
iv. Relatively less distance between the most hydrophobic and hydrophilic group at the R2 position may favor the inhibitory activity against the AChE enzyme,

v. An increase in the positive electrostatic potential of fragment R2 may lead to an increase in the inhibitory activity against the AChE enzyme,
vi. Reduction in the number of –CH2 groups which are connected with the help of two single bonds at the R1 position would be better for the inhibitory activity of the compounds and,
vii. The absence of the N-CH3 group at the R3 position may favor the inhibitory activity against the AChE enzyme.

Also, pharmacophore mapping revealed that the RA, HBA, Hydrogen bond donor (HBD) and hydrophobic features are preliminary requirements for the inhibitory activity and these features are well corroborated with the NRS, +vePotentialSurfaceArea, MomInertiaX (moment of inertia at X-axis), F06 [C-N], and X2A descriptors of the 2D-QSAR and GQSAR models. Moreover, molecular docking analysis revealed that hydrogen and hydrophobic bonding can be correlated with X2A, R2-+vePotentialSurfaceArea, R3-MomInertiaX, and NRS descriptors in the QSAR models. X2A, R2-+vePotentialSurfaceArea, and R3-MomInertiaX are related to hydrogen bonding and electrostatic interactions between protein and ligand. The descriptor, NRS, gives evidence of hydrophobic interaction. Thus, from the above-stated information, authors concluded that hydrogen bonding, hydrophobicity, electrostatic interactions, and unsaturation (π-π interaction) feature as obtained from both QSAR and docking study are essential for the inhibitory activity against the AChE enzyme. The developed QSAR models thus may be helpful for the prediction of the activity of new analogs even before their synthesis and evaluation.

5.5. Study 5- In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease

In the present investigation, authors have employed chemoinformatic tools to investigate the datasets of 997 and 761 heterocyclic compounds with defined AChE and BuChE enzyme inhibitory activities, respectively, to investigate the important structural features for enzyme inhibition. Additionally, 198 heterocyclic compounds from the same datasets having dual inhibitory activity against AChE and BuChE enzymes have been considered for exploring selectivity patterns. Significant and easily interpretable 2D descriptors are calculated for model development. Before the development of the final models, a multilayered variable selection strategy was employed for the selection of significant descriptors. The PLS regression-based methodology was used for the developed final models following the OECD guidelines. The statistical results obtained from the developed models exhibited acceptable quality in terms of both internal and external validation matrices. From the insights obtained from generated PLS models, authors have concluded that: a higher number of -CH_2- groups (corresponding SssCH2 descriptor values not more than 40), number of secondary aromatic amines (corresponding descriptor values not more than 4), smaller ring size (size of the ring corresponding to ring index not more than order 8), branching in the cationic structure (one central atom is attached to three other non-hydrogen atoms) and number of aromatic ketone groups (not less than 2 fragments) may be more favorable for the inhibitory activity against AChE enzyme. In the case of the BuChE inhibitor model, a central carbon atom (C) on an aromatic ring that has one carbon neighbor (R) and one heteroatom neighbor (X) on the same aromatic ring and the third neighbor outside this ring is a carbon (R), a sum of topological distances between two nitrogen atoms, number of fragments containing C(sp2) atoms that are attached with two electronegative atoms (O, N, S, Se, and halogens), i.e., one by a single bond and another by a double bond and the number of secondary aromatic amides may influence the inhibitory activity of BuChE enzyme inhibitors, whereas a higher number of thiazole rings, CRX fragments and H atoms connected to $C^0(sp^3)$ with 2X attached to next 'C' are detrimental to the BuChE enzyme inhibitory activity. The features obtained from the selectivity-based model suggest that the number of aromatic ethers and unsaturation content related to their molecular size and molecular shape may

be more specific for the inhibition AChE enzyme in comparison to the BuChE enzyme, whereas the number of total quaternary carbons (sp3) may be more specific for BuChE inhibitory activity. The identified features are responsible for increasing brain permeability and an entropically more favorable binding to the receptor and intermolecular interactions by strong H-bonds for the improvement of the inhibitory activity of both enzymes (AChE and BuChE). Finally, molecular docking analysis has performed to identify the interactions between target proteins (AChE and BuChE enzyme) and inhibitors in this dataset., and the results showed the active compounds (compounds **15** and **19** in the case of AChE inhibitors and **13** and **547** in case of BuChE inhibitors), formed hydrogen bond and hydrophobic π interactions with amino acid residues that lead to the identification of the active binding site of the target protein. Moreover, the information obtained from molecular docking analysis well supported the features obtained from the 2D-QSAR analysis results. The validated models might be supportive for the estimation of the inhibitory activity of novel compounds against the AChE and BuChE enzymes, and the information obtained from the 2D-QSAR analysis and molecular docking studies can be useful for the development of new analogs.

5.6. Study 6- Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease

The present investigation includes the development of *in silico*-based predictive 2D-QSAR models against the twelve major targets of AD using a PLS regression approach for the exploration of the structural features responsible for the inhibitory activity toward respective targets using simple and easily interpretable 2D descriptors. The authors have also developed the 17 QSAAR and 10 selectivity-based models by screening the common inhibitors from the primary datasets. All developed models are extensively validated and found to be robust enough to satisfy the acceptance criteria. From the insights obtained from the developed models, it can be inferred that features contributing to hydrophobicity as the significant parameter for the AD drugs, molecular bulk, and electronegativity of the compounds may enhance the inhibitory activity of the compounds. The identified features/properties are responsible for enhancing brain permeability and an entropically more favorable binding to the receptor and intermolecular interactions by strong H bonds for the improvement of the inhibitory activity against the targets. The developed individual models were used to check the applicability domain of the four chemical druglike databases (18,963,690 compounds) for the search for novel inhibitors with multitarget inhibitory activity. In a detailed analysis of the predicted chemical databases, authors have identified the top 56 lead compounds with multitarget inhibitory activity that could act as inhibitors for the treatment of AD. Furthermore, authors have also implemented chemical Read-Across predictions using the Read-Across-v3.1 tool (https://dtclab.webs.com/software-tools); the results for the external validation parameters were found to be better than the 2D-QSAR-derived predictions. Finally, molecular docking analysis has been implemented to identify the interactions between target proteins and inhibitors, and the results showed the active compounds formed hydrogen bonding and hydrophobic π interactions with active amino acid residues that lead to explaining the influences of different features which appeared in the 2D-QSAR models. Moreover, the evidence obtained from molecular docking analysis well corroborated with the features obtained from the 2D-QSAR analysis. Lastly, the proposed models and read-across hypotheses might be enormously valuable as guides for investigators to predict the inhibitory activity of novel compounds against respective targets and the evidence obtained from the 2D-QSAR analysis and molecular docking studies can be useful for the design of next-generation multi-target inhibitors for Alzheimer therapy.

5.7. Overall conclusion

In this thesis work, various *in silico* techniques were employed to study the potential leads against AD. The main rationale was to use different in silico approaches to find and improve potential anti-Alzheimer's leads against several crucial targets involved in AD. Along with the single-target drug designing approach (**study 1-4**), authors have also focused on identifying or designing dual-binding site AChE inhibitors (**study 5**), as well as multi-target inhibitors (**study 6**). Further, authors have explored the selectivity issue of inhibitors against AChE over BuChE (**study 5**), which is a commonly observed issue while designing molecules against enzymes. Even though we used a variety of *in silico* methods, such as QSAR, molecular docking, pharmacophore modeling, virtual screening, and so on, the majority of our work is focused on developing predictive and statistically robust QSAR models. The QSAR approach is used extensively in the lead optimization step of any drug development effort to reduce time, money, and, most importantly, animal sacrifice. A QSAR model is used to identify the structural features responsible for the activity as well as to achieve selectivity. Additionally, authors have also developed the quantitative structure activity-activity relationship (QSAAR) and selectivity-based models to explore the most important features contributing to the dual inhibition against the respective targets. Furthermore, the model provides significant information for designing new compounds with improved activity, and it is used to predict the activity of a query or newly designed compound.

REFERENCES

References

1. Samanta S, Ramesh M, Govindaraju T (2022) Chapter 1:Alzheimer's is a Multifactorial Disease , in *Alzheimer's Disease: Recent Findings in Pathophysiology. Diagnostic and Therapeutic Modalities.* 1-34. DOI: 10.1039/9781839162732-00001
2. Alzheimer's Association. (2021) Alzheimer's disease Facts and Figures. Alzheimers Dement.17(3).
3. Isaac GO, V. Jauregui G, Čarná M, P. Bennett J, Stokin GB (2021) Neuroinflammation in Alzheimer's disease. Biomedicines 9(5): 524. https://doi.org/10.3390/biomedicines9050524
4. Tatulian, SA (2022) Challenges and hopes for Alzheimer's disease." Drug Discovery Today. https://doi.org/10.1016/j.drudis.2022.01.016
5. Gauthier S, Rosa-Neto P, Morais JA, Webster C. (2021) World Alzheimer Report 2021: Journey through the diagnosis of dementia. London, England: Alzheimer's disease International.
6. Srivastava S, Ahmad R, Khare SK (2021) Alzheimer's disease and its treatment by different approaches: A review. Eur. J. Med. Chem. 216:113320. https://doi.org/10.1016/j.ejmech.2021.113320
7. Rajasekhar, K, Thimmaiah Govindaraju (2018) Current progress, challenges and future prospects of diagnostic and therapeutic interventions in Alzheimer's disease. RSC advances 8 (42): 23780-23804. DOI: 10.1039/C8RA03620A
8. Lei G, Zhong MB, Zhang Larry, Zhang Bin, Cai Dongming (2022) Sex differences in Alzheimer's disease: Insights from the multiomics landscape. Biol. Psychiatry 91(1):61-71. https://doi.org/10.1016/j.biopsych.2021.02.968
9. Konstantina GY, Papageorgiou SG (2020) Current and future treatments in Alzheimer disease: an update. J. Cent. Nerv. Syst. 12: 1-12. https://doi.org/10.1177/1179573520907397
10. Vaz M, Silva V, Monteiro C, Silvestre S (2022) Role of Aducanumab in the Treatment of Alzheimer's Disease: Challenges and Opportunities. Clinical Interventions in Aging, 17:797. doi: 10.2147/CIA.S325026
11. Liston DR, Nielsen JA, Villalobos A et al. Pharmacology of selective acetylcholinesterase inhibitors: implications for use in Alzheimer's disease. Eur. J. Pharmacol. 2004;486(1):9-17.
12. Nordberg A, Svensson AL. Cholinesterase inhibitors in the treatment of Alzheimer's disease. Drug Safety. 1998;19(6):465-80.
13. Awasthi M, Upadhyay AK, Singh S et al. Terpenoids as promising therapeutic molecules against Alzheimer's disease: amyloid beta-and acetylcholinesterase-directed pharmacokinetic and molecular docking analyses. Mol. Simul. 2018;44(1):1-11.
14. S.W. Snyder, U.S. Ladror, W.S. Wade, G.T. Wang, L.W. Barrett, E.D. Matayoshi, H.J. Huffaker, G.A. Krafft, T.F. Holzman, Amyloid-beta aggregation: selective inhibition of aggregation in mixtures of amyloid with different chain lengths, Biophys. J. 67 (3) (1994) 1216–1228. doi:10.1016/S0006-3495(94)80591-0.
15. Vassar R, Kovacs D M, Yan R and Wong P C (2009) The β-secretase enzyme BACE in health and Alzheimer's disease: regulation, cell biology, function, and therapeutic potential. Journal of Neuroscience 29(41) pp.12787-12794.
16. Vassar R J (2016) BACE1, the beta-secretase enzyme: a leading therapeutic target for Alzheimer's disease. Alzheimer's & Dementia. The Journal of the Alzheimer's Association 12(7) p.P162.
17. Aaldijk E, Vermeiren Y (2022) The role of serotonin within the microbiota-gut-brain axis in the development of Alzheimer's disease: A narrative review. Ageing Res. Rev., 101556. https://doi.org/10.1016/j.arr.2021.101556
18. Nykamp MJ, Zorumski CF, Reiersen AM, Nicol GE, Cirrito J, Lenze EJ (2022) Opportunities for Drug Repurposing of Serotonin Reuptake Inhibitors: Potential Uses in Inflammation, Infection, Cancer, Neuroprotection, and Alzheimer's Disease Prevention. Pharmacopsychiatry, 55(01):24-29. DOI: 10.1055/a-1686-9620
19. He, G., Luo, W., Li, P. et al. Gamma-secretase activating protein is a therapeutic target for Alzheimer's disease. Nature 467, 95–98 (2010). https://doi.org/10.1038/nature09325

20. Tate, B., McKee, T. D., Loureiro, R., Dumin, J. A., Xia, W., Pojasek, K., ... & Bronk, B. S. (2012). Modulation of gamma-secretase for the treatment of Alzheimer's disease. International journal of Alzheimer's disease, 2012. https://doi.org/10.1155/2012/210756
21. Hielscher-Michael, S., Griehl, C., Buchholz, M., Demuth, H. U., Arnold, N., & Wessjohann, L. A. (2016). Natural products from microalgae with potential against Alzheimer's disease: Sulfolipids are potent glutaminyl cyclase inhibitors. Marine drugs, 14(11), 203. https://doi.org/10.3390/md14110203
22. Morawski, M., Schilling, S., Kreuzberger, M., Waniek, A., Jaeger, C., Koch, B., ... & Roßner, S. (2014). Glutaminyl cyclase in human cortex: correlation with (pGlu)-amyloid-β load and cognitive decline in Alzheimer's disease. Journal of Alzheimer's Disease, 39(2), 385-400. DOI: 10.3233/JAD-131535
23. Nabavi, S. M., Talarek, S., Listos, J., Nabavi, S. F., Devi, K. P., de Oliveira, M. R., ... & Farzaei, M. H. (2019). Phosphodiesterase inhibitors say NO to Alzheimer's disease. Food and Chemical Toxicology, 134, 110822. https://doi.org/10.1016/j.fct.2019.110822
24. Sheng, J., Zhang, S., Wu, L., Kumar, G., Liao, Y., Pratap, G. K., & Fan, H. (2022). Inhibition of phosphodiesterase: A novel therapeutic target for the treatment of mild cognitive impairment and Alzheimer's disease. Frontiers in Aging Neuroscience, 14. https://doi.org/10.3389/fnagi.2022.1019187
25. Heckman, P. R., Blokland, A., & Prickaerts, J. (2017). From age-related cognitive decline to Alzheimer's disease: a translational overview of the potential role for phosphodiesterases. Phosphodiesterases: CNS Functions and Diseases, 135-168.
26. Abd-Elrahman KS, Ferguson SS (2022) Noncanonical metabotropic glutamate receptor 5 signaling in Alzheimer's disease. Annual Review of Pharmacology and Toxicology, 62:235-254. https://doi.org/10.1146/annurev-pharmtox-021821-091747
27. Pinky PD, Pfitzer JC, Senfeld J, Hong H, Bhattacharya S, Suppiramaniam V, Reed MN (2022) Recent Insights on Glutamatergic Dysfunction in Alzheimer's Disease and Therapeutic Implications. The Neuroscientist, 10738584211069897. https://doi.org/10.1177/10738584211069897
28. Bhounsule, A. S., Bhatt, L. K., Prabhavalkar, K. S., & Oza, M. (2017). Cyclin dependent kinase 5: A novel avenue for Alzheimer's disease. Brain research bulletin, 132, 28-38. https://doi.org/10.1016/j.brainresbull.2017.05.006
29. Cruz, J. C., & Tsai, L. H. (2004). Cdk5 deregulation in the pathogenesis of Alzheimer's disease. Trends in molecular medicine, 10(9), 452-458. https://doi.org/10.1016/j.molmed.2004.07.001
30. Balaraman, Y., Limaye, A. R., Levey, A. I., & Srinivasan, S. (2006). Glycogen synthase kinase 3β and Alzheimer's disease: pathophysiological and therapeutic significance. Cellular and Molecular Life Sciences CMLS, 63(11), 1226-1235. https://doi.org/10.1007/s00018-005-5597-y
31. Leroy, K., Boutajangout, A., Authelet, M., Woodgett, J. R., Anderton, B. H., & Brion, J. P. (2002). The active form of glycogen synthase kinase-3β is associated with granulovacuolar degeneration in neurons in Alzheimer's disease. Acta neuropathologica, 103(2), 91-99. https://doi.org/10.1007/s004010100435
32. Jin, N., et al., Truncation and activation of GSK-3β by calpain I: a molecular mechanism links to tau hyperphosphorylation in Alzheimer's disease. Sci. Rep. 2015, 5, 8187.
33. https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=4129.
34. Benzi, G.; Moretti, A., Are reactive oxygen species involved in Alzheimer's disease? *Neurobiol. Aging* **1995,** 16, (4), 661-674.
35. Kumar, M. J.; Andersen, J. K., Perspectives on MAO-B in aging and neurological disease. Where do we go from here? *Mol. Neurobiol.* **2004,** 30, (1), 77-89.
36. Ambure, P., et al., Pharmacophore mapping-based virtual screening followed by molecular docking studies in search of potential acetylcholinesterase inhibitors as anti-Alzheimer's agents. *Biosystems* **2014,** 116, 10-20.
37. Munoz-Ruiz, P., et al., Design, synthesis, and biological evaluation of dual binding site acetylcholinesterase inhibitors: new disease-modifying agents for Alzheimer's disease. *J. Med. Chem.* **2005,** 48, (23), 7223-7233.
38. Piazzi, L., et al., 3-(4-{[Benzyl (methyl) amino] methyl} phenyl)-6, 7-dimethoxy-2 H-2-chromenone (AP2238) Inhibits Both Acetylcholinesterase and Acetylcholinesterase-Induced Beta-Amyloid Aggregation: A Dual Function Lead for Alzheimer's Disease Therapy. *J. Med. Chem.* **2003,** 46, (12), 2279-2282.

39. Camps, P., et al., Tacrine-based dual binding site acetylcholinesterase inhibitors as potential disease-modifying anti-Alzheimer drug candidates. *Chem. Biol. Interact.* **2010,** 187, (1), 411-415.
40. Alonso, D., et al., Donepezil-tacrine hybrid related derivatives as new dual binding site inhibitors of AChE. *Bioorg. Med. Chem.* **2005,** 13, (24), 6588-6597.
41. Catto, M., et al., Design, synthesis and biological evaluation of coumarin alkylamines as potent and selective dual binding site inhibitors of acetylcholinesterase. *Bioorg. Med. Chem.* **2013,** 21, (1), 146-152.
42. Ignasik, M., et al., Design, Synthesis and Evaluation of Novel 2-(Aminoalkyl)-isoindoline-1, 3-dione Derivatives as Dual-Binding Site Acetylcholinesterase Inhibitors. *Arch. Pharm. (Weinheim)* **2012,** 345, (7), 509-516.
43. Yar, M., et al., Design and Synthesis of New Dual Binding Site Cholinesterase Inhibitors: in vitro Inhibition Studies with in silico Docking. *Lett. Drug Des. Discov.* **2014,** 11, (3), 331.
44. Gupta, S., et al., Discovery of dual binding site acetylcholinesterase inhibitors identified by pharmacophore modeling and sequential virtual screening techniques. *Bioorg. Med. Chem. Lett.* **2011,** 21, (4), 1105-1112.
45. Rodriguez-Franco, M. I., et al., Novel tacrine-melatonin hybrids as dual-acting drugs for Alzheimer disease, with improved acetylcholinesterase inhibitory and antioxidant properties. J. Med. Chem. 2006, 49, (2), 459-462.
46. Fang, L., et al., Design and synthesis of tacrine-ferulic acid hybrids as multi-potent anti-Alzheimer drug candidates. Bioorg. Med. Chem. Lett. 2008, 18, (9), 2905-2909.
47. Fernandez-Bachiller, M. I., et al., Novel tacrine- 8-hydroxyquinoline hybrids as multifunctional agents for the treatment of Alzheimer's disease, with neuroprotective, cholinergic, antioxidant, and copper-complexing properties. J. Med. Chem. 2010, 53, (13), 4927-4937.
48. Rosini, M., et al., Rational approach to discover multipotent anti-Alzheimer drugs. J. Med. Chem. 2005, 48, (2), 360-363.
49. Sterling, J., et al., Novel dual inhibitors of AChE and MAO derived from hydroxy aminoindan and phenethylamine as potential treatment for Alzheimer's disease. J. Med. Chem. 2002, 45, (24), 5260-5279.
50. Vina, D., et al., 3-Substituted coumarins as dual inhibitors of AChE and MAO for the treatment of Alzheimer's disease. MedChemComm 2012, 3, (2), 213-218.
51. Ucar, G., et al., 1-N-Substituted thiocarbamoyl-3-phenyl-5-thienyl-2-pyrazolines: A novel cholinesterase and selective monoamine oxidase B inhibitors for the treatment of Parkinson's and Alzheimer's diseases. Neurosci. Lett. 2005, 382, (3), 327-331.
52. Piazzi, L., et al., Multi-target-directed coumarin derivatives: hAChE and BACE1 inhibitors as potential anti-Alzheimer compounds. Bioorg. Med. Chem. Lett. 2008, 18, (1), 423-426.
53. Zhang P, Xu S, Zhu Z, Xu J (2019) Multi-target design strategies for the improved treatment of Alzheimer's disease. Eur. J. med. Chem., 176:228-247. https://doi.org/10.1016/j.ejmech.2019.05.020
54. Benek O, Korabecny J, Soukup O (2020) A perspective on multi-target drugs for Alzheimer's disease. Trends Pharmacol Sci., 41(7) 434-445. https://doi.org/10.1016/j.tips.2020.04.008
55. Zhang B, Zhao J, Wang Z, Guo P, Liu A, Du G (2021) Identification of multi-target anti-AD chemical constituents from traditional Chinese medicine formulae by integrating virtual screening and in vitro validation, Front. Pharmacol. 12: 709607, https://doi.org/10.3389/fphar.2021.709607
56. Nadeem MS, J.A., Rashid U (2021) Fluoxetine and sertraline based multitarget inhibitors of cholinesterases and monoamine oxidase-A/B for the treatment of Alzheimer's disease: Synthesis, pharmacology and molecular modeling studies. Int. J. Biol. Macromol. 193:19-26, https://doi.org/10.1016/j.ijbiomac.2021.10.102
57. Brunetti L, Leuci R, Carrieri A, Catto M, Occhineri S, Vinci G, Piemontese L (2022) Structure-based design of novel donepezil-like hybrids for a multi-target approach to the therapy of Alzheimer's disease. Eur. J. Med. Chem. 237: 114358, https://doi.org/10.1016/j.ejmech.2022.114358
58. Ajala A, Uzairu A, Shallangwa GA, Abechi SE (2022) 2D QSAR, design, docking study and ADMET of some N-aryl derivatives concerning inhibitory activity against Alzheimer disease, Future J. Pharm. Sci. 8(1): 1-14, https://doi.org/10.1186/s43094-022-00420-w

59. Ambure P, Bhat J, Puzyn T, Roy K (2019) Identifying natural compounds as multi-target-directed ligands against Alzheimer's disease: an in silico approach, J. Biomol. Struct. Dyn. 37(5): 1282-1306, https://doi.org/10.1080/07391102.2018.1456975
60. Kumar V, Saha A, Roy K (2020) In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease, Comput Biol Chem. 88: 107355, https://doi.org/10.1016/j.compbiolchem.2020.107355
61. Kumar A, Nisha CM, Silakari C, Sharma I, Anusha K, Gupta N, Kumar A (2016). Current and novel therapeutic molecules and targets in Alzheimer's disease. JFMA, 115(1):3-10. https://doi.org/10.1016/j.jfma.2015.04.001
62. K. Roy, S. Kar and R. N. Das, Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment, Academic press, 2015.
63. OECD, 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment, No. 69, OECD Publishing, Paris, Available at http://dx.doi.org/10.1787/9789264085442-en.).
64. Baumann, Désirée, and Knut Baumann. "Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation." Journal of cheminformatics 6.1 (2014): 1-19.
65. Roy, Kunal, and Pravin Ambure. "The "double cross-validation" software tool for MLR QSAR model development." Chemometrics and Intelligent Laboratory Systems 159 (2016): 108-126.
66. Roy, K., Das, R. N., Ambure, P., & Aher, R. B. (2016). Be aware of error measures. Further studies on validation of predictive QSAR models. Chemometrics and Intelligent Laboratory Systems, 152, 18-33.
67. Hu, Xiaoying, et al. "MMP-cliffs: systematic identification of activity cliffs on the basis of matched molecular pairs." Journal of chemical information and modeling 52.5 (2012): 1138-1145.
68. Kar, Supratik, et al. "Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach." Ecotoxicology and environmental safety 107 (2014): 162-169.
69. R. Todeschini and V. Consonni, Methods and principles in medicinal chemistry. Kubinyi H, Timmerman H (Series eds) Handbook of molecular descriptors. Wiley-VCH, Weinheim., 2000.
70. Golmohammadi, Hassan, Zahra Dashtbozorgi, and William E. Acree Jr. "Quantitative structure–activity relationship prediction of blood-to-brain partitioning behavior using support vector machine." European Journal of Pharmaceutical Sciences 47.2 (2012): 421-429.
71. Kennard, Ronald W., and Larry A. Stone. "Computer aided design of experiments." Technometrics 11.1 (1969): 137-148.
72. Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." Journal of the royal statistical society. series c (applied statistics) 28.1 (1979): 100-108.
73. Roy, Kunal. "Quantitative structure-activity relationships (QSARs): a few validation methods and software tools developed at the DTC laboratory." Journal of the Indian Chemical Society 95.12 (2018): 1497-1502.
74. Khan, Pathan Mohsin, and Kunal Roy. "Current approaches for choosing feature selection and learning algorithms in quantitative structure–activity relationships (QSAR)." Expert opinion on drug discovery 13.12 (2018): 1075-1089.
75. De, Priyanka, Rahul B. Aher, and Kunal Roy. "Chemometric modeling of larvicidal activity of plant derived compounds against zika virus vector Aedes aegypti: application of ETA indices." RSC advances 8.9 (2018): 4662-4670.
76. De, Priyanka, et al. "Second generation periodic table-based descriptors to encode toxicity of metal oxide nanoparticles to multiple species: QSTR modeling for exploration of toxicity mechanisms." Environmental Science: Nano 5.11 (2018): 2742-2760.
77. A. Tropsha, (2010). Best practices for QSAR model development, validation, and exploitation, Mol. Inform. 29(6-7) (2010) 476-488, https://doi.org/10.1002/minf.201000061
78. R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Validation of QSAR models-strategies and importance, Int. J. Drug Des. Discov. 3 (2011) 511-519.
79. SIMCA-P 10.0, Umetrics. Available from https://umetrics.com/products/simca (accessed January 10, 2022).

80. K. Roy, Multi-Target Drug Design Using Chem-Bioinformatic Approaches, Springer, 2019.
81. Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., & Andrade, C. H. (2018). QSAR-based virtual screening: advances and applications in drug discovery. Frontiers in pharmacology, 9, 1275. https://doi.org/10.3389/fphar.2018.01275
82. Saxena, A. K., & Prathipati, P. (2003). Comparison of mlr, pls and ga-mlr in qsar analysis. SAR and QSAR in Environmental Research, 14(5-6), 433-445. https://doi.org/10.1080/10629360310001624015
83. Golbraikh, A. (2000). Molecular dataset diversity indices and their applications to comparison of chemical databases and QSAR analysis. Journal of Chemical Information and Computer Sciences, 40(2), 414-425. https://doi.org/10.1021/ci990437u
84. Choudhury C, Sastry GN (2019) Pharmacophore Modelling and Screening: Concepts, Recent Developments and Applications in Rational Drug Design. In Structural Bioinformatics: Applications in Preclinical Drug Discovery Process (pp. 25-53). Springer, Cham. https://doi.org/10.1007/978-3-030-05282-9_2
85. Yang SY (2010). Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug discover today, *15*(11-12), 444-450. doi:10.1016/j.drudis.2010.03.013
86. Schaller D, Dora Š, Theresa N, Lihua D, Trung NN, Szymon P, David M, Marcel B, Gerhard W (2020) Next generation 3D pharmacophore modeling." Wiley Interdisciplinary Reviews: Computational Molecular Science e1468. DOI: 10.1002/wcms.1468
87. Systèmes, D. (2016). Biovia, discovery studio modeling environment. Dassault Systèmes Biovia: San Diego, CA, USA.
88. Pryce, J. D., & Marletta, M. (1991). A new multi-purpose software package for Schrödinger and Sturm-Liouville computations. Comput. Phys. Commun. 62(1), 42-52. https://doi.org/10.1016/0010-4655(91)90119-6
89. Wolber, G., & Langer, T. (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J. chem. Inform. Model. 45(1), 160-169. https://doi.org/10.1021/ci049885e
90. Dar AM, Mir S (2017). Molecular docking: approaches, types, applications and basic challenges. J Anal Bioanal Tech, *8*(2), 1-3. DOI: 10.4172/2155-9872.1000356
91. Morris GM, Lim-Wilby M (2008) Molecular docking. In Molecular modeling of proteins (pp. 365-382). Humana Press, Totowa, (New Jersey, United States).
92. Meng XY, Zhang HX., Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr comp-aid drug des, *7*(2), 146-157. DOI: https://doi.org/10.2174/157340911795677602
93. Forli, W., Halliday, S., Belew, R., & Olson, A. J. (2012). AutoDock Version 4.2.
94. Bitencourt-Ferreira, G., & de Azevedo, W. F. (2019). Molegro Virtual Docker for Docking. In Docking Screens for Drug Discovery (pp. 149-167). Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9752-7_10
95. Schellhammer, I., & Rarey, M. (2004). FlexX-Scan: Fast, structure-based virtual screening. PROTEINS: Structure, Function, and Bioinformatics, 57(3), 504-517. https://doi.org/10.1002/(SICI)1097-0134(19991101)37:2
96. Moro, S., Bacilieri, M., & Deflorian, F. (2007). Combining ligand-based and structure-based drug design in the virtual screening arena. Expert Opinion on Drug Discovery, 2(1), 37-49. https://doi.org/10.1517/17460441.2.1.37
97. Tong, W., Collantes, E. R., Chen, Y., & Welsh, W. J. (1996). A comparative molecular field analysis study of N-benzylpiperidines as acetylcholinesterase inhibitors. Journal of medicinal chemistry, 39(2), 380-387.
98. Recanatini, M., Cavalli, A., & Hansch, C. (1997). A comparative QSAR analysis of acetylcholinesterase inhibitors currently studied for the treatment of Alzheimer's disease. Chemico-biological interactions, 105(3), 199-228.
99. Kaur, J., & Zhang, M. Q. (2000). Molecular modelling and QSAR of reversible acetylcholines-terase inhibitors. Current Medicinal Chemistry, 7(3), 273-294.
100. Nicolotti, O., Pellegrini-Calace, M., Altomare, C., Carotti, A., Carrieri, A., & Sanz, F. (2002). Ligands of neuronal nicotinic acetylcholine receptor (nAChR): inferences from the Hansch and 3-D quantitative structure-activity relationship (QSAR) Models. Current medicinal chemistry, 9(1), 1-29.

101. Shen, L. L., Liu, G. X., & Tang, Y. (2007). Molecular docking and 3D-QSAR studies of 2-substituted 1-indanone derivatives as acetylcholinesterase inhibitors. Acta Pharmacologica Sinica, 28(12), 2053-2063.
102. Solomon, K. A., Sundararajan, S., & Abirami, V. (2009). QSAR studies on N-aryl derivative activity towards Alzheimer's disease. Molecules, 14(4), 1448-1455.
103. Gupta, S., Fallarero, A., Vainio, M. J., Saravanan, P., Santeri Puranen, J., Järvinen, P., ... & Mohan, C. G. (2011). Molecular docking guided comparative GFA, G/PLS, SVM and ANN models of structurally diverse dual binding site acetylcholinesterase inhibitors. Molecular informatics, 30(8), 689-706.
104. De Souza, S. D., De Souza, A. M. T., De Sousa, A. C. C., Sodero, A. C. R., Cabral, L. M., Albuquerque, M. G., ... & Rodrigues, C. R. (2012). Hologram QSAR models of 4-[(diethylamino) methyl]-phenol inhibitors of acetyl/butyrylcholinesterase enzymes as potential anti-Alzheimer agents. Molecules, 17(8), 9529-9539.
105. Kumar, V., De, P., Ojha, P. K., Saha, A., & Roy, K. (2020). A multi-layered variable selection strategy for QSAR modeling of butyrylcholinesterase inhibitors. Current Topics in Medicinal Chemistry, 20(18), 1601-1627.
106. Kumar, V., & Saha, A. (2020). Chemometric Modeling of Structurally Diverse Carbamates for the Inhibition of Acetylcholinesterase (AChE) Enzyme in Alzheimer's Disease. International Journal of Quantitative Structure-Property Relationships (IJQSPR), 5(3), 6-60.
107. Keerti, A. R., Kumar, B. A., Parthasarathy, T., & Uma, V. (2005). QSAR studies—potent benzodiazepine γ-secretase inhibitors. Bioorganic & medicinal chemistry, 13(5), 1873-1878.
108. Al-Nadaf, A., Sheikha, G. A., & Taha, M. O. (2010). Elaborate ligand-based pharmacophore exploration and QSAR analysis guide the synthesis of novel pyridinium-based potent β-secretase inhibitory leads. Bioorganic & medicinal chemistry, 18(9), 3088-3115.
109. Meek, A. R., Simms, G. A., & Weaver, D. F. (2012). In silico search for an endogenous anti-Alzheimer's molecule—Screening amino acid metabolic pathways. Canadian Journal of Chemistry, 90(10), 865-873.
110. Valasani, K. R., Hu, G., Chaney, M. O., & Yan, S. S. (2013). Structure-based design and synthesis of benzothiazole phosphonate analogues with inhibitors of human ABAD-Aβ for treatment of Alzheimer's disease. Chemical biology & drug design, 81(2), 238-249.
111. Kumar, V., Ojha, P. K., Saha, A., & Roy, K. (2020). Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease. SAR and QSAR in Environmental Research, 31(2), 87-133.
112. Kumar, V., Ojha, P. K., Saha, A., & Roy, K. (2020). Cheminformatic modelling of β-amyloid aggregation inhibitory activity against Alzheimer's disease. Computers in Biology and Medicine, 118, 103658.
113. Martinez, A., Alonso, M., Castro, A., Dorronsoro, I., Gelpí, J. L., Luque, F. J., ... & Moreno, F. J. (2005). SAR and 3D-QSAR studies on thiadiazolidinone derivatives: exploration of structural requirements for glycogen synthase kinase 3 inhibitors. Journal of medicinal chemistry, 48(23), 7103-7112.
114. Lather, V., Kristam, R., Saini, J. S., Kristam, R., Karthikeyan, N. A., & Balaji, V. N. (2008). QSAR models for prediction of glycogen synthase kinase-3β inhibitory activity of indirubin derivatives. QSAR & Combinatorial Science, 27(6), 718-728.
115. Park, H. R., Kim, M. K., Kim, D. W., Choo, I. H., & Chong, Y. H. (2010). 3D QSAR CoMFA Study on Phenylthiazolylhydrazide (PTH) Derivataives as Tau Protein Aggregation Inhibitors. Bulletin of the Korean Chemical Society, 31(12), 3838-3841.
116. Fang, J., Huang, D., Zhao, W., Ge, H., Luo, H. B., & Xu, J. (2011). A new protocol for predicting novel GSK-3β ATP competitive inhibitors. Journal of chemical information and modeling, 51(6), 1431-1438.
117. Ul Haq, Z., Uddin, R., Wai, L. K., Wadood, A., & Lajis, N. H. (2011). Docking and 3D-QSAR modeling of cyclin-dependent kinase 5/p25 inhibitors. Journal of molecular modeling, 17(5), 1149-1161.
118. Wang, W., Zhang, J., & Liu, B. (2005). QSAR study of 125I-labeled 2-(4-aminophenyl) benzothiazole derivatives as imaging agents for β-amyloid in the brain with Alzheimer's disease. Journal of Radioanalytical and Nuclear Chemistry, 266(1), 107-111.
119. Kim, M. K., Choo, I. H., Lee, H. S., Woo, J. I., & Chong, Y. H. (2007). 3D-QSAR of PET Agents for Imaging β-Amyloid in Alzheimer's Disease. Bulletin of the Korean Chemical Society, 28(7), 1231-1234.

120. Kovac, M., Mavel, S., Deuther-Conrad, W., Méheux, N., Glöckner, J., Wenzel, B., ... & Emond, P. (2010). 3D QSAR study, synthesis, and in vitro evaluation of (+)-5-FBVM as potential PET radioligand for the vesicular acetylcholine transporter (VAChT). Bioorganic & medicinal chemistry, 18(21), 7659-7667.
121. Cisek, K., & Kuret, J. (2012). QSAR studies for prediction of cross-β sheet aggregate binding affinity and selectivity. Bioorganic & medicinal chemistry, 20(4), 1434-1441.
122. Zhu, Y. Q., Pei, J. F., Liu, Z. M., Lai, L. H., Cui, J. R., & Li, R. T. (2006). 3D-QSAR studies on tripeptide aldehyde inhibitors of proteasome using CoMFA and CoMSIA methods. Bioorganic & medicinal chemistry, 14(5), 1483-1496.
123. Pripp, A. H. (2006). Quantitative structure− activity relationship of prolyl oligopeptidase inhibitory peptides derived from β-casein using simple amino acid descriptors. Journal of agricultural and food chemistry, 54(1), 224-228.
124. Dastmalchi, S., Hamzeh-Mivehroud, M., Ghafourian, T., & Hamzeiy, H. (2008). Molecular modeling of histamine H3 receptor and QSAR studies on arylbenzofuran derived H3 antagonists. Journal of Molecular Graphics and Modelling, 26(5), 834-844.
125. Firoozpour, L., Sadatnezhad, K., Dehghani, S., Pourbasheer, E., Foroumadi, A., Shafiee, A., & Amanlou, M. (2012). An efficient piecewise linear model for predicting activity of caspase-3 inhibitors. DARU Journal of Pharmaceutical Sciences, 20(1), 1-6.
126. Hajjo, R., Setola, V., Roth, B. L., & Tropsha, A. (2012). Chemocentric informatics approach to drug discovery: identification and experimental validation of selective estrogen receptor modulators as ligands of 5-hydroxytryptamine-6 receptors and as potential cognition enhancers. Journal of medicinal chemistry, 55(12), 5704-5719.
127. B Bharate, S., R Yadav, R., & A Vishwakarma, R. (2013). QSAR and pharmacophore study of Dyrk1A inhibitory meridianin analogs as potential agents for treatment of neurodegenerative diseases. Medicinal Chemistry, 9(1), 152-161.
128. Prado-Prado, F., Garcia-Mera, X., Escobar, M., Alonso, N., Caamano, O., Yanez, M., & Gonzalez-Diaz, H. (2012). 3D MI-DRAGON: new model for the reconstruction of US FDA drug-target network and theoretical-experimental studies of inhibitors of rasagiline derivatives for AChE. Current topics in medicinal chemistry, 12(16), 1843-1865.
129. Speck-Planche, A., V Kleandrova, V., Luan, F., & Cordeiro, N. D. (2013). Multi-target inhibitors for proteins associated with Alzheimer: in silico discovery using fragment-based descriptors. Current Alzheimer Research, 10(2), 117-124.
130. Kumar, V., Saha, A., & Roy, K. (2020). In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease. Computational Biology and Chemistry, 88, 107355.
131. Sermboonpaisarn, T. and Sawasdee, P. Potent and selective butyrylcholinesterase inhibitors from Ficusfoveolata. *Fitoterapia*, 2012, *83*(4), pp.780-784.
132. Schulze, M., Siol, O.; Decker, M. and Lehmann, J. Bivalent 5, 8, 9, 13b-tetrahydro-6H-isoquino [1, 2-a] isoquinolines and-isoquinolinium salts: Novel heterocyclic templates for butyrylcholinesterase inhibitors. *Bioorg. Med. Chem.lett*, 2010, *20*(9), pp.2946-2949.
133. Yan, J.W.; Li, Y.P.; Ye, W.J.; Chen, S.B.; Hou, J.Q.; Tan, J.H.;Ou, T.M., Li, D.; Gu, L.Q. and Huang, Z.S. Design, synthesis and evaluation of isaindigotone derivatives as dual inhibitors for acetylcholinesterase and amyloid beta aggregation. *Bioorg. Med. Chem.* 2012, *20*(8), pp.2527-2534.
134. Takahashi, J.; Hijikuro, I.; Kihara, T.; Murugesh, M.G.; Fuse, S.; Tsumura, Y.; Akaike, A., Niidome, T.; Takahashi, T. and Sugimoto, H. Design, synthesis and evaluation of carbamate-modified (−)-N1-phenethylnorphysostigmine derivatives as selective butyrylcholinesterase inhibitors. *Bioorg. Med. Chem.Let.* 2010, *20*(5), pp.1721-1723.
135. Tang, H., Zhao, H.T.;Zhong, S.M.; Wang, Z.Y.; Chen, Z.F. and Liang, H. Novel oxoisoaporphine-based inhibitors of acetyl-and butyrylcholinesterase and acetylcholinesterase-induced beta-amyloid aggregation. *Bioorg. Med. Chem.Lett.* 2012, *22*(6), pp.2257-2261.

136. Chen, Y., Sun, J.; Huang, Z.; Liao, H., Peng, S.; Lehmann, J. and Zhang, Y. NO-donating tacrine derivatives as potential butyrylcholinesterase inhibitors with vasorelaxation activity. *Bioorg. Med. Chem.Lett.* 2013, *23*(11), pp.3162-3165.
137. Yu, L.; Cao, R.; Yi, W.; Yan, Q.; Chen, Z.; Ma, L.; Peng, W. and Song, H. Synthesis of 4-[(diethylamino) methyl]-phenol derivatives as novel cholinesterase inhibitors with selectivity towards butyrylcholinesterase. *Bioorg. Med. Chem. Lett.* 2010, *20*(11), pp.3254-3258.
138. Mutahir, S.; Jończyk, J.;Bajda, M.; Khan, I.U.; Khan, M.A.; Ullah, N.; Ashraf, M., Riaz, S.; Hussain, S. and Yar, M. Novel biphenyl bis-sulfonamides as acetyl and butyrylcholinesterase inhibitors: Synthesis, biological evaluation and molecular modeling studies. *Bioorg.Chem.* 2016, *64*, pp.13-20.
139. Mughal, E.U.; Sadiq, A.; Murtaza, S.; Rafique, H.; Zafar, M.N.;Riaz, T.; Khan, B.A.; Hameed, A. and Khan, K.M. Synthesis, structure–activity relationship and molecular docking of 3-oxoaurones and 3-thioaurones as acetylcholinesterase and butyrylcholinesterase inhibitors. *Bioorg. Med. Chem.* 2017, *25*(1), pp.100-106.
140. Tang, H.; Wei, Y.B.; Zhang, C.; Ning, F.X.; Qiao, W.; Huang, S.L.; Ma, L; Huang, Z.S. and Gu, L.Q. Synthesis, biological evaluation and molecular modeling of oxoisoaporphine and oxoaporphine derivatives as new dual inhibitors of acetylcholinesterase/butyrylcholinesterase. *Euro. J. med.Chem.* 2009, *44*(6), pp.2523-2532.
141. Tasso, B.; Catto, M.; Nicolotti, O.;Novelli, F.;Tonelli, M.; Giangreco, I.;Pisani, L.; Sparatore, A.;Boido, V.; Carotti, A. and Sparatore, F. Quinolizidinyl derivatives of bi-and tricyclic systems as potent inhibitors of acetyl-and butyrylcholinesterase with potential in Alzheimer's disease. *Euro. J. med.Chem.* 2011, *46*(6), pp.2170-2184.
142. da Costa, J.S.; Lopes, J.P.B.; Russowsky, D.; Petzhold, C.L.; de Amorim Borges, A.C.; Ceschi, M.A.; Konrath, E., Batassini, C.;Lunardi, P.S. and Gonçalves, C.A.S. Synthesis of tacrine-lophine hybrids via one-pot four component reaction and biological evaluation as acetyl-and butyrylcholinesterase inhibitors. *Euro. J. med.Chem.* 2013, *62*, pp.556-563.
143. Jabeen, F.; Oliferenko, P.V.; Oliferenko, A.A.; Pillai, G.G.; Ansari, F.L.; Hall, C.D. and Katritzky, A.R. Dual inhibition of the α-glucosidase and butyrylcholinesterase studied by Molecular Field Topology Analysis. *Euro. J. med.Chem.* 2014, *80*, pp.228-242.
144. Darvesh, S; Darvesh, KV; McDonald, RS; Mataija, D; Walsh, R; Mothana, S; Lockridge, O; Martin, E. Carbamates with differential mechanism of inhibition toward acetylcholinesterase and butyrylcholinesterase. *J. med. chem*. 2008, 24;51(14), pp.4200-12.
145. Yu, QS; Holloway, HW; Flippen-Anderson, JL; Hoffman, B; Bross,i A; Greig, NH. Methyl analogues of the experimental Alzheimer drug phenserine: synthesis and structure/activity relationships for acetyl-and butyrylcholinesterase inhibitory action. *J. Med. Chem.* 2001, 22;44(24), pp.4062-71.
146. Decker, M. Novel inhibitors of acetyl-and butyrylcholinesterase derived from the alkaloids dehydroevodiamine and rutaecarpine. *Eur. J. Med. Chem*. 2005, 40(3), 305-13.
147. Orhan, I.E.; Senol, F.S.;Shekfeh, S., Skalicka-Wozniak, K. andBanoglu, E. Pteryxin-A promising butyrylcholinesterase-inhibiting coumarin derivative from Mutellinapurpurea. *Food Chem.Toxicol.* 2017, *109*, pp.970-974.
148. Zhang, N. and Casida, J.E. Novel irreversible butyrylcholinesterase inhibitors: 2-chloro-1-(substituted-phenyl) ethylphosphonic acids. *Bioorg. Med.Chem.* 2002, *10*(5), pp.1281-1290.
149. Bolea, I.; Juárez-Jiménez, J.; de los Ríos, C.; Chioua, M., Pouplana, R.; Luque, F.J.; Unzeta, M.; Marco-Contelles, J. and Samadi, A. Synthesis, biological evaluation, and molecular modeling of donepezil and N-[(5-(benzyloxy)-1-methyl-1 H-indol-2-yl) methyl]-N-methylprop-2-yn-1-amine hybrids as new multipotent cholinesterase/monoamine oxidase inhibitors for the treatment of Alzheimer's disease. *J. Med. Chem*, 2011, *54*(24), pp.8251-8270.
150. Savini, L; Gaeta, A; Fattorusso, C; Catalanotti, B; Campiani, G; Chiasserini, L; Pellerano, C; Novellino, E; McKissic, D; Saxena, A. Specific targeting of acetylcholinesterase and butyrylcholinesterase recognition sites. Rational design of novel, selective, and highly potent cholinesterase inhibitors. *J. Med. Chem*. 2003, 46(1), pp.1-4.
151. Lam, Y., Berberine target key enzymes and amino acid inibitiors in AD treatment creation from berberine-based structure screening. Jul 2014, Vol 5 No 07.
152. Chen, X.; Tikhonova, I.G. and Decker, M. Probing the mid-gorge of cholinesterases with spacer-modified bivalent quinazolinimines leads to highly potent and selective butyrylcholinesterase inhibitors. *Bioorg. Med.Chem*. 2011, *19*(3), pp.1222-1235.

153. Maryamabadi, A.;Hasaninejad, A.;Nowrouzi, N.; Mohebbi, G. and Asghari, B. Application of PEG-400 as a green biodegradable polymeric medium for the catalyst-free synthesis of spiro-dihydropyridines and their use as acetyl and butyrylcholinesterase inhibitors. *Bioorg. Med.Chem*. 2016, *24*(6), pp.1408-1417.
154. Nadri, H.; Pirali-Hamedani, M.;Moradi, A.; Sakhteman, A.; Vahidi, A.; Sheibani, V.; Asadipour, A.; Hosseinzadeh, N.; Abdollahi, M.;Shafiee, A. and Foroumadi, A. 5, 6-Dimethoxybenzofuran-3-one derivatives: a novel series of dual Acetylcholinesterase/Butyrylcholinesterase inhibitors bearing benzyl pyridinium moiety. *DARU J. Pharm. Sci.* 2013, *21*(1), p.15.
155. Ilhami, G.;Malahat, A.; Taslimi, P.; Zubeyir, H.; Leyla, S.; Afsun, S.; Vagif, F.;Sukru, B.; Saleh H, A. and ClaudiuT, S. Synthesis and biological evaluation of aminomethyl and alkoxymethyl derivatives as carbonic anhydrase, acetylcholinesterase and butyrylcholinesterase inhibitors2017.
156. Sakkiah, S. and Lee, K.W. Pharmacophore-based virtual screening and density functional theory approach to identifying novel butyrylcholinesterase inhibitors. *ActaPharmacologicaSinica*, 2012, *33*(7), p.964.
157. Zeb, A.; Hameed, A.; Khan, L.; Khan, I.; Dalvandi, K.;IqbalChoudhary, M. and Z Basha, F. Quinoxaline derivatives: Novel and selective butyrylcholinesterase inhibitors. *Med. Chem.* 2014, *10*(7), pp.724-729.
158. Chen, Y.; Lin, H.; Yang, H.; Tan, R.; Bian, Y.; Fu, T.; Li, W.;Wu, L.; Pei, Y. and Sun, H. Discovery of new acetylcholinesterase and butyrylcholinesterase inhibitors through structure-based virtual screening. *RSC advances*, 2017, *7*(6), pp.3429-3438.
159. Guido, R.V.; Castilho, M.S.; Mota, S.G.;Oliva, G. and Andricopulo, A.D. Classical and Hologram QSAR Studies on a Series of Inhibitors of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase. *QSAR Comb. Sci.* 2008, *27*(6), pp.768-781.
160. Cherif, O.; Allouche, F.;Chabchoub, F.; Chioua, M.; Soriano, E.;Yañez, M.; Cacabelos, R.; Romero, A.;López, M.G. and Marco-Contelles, J. Isoxazolotacrines as non-toxic and selective butyrylcholinesterase inhibitors for Alzheimer's disease. *Future med.Chem.* 2014, *6*(17), pp.1883-1891.
161. Silva, D; Chioua, M; Samadi, A; Agostinho, P; Garção, P; Lajarín-Cuesta, R; de los Ríos C; Iriepa, I; Moraleda, I; Gonzalez-Lafuente, L; Mendes, E. Synthesis, pharmacological assessment, and molecular modeling of acetylcholinesterase/butyrylcholinesterase inhibitors: effect against amyloid-β-induced neurotoxicity. *ACS Chem.Neurosci*. 2013, 17, 4(4), pp.547-65.
162. Decker, M.; Kraus, B. and Heilmann, J. Design, synthesis and pharmacological evaluation of hybrid molecules out of quinazolinimines and lipoic acid lead to highly potent and selective butyrylcholinesterase inhibitors with antioxidant properties. *Bioorg. Med. Chem.* 2008, *16*(8), pp.4252-4261.
163. Schott, Y., Decker, M.;Rommelspacher, H. and Lehmann, J. 6-Hydroxy-and 6-methoxy-β-carbolines as acetyl-and butyrylcholinesterase inhibitors. *Bioorg. Med. Chem..lett.* 2006, *16*(22), pp.5840-5843.
164. Dolles, D.; Nimczick, M.; Scheiner, M.; Ramler, J.; Stadtmüller, P.;Sawatzky, E.;Drakopoulos, A., Sotriffer, C., Wittmann, H.J., Strasser, A. and Decker, M. Aminobenzimidazoles and Structural Isomers as Templates for Dual-Acting Butyrylcholinesterase Inhibitors and hCB2R Ligands To Combat Neurodegenerative Disorders. *Chem. Med. Chem*. 2016, *11*(12), pp.1270-1283.
165. Granica, S.; Kiss, A.K.; Jarończyk, M.; Maurin, J.K.;Mazurek, A.P. and Czarnocki, Z. Synthesis of imperatorin analogs and their evaluation as acetylcholinesterase and butyrylcholinesterase inhibitors. *Archiv der Pharmazie*, 2013, *346*(11), pp.775-782.
166. Woo, Y.J.; Lee, B.H.; Yeun, G.H.; Kim, H.J.; Won, M.H.; Kim, S.H.; Lee, B.H. and Park, J.H. Selective butyrylcholinesterase inhibitors using polyphenol-polyphenol hybrid molecules. *B. Korean Chem. Soc*. 2011, *32*(8), pp.2593-2598.
167. Decker, M. Homobivalentquinazolinimines as novel nanomolar inhibitors of cholinesterases with dirigible selectivity toward butyrylcholinesterase. *J. med.Chem.* 2006, *49*(18), pp.5411-5413.
168. Conejo-García, A.; Pisani, L.; del Carmen Núnez, M.; Catto, M.; Nicolotti, O.; Leonetti, F.; Campos, J.M.; Gallo, M.A.; Espinosa, A. and Carotti, A. Homodimericbis-quaternary heterocyclic ammonium salts as potent acetyl-and butyrylcholinesterase inhibitors: a systematic investigation of the influence of linker and cationic heads over affinity and selectivity. *J. med.Chem.*, 2011, *54*(8), pp.2627-2645.
169. Chen, Y.; Sun, J.; Fang, L.; Liu, M.;Peng, S.; Liao, H.;Lehmann, J. and Zhang, Y. Tacrine–ferulic acid–nitric oxide (NO) donor trihybrids as potent, multifunctional acetyl-and butyrylcholinesterase inhibitors. *J. med.Chem.* 2012, *55*(9), pp.4309-4321.
170. Rivera-Becerril, E.; Joseph-Nathan, P.; Perez-Alvarez, V.M. and Morales-Rios, M.S. Synthesis and biological evaluation of (−)-and (+)-Debromoflustramine B and its analogues as selective butyrylcholinesterase inhibitors. *J. med.Chem.* 2008, *51*(17), pp.5271-5284.

171. Yu, Q.S.; Holloway, H.W.;Utsuki, T.; Brossi, A. and Greig, N.H. Synthesis of novel phenserine-based-selective inhibitors of butyrylcholinesterase for Alzheimer's disease. *J. med.Chem.*1999, *42*(10), pp.1855-1861.
172. Makhaeva, G.F.;Lushchekina, S.V.;Boltneva, N.P.; Sokolov, V.B.; Grigoriev, V.V.; Serebryakova, O.G.; Vikhareva, E.A.; Aksinenko, A.Y.;Barreto, G.E.; Aliev, G. and Bachurin, S.O. Conjugates of γ-Carbolines and Phenothiazine as new selective inhibitors of butyrylcholinesterase and blockers of NMDA receptors for Alzheimer Disease. *Scientific reports*, 2015, *5*, p.13164.
173. Jones, M.; Wang, J.; Harmon, S.; Kling, B.; Heilmann, J. and Gilmer, J. Novel selective butyrylcholinesterase inhibitors incorporating antioxidant functionalities as potential bimodal therapeutics for Alzheimer's disease. *Molecules*, 2016, *21*(4), p.440.
174. Delogu, G.L.; Matos, M.J.;Fanti, M.;Era, B., Medda, R., Pieroni, E.;Fais, A., Kumar, A. and Pintus, F. 2-Phenylbenzofuran derivatives as butyrylcholinesterase inhibitors: synthesis, biological activity and molecular modeling. *Bioorg. Med.Chem.Lett.* 2016, *26*(9), pp.2308-2313.
175. Shan, W.J.; Huang, L.; Zhou, Q.; Meng, F.C. and Li, X.S. Synthesis, biological evaluation of 9-N-substituted berberine derivatives as multi-functional agents of antioxidant, inhibitors of acetylcholinesterase, butyrylcholinesterase and amyloid-β aggregation. *Euro. J. med.Chem.* 2011, *46*(12), pp.5885-5893.
176. Liu, H., Liu, L.; Gao, X.; Liu, Y.;Xu, W.; He, W.; Jiang, H.; Tang, J.; Fan, H. and Xia, X. Novel ferulic amide derivatives with tertiary amine side chain as acetylcholinesterase and butyrylcholinesterase inhibitors: the influence of carbon spacer length, alkylamine and aromatic group. *Euro. J. med.Chem.* 2017, *126*, pp.810-822.
177. Sawatzky, E., Wehle, S., Kling, B., Wendrich, J., Bringmann, G., Sotriffer, C.A., Heilmann, J. and Decker, M. Discovery of highly selective and nanomolarcarbamate-based butyrylcholinesterase inhibitors by rational investigation into their inhibition mode. *Euro. J. med.Chem.* 2016, *59*(5), pp.2067-2082.
178. Khoobi, M.;Alipour, M.;Sakhteman, A.;Nadri, H., Moradi, A.;Ghandi, M.;Emami, S.; Foroumadi, A. and Shafiee, A. Design, synthesis, biological evaluation and docking study of 5-oxo-4, 5-dihydropyrano [3, 2-c] chromene derivatives as acetylcholinesterase and butyrylcholinesterase inhibitors. *Euro. J. med.Chem.* 2013, *68*, pp.260-269.
179. Uddin, R.; Yuan, H.; Petukhov, P.A.;Choudhary, M.I. and Madura, J.D. Receptor-based modeling and 3D-QSAR for a quantitative production of the butyrylcholinesterase inhibitors based on genetic algorithm. *J. chem..inform. Model.* 2008, *48*(5), pp.1092-1103.
180. Asadipour, A.;Alipour, M.; Jafari, M.;Khoobi, M.;Emami, S.;Nadri, H.;Sakhteman, A., Moradi, A.;Sheibani, V.; Moghadam, F.H. and Shafiee, A. Novel coumarin-3-carboxamides bearing N-benzylpiperidine moiety as potent acetylcholinesterase inhibitors. *Euro. J. med.Chem.* 2013, *70*, pp.623-630.
181. Rahim, F.; Javed, M.T.;Ullah, H.; Wadood, A.; Taha, M.; Ashraf, M.; Khan, M.A.; Khan, F.; Mirza, S. and Khan, K.M. Synthesis, molecular docking, acetylcholinesterase and butyrylcholinesterase inhibitory potential of thiazole analogs as new inhibitors for Alzheimer disease. *Bioorg.Chem..*, 2015, *62*, pp.106-116.
182. Bolognesi, M.L.;Bartolini, M.;Cavalli, A.;Andrisano, V., Rosini, M.;Minarini, A. and Melchiorre, C. Design, synthesis, and biological evaluation of conformationally restricted rivastigmine analogues. *J. med.Chem..* 2004, *47*(24), pp.5945-5952.
183. Pang, X.; Fu, H.; Yang, S.; Wang, L.; Liu, A.L.; Wu, S. and Du, G.H. Evaluation of novel dual acetyl-and butyrylcholinesterase inhibitors as potential anti-Alzheimer's disease agents using pharmacophore, 3D-QSAR, and molecular docking approaches. *Molecules*, 2017, *22*(8), p.1254.
184. Gogoi, D.;Chaliha, A.K.;Sarma, D.;Kakoti, B.B. and Buragohain, A.K. Novel butyrylcholinesterase inhibitors through pharmacophore modeling, virtual screening and DFT-based approaches along-with design of bioisosterism-based analogues. *Biomed. Pharmacother.* 2017, *85*, pp.646-657.
185. M.K. Gilson, T. Liu, M.G. Baitaluk, G. Nicola, L. Hwang, and J. Chong, BindingDB: A public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Res. 19 (2015), pp. 1045–1053. Available at https://www.bindingdb.org/bind/index.jsp
186. Sterling, J., Herzig, Y., Goren, T., Finkelstein, N., Lerner, D., Goldenberg, W., & Toth, G. et al. (2002). Novel dual inhibitors of AChE and MAO derived from hydroxy aminoindan and phenethylamine as potential treatment for Alzheimer's disease. Journal of Medicinal Chemistry, 45(24), 5260–5279. doi:10.1021/jm020120c

187. Jaén, J.C.; Gregor, V.E.; Lee, C. et al. (1996) 'Acetylcholinesterase inhibition by fused dihydroquinazoline compounds', *Bioorg. Med. Chem. Lett.* 6(6), pp. 737-42.
188. Högenauer, K.; Baumann, K; EnzA et al. (2001) 'Synthesis and acetylcholinesterase inhibition of 5-desamino huperzineA derivatives', *Bioorg. Med. Chem. Lett.* 11(19), pp. 2627-30.
189. Andreani, A.; Burnelli, S.; Granaiola, M. et al. (2008) 'Chemiluminescent high-throughput microassay applied to imidazo [2, 1-b] thiazole derivatives as potential acetylcholinesterase and butyrylcholinesterase inhibitors', *Eur. J. Med. Chem*. 43(3), pp. 657-61.
190. Carlier, P.R.; Chow, E.S.; Han, Y. (1999) 'Heterodimerictacrine-based acetylcholinesterase inhibitors: investigating ligand− peripheral site interactions', *J. Med. Chem*. 42(20), pp. 4225-31.
191. Ceschi, M.A.; da Costa, J.S.; Lopes, J.P. et al. (2016) 'Novel series of tacrine-tianeptine hybrids: Synthesis, cholinesterase inhibitory activity, S100B secretion and a molecular modeling approach', *Eur. J. Med. Chem*. 121, pp. 758-72.
192. Contreras, J.M.; Rival, Y.M.; Chayer, S. et al. (1999) 'Aminopyridazines as acetylcholinesterase inhibitors', *J. Med. Chem.* 42(4), pp. 730-41.
193. da Costa, J.S.; Lopes, J.P.; Russowsky, D. et al. (2013) 'Synthesis of tacrine-lophine hybrids via one-pot four component reaction and biological evaluation as acetyl-and butyrylcholinesterase inhibitors', *Eur. J. Med. Chem.* 62, pp. 556-63.
194. DeBernardis, J.F.; Gifford, P.; Rizk, M. (1988) 'Evaluation of the side arm of (naphthylvinyl) pyridinium inhibitors of choline acetyltransferase', *J. Med. Chem*. 31(1), pp. 117-21.
195. Anand, P.; Singh, B. (2012) 'Synthesis and evaluation of novel 4-[(3H, 3aH, 6aH)-3-phenyl)-4, 6-dioxo-2-phenyldihydro-2H-pyrrolo [3, 4-d] isoxazol-5 (3H, 6H, 6aH)-yl] benzoic acid derivatives as potent acetylcholinesterase inhibitors and anti-amnestic agents', *Bioorg. Med. Chem.* 20(1), pp. 521-30.
196. Barreiro, E.J.; Camara, C.A.; Verli, H. (2003) 'Design, synthesis, and pharmacological profile of novel fused pyrazolo [4, 3-d] pyridine and pyrazolo [3, 4-b][1, 8] naphthyridineisosteres: a new class of potent and selective acetylcholinesterase inhibitors', *J. Med. Chem*. 46(7), pp. 1144-52.
197. Cardoso, C.L.; Castro-Gamboa, I.; Silva, D.H. (2004) 'Indoleglucoalkaloids from Chimarrhisturbinata and their evaluation as antioxidant agents and acetylcholinesterase inhibitors', *J. Nat. Prod.* 67(11), pp. 1882-5.
198. Li, X.; Wang, H.; Lu, Z. (2016) 'Development of multifunctional pyrimidinylthiourea derivatives as potential anti-Alzheimer agents', *J Med. Chem*. 59(18), pp. 8326-44.
199. Erlanson, D.A.; McDowell, R.S.; O'Brien, T. (2004) 'Fragment-based drug discovery', *J. Med. Chem*. 47(14), pp. 3463-82.
200. Feng, S.; Xia, Y.; Han, D. (2005) 'Synthesis and acetylcholinesterase inhibition of derivatives of huperzine B', *Bioorg. Med. Chem. Lett.* 15(3), pp. 523-6.
201. Fink, D.M.; Bores, G.M.; Effland, R.C. et al. (1995) 'Synthesis and Evaluation of 5-Amino-5, 6, 7, 8-tetrahydroquinolinones as Potential Agents for the Treatment of Alzheimer's Disease', *J. Med. Chem*. 38(18), pp. 3645-51.
202. Girisha, H.R.; Chandra, J.N.; Boppana, S. et al. (2009) 'Active site directed docking studies: synthesis and pharmacological evaluation of cis-2, 6-dimethyl piperidine sulfonamides as inhibitors of acetylcholinesterase', *Eur. J. Med. Chem.* 44(10), pp. 4057-62.
203. Gray, A.P.; Platz, R.D.; Chang, T.C. et al. (1985) 'Synthesis of some quaternary ammonium alkylating agents and their effects on soman-inhibited acetylcholinesterase', *J. Med. Chem.* 28(1), pp. 111-16.
204. Han, S.Y.; Mayer, S.C.; Schweiger, E.J. (1991) 'Synthesis and biological activity of galanthamine derivatives as acetylcholinesterase (AChE) inhibitors', *Bioorg. Med. Chem. Lett.* 1(11), pp. 579-80.
205. He, X.C.; Feng, S.; Wang, Z.F. (2007) 'Study on dual-site inhibitors of acetylcholinesterase: Highly potent derivatives of bis-and bifunctionalhuperzine B', *Bioorg. Med. Chem.* 15(3), pp. 1394-408.
206. Hu, M.K.; Wu, L.J.; Hsiao, G. et al. (2002) 'Homodimerictacrine congeners as acetylcholinesterase inhibitors', *J. Med. Chem*. 45(11), pp. 2277-82.

207. Hu, Y.; Zhang, J.; Chandrashankra, O. (2013) 'Design, synthesis and evaluation of novel heterodimers of donepezil and huperzine fragments as acetylcholinesterase inhibitors', *Bioorg. Med. Chem*. 21(3), pp. 676-83.
208. Huang, W.; Tang, L.; Shi, Y. (2011) 'Searching for the multi-target-directed ligands against Alzheimer 's disease: discovery of quinoxaline-based hybrid compounds with AChE, H3R and BACE 1 inhibitory activities', *Bioorg Med. Chem*. 19(23), pp. 7158-67.
209. Ishihara, Y.; Hirai, K.; Miyamoto, M. (1994) 'Central cholinergic agents. 6. Synthesis and evaluation of 3-[1-(phenylmethyl)-4-piperidinyl]-1-(2, 3, 4, 5-tetrahydro-1H-1-benzazepin-8-yl)-1-propanones and their analogs as central selective acetylcholinesterase inhibitors', *J. Med. Chem*. 37(15), pp. 2292-99.
210. Jia, P.; Sheng, R.; Zhang, J. (2009) 'Design, synthesis and evaluation of galanthamine derivatives as acetylcholinesterase inhibitors', *Eur. J. Med. Chem.* 44(2), pp. 772-84.
211. Kapples, K.J.; Shutske, G.M.; Bores, G.M. et al. (1993) 'Synthesis and in vitro acetylcholinesterase inhibitory activity of some 1-substituted analogues of velnacrine', *Bioorg. Med. Chem. Lett.* 3(12), pp. 2789-92.
212. Kavitha, C.V.; Gaonkar, S.L.; Chandra, J.N. (2007) 'Synthesis and screening for acetylcholinesterase inhibitor activity of some novel 2-butyl-1, 3-diaza-spiro [4, 4] non-1-en-4-ones: derivatives of irbesartan key intermediate', *Bioorg. Med. Chem.* 5(23), pp. 7391-98.
213. Li, Y.; Peng, P.; Tang, L. (2014) 'Design, synthesis and evaluation of rivastigmine and curcumin hybrids as site-activated multitarget-directed ligands for Alzheimer's disease therapy', *Bioorg. Med. Chem.* 22(17), pp. 4717-25.
214. Li, Y.; Qiang, X. and Luo L et al. (2017) 'Multitarget drug design strategy against Alzheimer's disease: HomoisoflavonoidMannich base derivatives serve as acetylcholinesterase and monoamine oxidase B dual inhibitors with multifunctional properties', *Bioorg. Med. Chem*. 2017;25(2):714-26.
215. Liu, H.R.; Huang, X.Q.; Lou, D.H. et al. (2014) 'Synthesis and acetylcholinesterase inhibitory activity of Mannich base derivatives flavokawain B', *Bioorg. Med. Chem. Lett*. 24(19), pp. 4749-53.
216. McKenna, .M..; Proctor, G.R.; Young, L.C. et al. (1997) 'Novel tacrine analogues for potential use against Alzheimer's disease: potent and selective acetylcholinesterase inhibitors and 5-HT uptake inhibitors', *J. Med. Chem.* 40(22), pp. 3516-23.
217. Morini, G.; Comini, M.; Rivara, M. et al. (2008) 'Synthesis and structure–activity relationships for biphenyl H3 receptor antagonists with moderate anti-cholinesterase activity', *Bioorg. Med. Chem*. 16(23), pp. 9911-24.
218. Pool, W.F., Woolf, T.F.; Reily, M.D. et al. (1996) 'Identification of a 3-hydroxylated tacrine metabolite in rat and man: metabolic profiling implications and pharmacology', *J. Med. Chem.* 39(15), pp. 3014-18.
219. Anand, P.; Singh, B.; Singh, N. (2012) 'A review on coumarins as acetylcholinesterase inhibitors for Alzheimer's disease', *Bioorg. Med. Chem.* 20(3), pp. 1175-80.
220. Sadashiva, C.T.; Chandra, J.N.; Ponnappa, K.C. (2006) 'Synthesis and efficacy of 1-[bis (4-fluorophenyl)-methyl] piperazine derivatives for acetylcholinesterase inhibition, as a stimulant of central cholinergic neurotransmission in Alzheimer's disease', *Bioorg. Med. Chem. Lett.* 16(15), pp. 3932-36.
221. Sang, Z.; Qiang, X.' Li, Y. et al. (2015) 'Design, synthesis and evaluation of scutellarein-O-alkylamines as multifunctional agents for the treatment of Alzheimer's disease', *Eur. J. Med. Chem.* 94, pp. 348-66.
222. Sang, Z.; Li, Y.; Qiang, X. et al. (2015) 'Multifunctional scutellarin–rivastigmine hybrids with cholinergic, antioxidant, biometal chelating and neuroprotective properties for the treatment of Alzheimer's disease', *Bioorg. Med. Chem*. 23(4), pp. 668-80.
223. Sang, Z.P.; Qiang, X.M.; Li, Y. et al. (2015) 'Design, synthesis, and biological evaluation of scutellareincarbamate derivatives as potential multifunctional agents for the treatment of alzheimer's disease', *Chem. Bio. Drug Des.* 86(5), pp. 1168-77.
224. dos Santos, P.D.; da Costa, J.S.; Gamba, D. et al. (2010) 'Synthesis and AChE inhibitory activity of new chiral tetrahydroacridine analogues from terpeniccyclanones', *Eur. J. Med. Chem*. 45(2), pp. 526-35.

225. Shao, D.; Zou, C.; Luo, C. et al. (2004) 'Synthesis and evaluation of tacrine–E2020 hybrids as acetylcholinesterase inhibitors for the treatment of alzheimer's disease', *Bioorg. Med. Chem. Lett.* 14(18), pp. 4639-42.
226. Shen, Y.; Sheng, R.; Zhang, J. (2008) '2-Phenoxy-indan-1-one derivatives as acetylcholinesterase inhibitors: A study on the importance of modifications at the side chain on the activity', *Bioorg. Med. Chem*. 16(16), pp. 7646-53.
227. Sheng, R.; Lin, X.; Li, J. (2005) 'Design, synthesis, and evaluation of 2-phenoxy-indan-1-one derivatives as acetylcholinesterase inhibitors', *Bioorg. Med. Chem. Lett.* 15(17), pp. 3834-37.
228. Sheng, R.; Xu, Y.; Hu, C. (2009) 'Design, synthesis and AChE inhibitory activity of indanone and aurone derivatives', *Eur. J. Med. Chem.* 44(1), pp. 7-17.
229. Sheng, R.; Lin, X.; Zhang, J. et al. (2009) 'Design, synthesis and evaluation of flavonoid derivatives as potent AChE inhibitors', *Bioorg. Med. Chem.* 17(18), pp. 6692-98.
230. Shi, D.H.; Huang, W.; Li, C. et al. (2013) 'Synthesis, biological evaluation and molecular modeling of aloe-emodin derivatives as new acetylcholinesterase inhibitors', *Bioorg. Med. Chem*. 21(5), pp. 1064-73.
231. Shinada, M.; Narumi, F.; Osada, Y. (2012) 'Synthesis of phenserine analogues and evaluation of their cholinesterase inhibitory activities', *Bioorg. Med. Chem*. 20(16), pp. 4901-14.
232. Shutske, G.M.; Pierrat, F.A.; Kapples, K.J. (1989) '9-Amino-1, 2, 3, 4-tetrahydroacridin-1-ols. Synthesis and evaluation as potential Alzheimer's disease therapeutics', *J. Med. Chem.* 32(8), pp. 1805-13.
233. Simoni, E.' Daniele, S.; Bottegoni, G. et al. (2012) 'Combining galantamine and memantine in multitargeted, new chemical entities potentially useful in Alzheimer's disease', *J. Med. Chem.* 55(22), pp. 9708-21.
234. Sugimoto, H.; Iimura, Y.; Yamanishi, Y. et al. (1992) 'Synthesis and anti-acetylcholinesterase activity of 1-benzyl-4-[(5, 6-dimethoxy-1-indanon-2-yl) methyl] piperidine hydrochloride (E2020) and related compounds', *Bioorg. Med. Chem. Lett*. 2(8), pp. 871-76.
235. Sugimoto, H.; Iimura, Y.; Yamanishi, Y. et al. (1995) 'Synthesis and structure-activity relationships of acetylcholinesterase inhibitors: 1-benzyl-4-[(5, 6-dimethoxy-1-oxoindan-2-yl) methyl] piperidine hydrochloride and related compounds', *J. Med. Chem*. 38(24), pp. 4821-29.
236. Valenti, P.; Rampa, A.; BisiA. et al. (1997) 'Acetylcholinesterase inhibition by tacrine analogues', *Bioorg. Med. Chem. Lett*. 7(20), pp. 2599-602.
237. Vidaluc, J.L.; Calmel, F.; Bigg, D. et al. (1994) 'Novel [2-(4-piperidinyl) ethyl](thio) ureas: synthesis and antiacetylcholinesterase activity', *J. Med. Chem*. 37(5), pp. 689-95.
238. Vidaluc, J.L.; Calmel. F.; Bigg, D.C. et al. (1995) 'Flexible 1-[(2-aminoethoxy) alkyl]-3-ar (o) yl (thio) ureas as novel acetylcholinesterase inhibitors. Synthesis and biochemical evaluation', *J. Med. Chem.* 38(15), pp. 2969-73.
239. Wong, D.M.; Greenblatt, H.M.; Dvir, H. (2003) 'Acetylcholinesterasecomplexed with bivalent ligands related to huperzine A: Experimental evidence for species-dependent protein− ligand complementarity', *ACS*. 125(2), pp. 363-73.
240. Yang, X.; Qiang, X.; Li, Y. (2017) 'Pyridoxine-resveratrol hybrids Mannich base derivatives as novel dual inhibitors of AChE and MAO-B with antioxidant and metal-chelating properties for the treatment of alzheimer's disease', *Bioorg. Chem*. 71, pp. 305-14.
241. Zeng, F.; Jiang, H.; Zhai, Y. et al. (1999) 'Synthesis and acetylcholinesterase inhibitory activity of huperzine A—E2020 combined compound', *Bioorg. Med. Chem. Lett.* 9(23), pp. 3279-84.
242. Zhan, Z.J.; Bian, H.L.; Wang, J.W. et al. (2010) 'Synthesis of physostigmine analogues and evaluation of their anticholinesterase activities', *Bioorg. Med. Chem. Lett.* 20(5), pp. 1532-34.
243. Zheng, H.; Youdim, M.B.; Fridkin, M. (2010) 'Site-activated chelators targeting acetylcholinesterase and monoamine oxidase for alzheimer's therapy', *ACS Chem. Bio.* 5(6), pp. 603-10.
244. Zhu, Y.; Xiao, K.; Ma, L. et al. (2009) 'Design, synthesis and biological evaluation of novel dual inhibitors of acetylcholinesterase and β-secretase', *Bioorg. Med. Chem.* 17(4), pp. 1600-613.

245. Szymański, P.; Markowicz, M.; Mikiciuk-Olasik, E. (2011) 'Synthesis and biological activity of derivatives of tetrahydroacridine as acetylcholinesterase inhibitors', *Bioorg. Chem.* 39(4), pp. 138-42.
246. Krátký, M.; Štěpánková, Š.; Vorčáková, K. et al. (2015) 'Salicylanilide diethyl phosphates as cholinesterases inhibitors', *Bioorg. Chem.* 58, pp. 48-52.
247. Krátký, M.; Štěpánková, Š.; Vorčáková, K. et al. (2017) 'Synthesis of readily available fluorophenylalanine derivatives and investigation of their biological activity', *Bioorg. Chem.* 71, pp. 244-56.
248. Rodríguez-Franco, M.I.; Fernández-Bachiller, M.I.; Pérez, C. et al. (2006) 'Novel tacrine− melatonin hybrids as dual-acting drugs for Alzheimer disease, with improved acetylcholinesterase inhibitory and antioxidant properties', *J. Med. Chem.* 49(2), pp. 459-62.
249. Conejo-García, A., Pisani, L., del Carmen Núnez, M. et al. (2011) 'Homodimericbis-quaternary heterocyclic ammonium salts as potent acetyl-and butyrylcholinesterase inhibitors: a systematic investigation of the influence of linker and cationic heads over affinity and selectivity', *J. Med. Chem*. 54(8), pp.2627-45.
250. Rodríguez, Y.A.; Gutiérrez, M.; Ramírez, D. et al. (2016) 'Novel N-allyl/propargyltetrahydroquinolines: Synthesis via three-component cationic iminodiels–alder reaction, binding prediction, and evaluation as cholinesterase inhibitors', *Chem. Bio. Drug Des.* 88(4), pp. 498-510.
251. Yurttaş, L.; Kaplancıklı, Z.A.; Özkay, Y. (2013) 'Design, synthesis and evaluation of new thiazole-piperazines as acetylcholinesterase inhibitors', *J. Enzyme Inhib.* 28(5), pp. 1040-47.
252. Zelík, P.; Lukešová, A.; Čejka, J. et al. (2010) 'Nostotrebin 6, a bis (cyclopentenedione) with cholinesterase inhibitory activity isolated from Nostoc sp. str. Lukešová 27/97', *J. Enzyme Inhib.* 25(3), pp. 414-20.
253. Ahmad, I.; Fatima, I. (2008) 'Butyrylcholinesterase, lipoxygenase inhibiting and antifungal alkaloids from Isatistinctoria', *J. Enzyme Inhib*. 23(3), pp. 313-16.
254. Ahmad, S,; Iftikhar, F.; Ullah, F. et al. (2016) 'Rational design and synthesis of dihydropyrimidine based dual binding site acetylcholinesterase inhibitors', *Bioorg. Chem.* 69, pp. 91-101.
255. Ahmed, E.; Nawaz, S.A.; Malik, A. et al. (2006) 'Isolation and cholinesterase-inhibition studies of sterols from Haloxylonrecurvum', *Bioorg. Med. Chem. Lett*. 16(3), pp. 573-80.
256. Bacalhau, P.; San Juan, A.A.; Marques, C.S. et al. (2016) 'New cholinesterase inhibitors for Alzheimer's disease: Structure activity studies (SARs) and molecular docking of isoquinolone and azepanone derivatives', *Bioorg. Chem.* 67, pp. 1-8.
257. Bagheri, S.M.; Khoobi, M.; Nadri, H. et al. (2015) 'Synthesis and anticholinergic activity of 4-hydroxycoumarin derivatives containing substituted benzyl-1, 2, 3-triazole moiety', *Chem. Bio. Drug Des.* 86(5), pp. 1215-20.
258. Bolognesi, M.L.; Andrisano, V.; Bartolini, M. et al. (2005) 'Propidium-based polyamine ligands as potent inhibitors of acetylcholinesterase and acetylcholinesterase-induced amyloid-β aggregation', *J. Med. Chem*. 48(1), pp. 24-7.
259. Camps, P.; Formosa, X.; Galdeano, C. et al. (2008) 'Novel donepezil-based inhibitors of acetyl-and butyrylcholinesterase and acetylcholinesterase-induced β-amyloid aggregation', *J. Med. Chem*. 51(12), pp. 3588-98.
260. Carlier, P.R.; Du, D.M.; Han, Y. et al. (1999) 'Potent, easily synthesized huperzine A-tacrine hybrid acetylcholinesterase inhibitors' *Bioorg. Med. Chem. Letts*. 9(16), pp. 2335-38.
261. Cho, J.K.; Ryu, Y.B.; Curtis-Long, M.J. (2017) 'Cholinestrase inhibitory effects of geranylated flavonoids from Paulownia tomentosa fruits', *Bioorg. Med. Chem*. 20(8), pp. 2595-602.
262. Czarnecka, K.; Szymański, P.; Girek, M. et al. (2017) 'Tetrahydroacridine derivatives with fluorobenzoic acid moiety as multifunctional agents for alzheimer's disease treatment', *Bioorg. Chem*. 72, pp. 315-22.
263. Decker, M. (2006) 'Homobivalentquinazolinimines as novel nanomolar inhibitors of cholinesterases with dirigible selectivity toward butyrylcholinesterase', *J. Med. Chem.* 49(18), pp. 5411-13.

264. Decker, M.; Kraus, B.; Heilmann, J. (2008) 'Design, synthesis and pharmacological evaluation of hybrid molecules out of quinazolinimines and lipoic acid lead to highly potent and selective butyrylcholinesterase inhibitors with antioxidant properties', *Bioorg. Med. Chem.* 16(8), pp. 4252-61.
265. Fang, L.; Appenroth, D.; Decker, M. et al. (2008) 'NO-donating tacrine hybrid compounds improve scopolamine-induced cognition impairment and show less hepatotoxicity', *J. Med. Chem.* 51(24), pp. 7666-69.
266. Fang, L.; Kraus, B.; Lehmann, J. et al. (2008) 'Design and synthesis of tacrine–ferulic acid hybrids as multi-potent anti-Alzheimer drug candidates', *Bioorg. Med. Chem. Lett*. 18(9), pp. 2905-909.
267. Feng, S.; Wang, Z.; He, X. et al. (2005) 'Bis-huperzine B: highly potent and selective acetylcholinesterase inhibitors', *J. Med. Chem.* 48(3), pp. 655-57.
268. Lin, G.; Chen, GH.; Ho, HC. (1998) 'Conformationally restricted carbamate inhibitors of horse serum butyrylcholinesterase', *Bioorg. Med. Chem. Lett.* 8(19), pp. 2747-50.
269. Gregor, V.E.; Emmerling, M.R.; Lee, C. et al. (1992) 'The synthesis and in vitro acetylcholinesterase and butyrylcholinesterase inhibitory activity of tacrine (Cognex®) derivaties', *Bioorg. Med. Chem. Lett.* 2(8), pp. 861-64.
270. Hameed, A.; Zehra, S.T.; Shah, S.J. et al. (2015) 'Syntheses, cholinesterases inhibition, and molecular docking studies of pyrido [2, 3-b] pyrazine derivatives', *Chem. Bio. Drug Des.* 86(5), pp. 1115-20.
271. Hameed, A.; Zehra , S.T.; Abbas, S. et al. (2016) 'One-pot synthesis of tetrazole-1, 2, 5, 6-tetrahydronicotinonitriles and cholinesterase inhibition: Probing the plausible reaction mechanism via computational studies', *Bioorg. Chem*. 65, pp. 38-47.
272. Hasan, A.; Khan, K.M.; Sher, M. et al. (2005) 'Synthesis and inhibitory potential towards acetylcholinesterase, butyrylcholinesterase and lipoxygenase of some variably substituted chalcones', *J. Enzyme Inhib.* 20(1), pp. 41-47.
273. Huang, L.; Luo, Z.; He, F. et al. (2010) 'Berberine derivatives, with substituted amino groups linked at the 9-position, as inhibitors of acetylcholinesterase/butyrylcholinesterase', *Bioorg. Med. Chem. Lett.* 20(22), pp. 6649-52.
274. Jiang, H.; Wang, X.; Huang, L. et al. (2011) 'Benzenediol-berberine hybrids: Multifunctional agents for alzheimer's disease', *Bioorg. Med. Chem.* 19(23), pp. 7228-35.
275. Kanhed, A.M.; Sinha, A.; Machhi, J. et al. (2015) 'Discovery of isoalloxazine derivatives as a new class of potential anti-Alzheimer agents and their synthesis', *Bioorg. Chem.* 61, pp. 7-12.
276. Kurt, B.Z.; Gazioglu, I.; Sonmez, F. et al. (2015) 'Synthesis, antioxidant and anticholinesterase activities of novel coumarylthiazole derivatives', *Bioorg. Chem*. 59, pp. 80-90.
277. Leader, H.; Wolfe, A.D.; Chiang, P.K. et al. (2002) 'Pyridophens: binary pyridostigmine− aprophenprodrugs with differential inhibition of acetylcholinesterase, butyrylcholinesterase, and muscarinic receptors', *J. Med. Chem*. 45(4), pp. 902-10.
278. Leng, J.; Qin, H.L.; Zhu, K. et al. (2016) 'Evaluation of multifunctional synthetic tetralone derivatives for treatment of alzheimer's disease', *Chem. Bio. Drug Des*. 88(6), pp. 889-98.
279. Li, Z.; Wang, B.; Hou, J.Q. et al. (2013) '2-(2-indolyl-)-4 (3 H)-quinazolinesderivates as new inhibitors of AChE: design, synthesis, biological evaluation and molecular modelling', *J. Enzyme Inhib*. 28(3), pp. 583-92.
280. Li, Y. Qiang, X.; Luo, L. et al. (2017) 'AuroneMannich base derivatives as promising multifunctional agents with acetylcholinesterase inhibition, anti-β-amyloid aggragation and neuroprotective properties for the treatment of alzheimer's disease', *Eur. J. Med. Chem*. 126, pp. 762-75.
281. Luo, W.; Li, Y.P.; He, Y. et al. (2011) 'Design, synthesis and evaluation of novel tacrine-multialkoxybenzene hybrids as dual inhibitors for cholinesterases and amyloid beta aggregation', *Bioorg. Med. Chem.* 19(2), pp. 763-70.

282. Mohamed, T.; Zhao, X.; Habib, L.K. et al. (2011) 'Design, synthesis and structure–activity relationship (SAR) studies of 2, 4-disubstituted pyrimidine derivatives: Dual activity as cholinesterase and Aβ-aggregation inhibitors', *Bioorg. Med. Chem.* 19(7), pp. 2269-81.
283. Mohammadi, K.M.; Mahdavi, M.; Saeedi, M. et al. (2015) 'Design, Synthesis, Biological Evaluation, and Docking Study of Acetylcholinesterase Inhibitors: New Acridone-1, 2, 4-oxadiazole-1, 2, 3-triazole Hybrids', *Chem. Bio. Drug Des.* 86(6), pp. 1425-32.
284. Najafi, Z.; Saeedi, M.; Mahdavi, M. et al. (2016) 'Design and synthesis of novel anti-alzheimer's agents: Acridine-chromenone and quinoline-chromenone hybrids', *Bioorg. Chem*. 67, pp. 84-94.
285. Pouramiri, B.; Moghimi, S.; Mahdavi, M. et al. (2017) 'Synthesis and anticholinesterase activity of new substituted benzooxazole-based derivatives', *Chem. Bio. Drug Des.* 89(5), pp. 783-89.
286. Rydberg, E.H.; Brumshtein, B.; Greenblatt, H.M. et al. (2006) 'Complexes of Alkylene-linked Tacrine dimers with Torpedo c alifornicaacetylcholinesterase: Binding of Bis (5)-tacrine produces a dramatic rearrangement in the active-site gorge', *J. Med. Chem.* 49(18), pp. 5491-500.
287. Saeed, A.; Zaib, S.; Ashraf, S. et al. (2015) 'Synthesis, cholinesterase inhibition and molecular modelling studies of coumarin linked thiourea derivatives', *Bioorg. Chem*. 63, pp. 58-63.
288. Saeedi, M.; Safavi, M.; Karimpour, R.E. et al. (2017) 'Synthesis of novel chromenones linked to 1, 2, 3-triazole ring system: Investigation of biological activities against alzheimer's disease', *Bioorg. Chem.* 70, pp. 86-93.
289. Samadi, A.' Marco, C.J.; Soriano, E. et al. (2010) 'Multipotent drugs with cholinergic and neuroprotective properties for the treatment of alzheimer and neuronal vascular diseases. I. Synthesis, biological assessment, and molecular modeling of simple and readily available 2-aminopyridine-, and 2-chloropyridine-3, 5-dicarbonitriles', *Bioorg. Med. Chem.* 18(16), pp. 861-872.
290. Samadi, A.; de los Ríos, C.; Bolea, I. et al. (2012) 'Multipotent MAO and cholinesterase inhibitors for the treatment of alzheimer's disease: synthesis, pharmacological analysis and molecular modeling of heterocyclic substituted alkyl and cycloalkylpropargyl amine', *Eur. J. Med. Chem.* 52, pp. 251-62.
291. Sarfraz, M.; Sultana, N.; Rashid, U. et al. (2017) 'Synthesis, biological evaluation and docking studies of 2, 3-dihydroquinazolin-4 (1H)-one derivatives as inhibitors of cholinesterases', *Bioorg. Chem.* 70, pp. 237-44.
292. Shi, A.; Huang, L.; Lu, C. et al. (2011) 'Synthesis, biological evaluation and molecular modeling of novel triazole-containing berberine derivatives as acetylcholinesterase and β-amyloid aggregation inhibitors', *Bioorg. Med. Chem.* 19(7), pp. 2298-305.
293. Skrzypek, A.; Matysiak, J.; Karpińska, M.M. et al. (2013) 'Synthesis and anticholinesterase activities of novel 1, 3, 4-thiadiazole based compounds', *J. Enzyme Inhib*. 28(4), pp. 816-23.
294. Sterling, J.; Herzig, Y.; Goren, T. et al. (2002) 'Novel dual inhibitors of AChE and MAO derived from hydroxyaminoindan and phenethylamine as potential treatment for alzheimer's disease', *J. Med. Chem.* 45(24), pp. 5260-79.
295. Tang, H.; Ning, F.X.; Wei, Y.B. et al. (2007) 'Derivatives of oxoisoaporphine alkaloids: A novel class of selective acetylcholinesterase inhibitors', *Bioorg. Med. Chem. Lett.* 17(13), pp. 3765-68.
296. Villalobos, A.; Blake, J.F.; Biggers, C.K. et al. (1994) 'Novel benzisoxazole derivatives as potent and selective inhibitors of acetylcholinesterase', *J. Med. Chem*. 37(17), pp. 2721-34.
297. Wen, H.; Zhou, Y.; Lin, C. et al. (2007) 'Methyl 2-(2-(4-formylphenoxy) acetamido)-2-substituted acetate derivatives: A new class of acetylcholinesterase inhibitors', *Bioorg. Med. Chem. Lett.* 17(8), pp. 2123-25.
298. Yanovsky, I.; Finkin, G.E.; Zaikin, A. et al. (2012) 'Carbamate derivatives of indolines as cholinesterase inhibitors and antioxidants for the treatment of alzheimer's disease', *J. Med. Chem.* 55(23), pp. 10700-715.
299. Zakhari, J.S.; Kinoyama, I.; Hixon, M.S. et al. (2011) 'Formulating a new basis for the treatment against botulinum neurotoxin intoxication: 3, 4-Diaminopyridine prodrug design and characterization', *Bioorg. Med. Chem.* 19(21), pp. 6203-209.

300. Roy, K.; Kar, S. and Ambure, P. On a simple approach for determining applicability domain of QSAR models. Chemom. Intell. Lab. Syst. 2011, 2015, 145, pp.22-29.
301. Mauri A, Consonni V, Pavan M et al. Dragon software: An easy approach to molecular descriptor calculations. Match. 2006;56(2):237-48.
302. Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 2011;32(7):1466-74.
303. Roy, K., Das, R.N.; Ambure, P. and Aher, R.B. Be aware of error measures. Further studies on validation of predictive QSAR models. Chemom. Intell. Lab Syst. 2016, 152, pp.18-33.
304. Khan, Kabiruddin; Benfenati, Emilio; and Roy Kunal. Consensus QSAR modeling of toxicity of pharmaceuticals to different aquatic organisms: ranking and prioritization of the DrugBank database compounds, Ecotoxicol. Environ. Saf. 2019, (168), pp. 287-297.

305. Das, S.; Ojha, P.K. and Roy, K. Multilayered variable selection in QSPR: a case study of modeling melting point of bromide ionic liquids. IJQSPR, 2017, 2(1), pp.106-124.
306. Das, S.; Ojha, P.K. and Roy, K. Development of a temperature dependent 2D-QSPR model for viscosity of diverse functional ionic liquids. J. Mol. Liq. 2017, 240, pp.454-467.
307. Ojha, P.K. and Roy, K. Chemometric modeling of odor threshold property of diverse aroma components of wine. RSC adv. 2018, 8(9), pp.4750-4760.
308. Ojha, P.K. and Roy, K. PLS regression-based chemometric modeling of odorant properties of diverse chemical constituents of black tea and coffee. RSC adv. 2018, 8(5), pp.2293-2304.
309. Rücker, Christoph; Rücker Gerta; and Meringer Markus. y-Randomization and its variants in QSPR/QSAR. J. chem.. inform. Model. 2007, 6(47), pp. 2345-2357.
310. Veerasamy, Ravichandran; Rajak Harish; Jain Abhishek; Sivadasan Shalini; Varghese Christapher P and Agrawal R. K. Validation of QSAR models-strategies and importance. RRJoDDD. 2011, 2(3), pp. 511-519.
311. Khan, Kabiruddin; Khan P. M.; Lavado Giovanna; Valsecchi Cecile; Pasqualini Julia; Baderna Diego; Marzo Marco; Lombardo Anna; Roy Kunal and Benfenati Emilio. QSAR modeling of Daphnia magna and fish toxicities of biocides using 2D descriptors. Chemosphere,2019, (229), pp. 8-17.
312. De, Ruyck J; Brysbaert, G; Blossey, R; Lensink, MF. Molecular docking as a popular tool in drug design, an in silico travel. Adv Appl Bioinforma Chem, 2016, 9, pp. 1.
313. Pal, S.; Kumar, V.; Kundu, B., Bhattacharya, D.; Preethy, N.; Reddy, M.P. and Talukdar, A. Ligand-based Pharmacophore Modeling, Virtual Screening and Molecular Docking Studies for Discovery of Potential Topoisomerase I Inhibitors. CSBJ, 2019, 17, pp.291-310.
314. J. Jiaranaikulwanitch, C. Boonyarat, V.V. Fokin and O. Vajragupta, Triazolyl tryptoline derivatives as β-secretase inhibitors, Bioorg. Med. Chem. Lett. 20 (2010), pp. 6572-6. doi: 10.1016/j.bmcl.2010.09.043.
315. P. Nowak, D.C. Cole, A. Aulabaugh, J. Bard, R. Chopra, R. Cowling, K.Y. Fan, B. Hu, S. Jacobsen, M. Jani and G. Jin, Discovery and initial optimization of 5, 5′-disubstituted aminohydantoins as potent β-secretase (BACE1) inhibitors, Bioorg. Med. Chem. Lett. 20 (2010), pp. 632-5. doi: 10.1016/j.bmcl.2009.11.052.
316. M.S. Malamas, K. Barnes, Y. Hui, M. Johnson, F. Lovering, J. Condon, W. Fobare, W. Solvibile, J. Turner, Y. Hu and E.S. Manas, Novel pyrrolyl 2-aminopyridines as potent and selective human β-secretase (BACE1) inhibitors, Bioorg. Med. Chem. Lett. 20 (2010), pp. 2068-73. doi: 10.1016/j.bmcl.2010.02.075.
317. G. Larbig and B. Schmidt, Synthesis of tetramic and tetronic acids as β-secretase inhibitors, J. Comb. Chem. 8 (2006), pp. 480-90. doi: 10.1021/cc0600021.
318. Y. Niu, C. Ma, H. Jin, F. Xu, H. Gao, P. Liu, Y. Li, C. Wang, G. Yang and P. Xu, The Discovery of Novel β-Secretase Inhibitors: Pharmacophore Modeling, Virtual Screening, and Docking Studies, Chem. Bio. Drug Des. 79 (2012), pp. 972-80. doi: 10.1111/j.1747-0285.2012.01367.x.
319. D.C. Cole, E.S. Manas, J.R. Stock, J.S. Condon, L.D. Jennings, A. Aulabaugh, R. Chopra, R. Cowling, J.W. Ellingboe, K.Y. Fan and B.L. Harrison, Acylguanidines as small-molecule β-secretase inhibitors, J. Med. Chem. 49 (2006), pp. 6158-61. doi: 10.1021/jm0607451.

320. T.H. Al-Tel, R.A. Al-Qawasmeh, M.F. Schmidt, A. Al-Aboudi, S.N. Rao, S.S. Sabri and W. Voelter, Rational design and synthesis of potent dibenzazepine motifs as β-secretase inhibitors, J. Med. Chem. 52 (2009), pp. 6484-8. doi: 10.1021/jm9008482.
321. H.S. Park and C.H. Jun, A simple and fast algorithm for K-medoids clustering, Expert Syst. Appl. 36 (2009), pp. 3336-3341. doi: 10.1016/j.eswa.2008.01.039.
322. K. Roy, On some aspects of validation of predictive quantitative structure–activity relationship models, Expert opin. drug disco. 2 (2007), pp. 1567-1577. doi: 10.1517/17460441.2.12.1567.
323. R.B. Aher and K. Roy, QSAR and pharmacophore modeling of diverse aminothiazoles and aminopyridines for antimalarial potency against multidrug-resistant Plasmodium falciparum, Med. Chem. Res. 23 (2014), pp. 4238-4249. doi: 10.1007/s00044-014-0997-x.
324. H.J. Huang, C.C. Lee and C.Y.C. Chen, In silico design of BACE1 inhibitor for Alzheimer's disease by traditional Chinese medicine, BioMed Res. Int. (2014). doi: 10.1155/2014/741703.
325. A.G. Murzin, S.E. Brenner, T. Hubbard and C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol, 247 (1995) pp. 536-40. doi: 10.1016/S0022-2836(05)80134-2.
326. G. Wu, D.H. Robertson, C.L. Brooks and M. Vieth, Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm, J. Comput. Chem. 24 (2003), pp. 1549-1562. doi: 10.1002/jcc.10306.
327. H. Tang, H.T. Zhao, S.M. Zhong, Z.Y. Wang, Z.F. Chen, H. Liang, Novel oxoisoaporphine-based inhibitors of acetyl- and butyrylcholinesterase and acetylcholinesterase-induced beta-amyloid aggregation, Bioorganic Med. Chem. Lett. 22 (6) (2012) 2257–2261. doi:10.1016/j.bmcl.2012.01.090.
328. Y. Li, X. Qiang, Y. Li, X. Yang, L. Luo, G. Xiao, Z. Cao, Z. Tan, Y. Deng, Pterostilbene-O-acetamidoalkylbenzylamines derivatives as novel dual inhibitors of cholinesterase with anti-β-amyloid aggregation and antioxidant properties for the treatment of Alzheimer's disease, Bioorganic Med. Chem. Lett. 26 (8) (2016) 2035–2039. doi:10.1016/j.bmcl.2016.02.079.
329. D. Panek, A. Więckowska, T. Wichur, M. Bajda, J. Godyń, J. Jończyk, K. Mika, J. Janockova, O. Soukup, D. Knez, J. Korabecny, S. Gobec, B. Malawska, Design, synthesis and biological evaluation of new phthalimide and saccharin derivatives with alicyclic amines targeting cholinesterases, beta-secretase and amyloid beta aggregation, Eur. J. Med. Chem. 125 (2017) 676–695. doi:10.1016/j.ejmech.2016.09.078.
330. W. Luo, Y.P. Li, Y. He, S.L. Huang, J.H. Tan, T.M. Ou, D. Li, L.Q. Gu, Z.S. Huang, Design, synthesis and evaluation of novel tacrine-multialkoxybenzene hybrids as dual inhibitors for cholinesterases and amyloid beta aggregation, Bioorganic Med. Chem. 19 (2) (2011) 763–770. doi:10.1016/j.bmc.2010.12.022.
331. A. Shi, L. Huang, C. Lu, F. He, X. Li, Synthesis, biological evaluation and molecular modeling of novel triazole-containing berberine derivatives as acetylcholinesterase and β-amyloid aggregation inhibitors, Bioorganic Med. Chem. 19 (7) (2011) 2298–2305. doi:10.1016/j.bmc.2011.02.025.
332. J.W. Yan, Y.P. Li, W.J. Ye, S. Bin Chen, J.Q. Hou, J.H. Tan, T.M. Ou, D. Li, L.Q. Gu, Z.S. Huang, Design, synthesis and evaluation of isaindigotone derivatives as dual inhibitors for acetylcholinesterase and amyloid beta aggregation, Bioorganic Med. Chem. 20 (8) (2012) 2527–2534. doi:10.1016/j.bmc.2012.02.061.
333. F. Mao, L. Huang, Z. Luo, A. Liu, C. Lu, Z. Xie, X. Li, O-Hydroxyl- or o-amino benzylamine-tacrine hybrids: Multifunctional biometals chelators, antioxidants, and inhibitors of cholinesterase activity and amyloid-β aggregation, Bioorganic Med. Chem. 20 (19) (2012) 5884–5892. doi:10.1016/j.bmc.2012.07.045.
334. A. Więckowska, K. Więckowski, M. Bajda, B. Brus, K. Sałat, P. Czerwińska, S. Gobec, B. Filipek, B. Malawska, Synthesis of new N-benzylpiperidine derivatives as cholinesterase inhibitors with β-amyloid anti-aggregation properties and beneficial effects on memory in vivo, Bioorganic Med. Chem. 23 (10) (2015) 2445–2457. doi:10.1016/j.bmc.2015.03.051.
335. G.F. Zha, C.P. Zhang, H.L. Qin, I. Jantan, M. Sher, M.W. Amjad, M.A. Hussain, Z. Hussain, S.N.A. Bukhari, Biological evaluation of synthetic α,β-unsaturated carbonyl based cyclohexanone derivatives as neuroprotective novel inhibitors of acetylcholinesterase, butyrylcholinesterase and amyloid-β aggregation, Bioorganic Med. Chem. 24 (10) (2016) 2352–2359. doi:10.1016/j.bmc.2016.04.015.

336. M.L. Bolognesi, V. Andrisano, M. Bartolini, R. Banzi, C. Melchiorre, Propidium-Based Polyamine Ligands as Potent Inhibitors of Acetylcholinesterase and Acetylcholinesterase-Induced Amyloid-β Aggregation, J. Med. Chem. 48 (1) (2005) 24–27. doi:10.1021/jm049156q.
337. P. Camps, X. Formosa, C. Galdeano, T. Gómez, D. Muñoz-Torrero, M. Scarpellini, E. Viayna, A. Badia, M.V. Clos, A. Camins, M. Pallàs, M. Bartolini, F. Mancini, V. Andrisano, J. Estelrich, M. Lizondo, A. Bidon-Chanal, F.J. Luque, Novel Donepezil-Based Inhibitors of Acetyl- and Butyrylcholinesterase and Acetylcholinesterase-Induced β-Amyloid Aggregation, J. Med. Chem. 51 (12) (2008) 3588–3598. doi:10.1021/jm8001313.
338. S. Rizzo, C. Rivière, L. Piazzi, A. Bisi, S. Gobbi, M. Bartolini, V. Andrisano, F. Morroni, A. Tarozzi, J.-P. Monti, A. Rampa, Benzofuran-Based Hybrid Compounds for the Inhibition of Cholinesterase Activity, β Amyloid Aggregation, and Aβ Neurotoxicity, J. Med. Chem. 51 (10) (2008) 2883–2886. doi:10.1021/jm8002747.
339. M. Catto, A.A. Berezin, D. Lo Re, G. Loizou, M. Demetriades, A. De Stradis, F. Campagna, P.A. Koutentis, A. Carotti, Design, synthesis and biological evaluation of benzo[e][1,2,4]triazin-7(1H) -one and [1,2,4]-triazino[5,6,1-jk]carbazol-6-one derivatives as dual inhibitors of beta-amyloid aggregation and acetyl/butyryl cholinesterase, Eur. J. Med. Chem. 58 (2012) 84–97. doi:10.1016/j.ejmech.2012.10.003.
340. Y.P. Li, F.X. Ning, M.B. Yang, Y.C. Li, M.H. Nie, T.M. Ou, J.H. Tan, S.L. Huang, D. Li, L.Q. Gu, Z.S. Huang, Syntheses and characterization of novel oxoisoaporphine derivatives as dual inhibitors for cholinesterases and amyloid beta aggregation, Eur. J. Med. Chem. 46 (2011) 1572–1581. doi:10.1016/j.ejmech.2011.02.005.
341. F. Belluti, M. Bartolini, G. Bottegoni, A. Bisi, A. Cavalli, V. Andrisano, A. Rampa, Benzophenone-based derivatives: A novel series of potent and selective dual inhibitors of acetylcholinesterase and acetylcholinesterase-induced beta-amyloid aggregation, Eur. J. Med. Chem. 46 (2011) 1682–1693. doi:10.1016/j.ejmech.2011.02.019.
342. J. Rouleau, B.I. Iorga, C. Guillou, New potent human acetylcholinesterase inhibitors in the tetracyclic triterpene series with inhibitory potency on amyloid β aggregation, Eur. J. Med. Chem. 46 (2011) 2193–2205. doi:10.1016/j.ejmech.2011.02.073.
343. H. Tang, L.Z. Zhao, H.T. Zhao, S.L. Huang, S.M. Zhong, J.K. Qin, Z.F. Chen, Z.S. Huang, H. Liang, Hybrids of oxoisoaporphine-tacrine congeners: Novel acetylcholinesterase and acetylcholinesterase-induced β-amyloid aggregation inhibitors, Eur. J. Med. Chem. 46 (2011) 4970–4979. doi:10.1016/j.ejmech.2011.08.002.
344. W.J. Shan, L. Huang, Q. Zhou, F.C. Meng, X.S. Li, Synthesis, biological evaluation of 9-N-substituted berberine derivatives as multi-functional agents of antioxidant, inhibitors of acetylcholinesterase, butyrylcholinesterase and amyloid-β aggregation, Eur. J. Med. Chem. 46 (2011) 5885–5893. doi:10.1016/j.ejmech.2011.09.051.
345. N. Guzior, M. Bajda, M. Skrok, K. Kurpiewska, K. LewiÅski, B. Brus, A. Pišlar, J. Kos, S. Gobec, B. Malawska, Development of multifunctional, heterodimeric isoindoline-1,3-dione derivatives as cholinesterase and β-amyloid aggregation inhibitors with neuroprotective properties, Eur. J. Med. Chem. 92 (2015) 738–749. doi:10.1016/j.ejmech.2015.01.027.
346. N. García-Font, H. Hayour, A. Belfaitah, J. Pedraz, I. Moraleda, I. Iriepa, A. Bouraiou, M. Chioua, J. Marco-Contelles, M.J. Oset-Gasque, Potent anticholinesterasic and neuroprotective pyranotacrines as inhibitors of beta-amyloid aggregation, oxidative stress and tau-phosphorylation for Alzheimer's disease, Eur. J. Med. Chem. 118 (2016) 178–192. doi:10.1016/j.ejmech.2016.04.023.
347. D. Fourches, E. Muratov, A. Tropsha, Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research, 50 (7) (2010) 1889-204. doi:10.1021/ci100176x.
348. G.M, Mariana, J. J. Naveja, N. S. Cruz, and J. L. Medina-Franco, Open chemoinformatic resources to explore the structure, properties and chemical space of molecules, RSC advances, 7(85) (2017) 54153-54163. DOI: 10.1039/c7ra11831g.

349. G. Domenico, A. Lombardo, C. Toma, and E. Benfenati, A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications, J. cheminformatics, 10(1) (2018) 1-13. https://doi.org/10.1186/s13321-018-0315-6.
350. S. Fabian P., C.L. Mellor, T. Meinl, and M.T.D. Cronin, Screening Chemicals for Receptor-Mediated Toxicological and Pharmacological Endpoints: Using Public Data to Build Screening Tools within a KNIME Workflow, Mol. Inform. 34(2-3) (2015): 171-178. DOI: 10.1002/minf.201400188.
351. G. Landrum, RDKit: Open-source cheminformatics, (2006).
352. Khan, K., & Roy, K. (2019). Ecotoxicological QSAR modelling of organic chemicals against Pseudokirchneriella subcapitata using consensus predictions approach. SAR and QSAR in Environmental Res., 30(9), 665-681.
353. Leonard, J.T.; Roy, K. (2000) 'On selection of training and test sets for the development of predictive QSAR models, Qsar & Combinat. Sci. Mar. 25(3), pp. 235-51.
354. P. Ambure, A. Gajewicz-Skretna, M.N.D. Cordeiro, K. Roy, New workflow for QSAR model development from small data sets: small dataset curator and small dataset modeler. integration of data curation, exhaustive double cross-validation, and a set of optimal model selection techniques, J. Chem. Inf. Model. 59(10) (2019) 4070-4076, https://doi.org/10.1021/acs.jcim.9b00476
355. P. De, K. Roy, QSAR and QSAAR modeling of nitroimidazole sulfonamide radiosensitizers: application of small dataset modeling, Struct. Chem. 32(2) (2021) 631-642, https://doi.org/10.1007/s11224-021-01734-w
356. R.K. Mukherjee, V. Kumar, K. Roy, Ecotoxicological QSTR and QSTTR modeling for the prediction of acute oral toxicity of pesticides against multiple avian species, Environ. Sci. Technol. 56(1) (2021) 335-348, https://doi.org/10.1021/acs.est.1c05732
357. P. De, V. Kumar, S. Kar, K. Roy, J. Leszczynski, Repurposing FDA approved drugs as possible anti-SARS-CoV-2 medications using ligand-based computational approaches: sum of ranking difference-based model selection, Struct Chem. (2022) 1-13, https://doi.org/10.1007/s11224-022-01975-3
358. M. Chatterjee, K. Roy, Application of cross-validation strategies to avoid overestimation of performance of 2D-QSAR models for the prediction of aquatic toxicity of chemical mixtures, SAR QSAR Environ Res. 33 (6) (2022) 463-484, https://doi.org/10.1080/1062936X.2022.2081255
359. A. Banerjee, M. Chatterjee, P. De, K. Roy, Quantitative predictions from chemical read-across and their confidence measures, Chemom. Intell. Lab. Syst. 227 (2022) 104613, DOI: 10.1016/j.chemolab.2022.104613
360. A. Banerjee, K. Roy, First report of q-RASAR modeling toward an approach of easy interpretability and efficient transferability, Mol. Divers. (2022), DOI: 10.1007/s11030-022-10478-6
361. V. Kumar, S. Kar, P. De, K. Roy, J. Leszczynski, Identification of potential antivirals against 3CLpro enzyme for the treatment of SARS-CoV-2: A multi-step virtual screening study, SAR QSAR Environ. Res. 33(5) (2022) 357-386, https://doi.org/10.1080/1062936X.2022.2055140

362. Ivanciuc, Ovidiu, Balaban; Teodor-Silviu, and Balaban, Alexandru T. Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices. J. Math. Chem.1993, 1(12), pp. 309-318.
363. Yu, Qian-sheng; Holloway, Harold W.; Utsuki, Tadanobu; Brossi, Arnold and Greig, Nigel H. Synthesis of novel phenserine-based-selective inhibitors of butyrylcholinesterase for Alzheimer's disease. J. med. Chem.. 42, 10(42), pp. 1855-1861.
364. Fang, Jiansong; Xiaocong Pang; Ping Wu, Rong Yan; Li Gao, Chao Li; WenwenLian; Qi Wang; Ai-lin Liu and Du Guan-hua. Molecular Modeling on Berberine Derivatives toward BuChE: An Integrated Study with Quantitative Structure–Activity Relationships Models, Molecular Docking, and Molecular Dynamics Simulations. Chem. biol drug des. 87, no. 5 (2016): 649-663.
365. Zheng, Fang; Max, Zhan; Xiaoqin, Huang; Mohamed; Diwan, Hameed M. Abdul and Zhan, Chang-Guo. Modeling in vitro inhibition of butyrylcholinesterase using molecular docking, multi-linear regression and artificial neural network approaches. Bioorg. Med.Chem.. 22, no. 1 (2014): 538-549.
366. Bitam, S.; M. Hamadache; and Hanini S. Prediction of therapeutic potency of tacrine derivatives as BuChE inhibitors from quantitative structure–activity relationship modelling. SQER. 29, no. 3 (2018): 213-230.

367. Solomon, KamalakaranAnand; Srinivasan, Sundararajan and Veluchamy, Abirami. QSAR studies on N-aryl derivative activity towards Alzheimer's disease. Molecules 14, no. 4 (2009): 1448-1455.
368. P.L.A. Popelier and P.J. Smith, QSAR models based on quantum topological molecular similarity, Eur. J. Med. Chem. 41 (2006), pp. 862-873. doi: 10.1016/j.ejmech.2006.03.004.
369. K. Roy and R.N. Das, The "ETA" Indices in QSAR/QSPR/QSTR Research, In Pharmaceutical Sciences: Breakthroughs in Research and Practice, IGI Global 2017, pp. 978-1011. doi: 10.4018/978-1-5225-1762-7.ch038.
370. Planche A. Speck, V.V. Kleandrova and Vargas J.A. Rojas, QSAR model toward the rational design of new agrochemical fungicides with a defined resistance risk using substructural descriptors, Mol. Divers. 15 (2011), pp. 901-909. doi: 10.1007/s11030-011-9320-7.
371. B. Lei, J. Li, J. Lu, J. Du, H. Liu and X. Yao, Rational Prediction of the Herbicidal Activities of Novel Protoporphyrinogen Oxidase Inhibitors by Quantitative Structure Activity Relationship Model Based on Docking-Guided Active Conformation, J. Agr. Food. Chem. 57 (2009), pp. 9593-9598. doi: 10.1021/jf902010g.
372. D. Gadaleta, G.F. Mangiatordi, M. Catto, A. Carotti and O. Nicolotti, Applicability domain for QSAR models: where theory meets reality, IJQSPR 1 (2016), pp.45-63. doi: 10.4018/IJQSPR.2016010102.
373. C, Yoo and M, Shahlaei, The applications of PCA in QSAR studies: A case study on CCR5 antagonists, Chem. Biol. Drug. Des. 9 (2018), pp. 137-152. doi: 10.1111/cbdd.13064.
374. P. Ambure and K. Roy, Understanding the structural requirements of cyclic sulfone hydroxyethylamines as hBACE1 inhibitors against Aβ plaques in Alzheimer's disease: a predictive QSAR approach, RSC Adv. 6 (2016), pp. 28171-28186. doi: 10.1039/C6RA04104C.
375. P. Jain and H.R. Jadhav, Quantitative structure activity relationship analysis of aminoimidazoles as BACE-I inhibitors, Med. Chem. Res. 22 (2013), pp.1740-1746. doi: 10.1007/s00044-012-0166-z.
376. T. Hossain, M.A. Islam, R. Pal and A. Saha, Exploring structural requirement and binding interactions of β-amyloid cleavage enzyme inhibitors using molecular modeling techniques, Med. Chem. Res. 22 (2013), pp. 4766-4774. doi: 10.1007/s00044-012-0166-z.
377. S. Chakraborty and S. Basu, Multi-functional activities of citrus flavonoid narirutin in Alzheimer's disease therapeutics: An integrated screening approach and in vitro validation, Int. J. Biol. Macromol. 103 (2017), pp. 733-743. doi: 10.1016/j.ijbiomac.2017.05.110.
378. L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs, Environ. Health Perspect. 111 (10) (2003) 1361–1375. doi:10.1289/ehp.5758.
379. P.K. Ojha, K. Roy, Comparative QSARs for antimalarial endochins: Importance of descriptor-thinning and noise reduction prior to feature selection, Chemom. Intell. Lab. Syst. 109 (2) (2011) 146–161. doi:10.1016/j.chemolab.2011.08.007.
380. D. Castillo-González, M.Á. Cabrera-Pérez, M. Pérez-González, A.M. Helguera, A. Durán-Martínez, Prediction of telomerase inhibitory activity for acridinic derivatives based on chemical structure, Eur. J. Med. Chem. 44 (12) (2009) 4826–4840. doi:10.1016/j.ejmech.2009.07.029.
381. L. B. Kier, L. H. Hall, The E-state in database analysis: the PCBs as an example, Il Farmaco. 54 (6) (1999) 346-53.
382. E.B. de Melo, A new quantitative structure-property relationship model to predict bioconcentration factors of polychlorinated biphenyls (PCBs) in fishes using E-state index and topological descriptors, Ecotoxicol. Environ. Saf. 75 (2012) 213–222. doi:10.1016/j.ecoenv.2011.08.026.
383. L. Eriksson, P.L. Andersson, E. Johansson, M. Tysklind, Megavariate analysis of environmental QSAR data. Part I - A basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD), Mol. Divers. 10 (2) (2006) 169–186. doi:10.1007/s11030-006-9024-6.
384. F. Leal, C. da Silva Lima, R. de Alencastro, H. Castro, C. Rodrigues, M. Albuquerque, Hologram QSAR Models of a Series of 6-Arylquinazolin-4-Amine Inhibitors of a New Alzheimer's Disease Target: Dual

Specificity Tyrosine-Phosphorylation-Regulated Kinase-1A Enzyme, Int. J. Mol. Sci. 16 (3) (2015) 5235–5253. doi:10.3390/ijms16035235.

385. L. Zhao, L. Zhang, M. Lei, 3D-QSAR and docking studies on 2-arylbenzoxazole and linker-Y transthyretin amyloidogenesis inhibitors, Sci. China Chem. 56 (11) (2013) 1550–1563. doi:10.1007/s11426-013-4894-9.
386. L. Aswathy, R.S. Jisha, V.H. Masand, J.M. Gajbhiye, I.G. Shibi, Design of novel amyloid β aggregation inhibitors using QSAR, pharmacophore modeling, molecular docking and ADME prediction, Silico Pharmacol. 6 (1) (2018) 12. doi:10.1007/s40203-018-0049-1.
387. H. Safarizadeh, Z. Garkani-Nejad, Molecular docking, molecular dynamics simulations and QSAR studies on some of 2-arylethenylquinoline derivatives for inhibition of Alzheimer's amyloid-beta aggregation: Insight into mechanism of interactions and parameters for design of new inhibitors, J. Mol. Graph. Model. 87 (2019) 129–143. doi:10.1016/j.jmgm.2018.11.019.
388. C. Xiangji, QSAR and primary docking studies of trans-stilbene (TSB) series of imaging agents for β-amyloid plaques, J. Mol. Struct., 763(1-3) (2006) 83-89. doi:10.1016/j.theochem.2006.01.028.
389. Y. Yang, L. Zhu, X. Chen, and H. Zhang, Binding research on flavones as ligands of β-amyloid aggregates by fluorescence and their 3D-QSAR, docking studies, J. Mol. Graph. Model, 29(4) (2010) 538-545. doi:10.1016/j.jmgm.2010.10.006.
390. M. Najmeh, and D. Ajloo, QSAR, docking, and Molecular dynamic studies on the polyphenolic as inhibitors of β-amyloid aggregation, Med. Chem. Res., 25(10) (2016) 2104-2118. DOI 10.1007/s00044-016-1620-0.
391. L. Sehan, and M. G. Barron, Development of 3D-QSAR model for acetylcholinesterase inhibitors using a combination of fingerprint, molecular docking, and structure-based pharmacophore approaches, Toxicol. Sci., 148(1) (2015) 60-70. doi: 10.1093/toxsci/kfv160.
392. Mannhold, R. (2006). Calculation of lipophilicity: A classification of methods, In Pharmacokinetic profiling in drug research. *Wiley-VCH*, 333-352, Zurich (Switzerland).
393. Randić, M. (2004). Wiener Hosoya Index A Novel Graph Theoretical Molecular Descriptor. *J. Chem. Inf. Comput. Sci*., 44 (2), 373-7. DOI: 10.1021/ci030425f.
394. Roy, J., Ghosh, S., Ojha, P.K. and Roy, K. (2019). Predictive quantitative structure–property relationship (QSPR) modeling for adsorption of organic pollutants by carbon nanotubes (CNTs). *Environ. Sci: Nano*, 6 (1), 224-47. DOI: 10.1039/C8EN01059E.
395. Brahmachari, G., Choo, C., Ambure, P. and Roy, K. (2015). In vitro evaluation and in silico screening of synthetic acetylcholinesterase inhibitors bearing functionalized piperidine pharmacophores. *Bioorg. Med. Chem*., 23 (15), 4567-75. DOI: 10.1016/j.bmc.2015.06.005.
396. Goyal, M., Grover, S., Dhanjal, J.K., Goyal, S., Tyagi, C. and Grover, A. (2014). Molecular modelling studies on flavonoid derivatives as dual site inhibitors of human acetyl cholinesterase using 3D-QSAR, pharmacophore and high throughput screening approaches. *Med. Chem. Res*., 23 (4), 2122-32. DOI: 10.1007/s00044-013-0810-2.
397. Karmakar, A., Ambure, P., Mallick, T., Das, S., Roy, K. and Begum, N.A. (2019). Exploration of synthetic antioxidant flavonoid analogs as acetylcholinesterase inhibitors: an approach towards finding their quantitative structure–activity relationship. *Med. Chem. Res.*, 28 (5), 723-41. DOI: 10.1007/s00044-019-02330-8.
398. B, Wellenzohn; S, Tonmunphean; A, Khalid; MI, Choudhary; BM, Rode (2003). 3D-QSAR Studies on natural acetylcholinesterase inhibitors of Sarcococca saligna by comparative molecular field analysis (CoMFA). *Bioorg. Med. Chem*., 13(24), 4375-80.
399. Sharma, M.C.; Sharma, S.; Sahu, N.K. et al. 'QSAR studies of some substituted imidazolinones angiotensin II receptor antagonists using Partial Least Squares Regression (PLSR) method based feature selection', *J. Saudi Chem. Soc.* 17(2), pp. 219-25.
400. Du, X; Wang, X; Geng, M. (2018) 'Alzheimer's disease hypothesis and related therapies', Transl. Neurodegener. 7(1), pp. 2.

401. Shrivastava, S.K.; Sinha, S.K.; Srivastava, P. et al. (2019) 'Design and development of novel p-aminobenzoic acid derivatives as potential cholinesterase inhibitors for the treatment of Alzheimer's disease', *Bioorg. chem*. 82(1), pp. 11-23.
402. Bukhari, S.N.; Jantan, I.; Masand, V.H. et al (2014) 'Synthesis of α, β-unsaturated carbonyl based compounds as acetylcholinesterase and butyrylcholinesterase inhibitors: characterization, molecular modeling, QSAR studies and effect against amyloid β-induced cytotoxicity', *Eur. J. med. chem*. 83(18), pp. 355-65.

403. A. Pirovano, S. Brandmaier, M.A. Huijbregts, A.M. Ragas, K. Veltman, A.J. Hendriks, The utilisation of structural descriptors to predict metabolic constants of xenobiotics in mammals, Environ. Toxicol. Pharmacol. 39(1) (2015) 247-258, https://doi.org/10.1016/j.etap.2014.11.025
404. J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, QSAR applicability domain estimation by projection of the training set in descriptor space: a review, ATLA 33(5) (2015) 445-459, https://doi.org/10.1177/026119290503300508
405. S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58(2) (2001) 109-130, https://doi.org/10.1016/S0169-7439(01)00155-1

APPENDIX

REPRINTS

FINAL

Send Orders for Reprints to reprints@benthamscience.net

Current Topics in Medicinal Chemistry, 2020, *20*, 000-000

RESEARCH ARTICLE

A Multi-layered Variable Selection Strategy for QSAR Modeling of Butyrylcholinesterase Inhibitors

Vinay Kumar[1], Priyanka De[1], Probir Kumar Ojha[1], Achintya Saha[2] and Kunal Roy[1,*]

[1]*Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India;* [2]*Department of Chemical Technology, University of Calcutta, 92 A P C Road, Kolkata 700 032, India*

ARTICLE HISTORY

Received: April 14, 2019
Revised: July 23, 2019
Accepted: October 28, 2019

DOI:
10.2174/1568026620666200616142753

Abstract: ***Background*:** Alzheimer's disease (AD), a neurological disorder, is the most common cause of senile dementia. Butyrylcholinesterase (BuChE) enzyme plays a vital role in regulating the brain acetylcholine (ACh) neurotransmitter, but in the case of Alzheimer's disease (AD), BuChE activity gradually increases in patients with a decrease in the acetylcholine (ACh) concentration *via* hydrolysis. ACh plays an essential role in regulating learning and memory as the cortex originates from the basal forebrain, and thus, is involved in memory consolidation in these sites.

***Methods*:** In this work, we have developed a partial least squares (PLS)-regression based two dimensional quantitative structure-activity relationship (2D-QSAR) model using 1130 diverse chemical classes of compounds with defined activity against the BuChE enzyme. Keeping in mind the strict Organization for Economic Co-operation and Development (OECD) guidelines, we have tried to select significant descriptors from the large initial pool of descriptors using multi-layered variable selection strategy using stepwise regression followed by genetic algorithm (GA) followed by again stepwise regression technique and at the end best subset selection prior to development of final model thus reducing noise in the input. Partial least squares (PLS) regression technique was employed for the development of the final model while model validation was performed using various stringent validation criteria.

***Results*:** The results obtained from the QSAR model suggested that the quality of the model is acceptable in terms of both internal (R^2= 0.664, Q^2= 0.650) and external (R^2_{Pred}= 0.657) validation parameters. The QSAR studies were analyzed, and the structural features (hydrophobic, ring aromatic and hydrogen bond acceptor/donor) responsible for enhancement of the activity were identified. The developed model further suggests that the presence of hydrophobic features like long carbon chain would increase the BuChE inhibitory activity and presence of amino group and hydrazine fragment promoting the hydrogen bond interactions would be important for increasing the inhibitory activity against BuChE enzyme.

***Conclusion*:** Furthermore, molecular docking studies have been carried out to understand the molecular interactions between the ligand and receptor, and the results are then correlated with the structural features obtained from the QSAR models. The information obtained from the QSAR models are well corroborated with the results of the docking study.

Keywords: QSAR, Butyrylcholinesterase inhibitors, Multi-layered variable selection, Validation, Alzheimer's disease.

1. INTRODUCTION

Alzheimer's disease (AD) is an age-related neurodegenerative disorder, which is the most common cause of dementia in elderly individuals. It is characterized by selective neuronal cell death that affects the brain area related to memory and learning [1, 2]. This disease was first described by the German psychiatrist and neuropathologist Dr. Alois Alzheimer, in 1906. According to the World Alzheimer's report 2010, there were 35.6 million people living with dementia worldwide, and it may increase to 65.7 million by 2050 [3]. According to the World Alzheimer's report 2015, 9.9 million new cases were reported and 46.8 million people were living with dementia worldwide, and this number is expected to almost double every 20 years [3]. The primary signs of the disease include that a patient may be unable to remember recent events or conversations, and as the disease grows, a person affected with AD will develop severe memory deterioration and lose the skill to carry out everyday activity. Pathologically, AD is often cerebral proteopathy [4, 5]. Patients become symptomatic before 50 years of age, but the prevalence of the disease rises with age, and this roughly

*Address correspondence to this author at the Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India; Tel: +91 98315 94140;
Fax: +91-33-2837-1078; E-mail: kunal.roy@jadavpuruniversity.in;
URL: http://sites.google.com/site/kunalroyindia/

doubles every 5 years. The disease grows from a level of 1% for the 60-64 years old population and attains 40% or more for the 85-89 years old cohort. The increase in the prevalence of the disease with age has given upsurge to major medical, social and economic problems in countries with a growing number of elderly individuals [6]. AD was first reported a century ago, but the research on its pathogenesis and treatment has started only in the last few decades. In recent years, a significant research has been devoted to develop drugs that slow down neurodegeneration, and this revealed a number of biological targets such as N-methyl-d-aspartate (NMDA) receptor, glycogen synthase kinase 3β (GSK3β), AChE enzyme, BuChE enzyme, cyclin-dependent kinase 5 (CDK5), β-secretase and γ-secretase *etc.* However, we are still far from finding a precise treatment strategy for AD [3]. The well-known "cholinergic hypothesis" relates the neuronal degeneration with the loss of cholinergic neurotransmission. This is the oldest hypothesis of AD progression, according to which, a reduced synthesis of neurotransmitters acetylcholine (terminated by acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE)) [7] results in neuroinflammation and large scale aggregation of β-amyloid. Even though the distinctive features of neurodegeneration in Alzheimer's brains are well known, one of the current difficulties is related to the lack of solid evidence regarding the crucial factors that give rise to the pathogenesis of this disease, creating a great challenge for the efficient treatments of AD.

The current treatment strategy for AD patients is the use of AChE enzyme inhibitors, which gives only a symptomatic relief. However, recent studies indicated a long-lasting effect in a certain percentage of patients. In fact, there are accumulating evidences that AChE and BuChE have secondary non-cholinergic functions including the processing and deposition of β-amyloid (Aβ). Aβ is a physiological peptide secreted from neurons under normal conditions both *in vitro* and *in vivo* [8]. Aβ is generated from amyloid precursor protein (APP) by sequential limited proteolysis conducted by β and γ-secretases [8, 9]. BuChE and AChE could play a role in the Aβ metabolism and during an early step in the development of the senile plaque, as revealed by the finding that AChE and BuChE accelerate Aβ deposition. Considering the non-classical BuChE and AChE functions, their relationships with AD hallmarks and the assumed role of peripheral anionic site in all these functions, the dual binding site ChE inhibitors may acquire importance for the AD treatment. On the other hand, the interference of AChE inhibitors with Aβ processing is not a general rule for this class of compounds with the involvement of other features such as chemical structure and/or genetic regulation [10]. The recent development of highly selective BuChE inhibitors (2-Phenylbenzofuran derivatives) [11] will allow to test these new agents in patients with AD in order to find out whether they represent an advantage or not for the treatment of patients with AD as compared with selective (Donepezil) or relatively non-selective (rivastigmine, galantamine) ChE inhibitors presently in use [12]. There are now 5 approved drugs for the treatment of cognitive symptoms of AD, four are acetylcholinesterase enzyme (AChE) inhibitors (Tacrine, Rivastigmine, Galantamine and Donepezil) and one is a non-competitive glutamate (NMDA) receptor antagonist (Memantine). The benefit from their use is symptomatic, and no medicine has been clearly shown to delay or halt the progress of the disease [10]. There is a rising evidence that both AChE and BuChE may be important in the development and progression of AD [13]. Structural features of the both enzymes suggest the differences in their substrate specificity; AChE is highly selective for ACh neurotransmitter hydrolysis, while BuChE is capable of metabolizing several different neuro active peptides. The substrate diversity is identified by the amino acid sequences of the ChE (AChE and BuChE) enzymes that determine the 3D size and shape of their respective receptors. In case of AChE, the available space for ligand binding is limited by the presence of two large amino acids, phenylalanines (Phe295 and Phe297), but in BuChE, these two amino acids are replaced by two smaller amino acids, valine and leucine, creating greater space that permits binding of various larger molecules. The substituted amino acids must affect the size and hydrophobic nature of the active site [13-14].

There are a number of computational studies reported so far targeting AD, but still we are far from finding a precise treatment strategy for AD [10]. Fang *et al.* [15] developed a QSAR model and reported molecular docking, and molecular dynamics using 67 berberine derivatives for the inhibition of butyrylcholinesterase (BuChE) enzyme. Zheng *et al.* [16] reported a non-linear (neural network (NN)) QSAR model, a multi-linear regression (MLR) model and a molecular docking study using a set of 93 small BuChE inhibitors with defined inhibitory activities (pIC_{50} values). The authors also compared the results obtained from the non-linear model with the MLR model and docking studies. The authors concluded that the results obtained from the non-linear model were better than the MLR model. Solomon *et al.* [17] attempted to develop a QSAR model using a series of 88 N-aryl derivatives with defined inhibitory activity towards both acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzyme in Alzheimer's drug discovery. Bitam *et al.* [18] performed QSAR studies and developed MLR and MLP models using 151 tacrine derivatives with defined BuChE enzyme inhibitory activity. It appeared of interest to us to carry out a 2D-QSAR study on the BuChEI using a large dataset (1130 compounds) for their selective inhibitory activity. In the current study, we have employed a multi-layered variable selection strategy prior to the development of the final model to reduce the noise in input. Feature selection was achieved using a multi-layered variable selection strategy, and the final model was developed using Partial Least Squares (PLS) regression in order to obviate the effect of inter-correlation among the descriptors. The variable selection approach helps in reducing the noise in the input [19]. The QSAR model was built to obey the guidelines of the OECD principles [20]. The developed model has been validated by taking into consideration various strict internal and external validation metrics [21]. Furthermore, we have performed molecular docking study [22] with the most active, moderately active and least active compounds from the whole dataset. The development of sophisticated molecular docking methodologies allows an accurate interpretation of the interactions responsible for the biological activity of molecules while the quantification of their structural information through QSAR models can be useful for virtual screening to design new drugs.

2. MATERIALS AND METHODS

2.1. QSAR Methodology

2.1.1. Dataset

We have developed a PLS-regression based QSAR model for BuChEI using diverse classes of compounds (n =1130) collected from the previously published papers [23-76] (see in supplementary material sheet **1**) to identify the structural requirements which are essential for BuChE enzyme inhibitory activity. The experimental activity values of the dataset compounds were expressed as IC_{50} values (nM) and converted to pIC_{50} values for model development purposes. The data taken from above-mentioned sources were checked and filtered by the criteria of a defined endpoint, and same experimental procedures, following the OECD guidelines [20]. We have also checked that the data we had taken were experimented at the same time of exposure. All the structures were drawn using the ChemDraw ultra 12.0 software [77]. Then, each molecular structure was cleaned and hydrogens were added using Marvin view ChemAxon tool [78] and the structures were saved as MDL.mol format.

2.1.2. Descriptor Calculation and Data Pretreatment

Molecular descriptors are mathematical representations of molecular structure information obtained by a well-specified algorithm. The descriptors were calculated using two software tools, namely, Dragon software version 7 [79] and PaDEL-descriptor 2.20 software [80]. In this work, we have calculated only 2D descriptors covering constitutional, ring descriptors, connectivity index, functional group counts, atom centered fragments, 2D atom pairs, atom type E-states, molecular properties (using Dragon software version 7) and extended topochemical atom (ETA) indices (using PaDEL-Descriptor software). We have performed data pretreatment to remove inter-correlated descriptors from dataset using the tools Pretreatment V-WSP version 1.2 (available at http://dtclab.webs.com/software-tools).

2.1.3. Dataset Division

In the present study, our aim was to develop a QSAR model, which is statistically robust and capable of making accurate and reliable predictions. Therefore, the developed QSAR model was validated using new chemical entities, *i.e.,* a test set to check the predictive capacity of the developed models. The whole data set was divided into an internal set (training set) and an external set (test set) using the "Modified *k*-medoid" clustering technique (available at http://dtclab.webs.com/software-tools). The clustering technique categorizes a set of compounds into clusters so that the compounds present in the same cluster are similar to each other. On the other hand, when two compounds belong to two different clusters, they are expected to be dissimilar in nature. The representative compounds within a cluster are called medoids. This technique tends to select *k* from most middle compounds as the initial medoid. Eight clusters were generated for the BuChEI dataset containing 1130 compounds [23-76]. We have selected approximately 25% of compounds from each cluster for the test set and the remaining 75% of compounds were selected for the training set. The training set was used for model development and the test set was used for model validation purposes.

2.1.4. Multi-layered Variable Selection and QSAR Model Development

The selection of important and meaningful descriptors from a large pool is a crucial step in the QSAR model development. As we know the selection of significant descriptors from a large initial pool is important to reduce the noise in the input, thus, we have employed a multi-layered variable selection strategy prior to the development of the final model using stepwise regression (using a suitable stepping criterion, *e.g.,* 'F-for-inclusion 'and 'F-for-exclusion' based on partial F-statistic) followed by genetic algorithm (GA) followed by again stepwise regression and finally, best subset selection. For this purpose, first, we have run stepwise regression using the whole pool of descriptors and kept the model descriptors aside. Next, we have run again stepwise regression using the remaining pool (after removing the descriptors obtained from first run stepwise regression) of descriptors and selected the model descriptors. In this way, we have selected 80 descriptors from the initial pool of 600 descriptors (Layer-I). After the first layer of descriptor selection, we have developed some models using a genetic algorithm (GA) (available at http://dtclab.webs.com/software-tools) and selected 57 descriptors (Layer-II) from 80 descriptors. After that, we have selected 20 descriptors using stepwise regression technique (Layer-III) again. Using these 20 descriptors, we have run the best subset selection using a tool developed in our laboratory (available at http://dtclab.webs.com/ software-tools) to develop a 15 descriptor model which was selected based on Mean Absolute Error (MAE) based criteria [81, 82]. Although many groups of authors reported different variable selection strategies, we have followed here stepwise regression followed by GA followed by stepwise regression and finally the best subset selection method as reported previously also [83-86]. The final model was developed by employing PLS-regression methodology to avoid intercorrelation among the modeled descriptors using Minitab software [87]. Multi-layered variable selection strategy is schematically represented in Fig. (**1**).

2.1.5. Statistical Validation Metrics

We have developed a PLS-regression based QSAR model for the inhibitory activity of BuChE enzyme. The developed model was validated using both internal and external validation parameters. The internal statistical parameters used in this study are the determination coefficient (R^2) and leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$). The determination coefficient (R^2) represents how much variability of a factor can be explained by its relationship to another factor, it is computed as a value between 0 (0 percent) and 1 (100 percent). The higher value of this parameter indicates a better fit of the model. However, these parameters are not good enough to evaluate the robustness and predictivity of a significant model. Thus, we have employed additional statistical validation parameters such as R^2_{Pred} (external prediction variance) or Q^2F_1 and Q^2F_2 to assure the significance of the developed model. Additionally, we have performed Y-randomization test, checking of applicability domain criteria, *etc.* to investigate the robustness of the developed model. The main objective of Y-randomization test is to ascertain whether the developed model is obtained by chance or not. The Y-randomization

test was performed using the Simca-P software [88] through randomly reordering (100 permutations) the dependent variable. The validation parameter of the model obtained under such conditions should be of poor quality. The value of the R^2_{yrand} intercept should not exceed 0.3 and the value of the Q^2_{yrand} intercept should not exceed 0.05 [89].

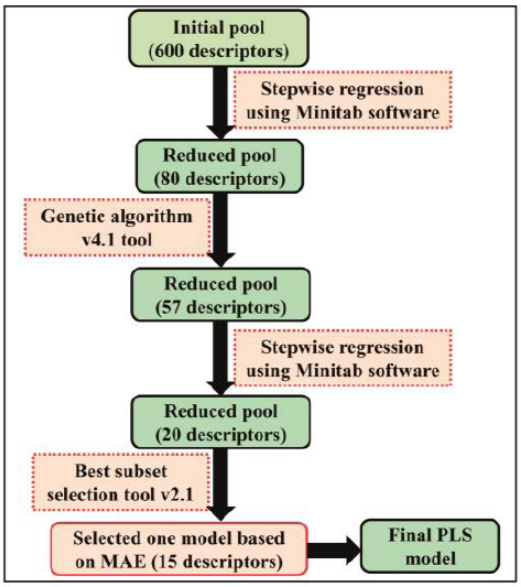


Fig. (1). Schematic representation of multi-layered variable selection strategy. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

2.1.6. Applicability Domain (AD)

The Applicability domain of a QSAR model has been described as the response and chemical structure space that is defined by the nature of the chemicals in the training set. If a new compound falls within the Applicability domain of the developed model, only then the developed model can predict the compound precisely. It is extremely useful for QSAR developers to have information about the applicability domain of the developed model to identify interpolation (true predictions) or extrapolation (less reliable predictions) [90-91]. Here, we have checked the applicability domain of the developed model employing the DModX (distance to model X) approach at 99% confidence level using SIMCA-P software [88].

2.1.7. Randomization of the PLS Model

The purpose of the Y-randomization test is to identify and quantify chance correlations between the dependent variable and the descriptors [89]. Here, the term chance correlation means that the real model may contain descriptors which are statistically well correlated to Y, but in reality there is no cause-effect relationship encoded in the respective correlations with Y because they are not related to the mechanism of action [89]. The Y-randomization test consists of several runs for which the original descriptor matrix X is kept fixed, and only the vector Y is randomized [89]. The validation parameter of the model obtained under such conditions should be of poor quality and without real meaning [89, 90]. The value of the R^2_{yrand} intercept should not exceed 0.3 and the value of the Q^2_{yrand} intercept should not exceed 0.05 [89]. In the present study, for the training set, the X data remained constant and the data Y were shuffled randomly. Here, we have generated the randomized model using 100 permutations. The schematic work flow for the development of QSAR model against BuChE inhibitors is shown in Fig. (**2**).

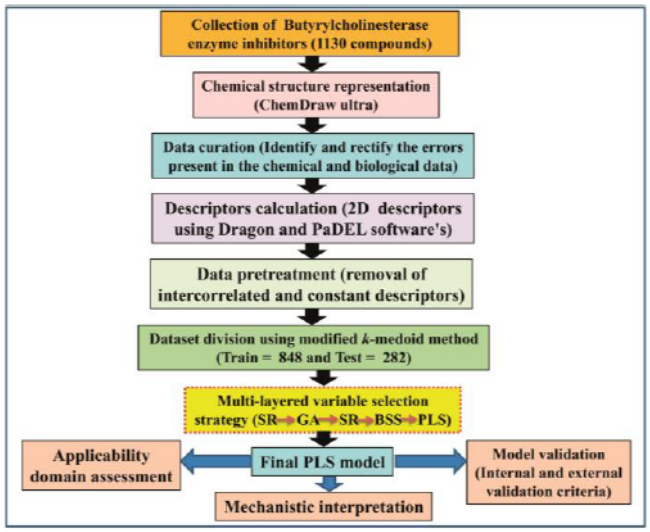


Fig. (2). Schematic work flow of QSAR model development against BuChE inhibitors (BuChEI).[PLS = Partial least squares, SR = Stepwise regression, BSS = Best subset selection]. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

2.2. Docking Studies

Molecular docking is a technique to understand drug-biomolecular interactions for the rational drug design and discovery as well as for the mechanistic study by placing a molecule (ligand) into the desired binding site of the target definite region of the protein/enzyme (receptor) mainly in a non-covalent manner to form a stable complex of potential efficacy and specificity [92]. The docking process includes two basic stages: prediction of the ligand conformation as well as its position and orientation within these sites (usually referred to as pose) and assessment of the binding affinity. The evidence obtained from the docking study can be used to suggest the binding orientation, binding energy, free energy, interaction energy and stability of complexes. In the current study, we have employed molecular docking studies to comprehend the interactions between the BuChE enzyme (the structure of the protein was retrieved from Protein Data Bank with PDB ID: 6EZ2 [93]) and the selected BuChE enzyme inhibitors. In this context, we have applied the CDOCKER module of receptor-ligand interaction available in BIOVIA Discovery Studio client 4.1 [94]. Before the docking experiment, we have defined the active site of the enzyme using the protocol Receptor-Ligand Interaction section using the option “define site from receptor cavities” in the BIOVIA Discovery Studio client 4.1 platform [94]. The selected inhibitors were subjected to ligand preparation to find a series of ligand conformers. Each orientation was used in the CDOCKER module for molecular docking using CHARMm based interaction energy using a rigid receptor [22]. The poses are sorted according to CHARMm interac-

tion energy, and the top scoring (most negative, thus favorable to binding) poses are kept.

3. RESULTS AND DISCUSSION

3.1. Mechanistic Interpretation of Modeled Descriptors

We have developed a PLS model for BuChE inhibitors using only 2D descriptors calculated from different software tools like Dragon and PaDEL-Descriptor software. Prior to the development of the final model, we have used a multi-layered variable selection strategy to reduce the initial number of descriptors in the pool thus reducing the noise in the input. The statistical quality of the model was determined by utilizing different internal and external validation metrics. The statistical quality of the model generated in this study with 2D descriptors for the large modeling set of BuChEI (1130 compounds) [23-76] appears better than those reported in the previous studies (Table **4**) [15-18]. Usually, a direct comparison with the previously reported study is difficult since different models were developed using different combinations of molecules, but it may be noted that our data set was much larger than the datasets dealt with previously [15-18]. The reported PLS model was developed by using 15 descriptors with corresponding latent variables of 6. Here, $n_{training}$ and n_{test} are the number of compounds present in the training and test sets. The R^2 (0.664), Q^2 (0.650) and R^2_{pred} values (0.657) of the PLS model were higher than 0.6 (Equation **1**), which indicated the acceptability and predictive ability of models. Thus, the results obtained from the PLS model (See in Table **1**) suggested that the model is acceptable in terms of fitness, stability and classical predictivity measures. The descriptors appearing in the model define the structural and functional requirements, which can improve the inhibitory activity of molecules against BuChE enzyme. The proximity of the observed and predicted values for the BuChE enzyme inhibitors in the data set can be further established from the scatter plot as shown in Fig. (**3**). The quantitative contributions of similar/dissimilar descriptors are given in loading plot (similar descriptors are placed in close proximity), the interrelationships between the X-variables and the Y-response are depicted in the loading plot **(**Fig. **4**). Additionally, we have also performed Y-Randomization test to check whether the model was obtained by any chance or not. The results (R^2_{rand} = -0.000805 and Q^2_{rand} = -0.147) obtained from the randomized model suggested that the developed model was not obtained by any chance correlation (Fig. **5**).

The descriptors in the PLS model are arranged according to their importance, and then described separately. The significance level of the modeled descriptors towards the BuChE inhibitory activity is computed based on the variable importance plot (VIP) (Fig. **6**) [95]. The VIP defines the importance of each variable obtained from the final PLS model that are responsible for regulating the BuChE inhibitory activity. As per the VIP plot, the significance level of the modeled descriptors are established to be in the following manner: RDCHI, F04[C-O], nPyridines (N..O), F06[C-O], C-041, B02[N-O], F03[C-C], nCrs, nN-N, B01[N-N], H-048, NRS, C-031 and nArCONHR. Among these descriptors, RDCHI, C-041, nN-N, T(N..O), nPyridines, nArCONHR, nCrs, H-048 and B02[N-O] contributed positively but C-031, F03[C-C], F06[C-O], B01[N-N], F04[C-O] and NRS descriptors contributed negatively towards the BuChE inhibitory activity as shown in (Fig. **6**) in the regression coefficient plot.

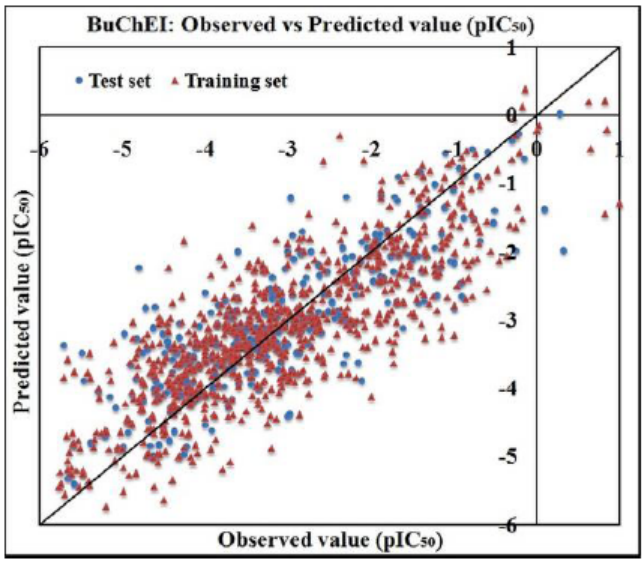


Fig. (3). The scatter plot of observed and predicted values of final PLS model against BuChE enzyme. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

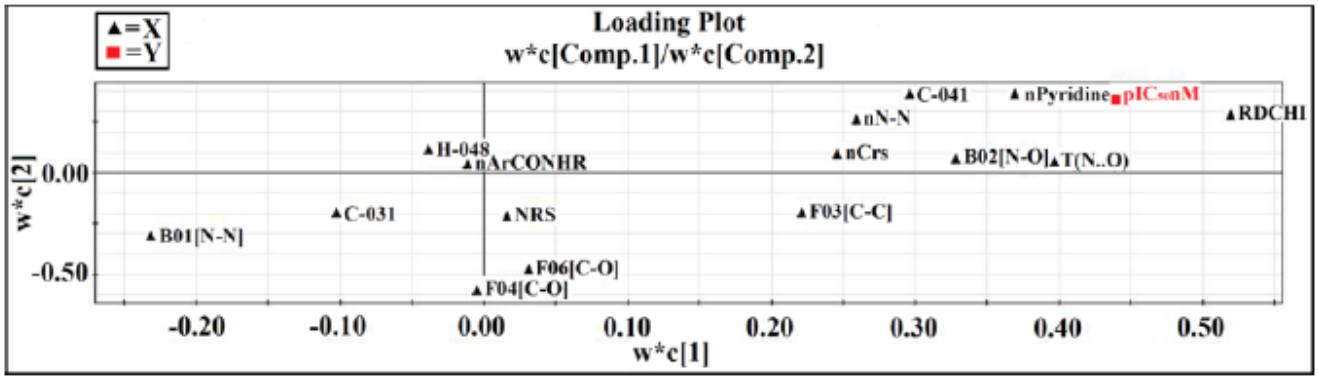


Fig. (4). Loading plot for final PLS model against BuChE enzyme. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

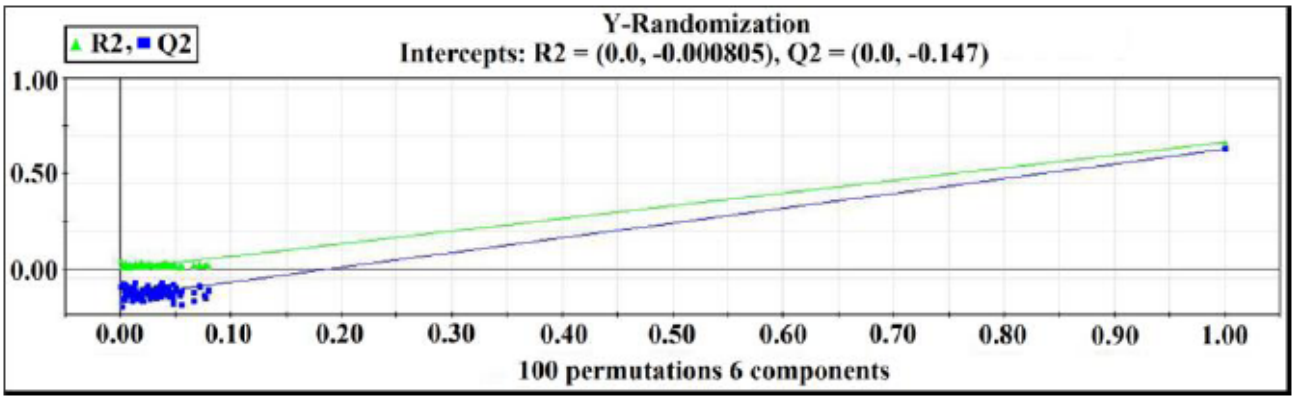


Fig. (5). Model Randomization plot for final PLS model against BuChE enzyme. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

The developed PLS regression model is shown below:

$$pIC50(nM) = -6.41 + 1.15 \times RDCHI - 0.08 \times F04[C-O] + 0.46 \times nPyridines + 0.004 \times T(N..O) - 0.07 \times F06[C-O] + 1.38 \times C-041 + 0.12 \times B02[N-O] - 0.015 \times F03[C-C] + 0.06 \times nCrs + 3.10 \times nN-N - 1.34 \times B01[N-N] + 0.18 \times H-048 - 0.43 \times NRS - 1.33 \times C-031 + 1.05 \times nArCONHR \quad (1)$$

$n_{training}$= 848, R^2= 0.664, $R^2_{adj=0.662}$, Q^2_{LOO}= 0.650, LV=6, n_{test}= 282, Q^2_{F1}= 0.657, Q^2_{F2}= 0.657

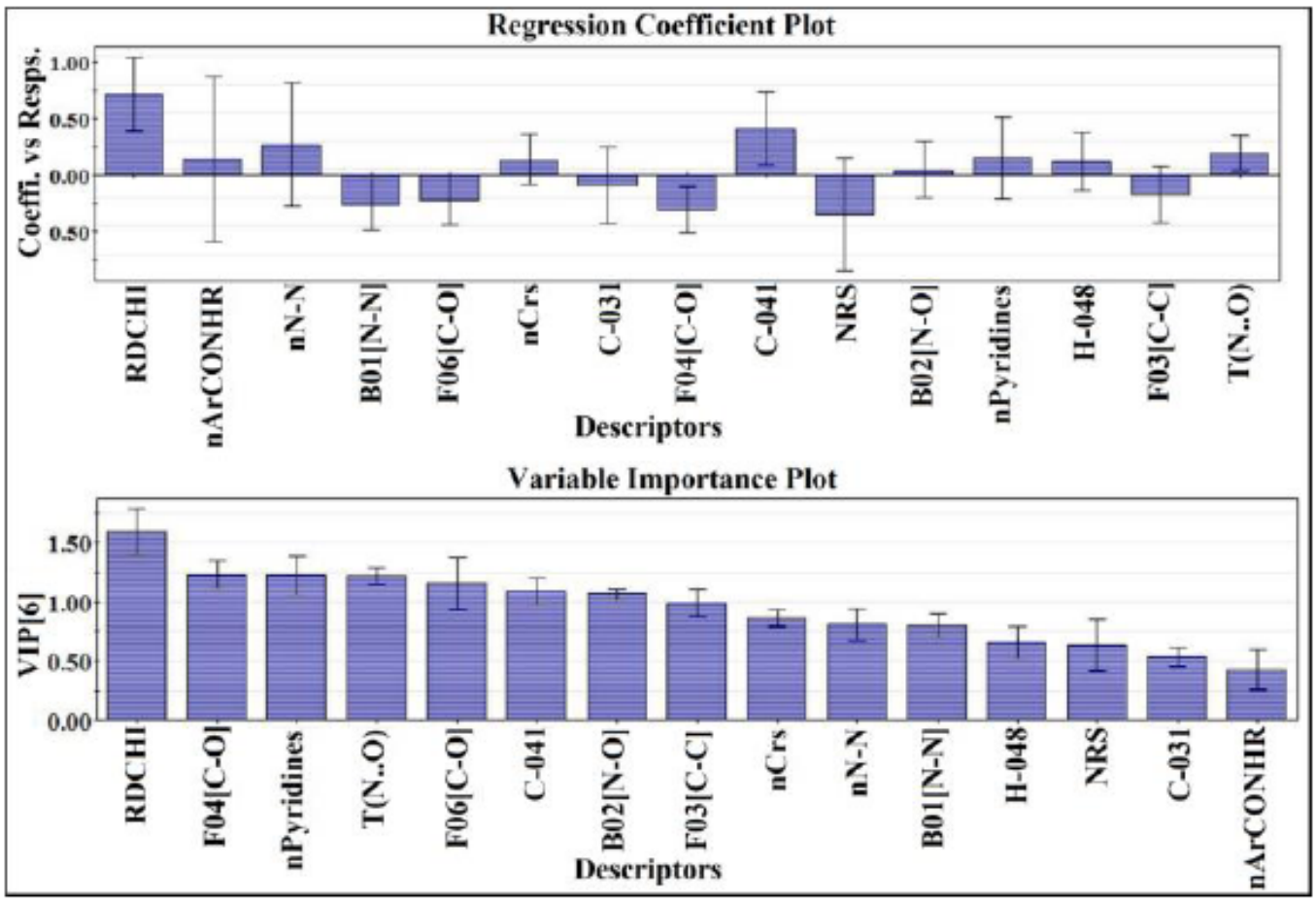


Fig. (6). Regression coefficient plot and variable importance plot (VIP) of the final PLS model against BuChE enzyme. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

Table 1. Statistical validation parameters obtained from the developed PLS model.

Dataset		Equation Length	LV	Training set statistics		Test set statistics	
Training set	Test set			R^2	$Q^2_{(LOO)}$	R^2_{Pred}	Q^2F_2
848	282	15	6	0.664	0.650	0.657	0.657

Table 2. Contribution, definition and mechanism of all the descriptors obtained from the PLS model.

S. No.	Name of Descriptors	Contribution	Family/Short description	Mechanism
1	RDCHI	+ve	Connectivity indices: Reciprocal distance sum Randic-like index	Hydrophobic interactions
2	nArCONHR	+ve	Functional group counts: Number of secondary amides (aromatic)	Hydrogen bonding and electrostatic interactions
3	nN-N	+ve	Functional group counts: Number of N hydrazines	Hydrogen bonding interaction
4	B01[N-N]	-ve	2D Atom Pairs: Presence/absence of N - N at topological distance 1	Hydrogen bonding interaction
5	F06[C-O]	-ve	2D Atom Pairs:- Frequency of C - O at topological distance 6	Hydrogen bonding interaction
6	nCrs	+ve	Functional group counts:- Number of ring secondary C(sp3)	π-π interactions
7	C-031	-ve	Atom-centred fragments:- X--CR--X	Electrostatic interactions
8	F04[C-O]	-ve	2D Atom Pairs:- Frequency of C - O at topological distance 4	Hydrogen bonding interaction
9	C-041	+ve	Atom-centred fragments:- X-C(=X)-X	Electrostatic interactions
10	NRS	-ve	Ring descriptors:-Number of ring systems	Hydrophobic interactions
11	B02[N-O]	+ve	2D Atom Pairs:- Presence/absence of N - O at topological distance 2	Hydrogen bonding and electrostatic interactions
12	nPyridines	+ve	Functional group counts:- Number of Pyridines	Hydrogen bonding and electrostatic interactions
13	H-048	+ve	Atom-centred fragments:- H attached to C2(sp3)/C1(sp2)/C0(sp)	Electrostatic interactions
14	F03[C-C]	-ve	2D Atom Pairs:- Frequency of C - C at topological distance 3	Hydrophobic interactions
15	T(N..O)	+ve	2D Atom Pairs:- Sum of topological distances between N..O	Electrostatics and π -π Interaction

The descriptor, RDCHI, simply characterizes the size and branching of molecules. Its value increases with molecular size but decreases with molecular branching. It can be calculated through Randic-like formula as shown below.

$$RDCHIindex = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} aij.(RDSi.RDSj) \quad (2)$$

Here, A is the number of vertices and a_{ij} is equal to 1 only for pairs of adjacent vertices and zero otherwise, RDS is a sum of reciprocal distance [96]. The descriptor contributes positively towards the BuChE inhibitory activity as suggested by the positive regression coefficient, as shown in (Fig. **7**) in compounds **264** (pIC_{50}: 1), **266** (pIC_{50}: 0.823) and **690** (pIC_{50}: 0.629) (containing descriptor values 5.482, 5.277 and 5.246 respectively)**.** These compounds have a large molecular size and less molecular branching. Conversely, compounds **292** (pIC_{50}: -5.63), **293** (pIC_{50}: -4.62) and **835** (pIC_{50}: -5.13) (containing descriptor values 2.351, 2.608 and 2.727 respectively) are low active and have higher branching **(**Fig. **7**).

The 2D atom pair descriptor, F04[C-O], denotes the frequency of C-O at the topological distance 4. The BuChE enzyme inhibitory activity is inversely correlated to the numerical value of this descriptor as indicated by its negative regression coefficient. The frequency of the C-O fragment at the topological distance 4 may reduce the inhibitory activity of BuChE enzyme. The higher number of C-O fragments

correlate to lower inhibitory activity as observed in (Fig. **7**) compounds **883** (pIC_{50}: -4.41) and **642** (pIC_{50}: -5.53) (containing descriptor values 8 and 15 respectively), while a lower numerical value of this descriptor correlates to higher inhibitory activity as observed in (Fig. **7**) compounds **757** (pIC_{50}: -0.819) and **752** (pIC_{50}: -0.991).

The functional group count descriptor, nPyridines, describes the number of Pyridine rings present in the compounds. The positive regression coefficient of this descriptor indicates that the presence of Pyridine rings in the compounds may enhance the inhibitory activity against BuChE enzyme as found in (Fig. **8**) compounds **690** (pIC_{50}: 0.621), **688** (pIC_{50}: 0.017), **697** (pIC_{50}: 0.85) and **695** (pIC_{50}: 0.645) (containing 1 pyridine ring in each cases) and vice versa in case of compounds **190** (pIC_{50}: -5.74), **204** (pIC_{50}: -5.02) and **849** (pIC_{50}: -4.91) (containing no such fragments) as given in Fig. (**8**). The pyridine ring with free electron-pairs and having no active atoms are "passive hydrophilic" moiety and are capable of forming "hydrogen-bridges" with other polar molecules. The structural polarity of pyridine makes it hydrophilic in nature [97].

Another 2D atom pair descriptor, T(N..O), stands for the sum of topological distances between N..O. This descriptor contributes positively towards the BuChE inhibitory activity as indicated by the positive regression coefficient. Thus, the molecules bearing higher topological distance between **N..O** fragment may have higher BuChE inhibitory activity as shown in (Fig. **8**) compounds **264** (pIC_{50}: 1) and **690** (pIC_{50}: 0.621) (containing descriptor values 202 and 237 respectively) whereas in contrary, compounds **190** (pIC_{50}: -5.74)

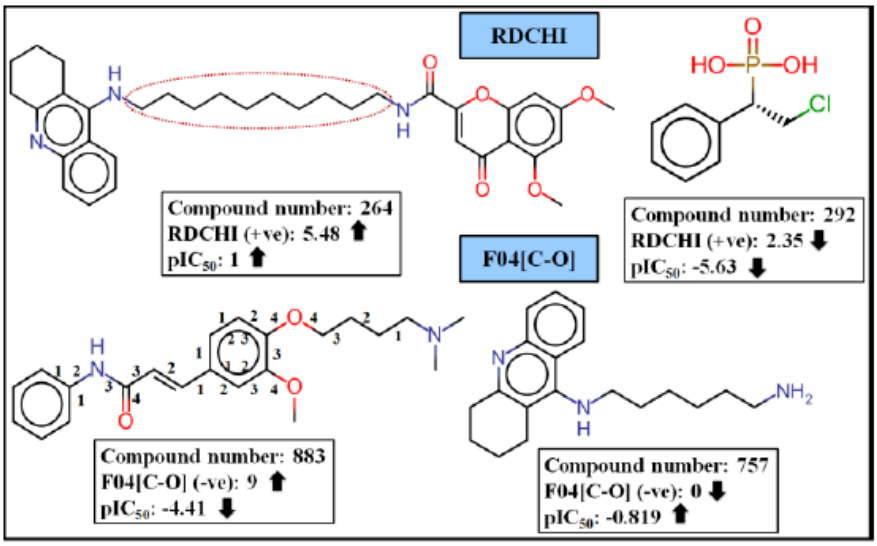


Fig. (7). Contribution of RDCHI and F04[C-O] descriptors on BuChE enzyme inhibition. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

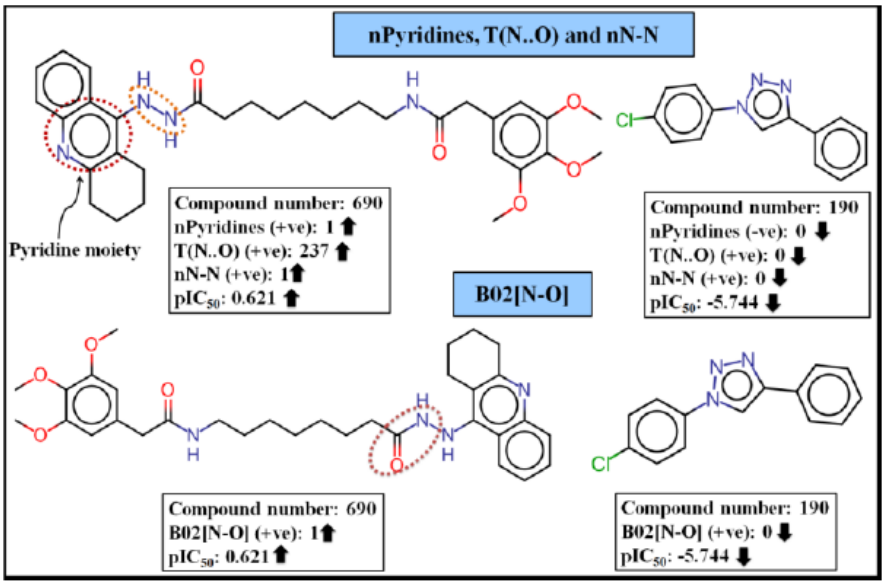


Fig. (8). Contribution of nPyridines, T(N..O), nN-N and B02[N-O] descriptors on BuChE enzyme inhibition. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

and **841** (pIC_{50}: -2.23) which do not contain any such fragment shows less BuChE inhibitory activity **(**Fig. **8**). From this observation, it can be concluded that the topological distances between nitrogen and oxygen atoms should be higher for better inhibitory activity against BuChE.

Another 2D atom pair descriptor, F06[C-O], indicates the frequency of C-O fragment at the topological distance 6. The negative regression coefficient of this descriptor suggests that the descriptor is inversely proportional to the BuChE inhibitory activity as observed in the case of compounds **757** (pIC_{50}: -0.819) and **758** (pIC_{50}: -0.44) (having higher enzyme inhibitory activity as the corresponding numerical descriptor value is in the lower range) whereas the reverse is observed in case of compounds **880** and **642** having lower enzyme inhibitory activity (pIC_{50}= -4.22 and -5.53 respectively) **(**Fig. **9**).

The atom centered fragments descriptor**,** C-041, represents the number of fragments containing C(sp2) atoms that are attached with two electronegative atoms (O, N, S, Se and halogens), *i.e.,* one by a single bond and another by a double bond. The positive regression coefficient suggests the influential effect of the feature containing C(sp2) atoms directly attached with two electronegative atoms towards BuChE inhibitory activity. This is observed in case of compound **269** (pIC_{50}: 0.823) (descriptor value 1), and the opposite is seen in compound **84** (pIC_{50}: -4.39) as depicted in Fig. (**9**). Thus, this descriptor provides us with an assumption that these fragments might involve in polar interaction with the binding pocket amino acid residues. This electrostatic bond may also help in stabilizing the π-π stacking binding and increase the affinity between BuChE enzyme and its inhibitors [98].

Another functional group count descriptor, B02[N-O], stands for the presence/absence of the N-O fragment at the topological distance 2. The positive regression coefficient of this descriptor indicates that the presence of the N-O fragment at topological distance 2 may favor the inhibitory activity of inhibitors against the BuChE enzyme as found in (Fig. **8**) compounds **690** (pIC_{50}: 0.621), **688** (pIC_{50}: 0.017), **697** (pIC_{50}: 0.850) and **695** (pIC_{50}: 0.645) (containing descriptor value of 1 for all the cases). On the other hand, compounds with a lower numerical value of this descriptor show lower inhibitory activity as observed in Fig. (**8**) compounds **190** (pIC_{50}: -5.74), **204** (pIC_{50}: -5.02) and **849** (pIC_{50}: -4.91). In the latter case, the compounds do not have such N-O fragments at the topological distance 2. Thus, N-O fragments at topological distance 2 are influential for the BuChE inhibitory activity.

Another 2D atom pair descriptor, F03[C-C], indicates the frequency of C-C fragment at the topological distance 3. The negative regression coefficient of this descriptor suggests that the presence of C-C fragment at the topological distance 3 inversely affects the BuChE inhibitory activity. This is observed in compounds **996** (pIC_{50}: -4.25) and **987** (pIC_{50}: -5.37) (containing higher descriptor values 90 and 81 respectively) (Fig. **10**). The opposite is observed in compounds **675** (pIC_{50}: -1.83) and **430** (pIC_{50}: -1.91) (containing descriptor values 9) show higher enzyme inhibitory activity due to lower numerical values of this descriptor.

The functional group count descriptor, nCrs, represents a number of ring secondary C (sp3) atoms present in the compounds. This descriptor positivity influences the activity of BuChE inhibitors as indicated by its positive regression coefficient. Thus, the compounds containing a higher number of ring secondary C (sp3) atoms may have high inhibitory activity against BuChE, as shown in (Fig. **10**) compounds **385** (pIC_{50}: -0.477), **386** (pIC_{50}: 0.9085) and **269** (pIC_{50}: 0.823) (containing 10, 10 and 9 ring secondary C (sp3) atoms respectively), whereas the compounds containing no such ring secondary C (sp3) atom have low inhibitory activity against BuChE as shown in (Fig. **10**) compounds **1041** (pIC_{50}: -5.33) and **1098** (pIC_{50}: -4.60). From this descriptor, it can be in-

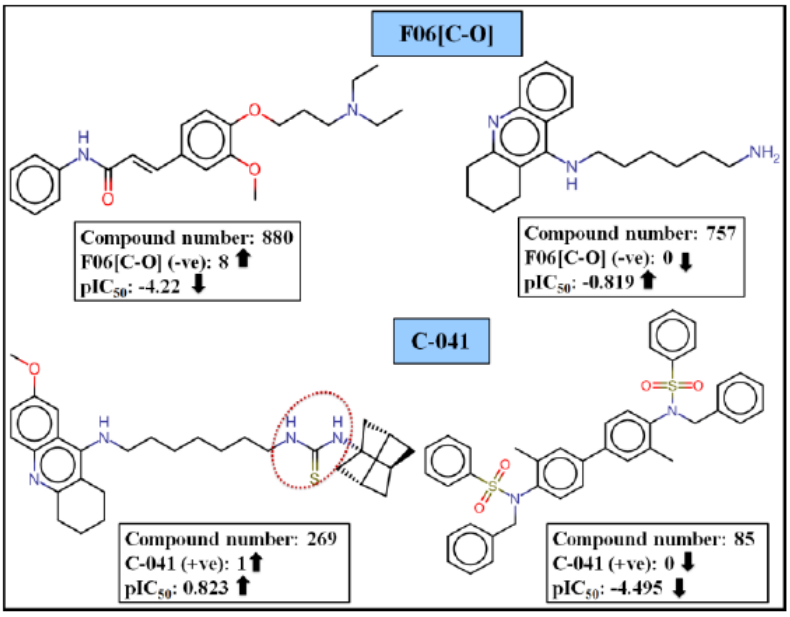


Fig. (9). Contribution of F06[C-O] and C-041 descriptors on BuChE enzyme inhibition. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

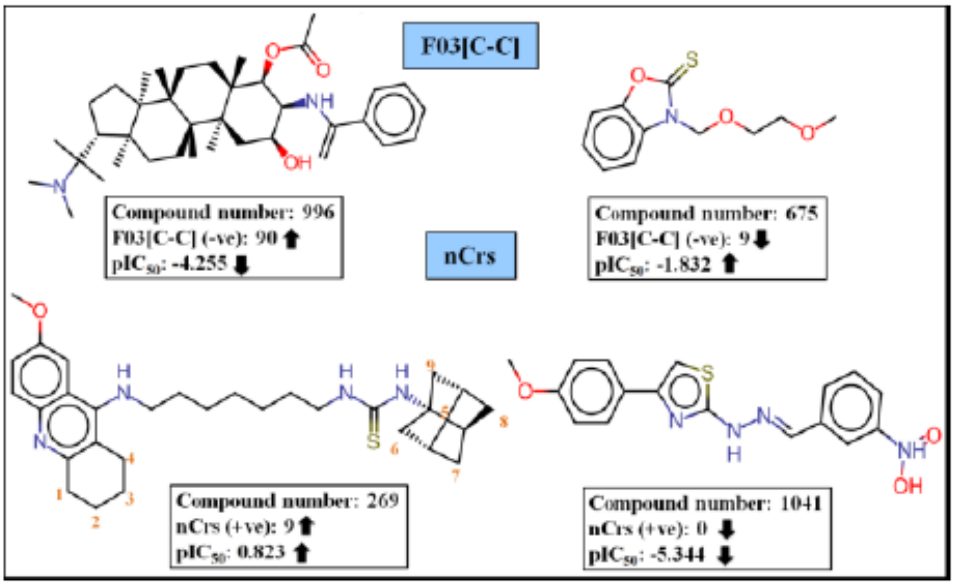


Fig. (10). Contribution of F03[C-C] and nCrs descriptors on BuChE enzyme inhibition. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

ferred that the cyclic ring containing carbon atom without any unsaturation may favor the BuChE inhibitory activity.

Another functional group count descriptor, nN-N, denotes the presence of a number of hydrazines moiety in the compounds. The positive regression coefficient of this descriptor indicates that the activity of inhibitors is directly proportional with the numerical value of the nN-N descriptor. Thus, the compounds having higher number of hydrazine moiety may have higher BuChE enzyme inhibitory activity as shown in (Fig. **8b** compounds **690** (pIC_{50}: 0.621), **688** (pIC_{50}: 0.017), **697** (pIC_{50}: 0.85) and **695** (pIC_{50}: 0.645)(containing descriptor values 1), whereas the compounds such as **190** (pIC_{50}: -5.74), **204** (pIC_{50}: -5.02) and **849** (pIC_{50}: -4.91) have less BuChE enzyme inhibitory activity due to the absence of such fragment as shown in Fig. (**8**). The hydrazine fragment may be involved in hydrogen bonding interactions with the surrounding amino acid residues in the binding pocket of BuChE enzyme. We have observed from docking studies (discussed later) that N-N fragments in the molecules form hydrogen bonds along with electrostatic interaction with their surrounding amino acid residues (Figs. **17** and **18**).

The 2D atom pair descriptors, B01[N-N], describes the presence/absence of N-N at topological distance 1. The negative regression coefficient of this descriptor indicates that compounds containing lower number of such fragments have good inhibitory activity against the BuChE enzyme as shown in (Fig. **11**) compounds **264** (pIC_{50}: 1) and **266** (pIC_{50}: 0.823)(containing descriptor value of 1), while a higher number of this fragment shows lower inhibitory activity as observed in (Fig. **11**) compounds **187** (pIC_{50}: -5.52) and **175** (pIC_{50}: -5.39).

We have found that descriptors nN-N and B01[N-N] are showing opposite effects on the inhibitory activity against BuChE. The descriptor nN-N representing the presence of a number of hydrazine moiety in the compounds has a positive regression coefficient. In a hydrazine group, two adjacent nitrogen atoms are attached with a single bond; there is no presence of unsaturation. On the other hand, the descriptor B01[N-N] denotes the presence/absence of N-N at topological distance 1 and has a negative regression coefficient. Here, the descriptor signifies the presence or absence of N-N at topological distance 1 without taking unsaturation effect into consideration (*i.e.,* unsaturation might be present or absent). In compounds **187** and **175**, there is an absence of hydrazine group (nN-N descriptor) but the presence of unsaturation between two adjacent nitrogen atoms, which can be explained by B01[N-N] but not by the N-N descriptor.

The atom centered fragment, H-048, denotes the number of H attached to C2(sp3)/C1(sp2)/C0(sp). This descriptor is defined as the number of specific atom type in a molecule and can be calculated by knowing the only molecular composition and atom connectivity. The count of hydrogen atoms of type H-048 discloses the importance of hydrogen bond interaction. The positive regression coefficient of this descriptor indicates that compounds containing a higher number of such hydrogen atoms have good inhibitory activity against BuChE enzyme as shown in (Fig. **11**) compounds **795** (pIC_{50}: -0.77) and **461** (pIC_{50}:-0.699) (containing descriptor values 1) while the compounds **927** (pIC_{50}: -5.29) and **243** (pIC_{50}:-5.70) show lower inhibitory activity due to the absence of such H atom (Fig. **11**).

Another functional group count descriptor, NRS, indicates the number of ring systems present in the compounds, which contributes negatively towards the BuChE enzyme inhibitory activity. As observed in the docking study (discussed later), the results properly corroborate with this observation. Hydrophobicity plays an important role in better BuChE inhibitory activity as we have observed in compounds such as **264** (pIC_{50}: 1) and **697** (pIC_{50}: 0.850) containing descriptor value 2 showing higher inhibitory activity, but according to regression coefficient plot, this descriptor contributes negatively. Thus, we have concluded that BuChE

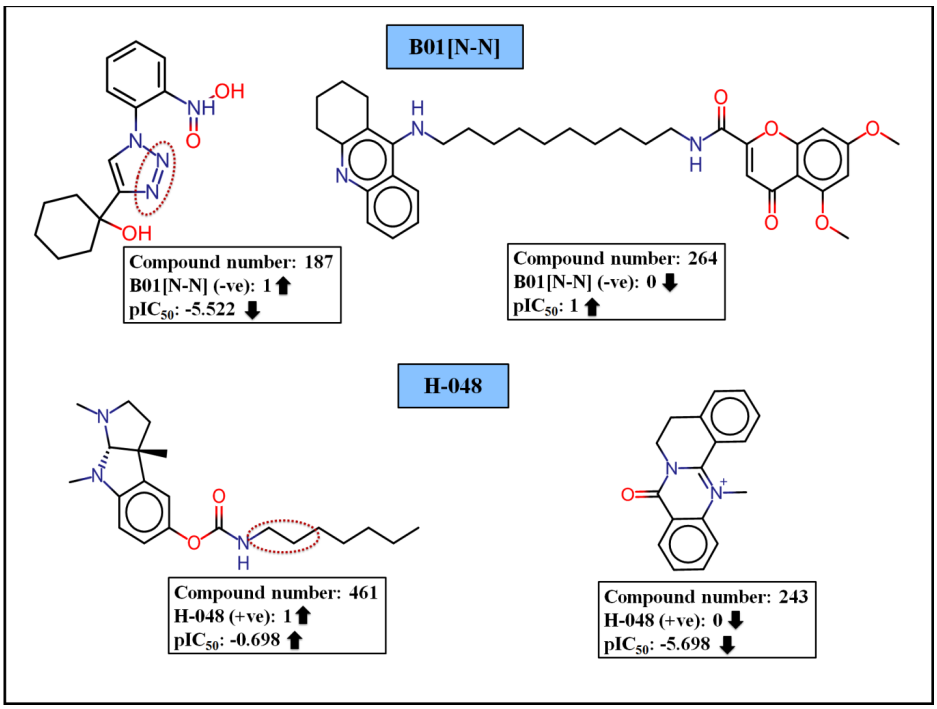


Fig. (11). Contribution of B01[N-N] and H-048 descriptors on BuChE enzyme inhibition. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

inhibitory activity decreases with increasing the descriptor value which could be due to some other features present in the molecules as shown in (Fig. **12**) Compounds **996** (pIC_{50}: -4.25) and **987** (pIC_{50}: -5.37) (containing descriptor values 2 and 1 respectively) and vice versa in case of compounds **675** (pIC_{50}: -1.83) and **430** (pIC_{50}: -1.91) **(**Fig. **12**).

The atom-centred fragment descriptor, C-031, simply refers to X-CR-X, where, R represents any group linked through carbon atom; X represents any heteroatom (O, N, S, P, Se, and halogens). This descriptor contributes negatively towards the BuChE enzyme as indicated by the negative regression coefficient. For example, compounds **306** (pIC_{50}: -4) and **303** (pIC_{50}:-4.22)(containing descriptor values 1) have lower BuChE inhibitory activity. On the contrary, the molecules which do not contain such feature have higher inhibitory activity as shown in compounds **264** (pIC_{50}: 1) and **697** (pIC_{50}: 0.850) as mentioned in Fig. (**12**).

The functional group count descriptor, nArCONHR, represents the presence of a number of secondary amides (aromatic) in the compounds. There are only 32 compounds **(261, 270, 272, 1061, 1068, 227, 229 and 231, etc)** out of 1130 molecules in the whole data set, which contain such fragments, and the frequency of this fragment in compounds is 1 (Fig. **12**). It may be assumed that this fragment contributes significantly to increase the intermolecular interactions by forming strong H-bonds. This descriptor contributed positively towards the BuChE inhibitory activity as indicated by the positive regression coefficient. Thus, the molecules bearing this fragment may enhance the BuChE inhibitory activity as shown in (Fig. **12**) compounds **261** (pIC_{50}: -1.04) and **272** (pIC_{50}: -1.25) (containing descriptor values 1). On the other hand, the compounds containing no such fragments have lower inhibitory activity as shown in compounds **179** (pIC_{50}: -5.61) and **175** (pIC_{50}: -5.39). It may be mentioned here that the entire dataset may not follow the exact pattern of correlation for all data points with respect to a single descriptor, since it is obvious that the property of any molecules is a function of multiple features. Here, we have given appropriate representative examples to understand the role of different features and descriptors in controlling the response values.

3.2. Applicability Domain of PLS Model

The applicability domain of a QSAR model is the structural, biological, physico-chemical information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds [90]. The applicability domain of a QSAR model should be described in terms of the most significant parameters that appeared in the developed model. Ideally, the QSAR should only be used to make predictions within that domain by interpolation, not extrapolation. The proposed PLS model was checked using the applicability domain at confidence level 99% according to the DModX (distance to model in the X-space) approach using SIMCA-P 10.0 software [88]. In case of the proposed model, (Fig. **13**) we found that 59 compounds (i.e. compounds number **99, 189, 221, 222, 225-231, 233, 261, 270, 272, 300-306, 512, 683, 684, 686, 688, 690, 691, 692, 693, 695, 696, 697, 778, 960, 975-981, 984, 988, 990, 995, 996, 999, 1003, 1006, 1011, 1057, 1059, 1060, 1061, 1062, 1065, 1066, 1068** and **1069**) in the training set are located outside the critical DModX value (D-Crit=1.62) and in case of the test set, 16 compounds (*i.e.,* compounds number **223, 228, 301, 685, 687, 689, 694, 959, 989, 993, 997, 1002, 1058, 1063, 1064** and **1067**) are located outside the critical DModX value (D-Crit=1.62).

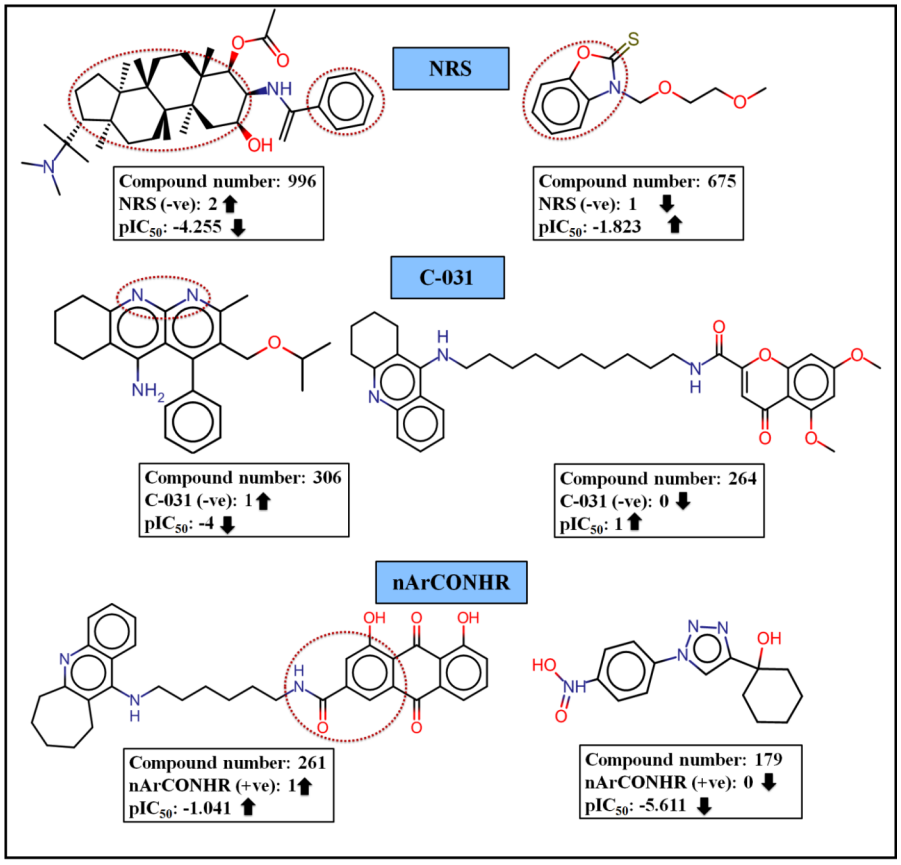


Fig. (12). Contribution of NRS, C-031 and nArCONHR descriptors on BuChE enzyme inhibition. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

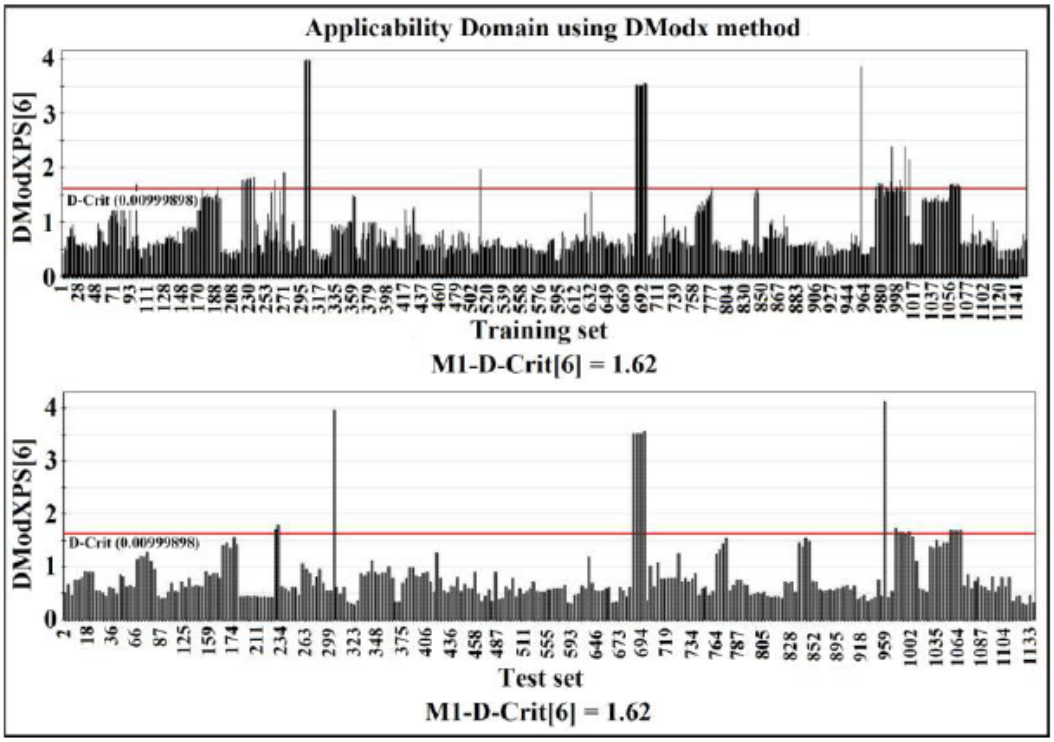


Fig. (13). Applicability domain DModX values of the training and test set compounds at 99% confidence level of the developed PLS model against BuChE enzyme. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

3.3. Molecular Docking

3.3.1. Molecular Docking for the Most Active Compounds from the Dataset

In case of compounds **264**, **688**, **697**, **695** and **690** having higher activity values (pIC_{50} = 1, 0.01, 0.850, 0.645 and 0.621 respectively), the interaction forces include hydrogen bond interactions (carbon hydrogen bonds and conventional hydrogen bonds), π-interactions (π-sigma, π-anion, π-cation, π-alkyl bonds and alkyl hydrophobic) and others attractive forces. The amino acid residues involved in the interaction are ALA A:101, ASP A: 3, ASN A:17, ARG A:14, ILE A:99, ILE A:5, ILE A:4, ASP A:3, THR A:26, MET A:16 and LYS A:103 (Figs. **14, 15, 16, 17 and 18**).

From compound **264** (Fig. **14**), we can see the interacting residues include ALA A: 101, ASP A: 3, MET A: 16, ASN A: 17, ILE A: 4, LYS A: 103 and THR A: 26 of which ALA A: 101 is bound with the ligand *via* alkyl bonding, ASP A: 3, ASN A: 17, MET A: 16 and THR A: 26 interact with the ligand with hydrogen bonding, while ILE A:4 and LYS A:103 show π-alkyl interactions.

In compound **688**, the interacting amino acids include ILE A: 5, ASP A: 3 and ILE A: 4. The different interactions are shown in Fig. (**15**). Hydrogen bonding interactions are exhibited by ILE A: 5, ASP A: 3 while ILE A: 4 show π-sigma interaction. It is observed from compound **697** (Fig. **16**) that amino acids ILE A: 4 and ASP A: 3 interact with the ligand *via* hydrogen bonding interaction, ASP A: 3 and ARG A: 14 interact with ligand *via* π-anion, π-cation and some other attractive charges whereas LYS A: 103, Met A: 16 show π-alkyl and alkyl interaction with the ligand.

In compound **695**, the interacting amino acids include LYS A:103, ILE A:99, ASP A:3, ILE A:4 and ASN A:17. The different interactions are shown in Fig. (**17**). Conventional hydrogen bond interactions are formed with ASN A:17 and ASP A:3 while LYS A:103, ILE A:99 and ILE A:4 form π-alkyl and alkyl bonding.

It is observed from compound **690** (Fig. **18**) that amino acids ILE A:5, ILE A:4, ASP A:3 and THR A:26 interact with the ligand *via* hydrogen bonding interaction, whereas MET A:16 and LYS A:103 show π-alkyl interaction with the ligand.

3.3.2. Moderately Active Molecules from the Dataset

In case of compounds **35**, **118**, **867**,**45 and 300** (pIC_{50} = -2.99, -2.96, -2.91, -3.770 and -3.959 respectively) which are moderately active against the BuChE enzyme, the interaction forces include hydrogen bond (carbon hydrogen bonds and conventional hydrogen bonds), pi-interaction (π-cation, π-alkyl, π-lone pair, π-cation, π-anion and π-π stacking) and interacting amino acids residues include such as . LYS A: 103, ASN A: 17, ILE A: 4, ASP A: 3, HIS A:438, TRP A: 82, TRP A: 430, TYR A: 332, THR A: 120, MET A: 16, THR A: 59, ILE A:99, LEU A:125, ALA A:328 and GLY A:115.

It is observed from compound **35** (Fig. **19**) that amino acids ASN A: 17 and ASP A: 3 interact with the ligand *via* hydrogen bonding interaction, LYS A: 103 and ILE A: 4 interact with ligand *via* π-alkyl and alkyl bonding.

In compound **118**, the interacting amino acids include HIS A:438, TRP A: 82, TRP A: 430, TYR A: 332, THR A: 120. The different interactions are shown in Fig. (**20**). The amino acid residues HIS A;438 interact with ligand through hydrogen bonding, whereas HIS A;438, TRP A: 82, TRP A: 430, TYR A: 332 and THR A: 120 share their hydrophobic feature through π-interactions (π-alkyl, pi-lone pair and π-cation).

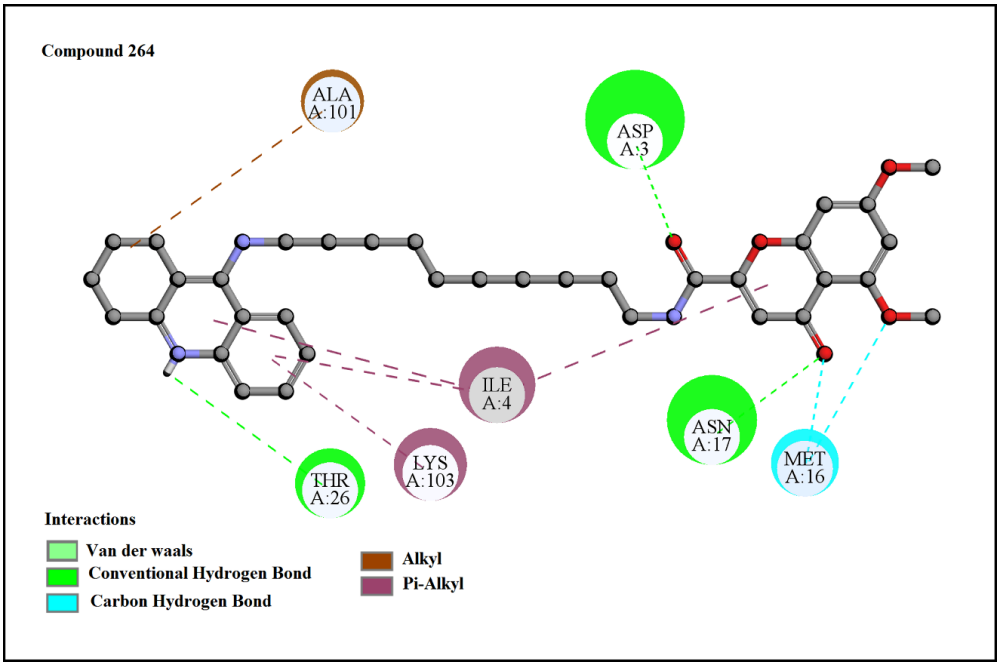


Fig. (14). Docking interactions of most active compound (compound **264**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

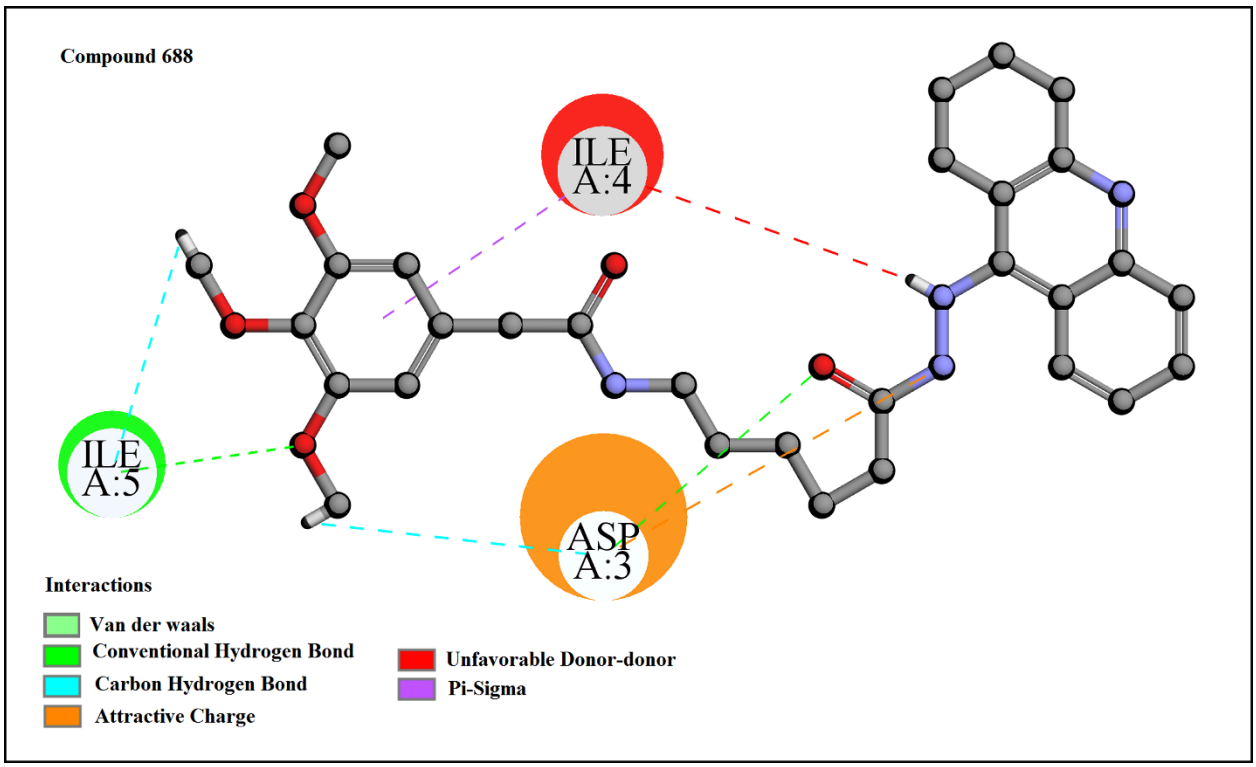


Fig. (15). Docking interactions of most active compound (compound **688**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

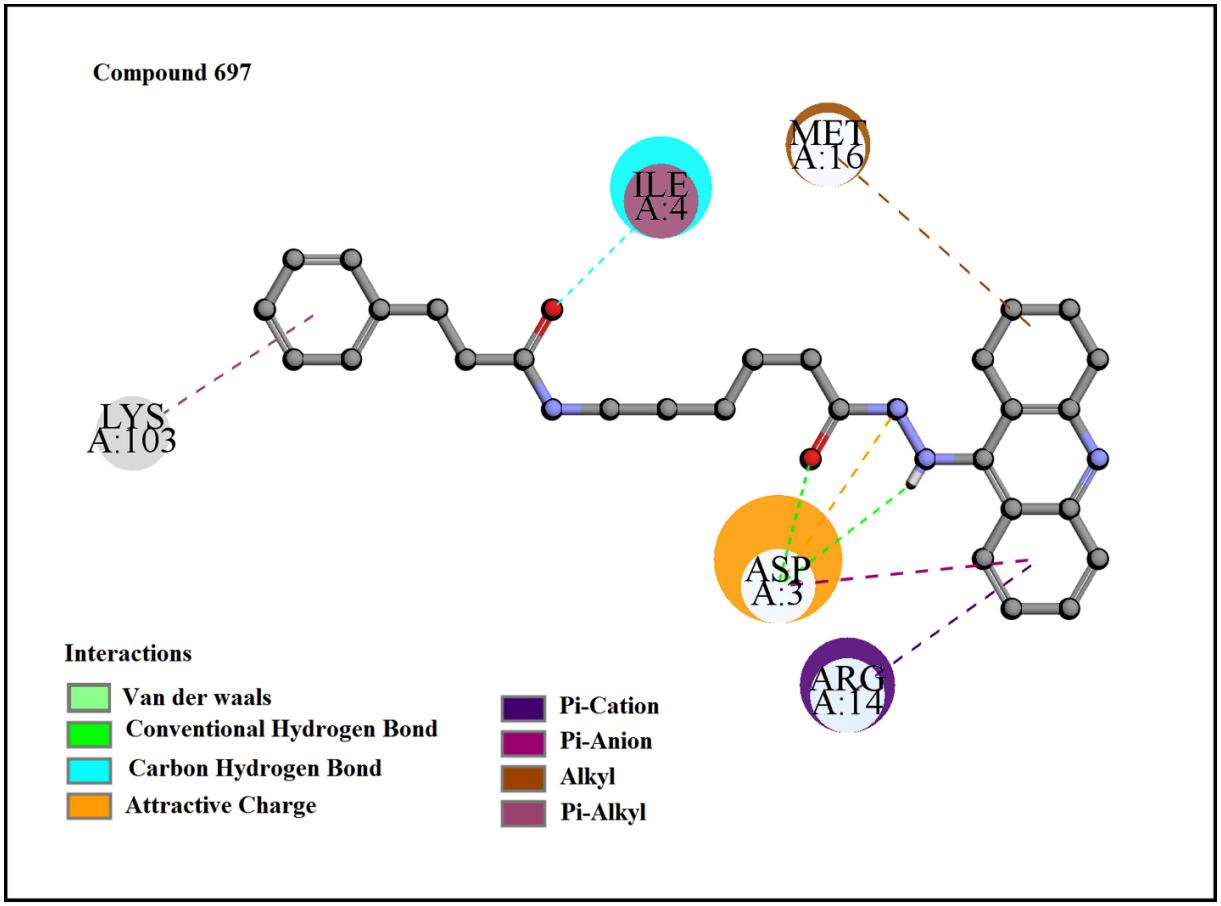


Fig. (16). Docking interactions of most active compound (compound **697**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

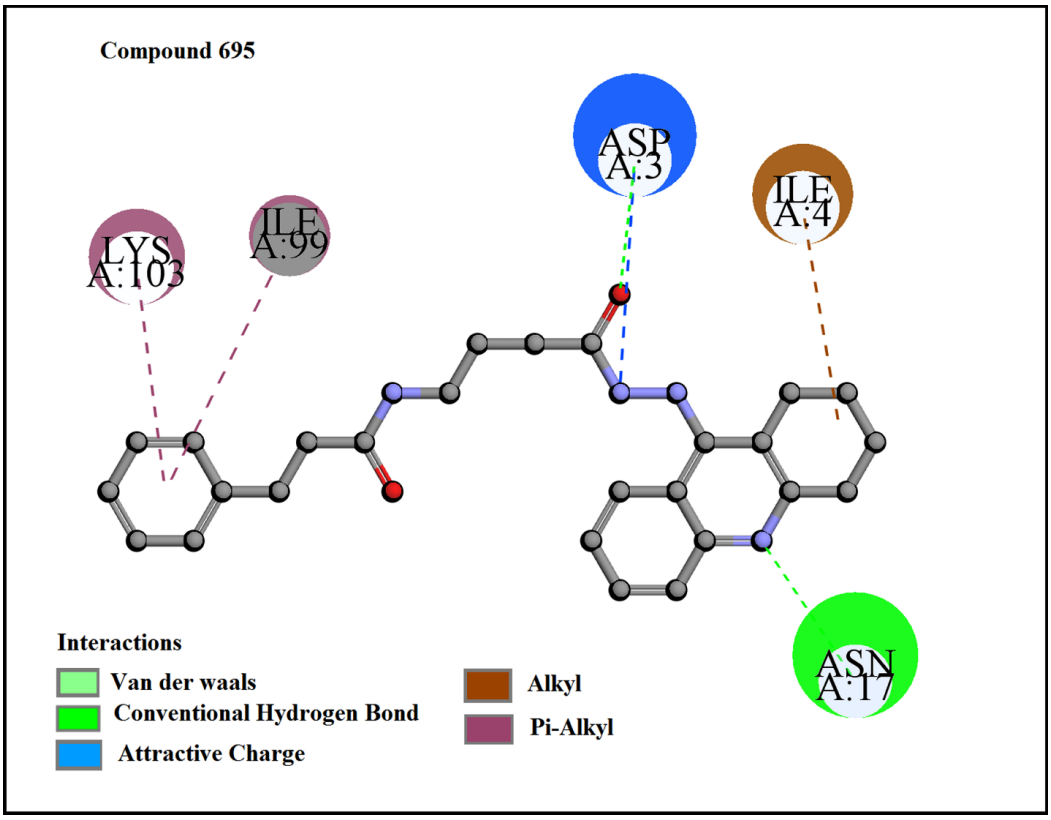


Fig. (17). Docking interactions of most active compound (compound **695**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

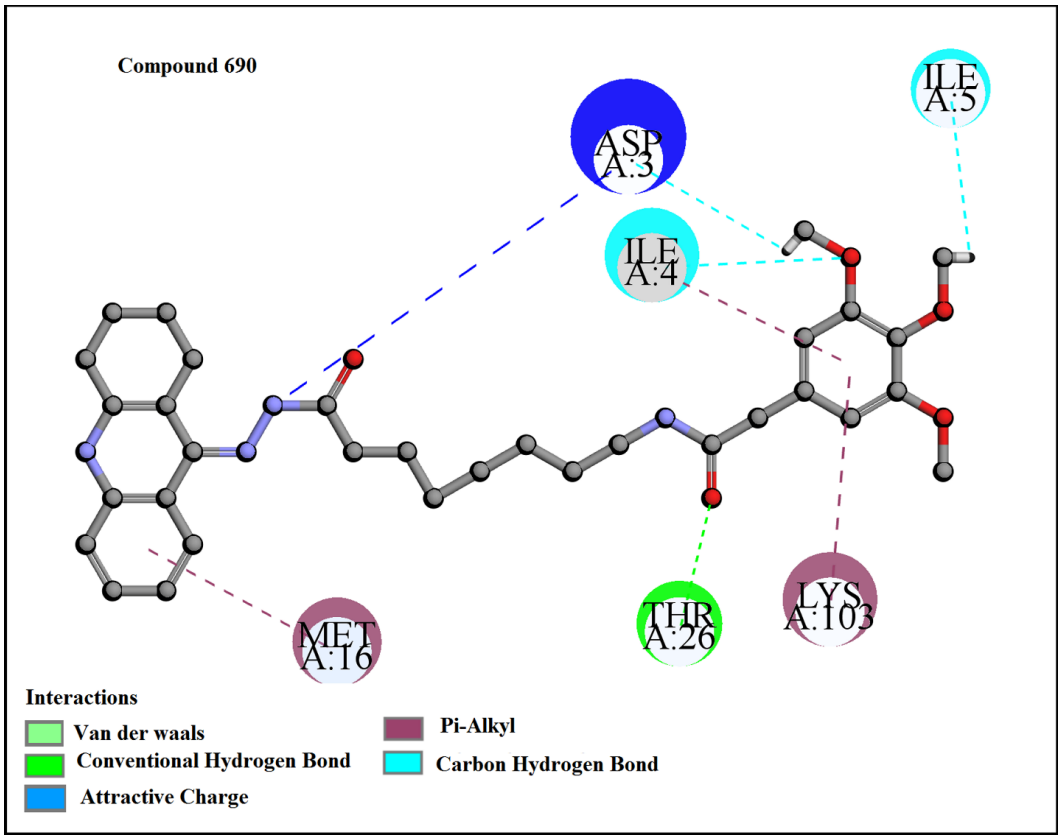


Fig. (18). Docking interactions of the most active compound (compound **690**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

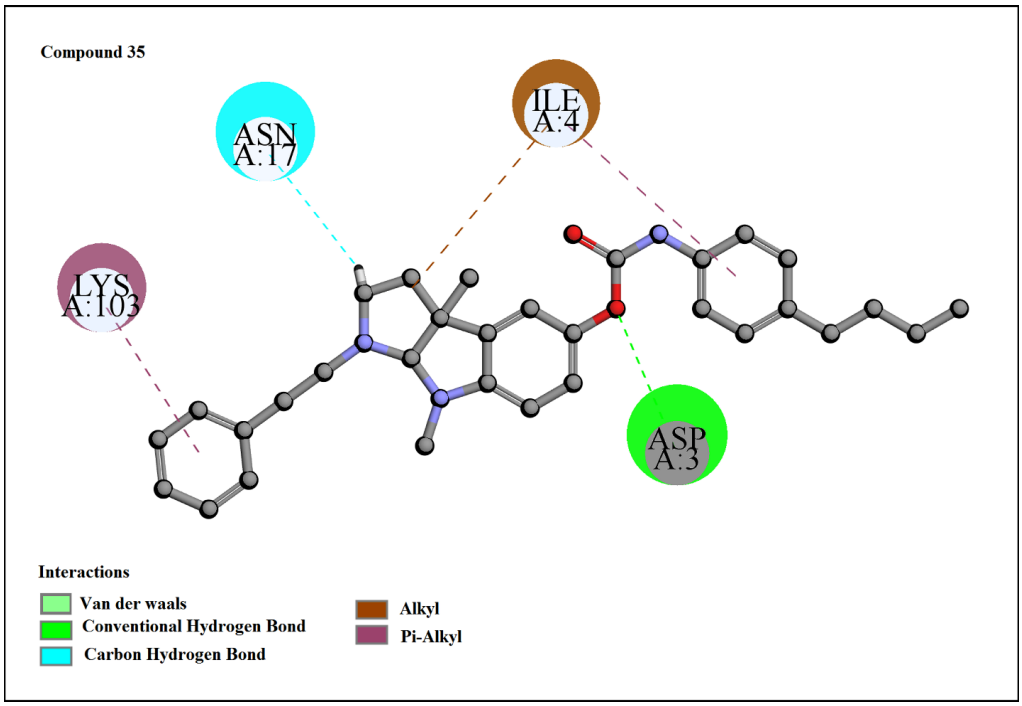


Fig. (19). Docking interactions of moderately active compound (compound **35**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

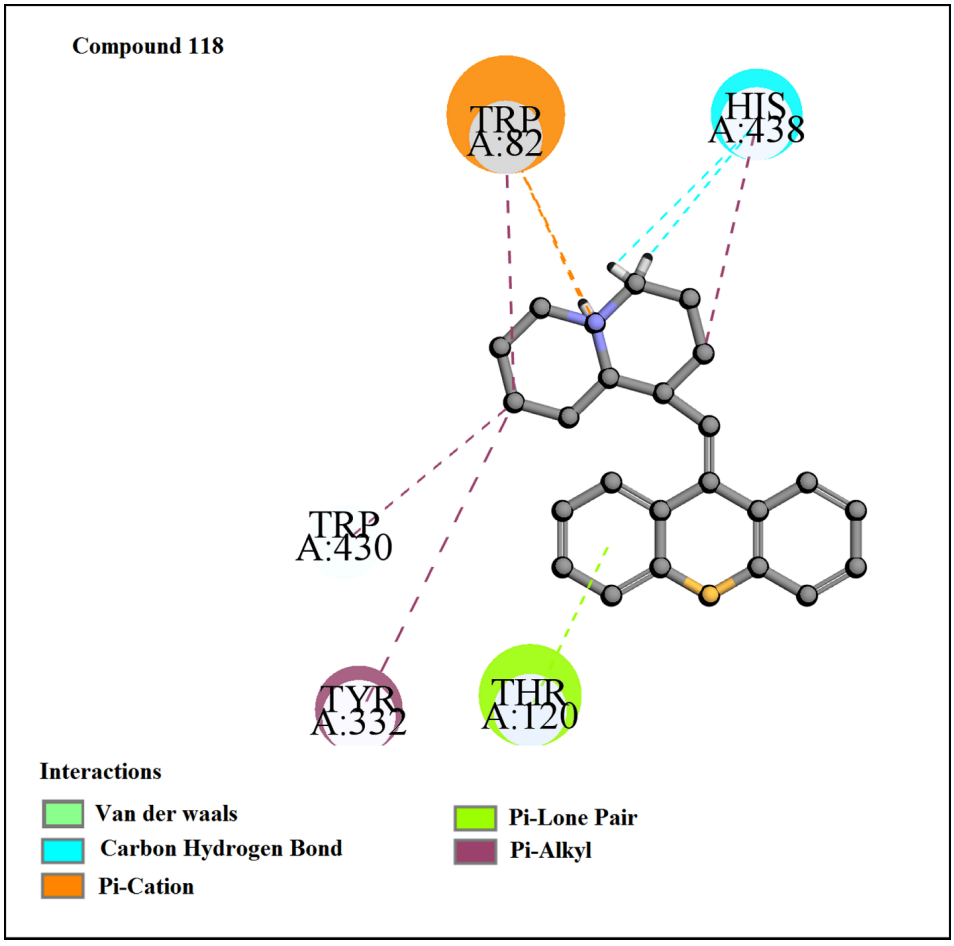


Fig. (20). Docking interactions of moderately active compound (compound **118**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

In case of compound **867 (**Fig. **21**), the interacting amino acid residues are ASP A: 3 and THR A: 59 which interact with the ligand making hydrogen bonding interaction while ILE A: 4, MET A: 16 bind with ligand *via* π-alkyl and alkyl bonds.

From compound **45** (Fig. **22**), we can see the interacting residues include ILE A:99, LYS A:103, ILE A:4 and ASP A:3, of which ASP A:3 is bound with the ligand *via* hydrogen bonding while ILE A:4, ILE A:99 and LYS A:103 share their hydrophobic feature *via* π-alkyl interactions.

In compound **300** (Fig. **23**), the interacting amino acids include LEU A:125, TYR A:332, THR A:120, TRP A:82, ALA A:328, TRY A:332, GLY A:115, TRP A:430 and HIS A:438. The different interactions are shown in Fig. (**15**). The amino acid residues HIS A;438, GLY A:115 and THR A:120 share their hydrophilic feature through hydrogen bonding interaction, whereas TRY A:332, ALA A:328, TRP A:82, TRY A:128, LEU A:125 and TRP A:430 are bound to the ligand *via* π-π interaction.

3.3.3. Least Active Molecules from the Dataset

The BuChE enzyme inhibitors **185**, **243**, **277**, **204** and **835** having a lower inhibitory activity (pIC_{50} = -5.65, -5.69, -5.60, -5.021 and -5.136 respectively) show similar kind of interactions (alkyl interaction, hydrogen and π interactions) as in case of higher inhibitory activity compounds, but the number of interacting amino acid residues are much less as shown in Figs. (**24, 25, 26, 27 and 28**).

From compound **185 (**Fig. **24**), we can see the interacting residues include TYR A: 332, TRP A: 82, TYR A: 128 and GLU A: 197, of which GLU A: 197 and TYR A: 128 are bound with the ligand *via* hydrogen bonding while TYR A: 332 and TRP A: 82 show π-sigma and π-π stacked interactions.

In case of compound **243** (Fig. **25**), the amino acids involved in the interaction are GLU A: 197, TRP A: 82, GLY A: 439, HIS A: 438, ALA A: 32. The amino acid residues HIS A:438 and GLU A:197 interact with the ligand *via* hydrogen bonding whereas TRP A:82, GLY A:439, ALA A:32 interact with the ligand through π interaction (π-alkyl, π-cation and π-π stacked).

From compound **277** (Fig. **26**), we can see the interacting residues include HIS A: 438, TYR A: 332, ALA A: 199 and GLY A: 117, in which GLY A: 117 and ALA A: 199 are bound with the ligand *via* hydrogen bonding while HIS A: 438 and TYR A: 332 show π-π T shaped interactions.

In case of compound **204** (Fig. **27**), the interacting amino acid residues are TRP A:231, GLU A:197, TRP A:128 and GLY A:115. GLU A:197 and TRP A:128, which interact with the ligand by hydrogen bonding while TRP A:231 bind with ligand *via* T-shaped π-π interaction, and GLY A:115 binds with the ligand through amide-π stacking interaction.

In case of compound **835** (Fig. **28**), the amino acids involved in the interaction are TRP A:82 and ALA A:328. The amino acid TRP A:82 interacts with the ligand through π-π stacking and π-sulphur interaction, whereas ALA A:328 binds with ligand *via* π-alkyl bonding interaction.

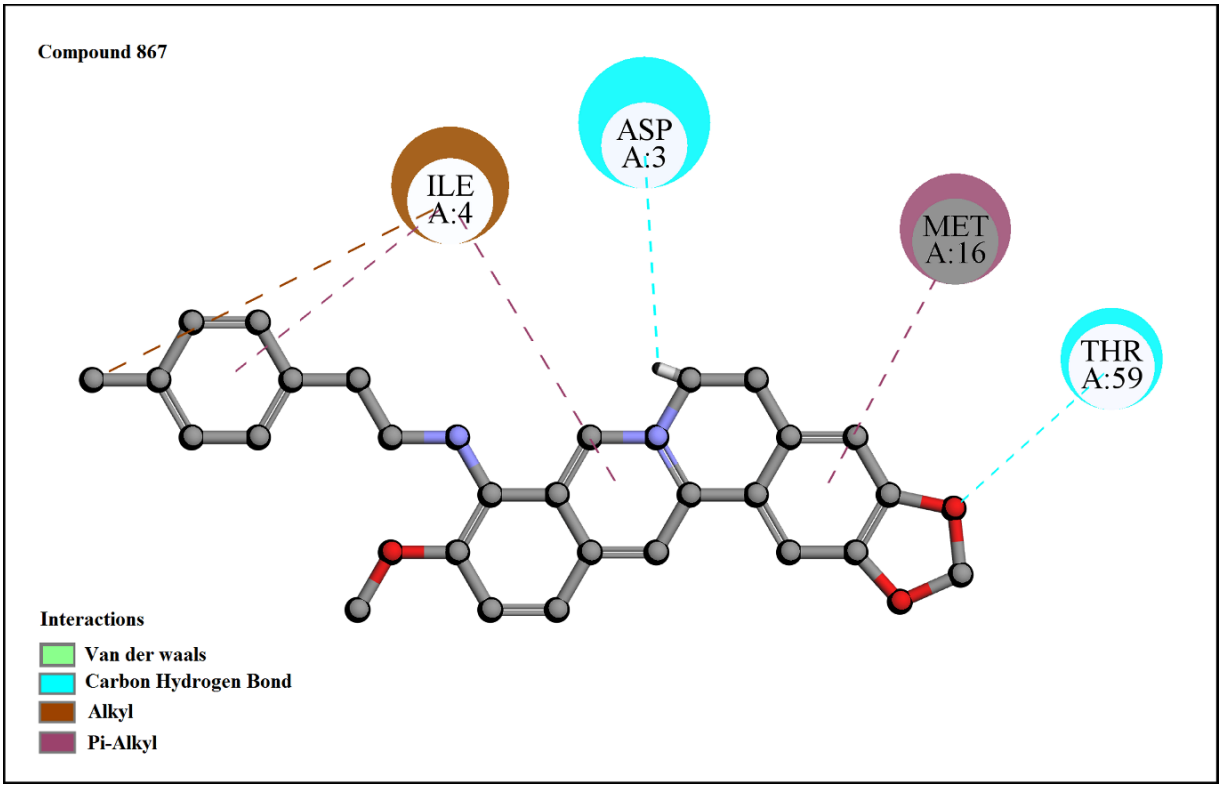


Fig. (21). Docking interactions of moderately active compound (compound **867**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

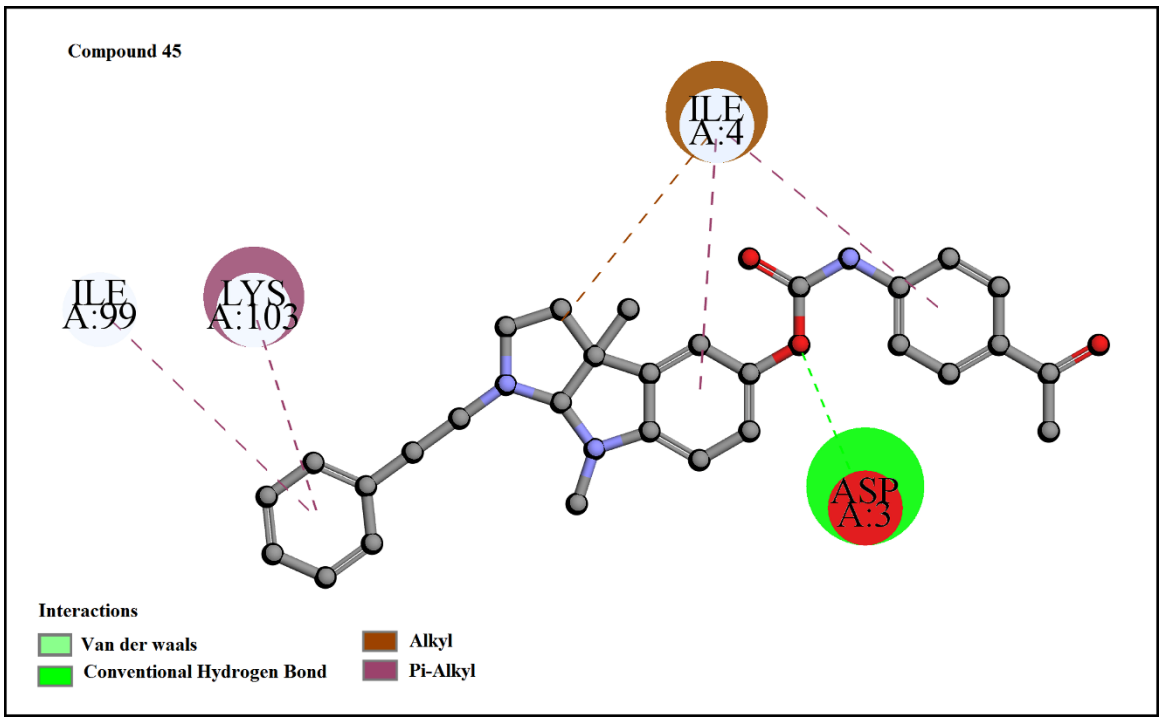


Fig. (22). Docking interactions of moderately active compound (compound **45**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

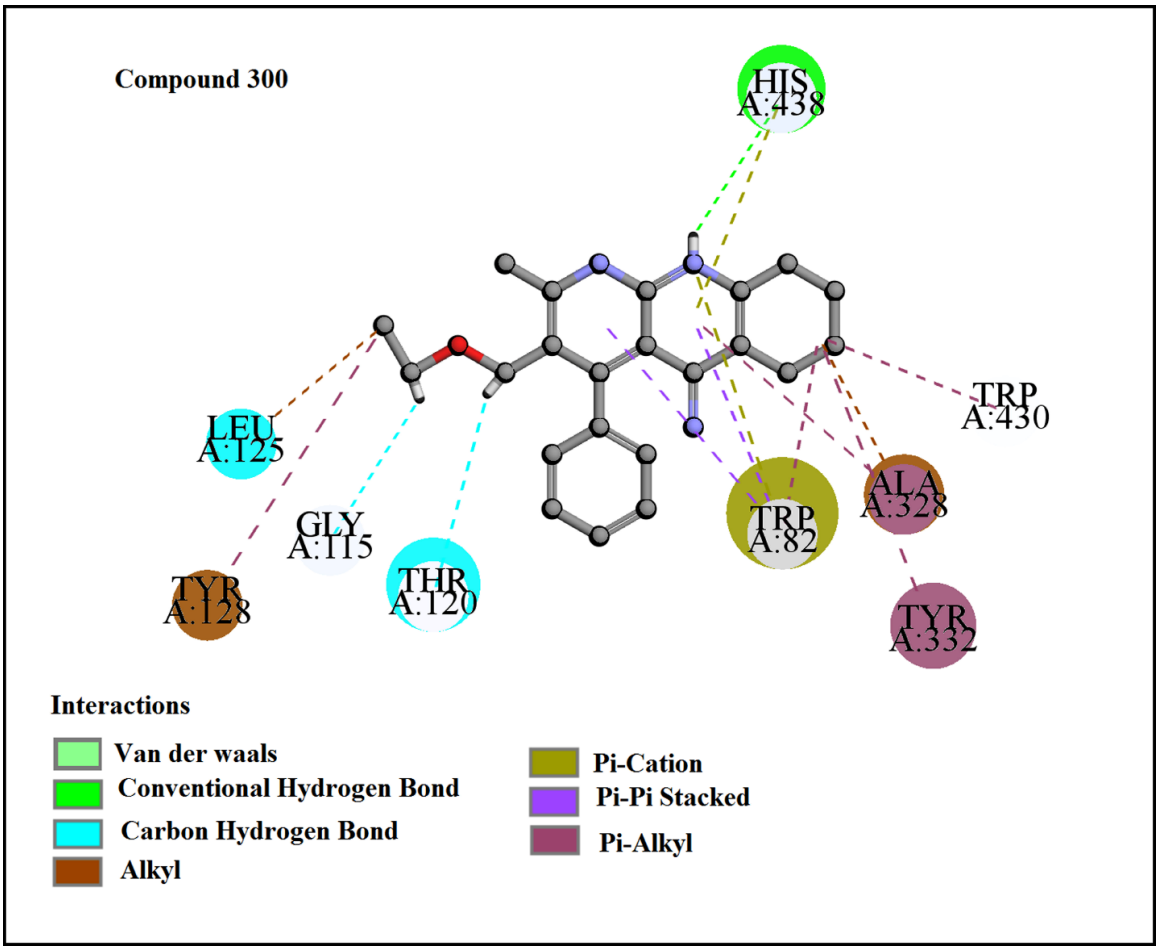


Fig. (23). Docking interactions of moderately active compound (compound **300**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

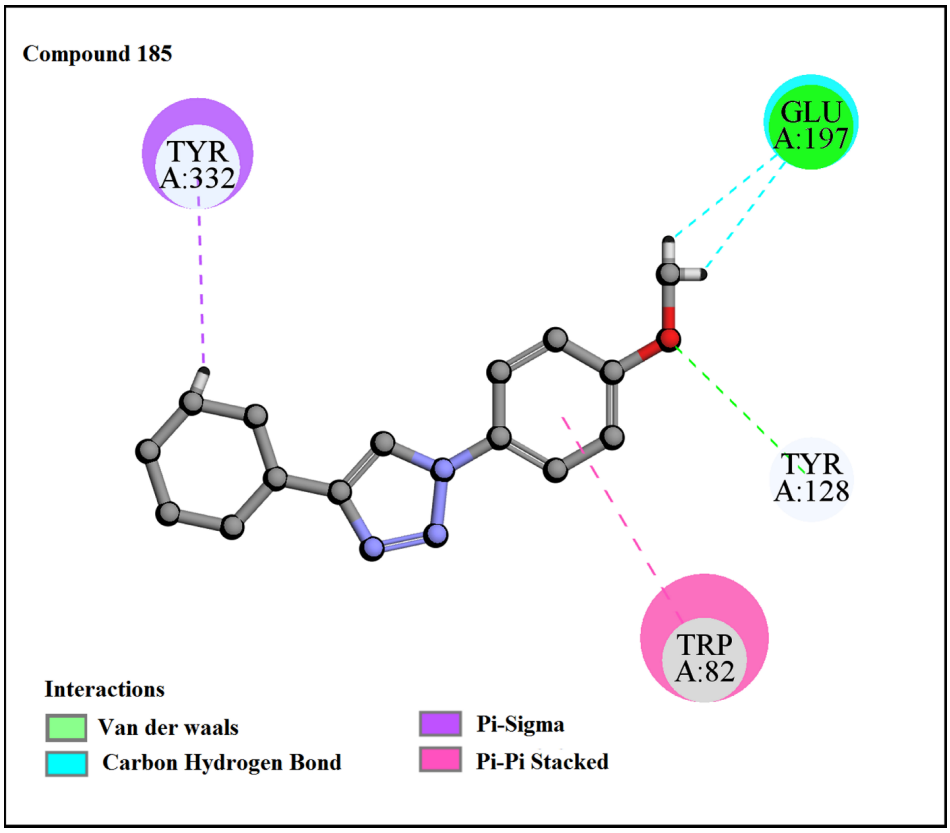


Fig. (24). Docking interactions of least active compound (compound **185**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

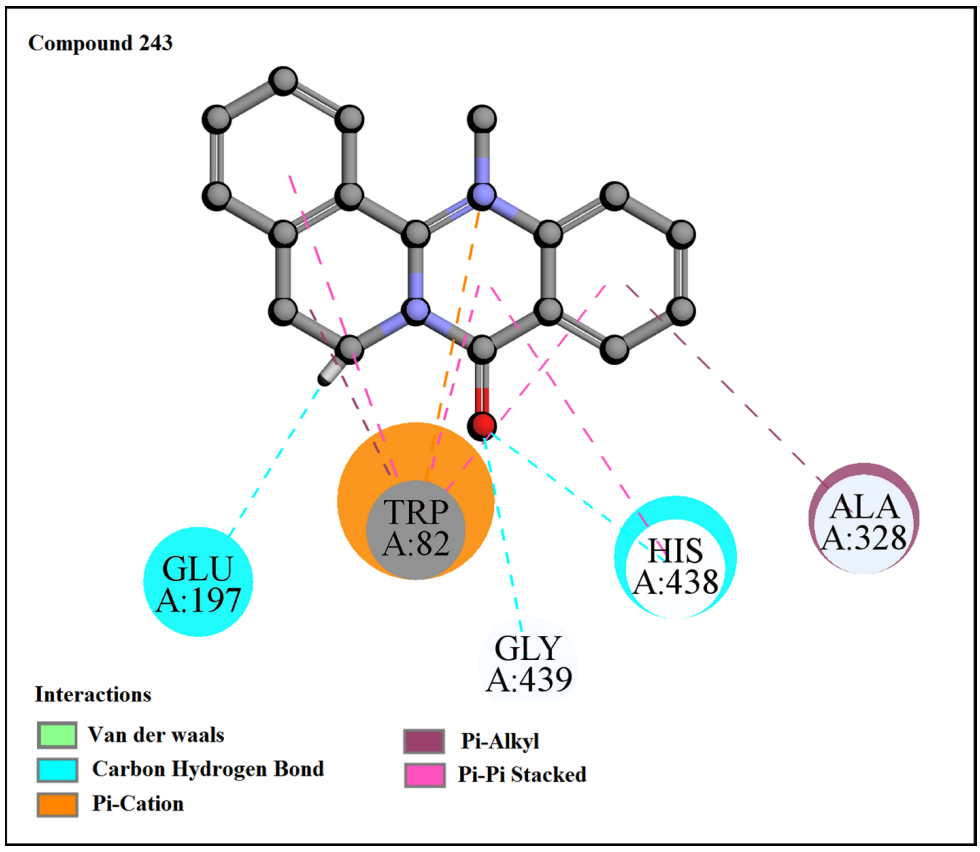


Fig. (25). Docking interactions of least active compound (compound **243**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

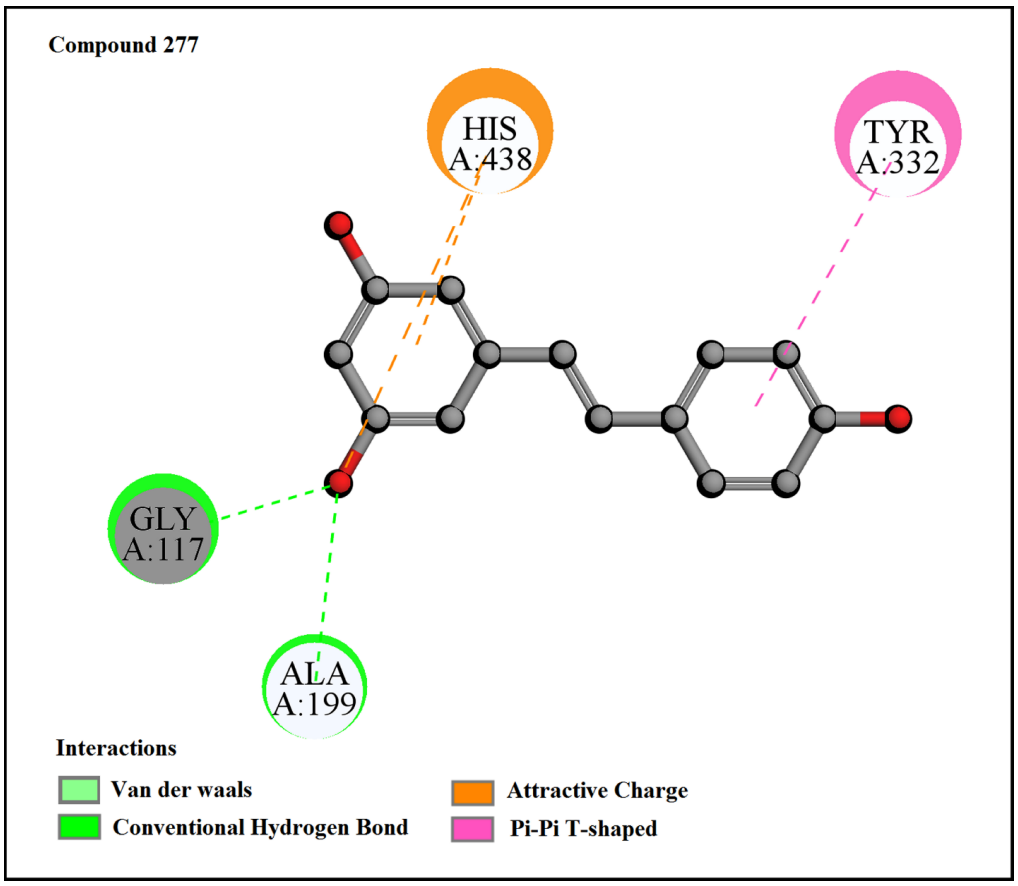


Fig. (26). Docking interactions of least active compound (compound **277**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

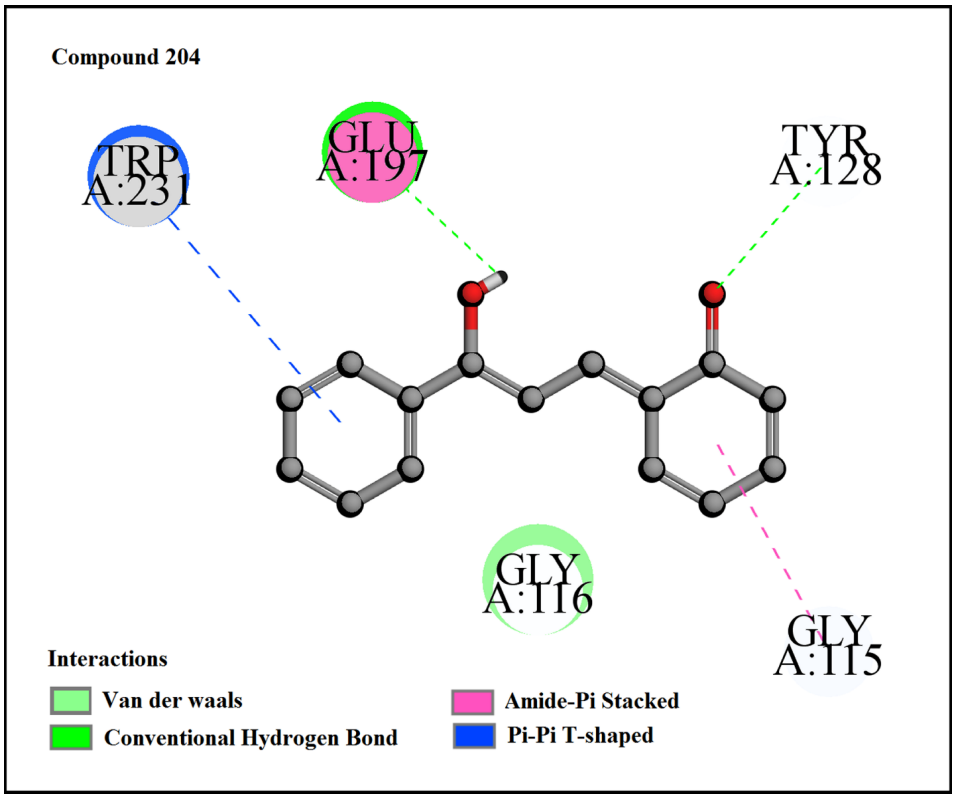


Fig. (27). Docking interactions of least active compound (compound **204**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

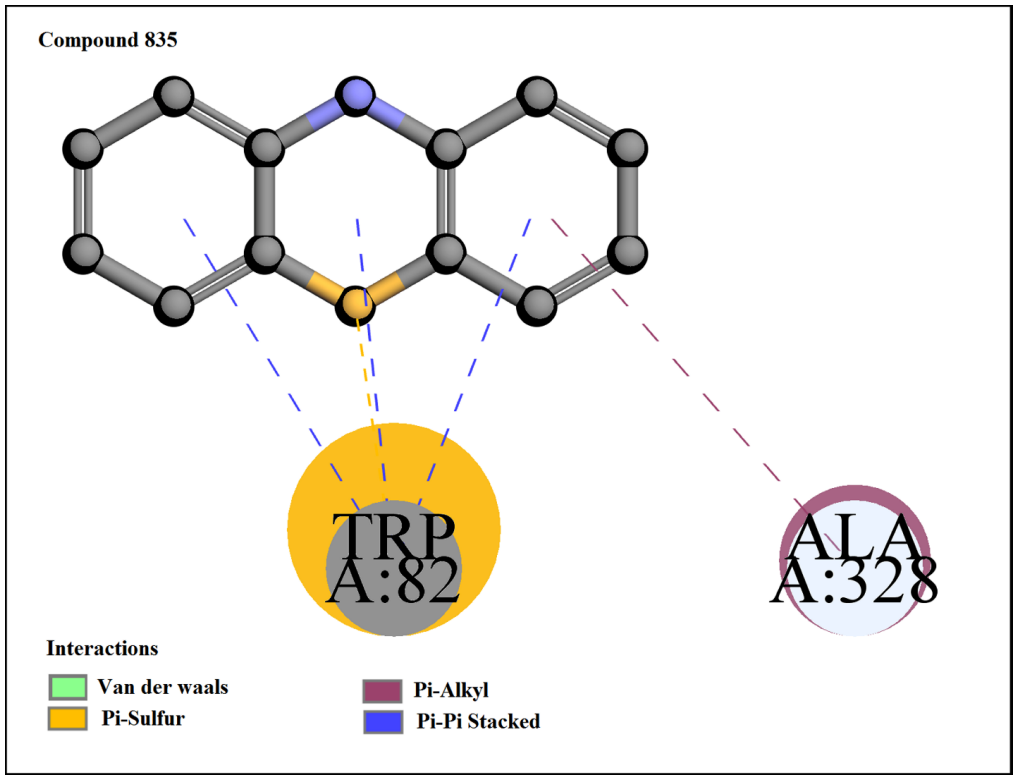


Fig. (28). Docking interactions of least active compounds (compound **835**). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

Table 3. Docking results and correlation with the final QSAR model.

S. No.	Compound Number	- CDocker interaction energy	Interacting residues	Interactions	Correlation with QSAR model
1	264 (high pIC_{50})	45.1	ALA A: 101, ASP A: 3, MET A: 16, ASN A: 17, ILE A: 4, LYS A: 103, THR A: 26	vdW, Hydrogen bonding, alkyl and pi-alkyl	nPyridines, nCrs and T(N..O)
2	688 (high pIC_{50})	47.20	ILE A: 5, ASP A: 3, ILE A: 4	vdW, Hydrogen bonding, attractive charges and pi-sigma	nPyridines, nN-N, and T(N..O)
3	697 (high pIC_{50})	42.67	Met A: 16, ILE A: 4, LYS A: 103, ASP A: 3, ARG A: 14	vdW, Hydrogen bonding, attractive charges, pi-cation, pi-anion, alkyl and pi-alkyl	nPyridines, RDCHI and nN-N, nCrs
4	695 (high pIC_{50})	40.56	LYS A:103, ILE A:99, ASP A:3, ILE A:4 and ASN A:17	vdW, Attractive charges, Hydrogen bonding, Pi-alkyl and alkyl	nPyridines, nN-N, nCrs and T(N..O)
5	690 (high pIC_{50})	48.38	Met A:16, ILE A: 4, ILE A: 5, LYS A:103, THR A: 26, ASP A:3	vdW,Hydrogen bonding, Attractive charge, pi-alkyl	nPyridines, RDCHI and nN-N
6	35 (moderate pIC_{50})	42.46	LYS A: 103, ASN A: 17, ILE A: 4, ASP A: 3	vdW, Hydrogen bonding, alkyl and pi-alkyl	RDCHI
7	118 (moderate pIC_{50})	42.09	HIS A:438, TRP A: 82, TRP A: 430, TYR A: 332, THR A: 120	vdW, Hydrogen bonding, Pi-cation, pi-alkyl, pi-lone pair	NRS and nCrs
8	867 (moderate pIC_{50})	28	ILE A: 4, ASP A: 3, MET A: 16, THR A: 59	vdW, Hydrogen bonding, alkyl, pi-alkyl	T(N..O)
9	45 (moderate pIC_{50})	40.59	ILE A:99, LYS A:103, ILE A:4 and ASP A:3	vdW, Hydrogen bonding, Pi-alkyl and alkyl	RDCHI, nCrs and C-041

(Table 3) contd….

S. No.	Compound Number	- CDocker interaction energy	Interacting residues	Interactions	Correlation with QSAR model
10	300 (moderate pIC_{50})	44.43	LEU A:125, TYR A:332, THR A:120, TRP A:82, ALA A:328, TRY A:332, GLY A:115, TRP A:430 and HIS A:438	vdW, Hydrogen bonding, Pi-alkyl, alkyl, pi-cation and pi-pi staking	nPyridines and nCrs
11	185 (low pIC_{50})	29.55	TYR A: 332, TRP A: 82, TYR A: 128, GLU A: 197	vdW, Hydrogen bonding, pi-pi stacked, pi-sigma,	B01[N-N] and NRS
12	243 (low pIC_{50})	37.65	GLU A: 197, TRP A: 82, GLY A: 439, HIS A: 438, ALA A: 328	vdW, Hydrogen bonding, pi-cation, pi-pi- stacked, pi alkyl	NRS
13	277 (low pIC_{50})	40	HIS A: 438, TYR A: 332, ALA A: 199, GLY A: 117	vdW, Hydrogen bonding, attractive charges, pi-cation, pi-pi-T-shaped	NRS
14	204 (low pIC_{50})	35.193	TRP A:231, GLU A:197, TRP A:128 and GLY A:115	vdW, Hydrogen bonding, pi-pi T-shaped and amide pi-stacked	NRS
15	835 (low pIC_{50})	25.128	TRP A:82 and ALA A:328	vdW, Pi-alkyl, pi-sulphur, pi-pi-stacked	NRS

3.3.4. Relation of the Docking Results with the QSAR Model

In the docking study, we have observed that the formation of hydrogen bonds and π-π stacking between the ligand and the target play a vital role in binding. Hydrogen bonding and π-π-interactions can be correlated with T(N..O) (sum of topological distances between N...O) and nCrs (number of ring secondary C(sp3)) descriptors in the QSAR model. T(N..O) is related to hydrogen bonding, electrostatic and π-donor hydrogen bonding interactions between protein and ligand. The descriptor, nCrs, gives evidence of π-π interaction. Furthermore, nPyridines (number of Pyridines) descriptor supports the evidence of π-π interaction (π-cation, π-anion, π-π stacking and π-alkyl) along with hydrogen bonding interaction as we have observed in compounds **264**, **688**, **697**,**695, 690** and **300** (Figs. **14, 15, 16, 17, 18** and **23**). The RDCHI (simply characterizes the size and branching of molecules) descriptor also supports the π-π interaction (π-alkyl and alkyl) while C-041 (X-C(=X)-X) descriptor signifies for hydrogen bonding interaction. The H-O48 (C2(sp3)/C1(sp2)/C0(sp)) descriptor characterizes both the hydrogen bonding and π-π (π-alkyl and alkyl) interactions in the QSAR model. Thus, from the above mentioned information, we can conclude that hydrogen bonding, hydrophobicity, electrostatic interactions and unsaturation (π-π interaction) features obtained from both QSAR and docking study are essential for the inhibitory activity against the BuChE enzyme.

3.4. Comparisons of the Performance of the Proposed Study with the Previously Published Studies

There are a number of previous QSAR models reported for the prediction of bioactivity of BuChE enzyme inhibitors. Here, we have performed a comparison of the best model currently derived from some previous models (Table **4**). The previously reported models were developed by the Multilinear Regression (MLR) analysis, partial least squares (PLS), Genetic function approximation (GFA), Multilayer perceptron (MLP) and Artificial neural network (ANN) method, which gave reliable predictions of bioactivity of BuChE enzyme inhibitors. However, the models reported previously were developed using a very low number of compounds covering a very narrow range of chemical diversity. However, in the current study, we have employed an extended list of compounds covering a wide range of chemicals and offering a larger chemical domain. We can see from Table **4** that Fang *et al.* [15] developed a PLS model against BuChE enzyme by using only 66 compounds; the model quality was good, but the equation length (12) was quite high compared to the number of data points. In this study, we have utilized a wide range of compounds and developed the model with 15 selected descriptors and 6 latent variables. We can also see from Table **4** that Zheng *et al.* [16] and Bitam *et al.* [18] developed QSAR models using very narrow groups of samples (151 and 93 compounds respectively) and developed MLR, ANN, and MLP based models. Solomon *et al.* [17] reported GFA models utilizing only 59 compounds. The details of different internal and external validation parameters obtained from our model and obtained from the previously reported models are given in Table **4**. The best model presented in this work is based on a larger group of samples and the validation parameters (both internal and external) of the training and test sets qualified the requisite thresholds. Prior to the development of the final model, we have performed multilayered variable selection strategy from a large pool of descriptors. The best model was selected based on different validation parameters and low equation length. The final model was built by using the PLS algorithm with latent variables of 6. The 15 selected descriptors reflect the fundamental structural characteristics of molecules, which are important in modeling the bioactivity of BuChE enzyme inhibitors. The docking results in this study also well collaborate with the descriptors obtained from the developed QSAR model and justify the significance of the developed model. In comparison with the previously reported models with respect to acceptability and reliability, the present work deals with diverse classes of compounds. Due to the wide applicability

Table 4. Comparison of the developed model with the previously published models.

Sources	Equation Length	Model	Training set			Test set	
			n	R^2	Q^2	n	R^2pred
Model in this study	15	PLS	848	0.664	0.650	282	0.657
Fang *et al.* 2016 [15]	12	MLR	48	0.883	0.726	18	0.731
Fang *et al.* 2016 [15]	12	PLS	48	0.883	0.777	18	0.775
Zheng *et al.* 2014 [16]	10	MLR	62	0.89	0.85	31	-
Zheng *et al.* 2014 [16]	10	ANN	62	0.950	0.900	31	-
Solomon *et al.* 2009 [17]	5	GFA	39	0.884	0.857	20	0.820
Bitam *et al.* 2018 [18]	8	MLR	121	0.879	0.857	30	0.847
Bitam *et al.* 2018 [18]	8	MLP	121	0.888	0.895	30	-

domain, the model reported in the present study may be used as a screening tool for the discovery and development of leads against BuChE enzyme.

4. OVERVIEW AND CONCLUSIONS

In the present study, we have developed a PLS-regression based 2D-QSAR model using multi-layered feature selection strategy for the prediction of inhibitory activity against BuChE enzyme using a large dataset containing 1130 diverse molecules [23-76]. The PLS model was developed by following strict OECD guidelines (a defined endpoint, unambiguous algorithm, acceptable quantitative metrics, applicability domain analysis and mechanistic interpretation). The 2D QSAR model gives us knowledge about the several structural and physicochemical features of inhibitors significantly contributing to the prediction of inhibitory activity against BuChE enzyme. From the developed model, it can be inferred that hydrogen bonding effect, hydrophobicity, electrostatic interactions and unsaturation (π-π interaction) play an essential role in regulating enzyme inhibitory activity. Furthermore, the results obtained from molecular docking studies are well corroborated with the results of the QSAR analysis. The developed 2D-QSAR model may thus be helpful for the chemists to easily identify the required features in the rational design of inhibitors for the treatment of AD.

LIST OF ABBREVIATIONS

AChE = Acetylcholinesterase enzyme

AChEI = Acetylcholinesterase enzyme inhibitors

BuChE = Butylcholinestrase enzyme

BuChEI = Butylcholinestrase enzyme inhibitors

AD = Alzheimer's disease

AD = Applicability domain

VIP = Variable importance plot

PLS = Partial least square

SW = Stepwise

NRS = Number of ring system

nCrs = Number of ring secondary C

nPyridines = Number of pyridines moiety

ETA = Extended topochemical atom

LOO = Leave one out

QSAR = Quantitative structure activity relationship

OECD = Organization for economic co-operation

NMDA = N-Methyl-D-aspartate

CDK5 = Cyclin dependent kinase 5

GSK3β = Glycogen synthase kinase 3

BSS = Best subset selection

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

Financial assistance from the Indian Council of Medical Research (ICMR), New Delhi in the form of a senior research fellowship (File No: 5/3/8/27/ITR-F/2018-ITR; dated: 18.05.2018) to VK is thankfully acknowledged.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Dinamarca, M.C.; Sagal, J.P.; Quintanilla, R.A. Godoy, J.A.; Arrázola, M.S.. and Inestrosa. Amyloid-β-Acetylcholinesterase complexes potentiate neurodegenerative changes induced by the Aβ peptide. Implications for the pathogenesis of Alzheimer's disease. *Mol. Neurodegener.,* **2010**, *5*(1), 4. http://dx.doi.org/10.1186/1750-1326-5-4 PMID: 20205793

[2] Zufferey, V.; Gunten, A.V.; Kherif, F. Interactions between personality, depression, anxiety and cognition to understand early stage of alzheimer's disease. *Curr. Top. Med. Chem.,* **2020**, *20*(9), 782-791. http://dx.doi.org/10.2174/1568026620666200211110545 PMID: 32066361

[3] Ambure, P.; Kar, S.; Roy, K. Pharmacophore mapping-based virtual screening followed by molecular docking studies in search of potential acetylcholinesterase inhibitors as anti-Alzheimer's agents. *Biosystems,* **2014**, *116*, 10-20. http://dx.doi.org/10.1016/j.biosystems.2013.12.002 PMID: 24325852

[4] Jucker, M. Pathogenic protein seeding in Alzheimer disease and other neurodegenerative disorders. *Annals of neurology, L.C,* **2011**, *70*(4), 532-540.

[5] Sharma, P.; Sharma, A.; Fayaz, F.; Wakode, S.; Pottoo, F.H. Biological Signatures of Alzheimer's Disease. *Curr. Top. Med. Chem.,* **2020**, *20*(9), 770-781. http://dx.doi.org/10.2174/1568026620666200228095553 PMID: 32108008

[6] McPhee, S.J. *Ganong pathophysiology of disease: an introduction to clinical medicine; The McGraw-Hill Companies, Inc: Pennsylvania,* **2010**, 493-500.

[7] Reid, G.A.; Chilukuri, N.; Darvesh, S. Butyrylcholinesterase and the cholinergic system. *Neuroscience,* **2013**, *234*, 53-68. http://dx.doi.org/10.1016/j.neuroscience.2012.12.054 PMID: 23305761

[8] Saido, T.C.; Iwata, N. Metabolism of amyloid β peptide and pathogenesis of Alzheimer's disease. Towards presymptomatic diagnosis, prevention and therapy. *Neurosci. Res.,* **2006**, *54*(4), 235-253. http://dx.doi.org/10.1016/j.neures.2005.12.015 PMID: 16457902

[9] Iwata, N.; Higuchi, M.; Saido, T.C. Metabolism of amyloid-β peptide and Alzheimer's disease. *Pharmacol. Ther.,* **2005**, *108*(2), 129-148. http://dx.doi.org/10.1016/j.pharmthera.2005.03.010 PMID: 16112736

[10] Colović, M.B.; Krstić, D.Z.; Lazarević-Pašti, T.D.; Bondžić, A.M.; Vasić, V.M. Acetylcholinesterase inhibitors: pharmacology and toxicology. *Curr. Neuropharmacol.,* **2013**, *11*(3), 315-335. http://dx.doi.org/10.2174/1570159X11311030006 PMID: 24179466

[11] Kumar, A.; Pintus, F.; Di Petrillo, A.; Medda, R.; Caria, P.; Matos, M.J.; Viña, D.; Pieroni, E.; Delogu, F.; Era, B.; Delogu, G.L.; Fais, A. Novel 2-pheynlbenzofuran derivatives as selective butyrylcholinesterase inhibitors for Alzheimer's disease. *Sci. Rep.,* **2018**, *8*(1), 4424. http://dx.doi.org/10.1038/s41598-018-22747-2 PMID: 29535344

[12] Giacobini, E. Selective inhibitors of butyrylcholinesterase: a valid alternative for therapy of Alzheimer's disease? *Drugs Aging,* **2001**, *18*(12), 891-898. http://dx.doi.org/10.2165/00002512-200118120-00001 PMID: 11888344

[13] Kandimalla, R.; Reddy, P.H. Therapeutics of neurotransmitters in Alzheimer's disease. *J. Alzheimers Dis.,* **2017**, *57*(4), 1049-1069. http://dx.doi.org/10.3233/JAD-161118 PMID: 28211810

[14] Greig, N.H.; Lahiri, D.K.; Sambamurti, K. Butyrylcholinesterase: an important new target in Alzheimer's disease therapy. *Int. Psychogeriatr.,* **2002**, *14*(S1)(Suppl. 1), 77-91. http://dx.doi.org/10.1017/S1041610203008676 PMID: 12636181

[15] Fang, J.; Pang, X.; Wu, P.; Yan, R.; Gao, L.; Li, C.; Lian, W.; Wang, Q.; Liu, A.L.; Du, G.H. Molecular modeling on berberine derivatives toward buche: an integrated study with quantitative structure-activity relationships models, molecular docking, and molecular dynamics simulations. *Chem. Biol. Drug Des.,* **2016**, *87*(5), 649-663. http://dx.doi.org/10.1111/cbdd.12700 PMID: 26648584

[16] Zheng, F.; Zhan, M.; Huang, X.; Abdul Hameed, M.D.; Zhan, C.G. Modeling *in vitro* inhibition of butyrylcholinesterase using molecular docking, multi-linear regression and artificial neural network approaches. *Bioorg. Med. Chem.,* **2014**, *22*(1), 538-549. http://dx.doi.org/10.1016/j.bmc.2013.10.053 PMID: 24290065

[17] Solomon, K.A.; Sundararajan, S.; Abirami, V. QSAR studies on N-aryl derivative activity towards Alzheimer's disease. *Molecules,* **2009**, *14*(4), 1448-1455. http://dx.doi.org/10.3390/molecules14041448 PMID: 19384276

[18] Bitam, S.; Hamadache, M.; Hanini, S. Prediction of therapeutic potency of tacrine derivatives as BuChE inhibitors from quantitative structure-activity relationship modelling. *SAR QSAR Environ. Res.,* **2018**, *29*(3), 213-230. http://dx.doi.org/10.1080/1062936X.2018.1423640 PMID: 29390887

[19] Ojha, P.K.; Roy, K. Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection. *Chemom. Intell. Lab. Syst.,* **2011**, *109*(2), 146-161. http://dx.doi.org/10.1016/j.chemolab.2011.08.007

[20] Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.,* **2011**, **2015**(145), 22-29.

[21] Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graph. Model.,* **2002**, *20*(4), 269-276. http://dx.doi.org/10.1016/S1093-3263(01)00123-1 PMID: 11858635

[22] Pal, S.; Kumar, V.; Kundu, B.; Bhattacharya, D.; Preethy, N.; Reddy, M.P.; Talukdar, A. Ligand-based Pharmacophore Modeling, Virtual Screening and Molecular Docking Studies for Discovery of Potential Topoisomerase I Inhibitors. *Comput. Struct. Biotechnol. J.,* **2019**, *17*, 291-310. http://dx.doi.org/10.1016/j.csbj.2019.02.006 PMID: 30867893

[23] Sermboonpaisarn, T.; Sawasdee, P. Potent and selective butyrylcholinesterase inhibitors from Ficus foveolata. *Fitoterapia,* **2012**, *83*(4), 780-784. http://dx.doi.org/10.1016/j.fitote.2012.03.009 PMID: 22450264

[24] Schulze, M.; Siol, O.; Decker, M.; Lehmann, J. Bivalent 5,8,9,13b-tetrahydro-6H-isoquino[1,2-a]isoquinolines and -isoquinolinium salts: novel heterocyclic templates for butyrylcholinesterase inhibitors. *Bioorg. Med. Chem. Lett.,* **2010**, *20*(9), 2946-2949. http://dx.doi.org/10.1016/j.bmcl.2010.03.011 PMID: 20350808

[25] Yan, J.W.; Li, Y.P.; Ye, W.J.; Chen, S.B.; Hou, J.Q.; Tan, J.H.; Ou, T.M.; Li, D.; Gu, L.Q.; Huang, Z.S. Design, synthesis and evaluation of isaindigotone derivatives as dual inhibitors for acetylcholinesterase and amyloid beta aggregation. *Bioorg. Med. Chem.,* **2012**, *20*(8), 2527-2534. http://dx.doi.org/10.1016/j.bmc.2012.02.061 PMID: 22444876

[26] Takahashi, J.; Hijikuro, I.; Kihara, T.; Murugesh, M.G.; Fuse, S.; Tsumura, Y.; Akaike, A.; Niidome, T.; Takahashi, T.; Sugimoto, H. Design, synthesis and evaluation of carbamate-modified (-)-N(1)-phenethylnorphysostigmine derivatives as selective butyrylcholinesterase inhibitors. *Bioorg. Med. Chem. Lett.,* **2010**, *20*(5), 1721-1723. http://dx.doi.org/10.1016/j.bmcl.2010.01.035 PMID: 20137941

[27] Tang, H.; Zhao, H.T.; Zhong, S.M.; Wang, Z.Y.; Chen, Z.F.; Liang, H. Novel oxoisoaporphine-based inhibitors of acetyl- and butyrylcholinesterase and acetylcholinesterase-induced beta-amyloid aggregation. *Bioorg. Med. Chem. Lett.,* **2012**, *22*(6), 2257-2261. http://dx.doi.org/10.1016/j.bmcl.2012.01.090 PMID: 22341944

[28] Chen, Y.; Sun, J.; Huang, Z.; Liao, H.; Peng, S.; Lehmann, J.; Zhang, Y. NO-donating tacrine derivatives as potential

butyrylcholinesterase inhibitors with vasorelaxation activity. *Bioorg. Med. Chem. Lett.,* **2013**, *23*(11), 3162-3165. http://dx.doi.org/10.1016/j.bmcl.2013.04.008 PMID: 23639542

[29] Yu, L.; Cao, R.; Yi, W.; Yan, Q.; Chen, Z.; Ma, L.; Peng, W.; Song, H. Synthesis of 4-[(diethylamino)methyl]-phenol derivatives as novel cholinesterase inhibitors with selectivity towards butyrylcholinesterase. *Bioorg. Med. Chem. Lett.,* **2010**, *20*(11), 3254-3258. http://dx.doi.org/10.1016/j.bmcl.2010.04.059 PMID: 20452769

[30] Mutahir, S.; Jończyk, J.; Bajda, M.; Khan, I.U.; Khan, M.A.; Ullah, N.; Ashraf, M.; Qurat-ul-Ain, ; Riaz, S.; Hussain, S.; Yar, M. Novel biphenyl bis-sulfonamides as acetyl and butyrylcholinesterase inhibitors: Synthesis, biological evaluation and molecular modeling studies. *Bioorg. Chem.,* **2016**, *64*, 13-20. http://dx.doi.org/10.1016/j.bioorg.2015.11.002 PMID: 26595185

[31] Mughal, E.U.; Sadiq, A.; Murtaza, S.; Rafique, H.; Zafar, M.N.; Riaz, T.; Khan, B.A.; Hameed, A.; Khan, K.M. Synthesis, structure-activity relationship and molecular docking of 3-oxoaurones and 3-thioaurones as acetylcholinesterase and butyrylcholinesterase inhibitors. *Bioorg. Med. Chem.,* **2017**, *25*(1), 100-106. http://dx.doi.org/10.1016/j.bmc.2016.10.016 PMID: 27780618

[32] Tang, H.; Wei, Y.B.; Zhang, C.; Ning, F.X.; Qiao, W.; Huang, S.L.; Ma, L.; Huang, Z.S.; Gu, L.Q. Synthesis, biological evaluation and molecular modeling of oxoisoaporphine and oxoaporphine derivatives as new dual inhibitors of acetylcholinesterase/butyrylcholinesterase. *Eur. J. Med. Chem.,* **2009**, *44*(6), 2523-2532. http://dx.doi.org/10.1016/j.ejmech.2009.01.021 PMID: 19243862

[33] Tasso, B.; Catto, M.; Nicolotti, O.; Novelli, F.; Tonelli, M.; Giangreco, I.; Pisani, L.; Sparatore, A.; Boido, V.; Carotti, A.; Sparatore, F. Quinolizidinyl derivatives of bi- and tricyclic systems as potent inhibitors of acetyl- and butyrylcholinesterase with potential in Alzheimer's disease. *Eur. J. Med. Chem.,* **2011**, *46*(6), 2170-2184. http://dx.doi.org/10.1016/j.ejmech.2011.02.071 PMID: 21459491

[34] da Costa, J.S.; Lopes, J.P.B.; Russowsky, D.; Petzhold, C.L.; Borges, A.C.; Ceschi, M.A.; Konrath, E.; Batassini, C.; Lunardi, P.S.; Gonçalves, C.A.S. Synthesis of tacrine-lophine hybrids *via* one-pot four component reaction and biological evaluation as acetyl- and butyrylcholinesterase inhibitors. *Eur. J. Med. Chem.,* **2013**, *62*, 556-563. http://dx.doi.org/10.1016/j.ejmech.2013.01.029 PMID: 23422935

[35] Jabeen, F.; Oliferenko, P.V.; Oliferenko, A.A.; Pillai, G.G.; Ansari, F.L.; Hall, C.D.; Katritzky, A.R. Dual inhibition of the α-glucosidase and butyrylcholinesterase studied by molecular field topology analysis. *Eur. J. Med. Chem.,* **2014**, *80*, 228-242. http://dx.doi.org/10.1016/j.ejmech.2014.04.018 PMID: 24780600

[36] Darvesh, S; Darvesh, KV; McDonald, RS; Mataija, D; Walsh, R; Mothana, S; Lockridge, O; Martin, E Carbamates with differential mechanism of inhibition toward acetylcholinesterase and butyrylcholinesterase. *J. med. chem,* **2008**, *24;51*(14), 4200-12. http://dx.doi.org/10.1021/jm8002075

[37] Yu, QS; Holloway, HW; Flippen-Anderson, JL; Hoffman, B Bross,i A; Greig, NH. . Methyl analogues of the experimental Alzheimer drug phenserine: synthesis and structure/activity relationships for acetyl-and butyrylcholinesterase inhibitory action. *J. Med. Chem.,* **2001**, *22;44*(24), 4062-71.

[38] Decker, M. Novel inhibitors of acetyl- and butyrylcholinesterase derived from the alkaloids dehydroevodiamine and rutaecarpine. *Eur. J. Med. Chem.,* **2005**, *40*(3), 305-313. http://dx.doi.org/10.1016/j.ejmech.2004.12.003 PMID: 15725500

[39] Orhan, I.E.; Senol, F.S.; Shekfeh, S.; Skalicka-Wozniak, K.; Banoglu, E. Pteryxin - A promising butyrylcholinesterase-inhibiting coumarin derivative from Mutellina purpurea. *Food Chem. Toxicol.,* **2017**, *109*(Pt 2), 970-974. http://dx.doi.org/10.1016/j.fct.2017.03.016 PMID: 28286309

[40] Zhang, N.; Casida, J.E. Novel irreversible butyrylcholinesterase inhibitors: 2-chloro-1-(substituted-phenyl)ethylphosphonic acids. *Bioorg. Med. Chem.,* **2002**, *10*(5), 1281-1290. http://dx.doi.org/10.1016/S0968-0896(01)00391-1 PMID: 11886791

[41] Bolea, I.; Juárez-Jiménez, J.; de Los Ríos, C.; Chioua, M.; Pouplana, R.; Luque, F.J.; Unzeta, M.; Marco-Contelles, J.; Samadi, A. Synthesis, biological evaluation, and molecular modeling of donepezil and N-[(5-(benzyloxy)-1-methyl-1H-indol-2-yl)methyl]-N-methylprop-2-yn-1-amine hybrids as new multipotent cholinesterase/monoamine oxidase inhibitors for the treatment of Alzheimer's disease. *J. Med. Chem.,* **2011**, *54*(24), 8251-8270. http://dx.doi.org/10.1021/jm200853t PMID: 22023459

[42] Savini, L.; Gaeta, A.; Fattorusso, C.; Catalanotti, B.; Campiani, G.; Chiasserini, L.; Pellerano, C.; Novellino, E.; McKissic, D.; Saxena, A. Specific targeting of acetylcholinesterase and butyrylcholinesterase recognition sites. Rational design of novel, selective, and highly potent cholinesterase inhibitors. *J. Med. Chem.,* **2003**, *46*(1), 1-4. http://dx.doi.org/10.1021/jm0255668 PMID: 12502352

[43] Lam, Y. *Berberine target key enzymes and amino acid inibitiors in AD treatment creation from berberine-based structure screening. IJPSR,* **2014**, *5*(07), 350-363.

[44] Chen, X.; Tikhonova, I.G.; Decker, M. Probing the mid-gorge of cholinesterases with spacer-modified bivalent quinazolinimines leads to highly potent and selective butyrylcholinesterase inhibitors. *Bioorg. Med. Chem.,* **2011**, *19*(3), 1222-1235. http://dx.doi.org/10.1016/j.bmc.2010.12.034 PMID: 21232964

[45] Maryamabadi, A.; Hasaninejad, A.; Nowrouzi, N.; Mohebbi, G.; Asghari, B. Application of PEG-400 as a green biodegradable polymeric medium for the catalyst-free synthesis of spiro-dihydropyridines and their use as acetyl and butyrylcholinesterase inhibitors. *Bioorg. Med. Chem.,* **2016**, *24*(6), 1408-1417. http://dx.doi.org/10.1016/j.bmc.2016.02.019 PMID: 26879857

[46] Nadri, H.; Pirali-Hamedani, M.; Moradi, A.; Sakhteman, A.; Vahidi, A.; Sheibani, V.; Asadipour, A.; Hosseinzadeh, N.; Abdollahi, M.; Shafiee, A.; Foroumadi, A. 5,6-Dimethoxybenzofuran-3-one derivatives: a novel series of dual Acetylcholinesterase/Butyrylcholinesterase inhibitors bearing benzyl pyridinium moiety. *Daru,* **2013**, *21*(1), 15. http://dx.doi.org/10.1186/2008-2231-21-15 PMID: 23445881

[47] Ilhami, G.; Malahat, A.; Taslimi, P.; Zubeyir, H.; Leyla, S.; Afsun, S.; Vagif, F.; Sukru, B.; Saleh, H, A.; Claudiu, T, S. *Synthesis and biological evaluation of aminomethyl and alkoxymethyl derivatives as carbonic anhydrase, acetylcholinesterase and butyrylcholinesterase inhibitors. J. Enzyme Inhib Med Chem.,* **2017**, *32*(1), 1174-1182.

[48] Sakkiah, S.; Lee, K.W. Pharmacophore-based virtual screening and density functional theory approach to identifying novel butyrylcholinesterase inhibitors. *Acta Pharmacol. Sin.,* **2012**, *33*(7), 964-978. http://dx.doi.org/10.1038/aps.2012.21 PMID: 22684028

[49] Zeb, A.; Hameed, A.; Khan, L.; Khan, I.; Dalvandi, K.; Choudhary, M.I.; Basha, F.Z. Quinoxaline derivatives: novel and selective butyrylcholinesterase inhibitors. *Med. Chem.,* **2014**, *10*(7), 724-729. http://dx.doi.org/10.2174/1573406410666140526145429 PMID: 24875826

[50] Chen, Y.; Lin, H.; Yang, H.; Tan, R.; Bian, Y.; Fu, T.; Li, W.; Wu, L.; Pei, Y.; Sun, H. Discovery of new acetylcholinesterase and butyrylcholinesterase inhibitors through structure-based virtual screening. *RSC Advances,* **2017**, *7*(6), 3429-3438. http://dx.doi.org/10.1039/C6RA25887E

[51] Guido, R.V.; Castilho, M.S.; Mota, S.G.; Oliva, G.; Andricopulo, A.D. Classical and Hologram QSAR Studies on a Series of Inhibitors of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase. *QSAR Comb. Sci.,* **2008**, *27*(6), 768-781. http://dx.doi.org/10.1002/qsar.200710139

[52] Cherif, O.; Allouche, F.; Chabchoub, F.; Chioua, M.; Soriano, E.; Yañez, M.; Cacabelos, R.; Romero, A.; López, M.G.; Marco-Contelles, J. Isoxazolotacrines as non-toxic and selective butyrylcholinesterase inhibitors for Alzheimer's disease. *Future Med. Chem.,* **2014**, *6*(17), 1883-1891. http://dx.doi.org/10.4155/fmc.14.115 PMID: 25495982

[53] Silva, D; Chioua, M; Samadi, A; Agostinho, P; Garção, P; Lajarín-Cuesta, R; de los Ríos, C; Iriepa, I; Moraleda, I; Gonzalez-Lafuente, L; Mendes, E Synthesis, pharmacological assessment, and molecular modeling of acetylcholinesterase/butyrylcholinesterase inhibitors: effect against amyloid-β-induced neurotoxicity. *ACS Chem. Neurosci. ,* **2013**, *17, 4*(4), 547-65.

[54] Decker, M.; Kraus, B.; Heilmann, J. Design, synthesis and pharmacological evaluation of hybrid molecules out of quinazolinimines and lipoic acid lead to highly potent and selective

butyrylcholinesterase inhibitors with antioxidant properties. *Bioorg. Med. Chem.,* **2008**, *16*(8), 4252-4261. http://dx.doi.org/10.1016/j.bmc.2008.02.083 PMID: 18343673

[55] Schott, Y.; Decker, M.; Rommelspacher, H.; Lehmann, J. 6-Hydroxy- and 6-methoxy-β-carbolines as acetyl- and butyrylcholinesterase inhibitors. *Bioorg. Med. Chem. Lett.,* **2006**, *16*(22), 5840-5843. http://dx.doi.org/10.1016/j.bmcl.2006.08.067 PMID: 16945529

[56] Dolles, D.; Nimczick, M.; Scheiner, M.; Ramler, J.; Stadtmüller, P.; Sawatzky, E.; Drakopoulos, A.; Sotriffer, C.; Wittmann, H.J.; Strasser, A.; Decker, M. Aminobenzimidazoles and structural isomers as templates for dual-acting butyrylcholinesterase inhibitors and hcb2 r ligands to combat neurodegenerative disorders. *ChemMedChem,* **2016**, *11*(12), 1270-1283. http://dx.doi.org/10.1002/cmdc.201500418 PMID: 26548365

[57] Granica, S.; Kiss, A.K.; Jarończyk, M.; Maurin, J.K.; Mazurek, A.P.; Czarnocki, Z. Synthesis of imperatorin analogs and their evaluation as acetylcholinesterase and butyrylcholinesterase inhibitors. *Arch. Pharm. (Weinheim),* **2013**, *346*(11), 775-782. http://dx.doi.org/10.1002/ardp.201300259 PMID: 24123207

[58] Woo, Y.J.; Lee, B.H.; Yeun, G.H.; Kim, H.J.; Won, M.H.; Kim, S.H.; Lee, B.H.; Park, J.H. Selective butyrylcholinesterase inhibitors using polyphenol-polyphenol hybrid molecules. *Bull. Korean Chem. Soc.,* **2011**, *32*(8), 2593-2598. http://dx.doi.org/10.5012/bkcs.2011.32.8.2593

[59] Decker, M. Homobivalent quinazolinimines as novel nanomolar inhibitors of cholinesterases with dirigible selectivity toward butyrylcholinesterase. *J. Med. Chem.,* **2006**, *49*(18), 5411-5413. http://dx.doi.org/10.1021/jm060682m PMID: 16942014

[60] Conejo-García, A.; Pisani, L.; Núñez, Mdel.C.; Catto, M.; Nicolotti, O.; Leonetti, F.; Campos, J.M.; Gallo, M.A.; Espinosa, A.; Carotti, A. Homodimeric bis-quaternary heterocyclic ammonium salts as potent acetyl- and butyrylcholinesterase inhibitors: a systematic investigation of the influence of linker and cationic heads over affinity and selectivity. *J. Med. Chem.,* **2011**, *54*(8), 2627-2645. http://dx.doi.org/10.1021/jm101299d PMID: 21417225

[61] Chen, Y.; Sun, J.; Fang, L.; Liu, M.; Peng, S.; Liao, H.; Lehmann, J.; Zhang, Y. Tacrine-ferulic acid-nitric oxide (NO) donor trihybrids as potent, multifunctional acetyl- and butyrylcholinesterase inhibitors. *J. Med. Chem.,* **2012**, *55*(9), 4309-4321. http://dx.doi.org/10.1021/jm300106z PMID: 22512543

[62] Rivera-Becerril, E.; Joseph-Nathan, P.; Pérez-Alvarez, V.M.; Morales-Ríos, M.S. Synthesis and biological evaluation of (-)- and (+)-debromoflustramine B and its analogues as selective butyrylcholinesterase inhibitors. *J. Med. Chem.,* **2008**, *51*(17), 5271-5284. http://dx.doi.org/10.1021/jm800277g PMID: 18686941

[63] Yu, Q.; Holloway, H.W.; Utsuki, T.; Brossi, A.; Greig, N.H. Synthesis of novel phenserine-based-selective inhibitors of butyrylcholinesterase for Alzheimer's disease. *J. Med. Chem.,* **1999**, *42*(10), 1855-1861. http://dx.doi.org/10.1021/jm980459s PMID: 10346939

[64] Makhaeva, G.F.; Lushchekina, S.V.; Boltneva, N.P.; Sokolov, V.B.; Grigoriev, V.V.; Serebryakova, O.G.; Vikhareva, E.A.; Aksinenko, A.Y.; Barreto, G.E.; Aliev, G.; Bachurin, S.O. Conjugates of γ-Carbolines and Phenothiazine as new selective inhibitors of butyrylcholinesterase and blockers of NMDA receptors for Alzheimer Disease. *Sci. Rep.,* **2015**, *5*, 13164. http://dx.doi.org/10.1038/srep13164 PMID: 26281952

[65] Jones, M.; Wang, J.; Harmon, S.; Kling, B.; Heilmann, J.; Gilmer, J.F. Novel selective butyrylcholinesterase inhibitors incorporating antioxidant functionalities as potential bimodal therapeutics for Alzheimer's disease. *Molecules,* **2016**, *21*(4), 440. http://dx.doi.org/10.3390/molecules21040440 PMID: 27534722

[66] Delogu, G.L.; Matos, M.J.; Fanti, M.; Era, B.; Medda, R.; Pieroni, E.; Fais, A.; Kumar, A.; Pintus, F. 2-Phenylbenzofuran derivatives as butyrylcholinesterase inhibitors: Synthesis, biological activity and molecular modeling. *Bioorg. Med. Chem. Lett.,* **2016**, *26*(9), 2308-2313. http://dx.doi.org/10.1016/j.bmcl.2016.03.039 PMID: 26995529

[67] Shan, W.J.; Huang, L.; Zhou, Q.; Meng, F.C.; Li, X.S. Synthesis, biological evaluation of 9-N-substituted berberine derivatives as multi-functional agents of antioxidant, inhibitors of acetylcholinesterase, butyrylcholinesterase and amyloid-β aggregation. *Eur. J. Med. Chem.,* **2011**, *46*(12), 5885-5893. http://dx.doi.org/10.1016/j.ejmech.2011.09.051 PMID: 22019228

[68] Liu, H.; Liu, L.; Gao, X.; Liu, Y.; Xu, W.; He, W.; Jiang, H.; Tang, J.; Fan, H.; Xia, X. Novel ferulic amide derivatives with tertiary amine side chain as acetylcholinesterase and butyrylcholinesterase inhibitors: The influence of carbon spacer length, alkylamine and aromatic group. *Eur. J. Med. Chem.,* **2017**, *126*, 810-822. http://dx.doi.org/10.1016/j.ejmech.2016.12.003 PMID: 27951489

[69] Sawatzky, E.; Wehle, S.; Kling, B.; Wendrich, J.; Bringmann, G.; Sotriffer, C.A.; Heilmann, J.; Decker, M. Discovery of highly selective and nanomolar carbamate-based butyrylcholinesterase inhibitors by rational investigation into their inhibition mode. *J. Med. Chem.,* **2016**, *59*(5), 2067-2082. http://dx.doi.org/10.1021/acs.jmedchem.5b01674 PMID: 26886849

[70] Khoobi, M.; Alipour, M.; Sakhteman, A.; Nadri, H.; Moradi, A.; Ghandi, M.; Emami, S.; Foroumadi, A.; Shafiee, A. Design, synthesis, biological evaluation and docking study of 5-oxo-4,5-dihydropyrano[3,2-c]chromene derivatives as acetylcholinesterase and butyrylcholinesterase inhibitors. *Eur. J. Med. Chem.,* **2013**, *68*, 260-269. http://dx.doi.org/10.1016/j.ejmech.2013.07.038 PMID: 23988409

[71] Zaheer-ul, H.; Uddin, R.; Yuan, H.; Petukhov, P.A.; Choudhary, M.I.; Madura, J.D. Receptor-based modeling and 3D-QSAR for a quantitative production of the butyrylcholinesterase inhibitors based on genetic algorithm. *J. Chem. Inf. Model.,* **2008**, *48*(5), 1092-1103. http://dx.doi.org/10.1021/ci8000056 PMID: 18444627

[72] Asadipour, A.; Alipour, M.; Jafari, M.; Khoobi, M.; Emami, S.; Nadri, H.; Sakhteman, A.; Moradi, A.; Sheibani, V.; Homayouni Moghadam, F.; Shafiee, A.; Foroumadi, A. Novel coumarin-3-carboxamides bearing N-benzylpiperidine moiety as potent acetylcholinesterase inhibitors. *Eur. J. Med. Chem.,* **2013**, *70*, 623-630. http://dx.doi.org/10.1016/j.ejmech.2013.10.024 PMID: 24211638

[73] Rahim, F.; Javed, M.T.; Ullah, H.; Wadood, A.; Taha, M.; Ashraf, M.; Qurat-ul-Ain, ; Khan, M.A.; Khan, F.; Mirza, S.; Khan, K.M. Synthesis, molecular docking, acetylcholinesterase and butyrylcholinesterase inhibitory potential of thiazole analogs as new inhibitors for Alzheimer disease. *Bioorg. Chem.,* **2015**, *62*, 106-116. http://dx.doi.org/10.1016/j.bioorg.2015.08.002 PMID: 26318401

[74] Bolognesi, M.L.; Bartolini, M.; Cavalli, A.; Andrisano, V.; Rosini, M.; Minarini, A.; Melchiorre, C. Design, synthesis, and biological evaluation of conformationally restricted rivastigmine analogues. *J. Med. Chem.,* **2004**, *47*(24), 5945-5952. http://dx.doi.org/10.1021/jm049782n PMID: 15537349

[75] Pang, X.; Fu, H.; Yang, S.; Wang, L.; Liu, A.L.; Wu, S.; Du, G.H. Evaluation of novel dual acetyl-and butyrylcholinesterase inhibitors as potential anti-Alzheimer's disease agents using pharmacophore, 3D-QSAR, and molecular docking approaches. *Molecules,* **2017**, *22*(8), 1254. http://dx.doi.org/10.3390/molecules22081254 PMID: 28933746

[76] Gogoi, D.; Chaliha, A.K.; Sarma, D.; Kakoti, B.B.; Buragohain, A.K. Novel butyrylcholinesterase inhibitors through pharmacophore modeling, virtual screening and DFT-based approaches along-with design of bioisosterism-based analogues. *Biomed. Pharmacother.,* **2017**, *85*, 646-657. http://dx.doi.org/10.1016/j.biopha.2016.11.076 PMID: 27903422

[77] CambridgeSoft. *Chemdraw ultra 12.0.*; Ohio: USA. Available on: https://chemistry.com.pk/software/free-download-chemdraw-ultra-12/

[78] ChemAxon. *Marvin*; Budapest: Hungary. Available on: https://chemaxon.com/products/marvin

[79] Mauri, A.; Consonni, V.; Pavan, M. Dragon software: An easy approach to molecular descriptor calculations. *Match (Mulh.),* **2006**, *56*(2), 237-248.

[80] Yap, C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.,* **2011**, *32*(7), 1466-1474. http://dx.doi.org/10.1002/jcc.21707 PMID: 21425294

[81] Roy, K.; Das, R.N.; Ambure, P.; Aher, R.B. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom. Intell. Lab. Syst.,* **2016**, *152*, 18-33. http://dx.doi.org/10.1016/j.chemolab.2016.01.008

[82] Khan, K.; Benfenati, E.; Roy, K. Consensus QSAR modeling of toxicity of pharmaceuticals to different aquatic organisms: Ranking and prioritization of the DrugBank database compounds. *Ecotoxicol. Environ. Saf.,* **2019**, *168*(168), 287-297. http://dx.doi.org/10.1016/j.ecoenv.2018.10.060 PMID: 30390527

[83] Das, S.; Ojha, P.K.; Roy, K. Multilayered variable selection in QSPR: a case study of modeling melting point of bromide ionic liquids. *IJQSPR,* **2017**, *2*(1), 106-124. http://dx.doi.org/10.4018/IJQSPR.2017010108

[84] Das, S.; Ojha, P.K.; Roy, K. Development of a temperature dependent 2D-QSPR model for viscosity of diverse functional ionic liquids. *J. Mol. Liq.,* **2017**, *240*, 454-467. http://dx.doi.org/10.1016/j.molliq.2017.05.113

[85] Ojha, P.K.; Roy, K. Chemometric modeling of odor threshold property of diverse aroma components of wine. *RSC adv,* **2018**, *8*(9), 4750-4760.

[86] Ojha, P.K.; Roy, K. PLS regression-based chemometric modeling of odorant properties of diverse chemical constituents of black tea and coffee. *RSC adv,* **2018**, *8*(5), 2293-2304.

[87] *Minitab*; . Pennsylvania: USA. Available on: http://www.minitab.com/en-US/default.aspx

[88] *SIMCA-P*; Umeå: Sweden. Available on: https://umetrics.com/kb/getting-started-simca-p

[89] Rücker, C.; Rücker, G.; Meringer, M. y-Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.,* **2007**, *47*(6), 2345-2357. http://dx.doi.org/10.1021/ci700157b PMID: 17880194

[90] Veerasamy, R.; Harish, R.; Abhishek, J.; Shalini, S. Varghese Christapher P and Agrawal R. K. Validation of QSAR models-strategies and importance. *RRJoDDD,* **2011**, *2*(3), 511-519.

[91] Khan, K.; Khan, P.M.; Lavado, G.; Valsecchi, C.; Pasqualini, J.; Baderna, D.; Marzo, M.; Lombardo, A.; Roy, K.; Benfenati, E. QSAR modeling of Daphnia magna and fish toxicities of biocides using 2D descriptors. *Chemosphere,* **2019**, *229*(229), 8-17. http://dx.doi.org/10.1016/j.chemosphere.2019.04.204 PMID: 31063877

[92] de Ruyck, J.; Brysbaert, G.; Blossey, R.; Lensink, M.F. Molecular docking as a popular tool in drug design, an in silico travel. *Adv. Appl. Bioinform. Chem.,* **2016**, *9*, 1-11. http://dx.doi.org/10.2147/AABC.S105289 PMID: 27390530

[93] Dighe, S.N.; de la Mora, E.; Chan, S.; Kantham, S.; McColl, M.; Veliyath, S.K.; Miles, J.A.; Nessar, Z.; McGeary, R.P.; Silman, I.; Parat, M.O.; Weik, M.; Brazzolotto, X.; Ros, B.P. Rivastigmine and metabolite analogues with putative Alzheimer's disease-modifying properties in a Caenorhabditis elegans model. *Commun. Chem.*, **2019**. Available from: https://www.rcsb.org/structure/6EZ2

[94] Dighe, S.N.; de la Mora, E.; Chan, S.; Kantham, S.; McColl, M.; Veliyath, S.K.; Miles, J.A.; Nessar, Z.; McGeary, R.P.; Silman, I.; Parat, M.O.; Weik, M.; Brazzolotto, X.; Ros, B.P. Rivastigmine and metabolite analogues with putative Alzheimer's disease-modifying properties in a Caenorhabditis elegans model. *Commun. Chem.*, **2019**. Available from: https://3dsbiovia.com/resource-center/downloads/

[95] Popelier, P.L.A.; Smith, P.J. QSAR models based on quantum topological molecular similarity. *Eur. J. Med. Chem.,* **2006**, *41*(7), 862-873. http://dx.doi.org/10.1016/j.ejmech.2006.03.004 PMID: 16697489

[96] Ivanciuc, O. Balaban; Teodor-Silviu, and Balaban, Alexandru T. Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices. *J. Math. Chem.,* **1993**, *1*(12), 309-318. http://dx.doi.org/10.1007/BF01164642

[97] National Center for Biotechnology Information. PubChem Database. Pyridine, CID=1049, Available from: https://pubchem.ncbi.nlm.nih.gov/compound/pyridine (Accessed on June 9, **2020**)

[98] Yu, Qian-sheng.; Holloway, Harold, W.; Utsuki, Tadanobu. Brossi, Arnold and Greig, Nigel H. Synthesis of novel phenserine-based-selective inhibitors of butyrylcholinesterase for Alzheimer's disease. *J. med. Chem,* **1993**, *42, 10*(42), 1855-1861.

SAR and QSAR in Environmental Research

ISSN: 1062-936X (Print) 1029-046X (Online) Journal homepage: https://www.tandfonline.com/loi/gsar20

Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease

V. Kumar, P.K. Ojha, A. Saha & K. Roy

To cite this article: V. Kumar, P.K. Ojha, A. Saha & K. Roy (2020) Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease, SAR and QSAR in Environmental Research, 31:2, 87-133

To link to this article: https://doi.org/10.1080/1062936X.2019.1695226

SAR AND QSAR IN ENVIRONMENTAL RESEARCH
2020, VOL. 31, NO. 2, 87–133
https://doi.org/10.1080/1062936X.2019.1695226

Taylor & Francis
Taylor & Francis Group

Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease

V. Kumar[a], P.K. Ojha[a], A. Saha[b] and K. Roy[a]

[a]Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India; [b]Department of Chemical Technology, University of Calcutta, Kolkata, India

ABSTRACT

We have developed a robust quantitative structure–activity relationship (QSAR) model employing a dataset of 98 heterocycle compounds to identify structural features responsible for BACE1 (beta-secretase 1) enzyme inhibition. We have used only 2D descriptors for model development purpose thus avoiding the conformational complications arising due to 3D geometry considerations. Following the strict Organization for Economic Co-operation and Development (OECD) guidelines, we have developed models using stepwise regression analysis followed by the best subset selection, while the final model was developed by partial least squares regression technique. The model was validated using various internationally accepted stringent validation parameters. From the insights obtained from the developed model, we have concluded that heteroatoms (nitrogen, oxygen, etc.) present within to an aromatic nucleus and the structural features such as hydrophobic, ring aromatic and hydrogen bond acceptor/donor are responsible for the enhancement of the BACE1 enzyme inhibitory activity. Moreover, we have performed the pharmacophore modelling to unveil the structural requirements for the inhibitory activity against the BACE1 enzyme. Furthermore, molecular docking studies were carried out to understand the molecular interactions involved in binding, and the results are then correlated with the requisite structural features obtained from the QSAR and pharmacophore models.

ARTICLE HISTORY
Received 25 September 2019
Accepted 17 November 2019

KEYWORDS
BACE1; Alzheimer's disease; QSAR; validation; OECD; Aβ; DModX

Introduction

Alzheimer's disease (AD) is an irreversible and progressive neurodegenerative disorder characterized by the presence of two abnormal peptides, amyloid-β (Aβ) and τ-proteins [1,2]. The early signs of the disease may be a continuous decline in loss of short-term memory and intellectual functions, frequently accompanied by strange behaviour such as aggression and depression [3]. Several hypotheses have been proposed to explain AD pathogenesis such as cholinergic hypothesis, tau hypothesis and amyloid hypothesis that only describe the basic cause of disease progression [4]. However, these hypotheses are still not sufficient to explain the root of cause of AD; they seem to be helpful to develop possible treatment strategies that slowdown the progression of AD [5]. Among these hypotheses, the most prominent is the amyloid hypothesis, which states that AD is caused

CONTACT K. Roy kunalroy_in@yahoo.com
Supplemental data for this article can be accessed at: https://doi.org/10.1080/1062936X.2019.1695226.

by the accumulation of Aβ in the brain [4]. It is reported previously in various literatures that several forms of Aβ are formed from the amyloid precursor protein (APP) [4]. However, it seems that Aβ42 (10% of all Aβs produced) is the major form of pathogenic and the most important component of amyloid plaques [4]. According to the amyloid hypothesis, Aβ accumulation is a basic cause of the AD [6]. In normal brain physiology, APP is cleaved by β, γ and α secretase enzyme yielding 40 soluble amino acid peptides [7]. But, in case of AD, a two-step proteolytic process is initiated by the Swedish double mutation at the BACE1 (β-secretase) followed by γ-secretase (catalytic subunits presenilin 1 and 2) yielding a 42 insoluble amino acid peptide called amyloid-β (Aβ), and consequently forming β-amyloid plaque [7]. β-secretase (BACE1) is the first protein that acts on amyloid precursor protein (APP) in the production of amyloid-β (Aβ) [8]. Due to its evident rate limiting function, BACE1 seems to be a prime target to prevent Aβ generation in AD [9]. The BACE1 enzyme has long been observed as an important therapeutic target for AD in the development of inhibitor drugs for reduction of Aβ [10]. The cloning and identification of β-secretase were first reported in 1999 which energized research on both the protease and its inhibitor drugs. Presently, β-secretase is a major drug target for AD, and the development of its inhibitor drugs is being pursued in many research laboratories around the world [11]. Heparan sulphite and its derivatives were reported as BACE1 enzyme inhibitors [12]. Furthermore, other oligosaccharides and their analogues have also been reported to inhibit BACE1 enzyme and reduce the Aβ deposition [13]. Enoxaparin has been reported to lower Aβ plaque deposition and recover cognitive function in AD transgenic mice [13].

Computational methods like quantitative structure–activity relationships (QSARs) have been successfully used to identify the essential structural features for their selective inhibitory activity. The QSAR is a methodology to find a consistent relationship between the biological activity of a compound and its structural arrangements and chemical property [14]. There are different regression and pattern recognition techniques which can be used for the variable selection and QSAR model development [14]. The developed methods for the QSAR are used particularly in chemoinformatics, drug discovery and to evaluate the biological activity of new chemical compounds, apart from toxicological and ecotoxicological evaluations of specific chemicals within the meaning of risk management [14,15]. There are number of computational studies reported for the identification of novel compounds for the treatment of AD, some of them we have discussed below but still, we do not have accurate treatment therapy for AD [16]. Ambure et al. [17] developed QSAR models using 74 compounds for the inhibition of beta-secretase 1 enzyme. Jain et al. [18] reported a QSAR model (multi-linear regression (MLR)) using a set of 27 aminoimidazoles derivatives as BACE1 inhibitors with defined inhibitory activity. Hossain et al. [19] attempted to develop 2D-QSAR and 3D-QSAR models along with molecular docking and pharmacophore modelling using 106 compounds with defined inhibitory activity against beta-secretase 1 enzyme. Chakraborty et al. [20] performed QSAR studies using linear heuristic method and developed a model using 30 compounds with defined BACE1 enzyme inhibitory activity. In the current study, we have utilized a dataset of 98 compounds (downloaded from BindingDB data base) [21] with BACE1 enzyme inhibitory activity for QSAR model development, using simple meaningful and easily interpretable 2D descriptors. The developed model is aimed at providing statistically robust predictions for the BACE1 enzyme inhibitory activity of the compounds, expressed as the negative log of half maximal inhibitory concentration (pIC_{50}).

The feature selection was done using a stepwise regression strategy followed by best subset (BSS) selection, and the final model was developed using Partial Least Squares (PLS) regression in order to obviate the effect of intercorrelation among the descriptors. The variable selection approach helps in reducing the noise in the input [22]. The QSAR model was built obeying the guidelines of the OECD principles [23]. The developed model has been validated by taking into consideration of various strict internal and external validation metrics [23]. Additionally, we have performed pharmacophore modelling to reveal the structural requirements for the inhibitory activity and to categorize the compounds into more active and less active classes against the BACE1 enzyme. Furthermore, we have performed molecular docking study with the most active and least active compounds from the whole dataset and tried to justify the contributions of different descriptors/features as evident in the QSAR/pharmacophore model.

Materials and methods

QSAR methodology

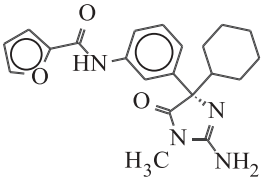
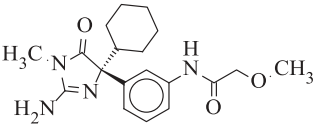
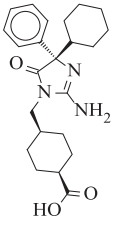
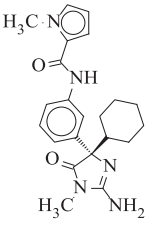
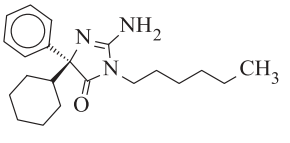
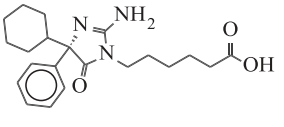
The dataset

In this study, we have collected 98 heterocyclic compounds (BACE1 enzyme inhibitors) (Table 1) from the BindingDB database [21] (see sheet 2 in Supplementary materials S1) for development of the QSAR model. The experimental IC_{50} values (nM) of the dataset compounds were converted in to pIC_{50} values for model development purpose. The reported assay (FRET bioassay) procedure of all the compounds used in this study followed the same protocol [24–30]. All the structures were drawn using Marvin ChemAxon tool [31] followed by cleaning of molecules, and finally, saved as MDL .mol format. All the compound structures were properly checked prior to the calculation of descriptors. The main purpose of this study was to identify the structural requirements which are essential for BACE1 enzyme inhibitory activity and to predict the activity of unknown compounds against the BACE1 enzyme.

Preliminary dataset preparation and data curation

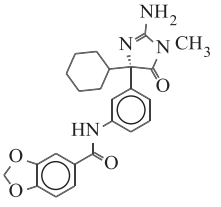
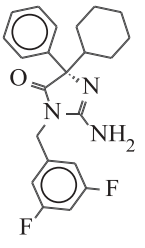
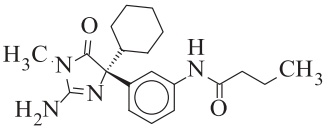
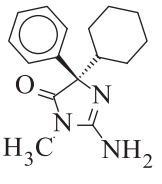
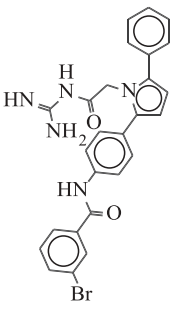
Prior to the development of QSAR models, we have performed dataset preparation and data curation (chemical and biological) steps. The dataset which we have downloaded in a structural data format (SDF) had both structural and biological activity information. The identifiers were given to all compounds present in the dataset that carry the following information, i.e., name of respective enzyme and a serial number. We have extracted the activity values from the dataset file to classify the compounds in four orders of magnitude. At last, we have submitted the dataset to chemical and biological curation. In this study, we have employed the most common steps (as detailed below) to perform the chemical curation that have been implemented in the in-house designed Konstanz Information Miner (KNIME) workflow (available at https://dtclab.webs.com/software-tools).

Table 1. The list of molecules present in the dataset with their names, structures and activity against BACE1 enzyme.

		Activity (pIC_{50})	
Nb.	*Molecular structure*	*Experimental*	*Predicted (2D QSAR analysis)*
1		−1.301	−1.729
2		−1.602	−2.877
3		−2.079	−3.193
4		−2.255	−2.169
5		−2.431	−3.682
6		−2.491	−3.041

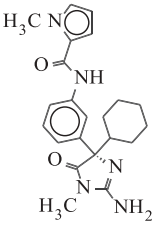
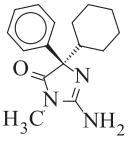
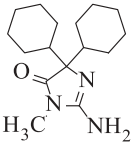
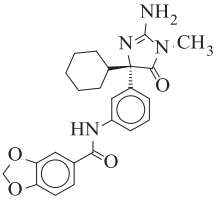
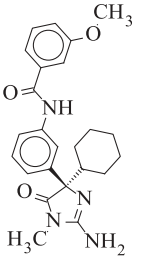
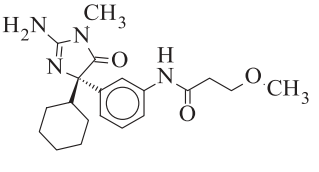
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
7		−2.579	−3.034
8		−2.612	−2.853
9		−2.690	−2.455
10		−2.707	−3.428
11		−2.778	−2.566

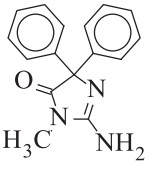
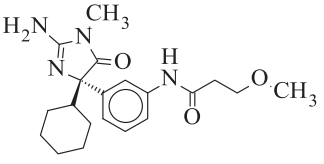
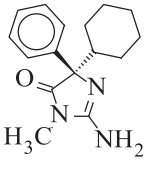
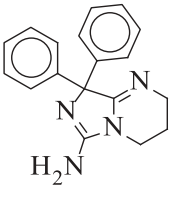
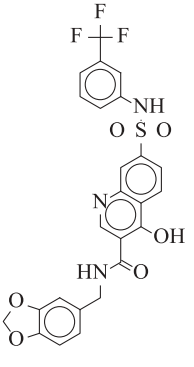
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	Predicted (2D QSAR analysis)
12		−2.934	−2.169
13		−2.995	−3.428
14		−3.033	−2.894
15		−3.060	−3.034
16		−3.133	−2.923
17		−3.267	−2.967

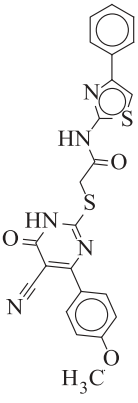
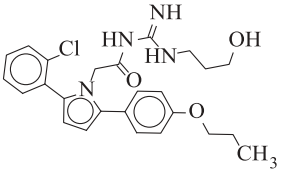
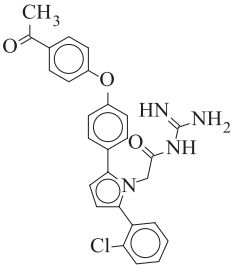
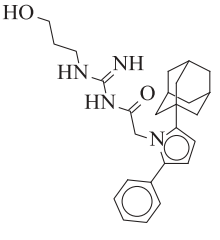
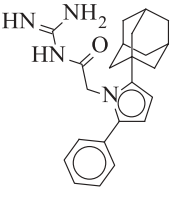
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
18		−3.531	−3.962
19		−3.895	−2.967
20		−4.380	−3.428
21		−4.579	−4.286
22		−2.491	−2.149

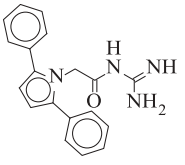
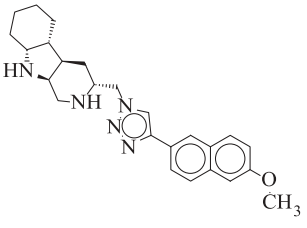
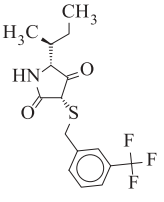
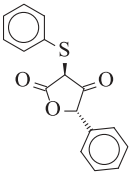
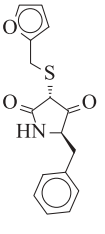
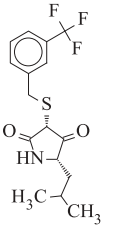
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
23		−3.677	−4.303
24		−2.041	−2.915
25		−2.380	−2.841
26		−3.505	−3.595
27		−3.770	−2.975

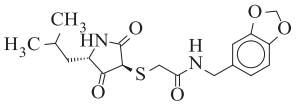
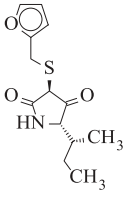
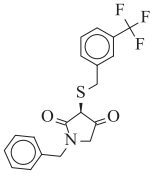
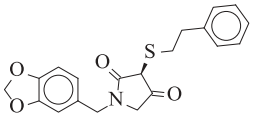
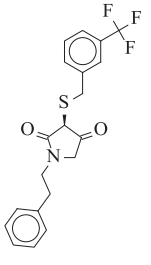
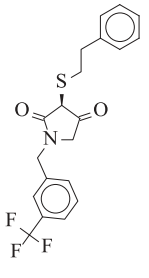
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
28		−4.287	−3.118
29		−3.173	−3.303
30		−4.778	−5.935
31		−5.021	−4.964
32		−5.161	−5.332
33		−5.287	−5.935

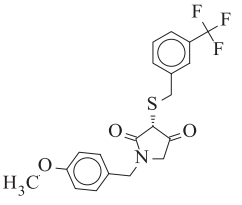
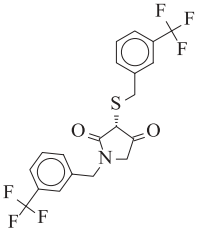
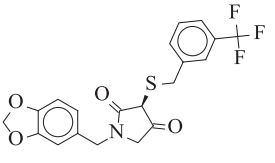
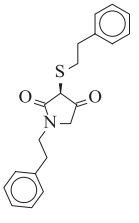
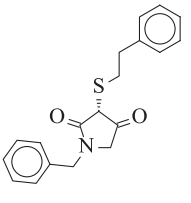
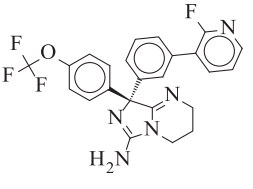
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
34		−5.301	−5.480
35		−5.301	−4.609
36		−5.326	−7.769
37		−8.143	−7.662
38		−8.264	−7.769
39		−8.326	−7.769

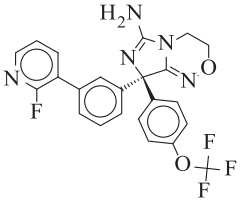
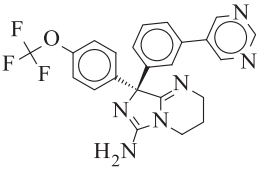
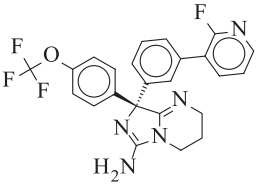
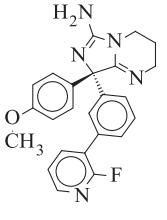
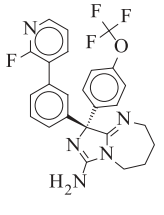
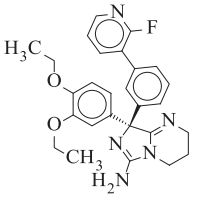
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
40		−8.382	−7.429
41		−8.411	−7.662
42		−8.423	−7.555
43		−8.627	−7.876
44		−8.661	−7.876
45		−1.477	−2.803

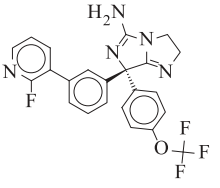
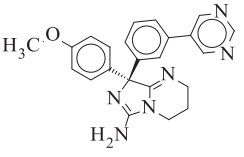
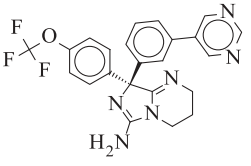
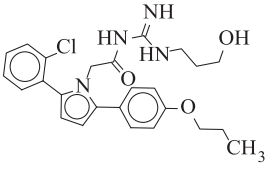
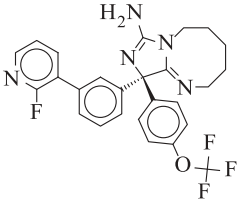
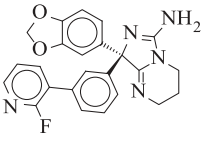
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	Predicted (2D QSAR analysis)
46		−1.778	−2.736
47		−1.903	−3.712
48		−1.903	−2.803
49		−2	−3.246
50		−2.113	−2.869
51		−2.278	−2.850

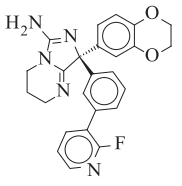
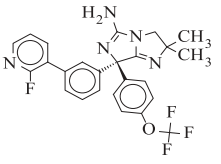
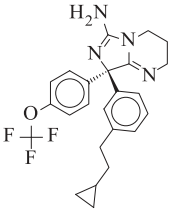
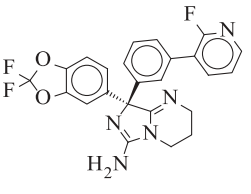
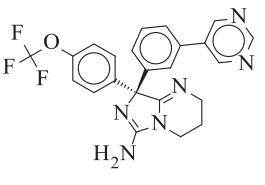
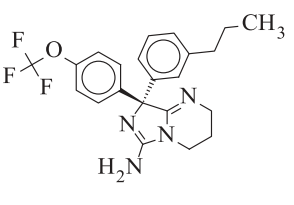
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
52		−2.897	−2.733
53		−3.309	−4.125
54		−3.380	−3.712
55		−3.577	−2.915
56		−3.819	−2.930
57		−3.845	−3.357

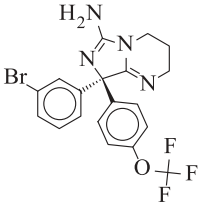
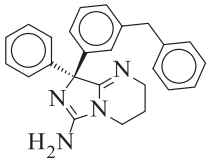
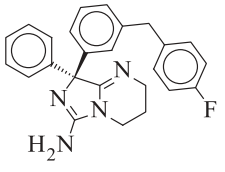
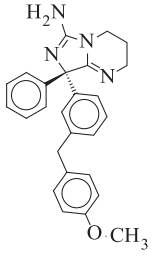
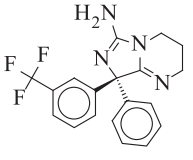
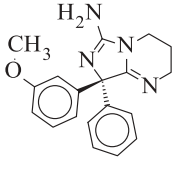
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
58		−3.982	−3.845
59		−4.029	−2.655
60		−4.037	−3.716
61		−4.086	−3.280
62		−4.215	−3.712
63		−4.238	−3.611

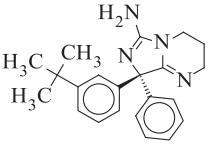
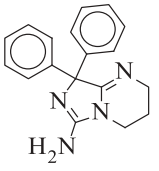
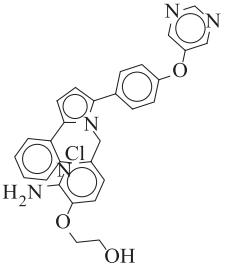
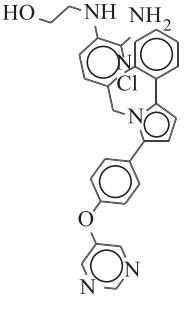
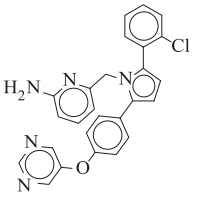
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
64		−4.318	−2.929
65		−4.387	−5.081
66		−4.480	−4.374
67		−4.579	−4.760
68		−4.583	−4.078
69		−4.755	−3.985

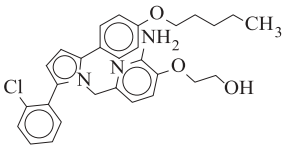
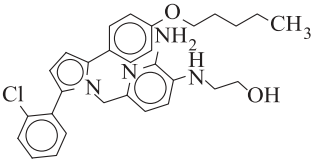
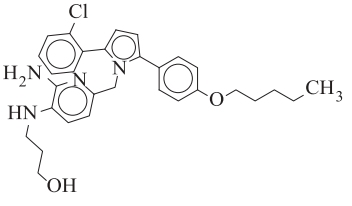
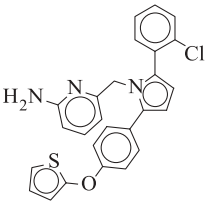
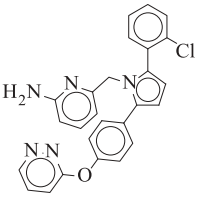
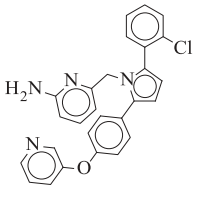
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
70		−4.895	−4.415
71		−5.176	−4.286
72		−1.602	−2.033
73		−1.845	−2.068
74		−2	−2.636

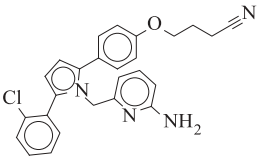
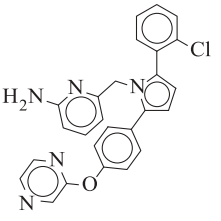
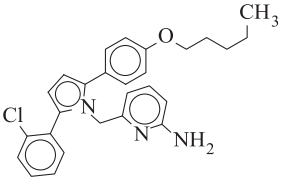
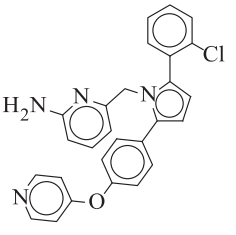
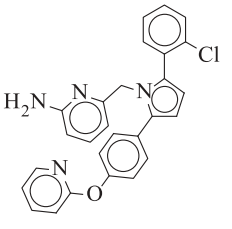
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
75		−2.041	−2.266
76		−2.230	−1.896
77		−2.342	−1.962
78		−2.462	−1.847
79		−2.612	−2.850
80		−2.612	−2.440

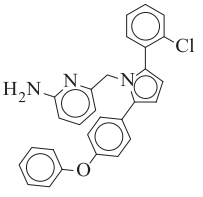
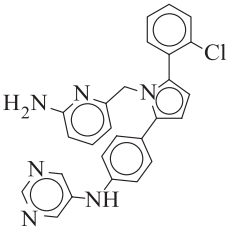
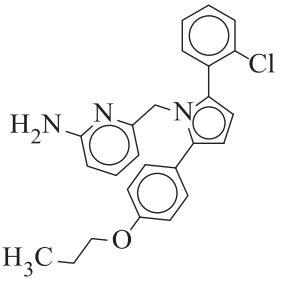
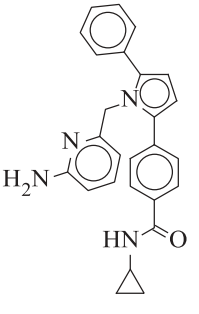
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	*Predicted (2D QSAR analysis)*
81		−2.623	−2.835
82		−2.672	−1.743
83		−2.740	−2.874
84		−2.770	−2.440
85		−2.770	−2.546

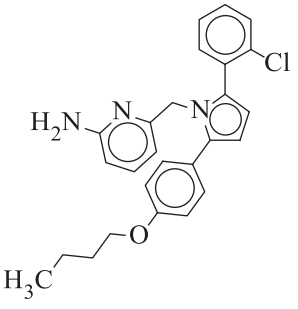
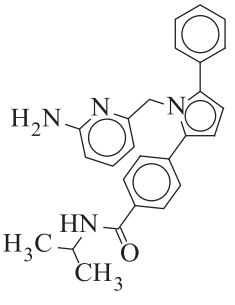
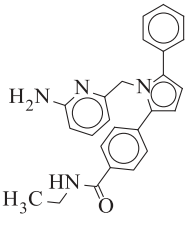
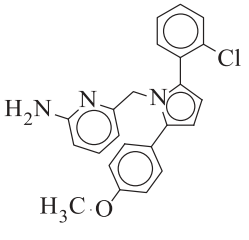
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	Predicted (2D QSAR analysis)
86		−2.908	−3.244
87		−2.991	−2.222
88		−3	−3.272
89		−3.060	−3.501

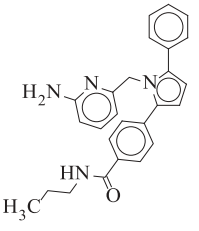
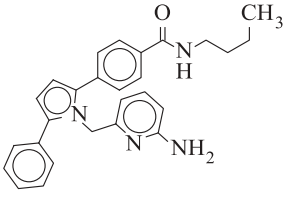
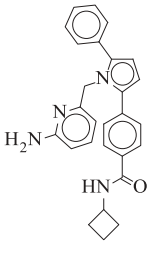
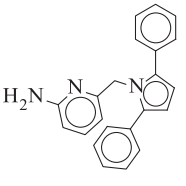
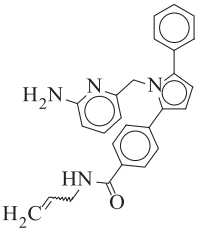
(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	Predicted (2D QSAR analysis)
90		−3.079	−2.821
91		−3.217	−3.501
92		−3.250	−4.008
93		−3.352	−2.743

(*Continued*)

Table 1. (Continued).

Nb.	Molecular structure	Activity (pIC_{50})	
		Experimental	Predicted (2D QSAR analysis)
94		−3.389	−3.608
95		−3.477	−3.673
96		−3.477	−3.567
97		−3.531	−3.738
98		−3.851	−3.929

Reading and storing the information present in the downloaded SDF file from BindingDB database

The dataset file (SDF) contains all the structural and topological information like molecule name, coordinates, bond counts, etc. Along with this necessary information, it also contains the biological information. For chemical data curation, structural information is enough to identify the correct chemical structure and remove the duplicates. Later, we have also performed the biological data curation to store other information related to the biological property, which is necessary for correct biological data curation. In the KNIME workflow, we have used 'SDF Reader' node for reading the input SDF file and storing the structural information. We have discarded the molecules with incorrect and incomplete information, and the molecules which have correct structural and biological information were stored as the revised SDF file for further use.

Removal of salts, mixtures, inorganics and organo-metallics

The molecular descriptors are generally computed for organic compounds and thus the majority of software tools can only consider organic compounds. Thus, the presence of salts, mixtures, inorganics and organo-metallics in the dataset may lead to incorrect descriptor values or such compounds are simply rejected by the software. So, in the present study, we have removed all the salts, mixtures, in-organics and organo-metallics before calculating the descriptors using 'RDKit Salt Stripper', 'Connectivity' node and 'Element Filter' nodes [32], respectively.

Normalization of chemical structures

There is a possibility of demonstrating the same functional group using different structural forms in dataset. Therefore, the normalization of chemical structure is necessary to remove different structural patterns. For example, nitro groups can be represented using two double bonds between nitrogen and oxygens, or one single bond linking the nitrogen and the protonated oxygen, or linking both nitrogen and oxygen atoms that are oppositely charged [32]. Different representations of the same chemical structure may create serious problems in a QSAR study because molecular descriptors calculated for these different representations of the same functional group could be significantly different. Thus, the transformation of all such functional groups to the standard forms is highly essential. In this study, 'RDKit Structure Normalizer' node was used for standardizing the chemical structures.

Biological curation

After chemical curation, we have employed biological curation of the screened compounds, and it includes two important steps, i.e., duplicate analysis and activity cliff analysis. In case of biological data curation, we have again employed the KNIME workflow (available at http://dtclab.webs.com/software-tools) to perform duplicate identification and activity cliff determination. First, we have simply performed the duplicate analysis based on the BindingDB Monomer ID and using a distance similarity index [32] (calculated using 3D

D-Similarity node available in KNIME), where the two compounds were considered identical or duplicates, only if both the BindingDB Monomer ID and distance similarity values are identical for both the compounds. Although the BindingDB Monomer ID might have been sufficient to identify duplicates, but to confirm that there is no error in the BindingDB Monomer ID itself, we have computed a distance similarity index using KNIME (available at https://dtclab.webs.com/software-tools). The list of molecules present in the dataset with their names, structures and activity against BACE1 enzyme is depicted in Table 1.

Descriptor calculation and data pretreatment

The molecular descriptors were calculated using two software tools, namely, Dragon software version 7 [33] (covering constitutional, ring descriptors, connectivity index, functional group counts, atom-centred fragments, 2D atom pairs, atom type E-states and molecular properties) and PaDEL-descriptor 2.20 software [34] (for extended topochemical atom or ETA indices). Molecular descriptors may be defined as the way of mathematical representation of a molecule by values associated with the chemical constitution for correlation of chemical structures with various chemical reactivity, biological activity or physical property [35]. After calculation of descriptors, we have implemented data pretreatment using the tool Pretreatment V-WSP version 1.2 (available at http://dtclab.webs.com/software-tools) to remove descriptors with at least one missing value, variables with constant or near constant values (standard deviation less than 0.0001), descriptors with all missing values and descriptors with (absolute) pair correlation larger than or equal to 0.95 from the initial pool of descriptors [35].

Dataset division

The whole dataset was divided into training and test sets based on k-Medoids clustering technique. Our aim was to develop a QSAR model which is statistically robust and capable of making accurate and reliable predictions. For this, we have employed a software tool 'Modified k-Medoids' (version 1.2) developed in our laboratory (available at http://dtclab.webs.com/software-tools). The clustering method classifies a set of compounds into clusters so that compounds belonging to the same cluster are similar to each other, while when two compounds belonging to two different clusters are expected to be dissimilar in nature [36]. The representative compounds within a cluster are called medoids. This technique tends to select k from most middle objects or compounds as the initial medoid. After clustering, we have sorted the whole dataset according to the cluster number followed by activity values. We have selected around 22% of compounds from each cluster as test set compounds ($n_{test} = 22$) and the remaining 78% as a training set ($n_{train} = 76$) compounds. Consequently, the developed QSAR model was validated using new chemical entities, i.e., the test set to check the predictive ability of the developed model.

Variable selection and QSAR model development

In this study, we have developed a QSAR model for inhibitory activity against BACE1 enzyme using pIC_{50} values as the response variable. The selection of important and meaningful descriptors from a large descriptors pool is a crucial step in QSAR model development. Thus, we have employed a variable selection strategy prior to the development of the final model using stepwise regression

(using a suitable stepping criterion, e.g., '*F*-for-inclusion 'and '*F*-for-exclusion' based on partial *F*-statistic) followed by the best subset selection. For this purpose, we have run stepwise regression using the whole pool of descriptors (393) to develop a stepwise regression model. In this case, stepwise regression has been applied using the initial pool of 393 descriptors, and the selected model descriptors were removed from the initial pool of descriptors and kept aside. Further, stepwise regression was run using the remaining pool of descriptors, and so on. Finally, we have clubbed the selected model descriptors in different cycles. In this way, we have reduced the initial pool of descriptors from 393 to 60 descriptors. Using these 60 descriptors, we ran the best subset selection tool v2.1 developed in our laboratory (available at http://dtclab.webs.com/software-tools) to generate a five descriptor model. Best subset regression is an investigative model building regression analysis technique. This technique compares all possible models using a specified set of predictors and displays the best-fitting models that contain one predictor, two predictors, and so on. The end result is a number of models and their summary statistics. It is up to us to compare them and choose one. Sometimes the results do not point to one best model and so our judgement is required to choose the best significant model. When selecting the best subset model, we are looking for the highest r^2, Q^2, r^2_{pred} and lowest Mean Absolute Error (MAE). Among the equations generated from the best subset selection, we selected one best model, based on the highest r^2, Q^2, r^2_{pred} and lowest MAE criteria. Among these models, we have selected one model based on MAE [37]. The final model was developed by employing PLS-regression methodology to avoid intercorrelation among the modelled descriptors using the Minitab software [38]. Information about the original variables is stored in latent variables (LV) generated by PLS. The final PLS model was developed with five selected descriptors using three latent variables (LV).

Statistical validation metrics

We have validated the developed PLS model using both internal and external validation parameters for measurement of the fitness, stability, robustness and predictivity. Various internal statistical parameters like determination coefficient (r^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$) and some r_m^2 metrics like average $r_m^2{}_{(LOO)}$ and $\Delta r_m^2{}_{(LOO)}$ were used to measure the robustness of the model. The determination coefficient (r^2) represents how much variability of a factor can be caused or explained by its relationship to another factor [39]. The higher value of this parameter indicated a better fit of the model. But these parameters are not good enough to evaluate the robustness and predictivity of a significant model. Thus, we have employed some other statistical validation parameters (external prediction) such as r^2_{Pred} or Q^2 F_1, Q^2F_2 to assure the significance of the developed model. Besides these parameters, r_m^2 metrics like average r_m^2 and Δr_m^2 and concordance correlation coefficient (CCC) were also calculated for external validation. The basic application of a predictive QSAR model is to judge the prediction errors for an external set, which should be within the chemical and response-based domain of the internal set (i.e., training set). The Q^2_{ext}-based metrics (i.e., r^2_{pred} and Q^2_{F2}) are not always able to provide a correct indication of the prediction quality because of the influence of the response range as well as the distribution of the values of response in both the training and test sets compounds. Thus, we have also validated the model using the mean absolute error (MAE) based criteria for both external and internal validation tests. The error-based metrics were used to determine the true indication of the prediction quality in terms of prediction error since they do not evaluate the performance of the model in comparison with the mean response [39]. Additionally, we have performed Y-randomization test and DModX approach (applicability

domain criteria) using the Simca-P software [40]. The Y-randomization test was performed to check whether the model was obtained by any chance or not, and the DModX approach was performed to check the whether the test set compounds lie within the applicability domain or outside the applicability domain of training set compounds. The Y-randomization test was performed using the Simca-P software [40] through randomly reordering (100 permutations) the dependent variable values. The validation parameter of the model obtained under such conditions should be of poor quality as compared to the selected model. The value of the r^2_{yrand} intercept should not exceed 0.3 and the value of the Q^2_{yrand} intercept should not exceed 0.05. The detailed methodology of work for the development of the 2D-QSAR model for BACE1 enzyme inhibitors is shown in Figure 1.

3D-QSAR pharmacophore modelling

In current study, we have performed 3D-QSAR pharmacophore modelling for the identification of essential pharmacophoric features, which are necessary for the inhibitory activity against BACE1 enzyme. The inhibitory activity against BACE1 enzyme expressed in term of IC_{50} values was used as the dependent variable for the pharmacophore model development. For the purpose of development of pharmacophore model, we have utilized the compounds (training set) which were selected as the test set compounds in case of 2D QSAR model whereas the training compounds in the 2D QSAR model were used as test compounds here [41]. The development of 3D QSAR pharmacophore model was carried out using the training set compounds and validated using the test set compounds. Prior to the development of pharmacophore models, we have performed conformation generation using training set compounds. After conformation generation, we have performed features mapping protocol for identifying the meaningful pharmacophoric features from the training set compounds using the BIOVIA Discovery Studio client 4.1 [42] platform and it resulted in hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrophobic (HYD), hydrophobic (aromatic and aliphatic) and ring aromatic (RA) features. In the development of pharmacophore model, different parameters were adjusted such as activity uncertainty value was kept 2, maximum five features containing hydrophobic (HYD), hydrophobic aliphatic, aromatic and ring aromatic (RA) and hydrogen bond acceptors (HBA) were selected and the final models were developed using FAST method of poling algorithm [41].

Validation of developed pharmacophore model

For the purpose of validation of the best pharmacophore model, we have used the different validation parameters such as quantitative and qualitative methods. In terms of test analysis, validation of the developed models was performed by mapping the whole test set molecules on the pharmacophore model. It was carried out in the BIOVIA Discovery Studio client 4.1 [42] platform with the same setting as we have used in pharmacophore model development. The predictive ability of a model to categorize both active and less active compounds has been determined by organizing the molecules with an activity threshold of 1000 nM. In the whole dataset, the training set (n = 22) consists of 12 most active and 10 least active compounds, whereas the test set (n = 76) consists of 32 most active and 44 least active compounds. To judge the quality of the developed pharmacophore model, we have performed the qualitative validation test calculating confusion matrix (validation parameters, namely, true positives, true negatives, false positives, and false negatives) based on the observed and predicted activity values

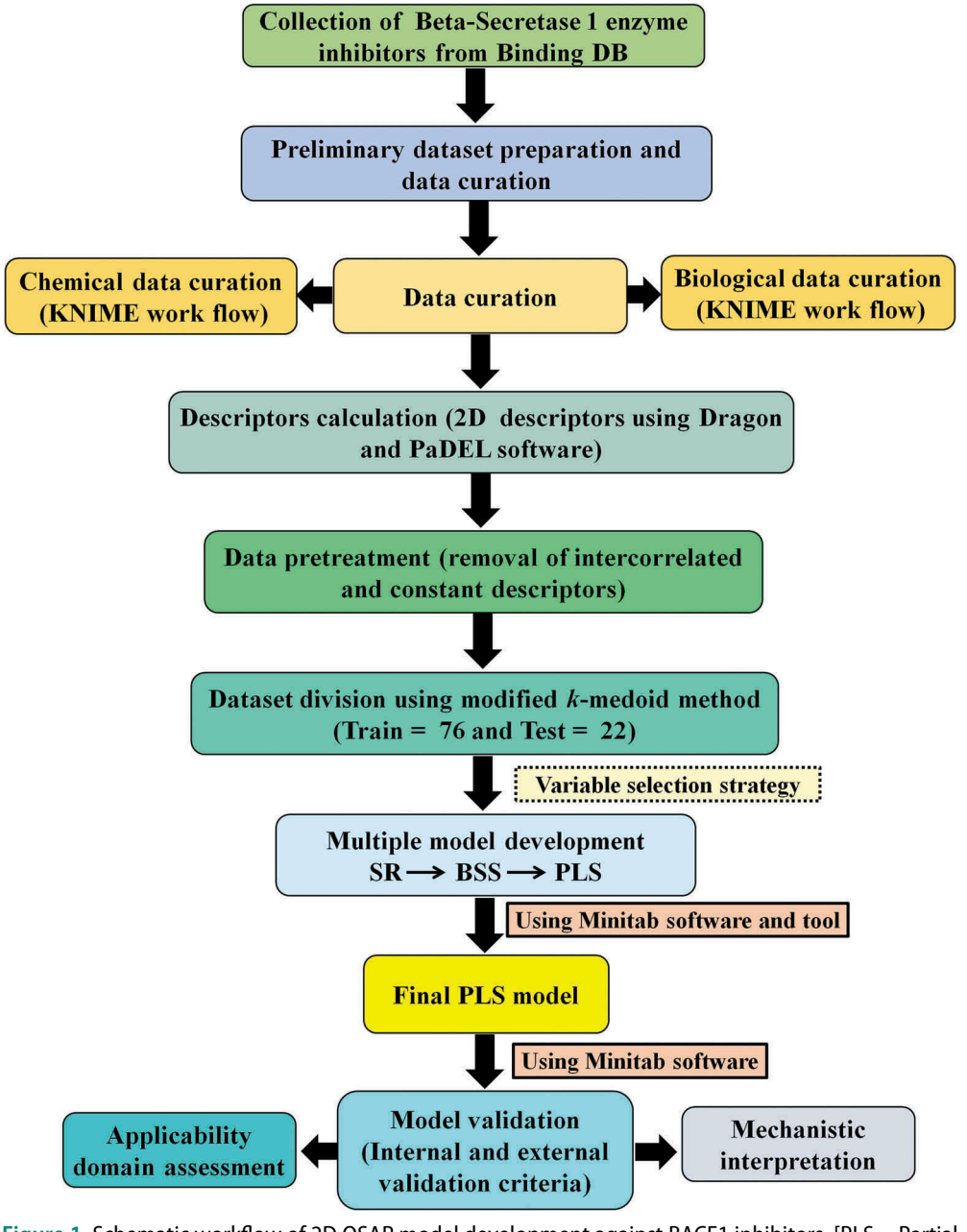


Figure 1. Schematic workflow of 2D QSAR model development against BACE1 inhibitors, [PLS = Partial least squares, SR = Stepwise regression, BSS = Best subset selection].

obtained from test set analysis [41]. The validation parameters in term of qualitative analysis used for pharmacophore model are sensitivity, specificity, accuracy, precision, *F*-measure and G-means. According to the Aher et al. [41], the selected model is measured to be robust if all the validation parameter values are more than 60% for both the sets (training and test set). In term of internal validation, we have performed cost analysis and selected the model based on the RMSD, correlation, fit values and cost difference values. We have also performed the Fischer

randomization test (F-test) at a confidence level 95%, to check whether the obtained model is by chance or not. It was carried out by randomly reordering the activity data of training set molecules and developed the model with the same settings as used for the actual pharmacophore model development [41]. According to F-test, the actual model is considered to be better model, if the results obtained from randomized models are bad quality than actual model [41]. The validated pharmacophore model could be further utilized for the prediction of inhibitory activity of the new compound against BACE1 enzyme (see Supplementary materials S2).

Molecular docking studies

Aβ is the main component of pathophysiology in Alzheimer's disease, and BACE1 enzyme is responsible for amyloidogenic cleavage of APP generating Aβ. Consequently, controlling the BACE1 enzyme activity to decrease Aβ is a rational therapeutic goal [43]. Molecular docking is a key tool in structural molecular biology and computer-assisted drug design [44]. The goal of ligand-protein docking is to predict the predominant binding mode of a ligand with a protein of known three-dimensional structure [44]. In the present study, we have employed molecular docking studies to understand the interaction pattern of BACE1 enzyme inhibitors within the active site of BACE1 enzyme (the structure of the protein was retrieved from Protein Data Bank with PDB ID: 4ivt) [45]. Molecular docking was performed by using the CDOCKER module of receptor–ligand interaction available in BIOVIA Discovery Studio client 4.1 [42]. Prior to the docking, we have defined the active site of the enzyme using the protocol Receptor–ligand Interaction section employing the option 'define site from receptor cavities' available in the BIOVIA Discovery Studio client 4.1 platform [42]. The selected inhibitors were subjected to ligand preparation to find a series of ligand conformers (maximum 255). Each orientation was used in the CDOCKER module for molecular docking using CHARMm-based interaction energy using a rigid receptor [46]. The poses are sorted according to CHARMm interaction energy and the top-scoring (most negative, thus favourable to binding) poses are kept.

Software tools used

A KNIME workflow was used for chemical and biological data curation (available at http://dtclab.webs.com/software-tools). Marvin ChemAxon software [31] was used for the representation of chemical structures. Descriptors were calculated by using the PaDEL-descriptor 2.20 [34] tool and Dragon software version 7 [33]. Data pretreatment was performed by the tool Pretreatment V-WSP version 1.2 (available at http://dtclab.webs.com/software-tools). Clustering of the dataset was performed by using the 'Modified K-Medoid' tool version 1.3 (available at http://dtclab.webs.com/software-tools) for its splitting into a training set and a test set. Variable selection and PLS model development were performed by using Minitab software version 14 [38]. Best subset selection (BSS) was performed by Best subset selection v2.1 software (available at http://dtclab.webs.com/software-tools). The docking study was performed by using CDOCKER module available in BIOVIA Discovery Studio client 4.1 [42]. The pharmacophore mapping was performed by using the HypoGen module of 3D pharmacophore modelling available in BIOVIA Discovery Studio client 4.1 [42].

Finally, several plots depicting the quality of the PLS model were developed by using Simca-P [40] software.

Results and discussion

Mechanistic interpretation of modelled descriptors

In this study, we have developed a PLS-regression model for BACE1 enzyme inhibitors using only 2D descriptors. Before the development of the final model, we have employed the stepwise regression method to reduce the initial number of descriptors to manageable one, followed by the best subset selection (BSS) (available at http://dtclab.webs.com/software-tools) as mentioned in materials and methods section. The derived model was validated carefully using different validation metrics (both internal and external). The statistical quality of the developed model is depicted below. The reported PLS model was developed with five descriptors using three LVs (latent variables). The statistical validation parameters like r^2 (0.826) and $Q^2_{(LOO)}$ (0.795) signifying the reliability of the model, and r^2_{Pred} or Q^2F_1 (0.846), Q^2F_2 (0.846) judge the good predictivity of the model. The descriptors appearing in the model describe the structural and functional requirements which can improve the inhibitory activity of molecules against BACE1 enzyme. The closeness of the observed and predicted values for the BACE1 enzyme inhibitors in the data set can be further recognized from the scatter plot as shown in Figure 2. The final PLS model with their validation parameters is depicted below:

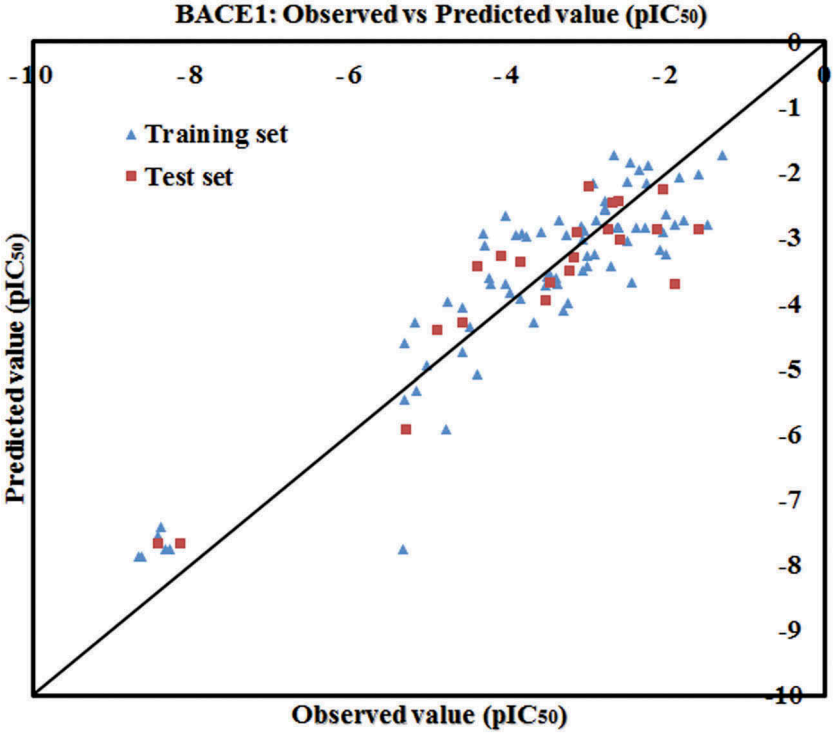


Figure 2. The scatter plot of observed and predicted values of the final PLS model against BACE1 enzyme.

BACE1 enzyme inhibitors: PLS model

$$\mathrm{pIC_{50}}(nM) = -4.235 + 36.67 \times \text{″}ETA\_dEpsilon\_D\text{″} - 1.82 \times \text{″}B03[N-S]\text{″} + 1.106 \times \text{″}ETA\_Beta\_ns\_d\text{″} - 0.107 \times \text{″}H-047\text{″} + 1 \times \text{″}C-033\text{″} \quad (1)$$

$$n_{training} = 76,\ r^2 = 0.826, Q^2 = 0.795,\ \text{Average } {r_m}^2 = 0.711,\ \Delta {r_m}^2 = 0.139,\ \text{MAE} = 0.625$$

$$\text{Fitting quality} = \text{Moderate}$$

$$n_{test} = 22, Q^2F_1 = 0.846, Q^2F_2 = 0.846,\ \text{Average } {r_m}^2 = 0.731,\ \Delta {r_m}^2 = 0.140,\ \text{MAE} = 0.544,\ \text{CCC} = 0.911$$

$$\text{Fitting quality} = \text{Good},\ \text{LV} = 3,\ \text{No. of descriptors} = 5$$

We have developed a simple but statistically robust PLS regression-based QSAR model against BACE1 enzyme. The regression coefficient plot (see Figure S1 in Supplementary materials S3) provides the information about the contribution of descriptors in the model towards the activity of the compounds. The positive regression coefficients of the descriptors indicated that the BACE1 inhibitory activity will increase with increasing their descriptor values as shown in the case of ETA_Beta_ns_d, ETA_dEpsilon_D and C-033 descriptors. In contrast, the negative regression coefficients of the descriptors suggested that the BACE1 enzyme inhibitory activity of the compounds will decrease with increasing the descriptor values as shown in case of H-047 and B03[N-S] descriptors. The significance level of the modelled descriptors towards the inhibitory activity against BACE1 enzyme is computed based on the variable importance plot (VIP) [47]. The variable importance plot (VIP) [47] defines the order of significance level among the model variables which are responsible to regulate the inhibitory activity towards the BACE1 enzyme. The descriptors contributing most (ETA_dEpsilon_D, B03[N-S] and ETA_Beta_ns_d) and least (H-047 and C-033) to the BACE1 inhibition can be identified with the help of this plot (see Figure S1 in Supplementary materials S3). The variables show higher statistical significance with VIP score >1 as compared to one with a low VIP score of 0.45. As suggested by the VIP plot [47], the significance level of the modelled descriptors is found to be in the following order: ETA_dEpsilon_D, B03[N-S], ETA_Beta_ns_d, H-047 and C-033. The details of statistical validation parameters in terms of both internal and external validation parameters are depicted in the model.

The most contributing descriptor as per the VIP plot [39] (see Figure S1 in Supplementary materials S3) is ETA_dEpsilon_D, an extended topochemical atom descriptor, denoting the measure of contribution of hydrogen bond donor atoms, i.e., the presence of groups such as –OH, –NH_2, –SH, etc., [48]. The positive regression coefficients of this descriptor indicated that the activity of inhibitors is directly proportional to the numerical value of ETA_dEpsilon_D. Thus, the compounds having higher number of hydrogen bond donor atoms may enhance the BACE1 enzyme inhibitory activity as shown in (Figure 3) compounds like 28 (*N*-carbamimidoyl-2-(2,5-diphenyl-1H-pyrrol-1-yl)acetamide) ($\mathrm{pIC_{50}}$: –4.28), 82 (6-((2-(2-chlorophenyl)-5-(4-(pyrazin-2-yloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-2-amine) ($\mathrm{pIC_{50}}$: –2.67)

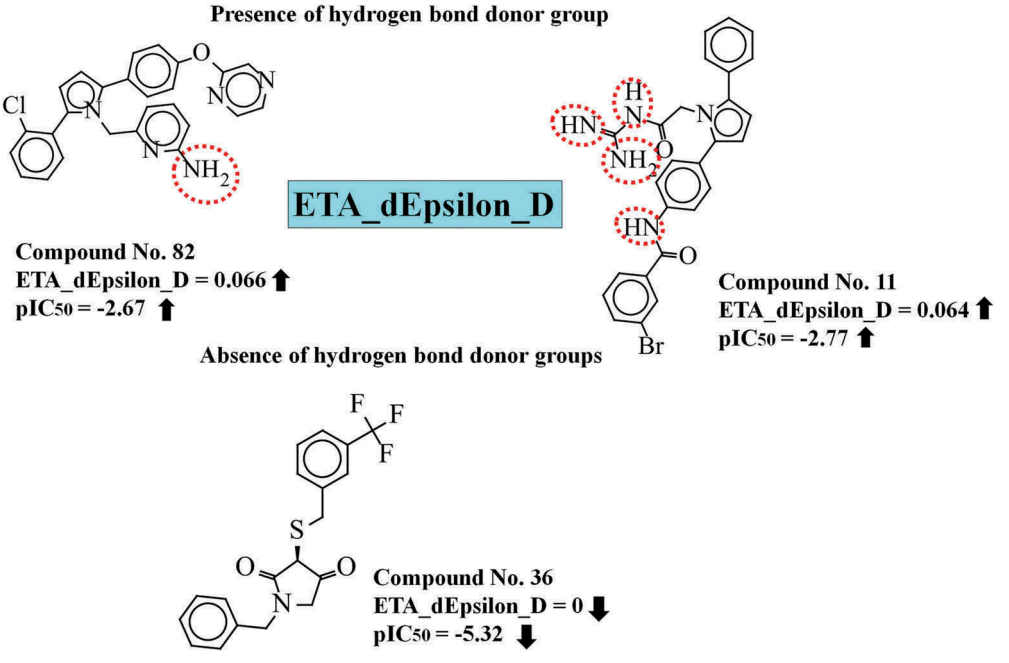


Figure 3. Impact of ETA_dEpsilon_D descriptor on the pIC_{50} values of the compounds.

and 11 (3-bromo-N-(4-(1-(2-guanidino-2-oxoethyl)-5-phenyl-1H-pyrrol-2-yl)phenyl)benzamide) (pIC_{50}: −2.77) and their corresponding descriptor values are 0.071, 0.066 and 0.0648, respectively. In contrast, compounds like 31 ((3S,5S)-5-phenyl-3-(phenylthio)furan-2,4(3H,5H)-dione), (pIC_{50}: −5.02) 36 ((S)-1-benzyl-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: −5.32) and 38 ((S)-1-phenethyl-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: −8.26) have no such atom to form a hydrogen bond, leading to lower inhibitory activity (Figure 3). From these observations, we have concluded that a hydrogen bond donor group is important for BACE1 inhibitory activity.

The next significant descriptor, B03[N-S] is an 2D atom pair descriptor that accounts for the presence/absence of N-S fragment at the topological distance 3 [49]. It contributes negatively towards the endpoint value which suggested that the numerical values of the descriptor are inversely proportional to the inhibitory activity. Thus, the compounds bearing such fragments show lower values of inhibitory activity as evidenced by (Figure 4) compounds 40 ((R)-1-(4-methoxybenzyl)-3-((3-(trifluoromethyl) benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: −8.38), 38 ((S)-1-phenethyl-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: −8.26) and 39 ((S)-3-(phenethylthio)-1-(3-(trifluoromethyl)benzyl)pyrrolidine-2,4-dione) (pIC_{50}: −8.32)) whereas, compounds having no such fragments show higher BACE1 inhibitory activity as shown in compounds 1 ((S)-N-(3-(2-amino-4-cyclohexyl-1-methyl-5-oxo-4,5-dihydro-1H-imidazol-4-yl) phenyl)furan-2-carboxamide) (pIC_{50}: −1.301) 45 (R)-8-(3-(2-fluoropyridin-3-yl)phenyl)-8-(4-(trifluoromethoxy)phenyl)-2,3,4,8-tetrahydroimidazo[1,5-a]pyrimidin-6-amine) (pIC_{50}: −1.477) and 72 (2-((2-amino-6-((2-(2-chlorophenyl)-5-(4-(pyrimidin-5-yloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-3-yl)oxy)ethanol)) (pIC_{50}: −1.602) (Figure 4).

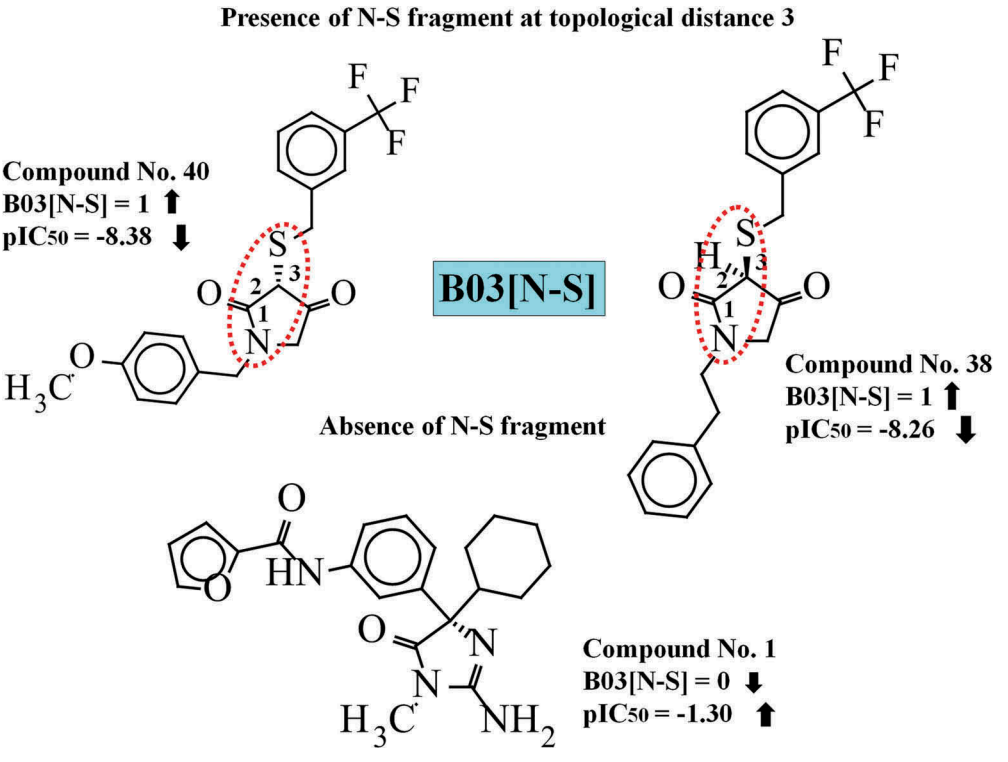


Figure 4. Impact of B03[N-S] descriptor on the pIC_{50} values of the compounds.

Another extended topochemical atom (ETA) descriptor, ETA_Beta_ns_d, represents the summed contribution of lone electron pairs capable of forming resonance interaction with an aromatic system. It is defined as sum of all $\beta_{ns(\delta)}$ values of all vertices [48]. The positive regression coefficient of this descriptor indicates that heteroatoms with a lone pair of electrons capable of resonance with an aromatic system are beneficial for the enzyme inhibitory activity as shown in compounds (Figure 5) 72 (2-((2-amino-6-((2-(2-chlorophenyl)-5-(4-(pyrimidin-5-yloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-3-yl) oxy)ethanol) (pIC_{50}: −1.60), 73 (2-((2-amino-6-((2-(2-chlorophenyl)-5-(4-(pyrimidin-5-yloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-3-yl)amino)ethanol) (pIC_{50}: −1.84) and 76 (2-((2-amino-6-((2-(2-chlorophenyl)-5-(4-(pentyloxy)phenyl)-1H-pyrrol-1-yl)methyl)pyridin-3-yl)amino)ethanol)) (pIC_{50}: −2.23) (all these compounds have descriptor value 2). In contrast, a low number of heteroatoms is detrimental to the enzyme inhibitory activity as we have observed in (Figure 5) compounds 42 ((S)-1-(benzo[d][1,3]dioxol-5-ylmethyl)-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: −8.42), 43 ((S)-1-phenethyl-3-(phenethylthio)pyrrolidine-2,4-dione) (pIC_{50}: −8.62) and 44 ((R)-1-benzyl-3-(phenethylthio)pyrrolidine-2,4-dione)) (pIC_{50}: −8.62).

An atom-centred fragment descriptor, H-047, stands for number of H atoms attached to C^1 (sp3)/C°(sp2); where the superscript represents the formal oxidation number (the formal oxidation number of a carbon atom which is equal to the sum of the conventional bond orders with electronegative atoms) [50]. This descriptor is defined as the number of specific atom types in a molecule which is calculated by knowing only the molecular composition and

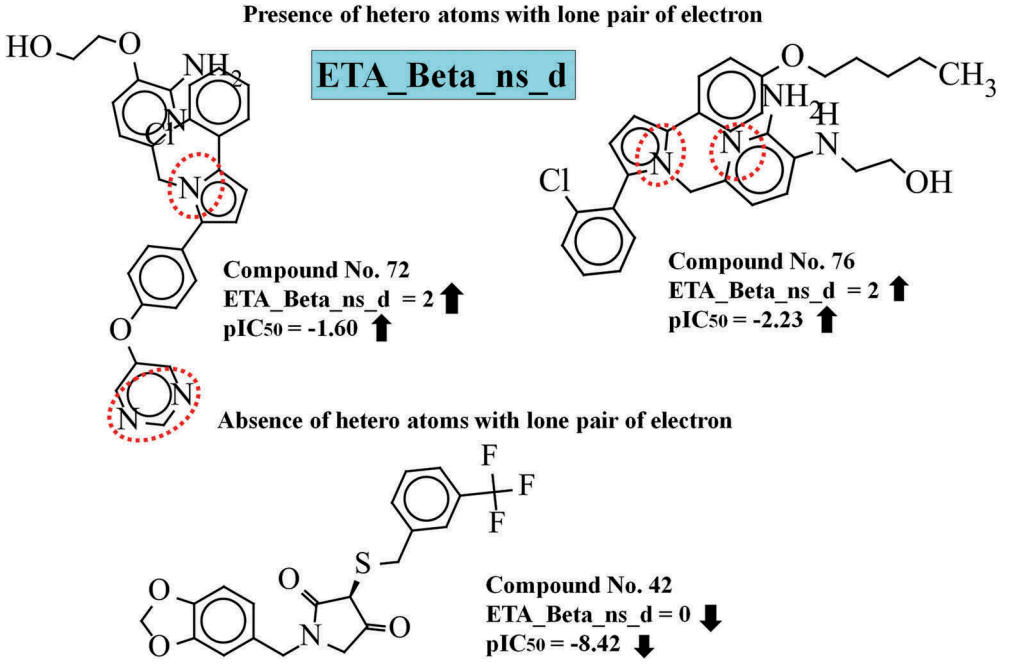


Figure 5. Contribution of ETA_Beta_ns_d descriptor to the pIC_{50} values of the compounds.

atom connectivity [51]. The negative regression coefficient of this descriptor suggests that higher numerical value of this descriptor leads to lower inhibitory activity as evidenced by the compounds 67 ((S)-8-(3-(4-methoxybenzyl)phenyl)-8-phenyl-2,3,4,8-tetrahydroimidazo[1,5-a] pyrimidin-6-amine) (pIC_{50}: −4.57), 40 ((R)-1-(4-methoxybenzyl)-3-((3-(trifluoromethyl)benzyl) thio)pyrrolidine-2,4-dione) (pIC_{50}: −8.38), and 43 ((S)-1-phenethyl-3-(phenethylthio)pyrrolidine-2,4-dione)) (pIC_{50}: −8.62) (corresponding descriptor values are 20, 18 and 17, respectively) (Figure 6). On the contrary, the compounds with lower descriptor values show higher BACE1 enzyme inhibitory activity as observed in case of compounds 1 ((S)-N-(3-(2-amino-4-cyclohexyl-1-methyl-5-oxo-4,5-dihydro-1H-imidazol-4-yl)phenyl)furan-2-carboxamide)(pIC_{50}: −1.30), 5 ((S)-2-amino-4-cyclohexyl-1-hexyl-4-phenyl-1H-imidazol-5(4H)-one) (pIC_{50}: −2.43) and 6 ((S)-6-(2-amino-4-cyclohexyl-5-oxo-4-phenyl-4,5-dihydro-1H-imidazol-1-yl)hexanoic acid) (pIC_{50}: −2.49) (there corresponding descriptor values are 9, 7 and 7, respectively) (Figure 6).

Another atom-centred fragment descriptor, C-033, stands for the fragment R–CH.X. It represents the number of the R–CH … X fragments in a molecule which means a central carbon atom (C) on an aromatic ring has a carbon neighbour (R), a heteroatom neighbour (X-any heteroatom (O, N, S, P, Se, and halogens)) and the third hydrogen (H) neighbour outside the ring. '–' and ' … ' stand for aromatic and aromatic single bonds, respectively [49,51]. For these β-secretase enzyme inhibitors, this fragment indeed plays an important role in the binding process and may influence the inhibitory activity prominently. The positive impact of this descriptor towards the inhibitory activity against the β-secretase enzyme was indicated by their positive regression coefficient. Thus, the information obtained from this descriptor suggested that the molecules containing R–CH.X fragment show higher inhibitory activity to β-secretase enzyme as shown in compounds 1 ((S)-N-(3-(2-amino-4-cyclohexyl-1-methyl-

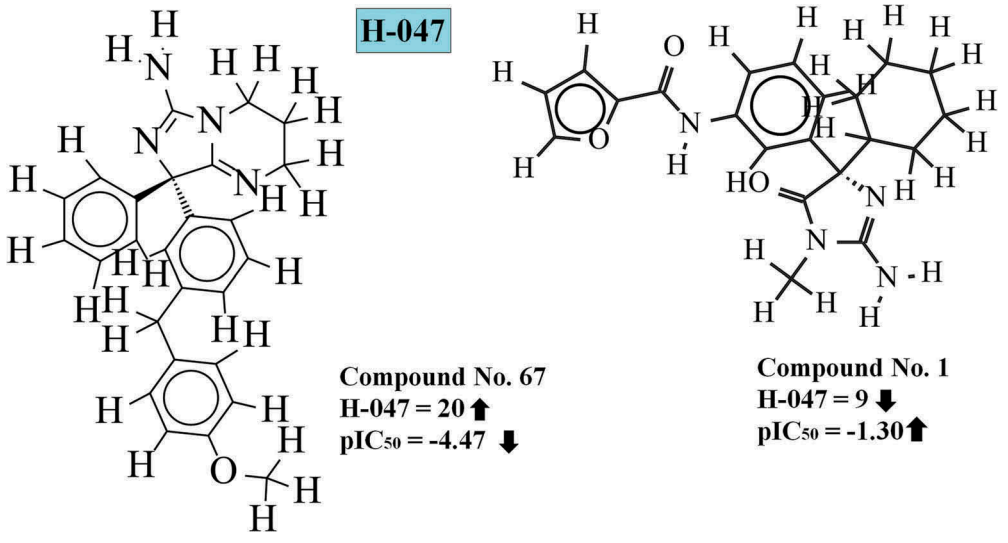


Figure 6. Impact of H-047descriptor on the pIC_{50} values of the compounds.

5-oxo-4,5-dihydro-1H-imidazol-4-yl)phenyl)furan-2-carboxamide) (pIC_{50}: −1.30), 35 ((3S,5S)-5-((R)-sec-butyl)-3-((furan-2-ylmethyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: −5.30) and 23 (2-((5-cyano-4-(4-methoxyphenyl)-6-oxo-1,6-dihydropyrimidin-2-yl)thio)-N-(4-phenylthiazol-2-yl)acetamide)) (pIC_{50}: −3.67) (Figure 7) while compounds 42 ((S)-1-(benzo[d][1,3]dioxol-5-ylmethyl)-3-((3-(trifluoromethyl)benzyl)thio)pyrrolidine-2,4-dione) (pIC_{50}: −8.42), 43 ((S)-1-phenethyl-3-(phenethylthio)pyrrolidine-2,4-dione) (pIC_{50}: −8.62) and 44 ((R)-1-benzyl-3-(phenethylthio)pyrrolidine-2,4-dione) (pIC_{50}: −8.66) show lower inhibitory activity due to the absence of this fragment (Figure 7).

Randomization model of the PLS model

The predictive quality of the developed model will be poor until the observations are not appropriately independent of each other. The randomization method is a way to test the robustness of the developed model [52]. The purpose of the development of a randomization plot is to identify that the selected descriptors are appropriate and the reported model is not due to chance correlation. In randomization method, many numbers of models are developed by several runs for which the original descriptor matrix X is kept fixed, and only the vector Y is randomized. The validation metrics of the developed model under such condition should be poor and the value of the r^2_{yrand} intercept should not more than 0.3 and the value of the Q^2_{yrand} intercept should not exceed 0.05 [52]. In the present study, for the training set, the X data were kept constant and the Y data were scrambled randomly using 100 permutations. The model obtained from such condition shows the intercepts as follows: (see Figure S2 in Supplementary materials S3) r^2_{yrand} = −0.0128 and Q^2_{yrand} = −0.26, which signify the validity of the model and confirm that the reported model was not obtained by any chance. The above results suggest that the developed model is non-

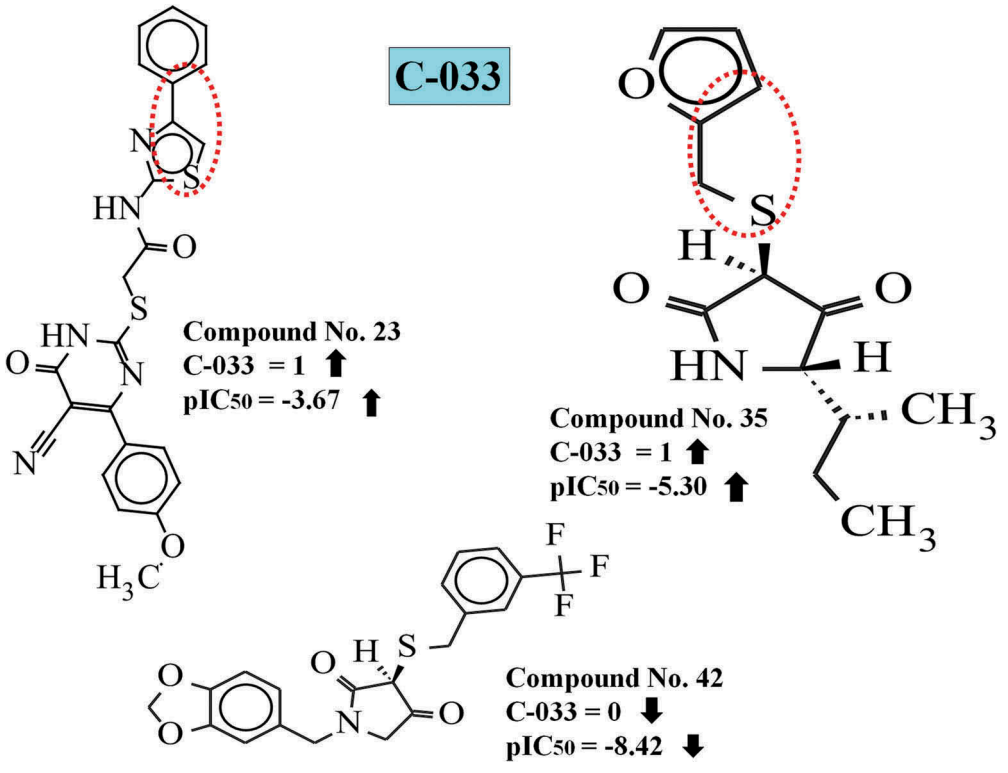


Figure 7. Impact of C-033 descriptor on the pIC_{50} values of the compounds.

random and robust, and suitable for prediction of the inhibitory activity against BACE1 enzyme.

Applicability domain of the PLS model

Prediction of the activity of entire space of chemicals is not possible by a robust and validated QSAR model until the compounds are predicted within the applicability domain of the model. The applicability domain (AD) gives a theoretical province in chemical space well-defined by the respective model descriptors and responses in which the predictions of activity are reliable [53]. In this study, we have checked the applicability domain of test set compounds at 95% confidence level using the DModX (distance to model in X-space) approach available within SIMCA-P 10.0 software [40]. In the plot (see Figure S3 in Supplementary materials S3), we have observed that all the test set compounds are within the critical DModX value (D-Crit = 2.584) except compound 29 ((3R,4aR,4bR,8aR,9aS)-3-((4-(6-methoxynaphthalen-2-yl)-1H-1,2,3-triazol-1-yl)methyl)dodecahydro-1H-pyrido [3,4-b]indole).

Loading plot of the PLS model

A loading plot of a PLS model (see Figure S4 in Supplementary materials S3) gives the information about the relationship among the X-variables and Y-variables. In the loading

plot, a descriptor which is close to zero is not well associated with the trends contained in the related scores [54]. In the loading plot (see Figure S4 in Supplementary materials S3), we can observe that the X-variable ETA_dEpsilon_D is significant for the Y-variable (pIC_{50}) because it is very much close to the Y-variable. It has also been observed that this descriptor is situated on the same side of Y-variable. From this observation, it can be concluded that this descriptor is directly proportional to the activity. Thus, the BACE1 enzyme inhibitory activity may increase with increasing the numerical value of this descriptor. On the other hand, the variable B03[N-S], which is situated on the opposite side of the plot origin with respect to the activity (Y-variable) contributes negatively towards the BACE1 enzyme inhibitory activity. The algebraic sign of the PLS loading is also taken into account, which gives important information about the correlation between the variables.

3D QSAR pharmacophore modelling

In the present work, we have developed 10 different pharmacophore hypotheses using a training set (22) of compounds. The best pharmacophore model (Hypo-1) was selected based on the different internal validation parameters such as high correlation coefficient (r = 0.912), lower root mean square deviation (rmsd: 1.320), Maximum Fit (8.393), total cost (100.355), configuration cost (16.122), error (83.054) and weight (1.177) was found to be acceptable. In term of the actual cost for Hypo-1, it is found much closer to the fixed cost with only a difference of 19.240 bits (mentioned in Table 2) which specifies an accurate correlation of the dataset. From Table 2, we can see that in Hypo 1 there is a large difference 78.229 bits between the actual cost and the null cost. Based on all validation matrices Hypo-1 was found to be the best 1 among the 10 hypotheses with one hydrogen bond acceptor (HBA), two hydrophobic (HYD), and one ring aromatic (RA) features (Figure 8). The results of 10 pharmacophore hypotheses against BACE1 enzyme are depicted in Table 2. The external validation of the model has been performed by mapping the test set molecules on the Hypo-1 using same parameters as we have used in the development of the pharmacophore model. After mapping, we have observed that 64 molecules from the data set of 76 compounds were mapped, only 10 compounds failed in absence of the features appeared in the developed

Table 2. Results of 3D-QSAR pharmacophore development against BACE1 enzyme.

Hypo.	*Total cost*	*ΔCost*[a]	*ΔCost*[b]	*Correlation (r)*	*RMS*	*Features*
1[c]	100.355	78.229	19.240	0.912563	1.320	HBA, HYD, HYD, RA
2	101.328	77.256	20.213	0.908511	1.349	HBA, $HYD_{(ali)}$, HYD, RA
3	114.143	64.441	33.028	0.851373	1.697	HBA, HBA, RA
4	119.013	59.571	37.898	0.818559	1.855	HBA, $HYD_{(ali)}$, HYD, RA
5	121.711	56.873	40.596	0.803815	1.921	HBA, HYD, RA
6	122.254	56.330	41.139	0.807127	1.907	HBA, $HYD_{(ali)}$, HYD, RA
7	123.086	55.498	41.971	0.811693	1.895	HBA, HBA, RA
8	123.516	55.068	42.401	0.79692	1.951	HBA, HYD, HYD, RA
9	124.065	54.519	42.950	0.8008	1.935	HBA, HBA, RA
10	124.133	54.451	43.018	0.795531	1.957	HBA, $HYD_{(ali)}$, HYD, RA

Cost difference[a] = Null cost – total cost, Cost difference[b] = Total cost – fixed cost, Null cost = 178.584, Fixed Cost = 81.1148, Best records in pass: 5, Config. Cost = 16.1229, c = Best Hypothesis, Note – RA: Ring aromatic, HYD: Hydrophobic, HYD (ali): Hydrophobic Aliphatic, HBA: Hydrogen bond acceptor.

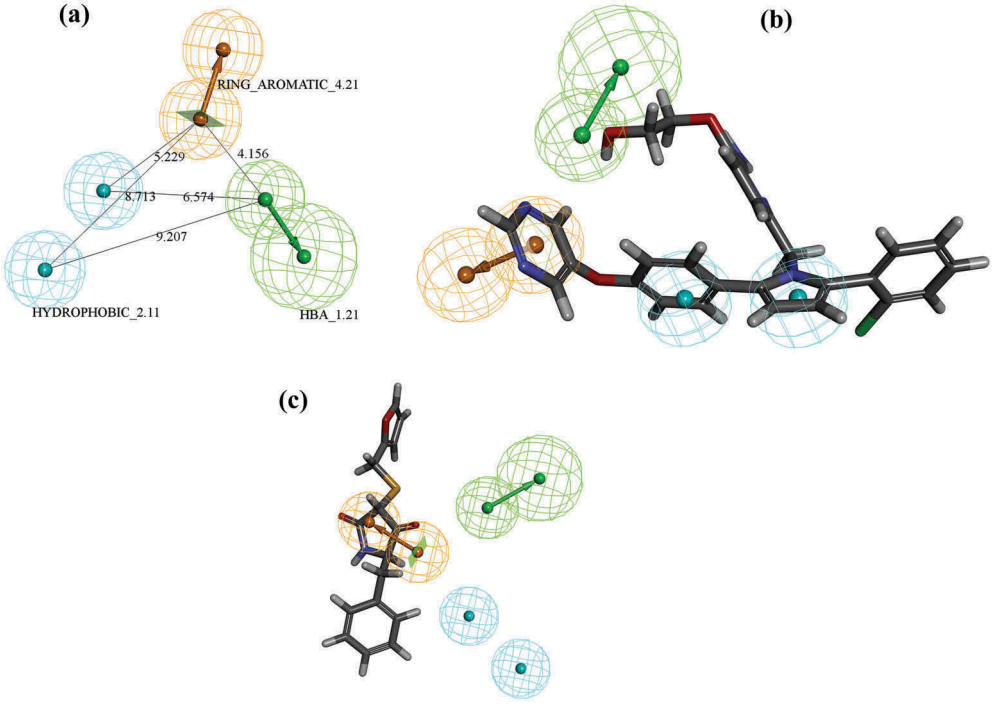


Figure 8. The best pharmacophore model (Hypo1) of BACE1 enzyme inhibitors generated by the HypoGen module: (a) the best pharmacophore model Hypo1 represented with distance constraints (Å), (b) Hypo1 mapping with one of the most active compounds 72 of test set compounds and (c) Hypo1 mapping with one of the least active compounds 32 of test set compounds. Pharmacophoric features are coloured as follows: hydrogen bond acceptor (green), hydrophobic (cyan), and ring aromatic (orange).

pharmacophore model. To judge the predictive quality of the selected model and to categorize the compounds into active and less active BACE1 enzyme inhibitors were analysed by comparing the observed activity with predicted activity by the classification-based technique. For this determination, the compounds with IC_{50} values ≤1000 nM were classified as active compounds and compounds with IC_{50} values >1000 nM as less actives. The observed and predicted activity of the training and test sets compounds based on Hypo-1 are given in S2 sheets 2 and 3 (Supplementary materials section), respectively. The values of different validation parameters for training and test sets are given in Table 3 (qualitative validation parameters). From the observation of activity predicted by the selected model, we have found that the model correctly classified 11 out of 12 compounds as more actives and 9 out of 10 compounds as less actives for the training set. For the test set, the model correctly classified 18 out of 25 compounds as most active and 30 out of 39 compounds as less actives. Aher et al. [41] suggested that if the values of different validation parameters for both the training and test sets are greater than 60%, it means that the model is following acceptability criteria and good enough to predict the activity of new compound of same chemical domain. From the above observation, we have concluded that Hypo-1 is best appropriate for the classification of more active BACE1 enzyme inhibitors.

Table 3. Different qualitative validation parameters (%) of Hypo-1 model obtained by classification of more active and less active compounds for the training and test sets of BACE1 enzyme inhibitors.

		Qualitative validation parameters (%)					
Dataset	*No. of compounds*	*Sensitivity*	*Specificity*	*Accuracy*	*Precision*	*F-measure*	*G-means*
Train	22	91.66	90	90.90	91.66	91.66	90.82
Test	76	68	76.92	73.43	65.38	66.66	72.32

*Compounds with IC_{50} < 1000 nM: more active (H) and IC_{50} > 1000 nM: less active (L).

Relation of the 3D-pharmacophore model with the 2D QSAR model

In the dataset, all compounds have at least one ring aromatic feature as we have observed from the both QSAR models (2D QSAR and 3D QSAR pharmacophore models). The RA feature is an initial necessity for the inhibitory activity against BACE1 enzyme. The RA feature from pharmacophore model is well corroborated with the ETA_Beta_ns_d and C-033 descriptors of the 2D-QSAR model (Equation 1). As we have observed from the most active compound of the training set (45, IC_{50}: 30nM) mapped correctly with all features appeared in Hypo 1 (Figure 9). One benzene ring lies in the RA region, a nitro group in the hydrogen bond acceptor region and a halogen atom of the aromatic ring lies in the hydrophobic region. Whereas, the least active compound (70, IC_{50}: 78,700 nM) of the training set does not map correctly with Hypo 1 because of the absence of RA and HBA features in the molecules (Figure 9). Hydrophobic feature from the developed pharmacophore model is well corroborated with the C-033 descriptor of the 2D-QSAR model (Equation 1). We have observed from the most active compound of the training set (45), the hydrophobic feature on Hyp1 mapped completely with these molecules. The most active compound of the test set (72, IC_{50}: 40nM) mapped correctly on Hypo-1 with all the three features (Figure 8) appeared in the developed pharmacophore model. The least active compound (32, IC_{50}: 145,000 nM) of the test set mapped partially with Hyo-1 (Figure 8). From the above observation, we have concluded that the absence of these three features appeared in developed pharmacophore model decreases the inhibitory activity of compounds against BACE1 enzyme. The results obtained from the F-test suggest that the selected pharmacophore model (Hypo-1) is not due to a chance. This observation was confirmed by a lower cost value (100.355) of the selected pharmacophore

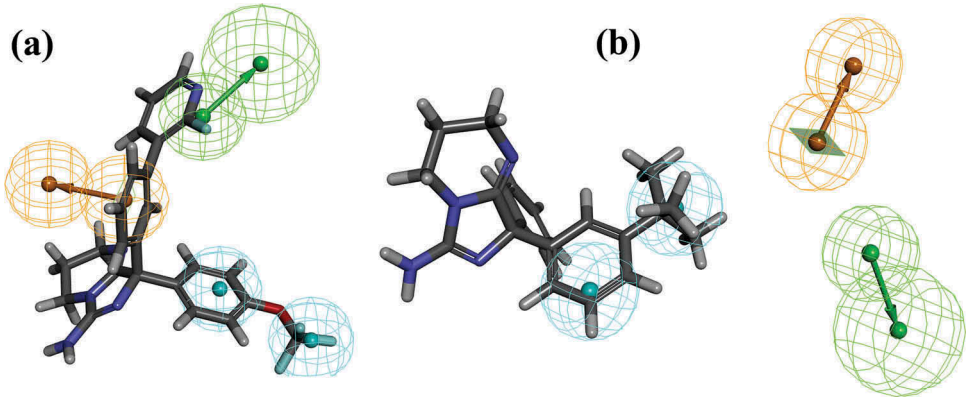


Figure 9. Pharmacophore mapping with training set compounds: (a) Hypo1 mapping with one of the most active compounds 45 of training set compounds and (b) Hypo1 mapping with one of the least active compounds 70 of training set compounds.

model than the average cost of randomized pharmacophore models (155.91) and higher correlation coefficient (r = 0.912) of the selected pharmacophore model than the average correlation coefficient of random models (r_r = 0.568). The actual and randomized total cost and correlation values of hypotheses for *F*-test are given in S2 sheet 4 and 5 of the Supplementary materials section.

Molecular docking

We have performed molecular docking study of most active and least active compounds of the dataset. The molecular docking study suggests that the molecules interacted with a pocket containing (ILE A:110, PHE A:108, VAL A:332, GLY A:11, GLY A:13, GLY A:230, VAL A:332, ILE A:226, LEU A:30 and GLY A:34 (hydrophobic nature), GLN A:73, SER A:229, THR A:231, THR A:329, THR A:71, GLN A:73, GLN A:12 and THR A:232 (hydrophilic nature), ARG A:235, ASP A:32, ASP A:228 and LYS A:107 (charged) and TYR A:71 (amphipathic nature)) amino acid residues. Docking results and correlation with the 2D-QSAR model are depicted in Table 4.

Molecular docking for the most active compounds from the dataset

Three most active compounds from the dataset (pIC_{50} = −1.301, −1.47 and −1.77, respectively) namely 1, 45 and 16 interacted with the active site amino acid residues through different interaction forces like hydrogen bonding interactions (carbon hydrogen bonds conventional hydrogen bonds and π-donor hydrogen bond), π-interactions (π-alkyl bonds, alkyl hydrophobic, π-lone pair and π-π-T-shaped), salt bridge interaction and halogen bonding (halogen bonding is an attractive, non-covalent interaction that can form between an electrophilic region of a halogen atom (fluorine) in a molecule and a nucleophilic region of a molecule). The amino acid residues involved in interaction with these compounds such as THR A:231, GLY A:11, GLN A:73, THR A:71, ASP A:32, ASP A:228, VAL A:332, GLY A:13, THR A:232, GLY A:230, SER A:229, GLY A:34, LYS A:107, ILE A:110, PHE A:108 and ARG A:235 (shown in Figure 10 and Figures S5 and S6 in Supplementary section S3).

Figure 10 shows that compound 1 (one of the most active compounds in dataset) interacts with GLN A:73, ASP A:32 and ASP A:228 amino acid residues through hydrogen bonding interaction, with VAL A:332, THR A:71 and THR A:231 amino acid residues through alkyl, π-alkyl and π-lone pair interactions, respectively, with ASP A:228 amino acid through salt bridge formation and with ASP A:32 amino acid through attractive charges (interaction between two oppositely charged atoms).

Another most active compound, 45, interacts with the amino acid residues through hydrogen bonding (GLNA:73, GLY A:13, PHE A:108, LYS A:107, THR A:232 and GLY A:34), halogen bonding (THR A:231, SER A:229 and GLY A:230), π-anion (ASP A:228), π-donor hydrogen bond (THR A:231) and alkyl bonding (and ILE A:110) interactions (see S3 Figure S5 in Supplementary materials). Figure S6 (see S3 Figure S6 in Supplementary materials) shows that compound 16 interacts with GLY A:11, ARG A:235, GLY A:34, THR A:231, ASP A:32, ASP A:228 (through hydrogen bonding), THR A:71 (π-π-T-shaped), VAL A:332 (π-alkyl bonds) ASP A:228 (salt bridge interaction) and ASP A:32 (attractive charges) amino acid residues.

Table 4. Docking results and correlation with the QSAR model.

No.	Compound number	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
1	**1** (high pIC_{50})	45.340	THR A:231, GLY A:11, GLN A:73, TYR A:71, ASP A:32, ASP A:228 and VAL A:332	Vdw, Hydrogen bonding, Pi-alkyl, alkyl, salt bridge, Attractive charge and π-lone pair	ETA_dEpsilon_D, ETA_Beta_ns_d, C-033 and H-047
2	**45** (high pIC_{50})	39.503	GLY A:13, THR A:232, GLY A:230, SER A:229, GLY A:34, ASP A:228, GLN A:73, THR A:231, LYS A:107, ILE A:110 and PHE A:108	Vdw, Hydrogen bonding, halogen (fluorine), π-anion, π-donor hydrogen bond and alkyl	ETA_dEpsilon_D, ETA_Beta_ns_d and H-047
3	**16** (high pIC_{50})	51.278	TYR A:71, ARG A:235, VAL A:332, ASP A:228, ASP A:32, GLY A:34, THR A:231 and GLY A:11	Vdw, salt bridge, Attractive charge, hydrogen bonding, π- π-T-shaped and π-alkyl	ETA_dEpsilon_D, ETA_Beta_ns_d and H-047
4	**10** (low pIC_{50})	46.118	VAL A:332, GLN A:73, GLY A:230, ASP A:32, TYR A:71, ASP A:228 and GLY A:34	Vdw, salt bridge, Attractive charge, hydrogen bonding, π- π-T-shaped, alkyl and π-alkyl	H-047 and ETA_dEpsilon_D
5	**84** (low pIC_{50})	43.144	GLN A:73, ASP A:228, ILE A:226 and VAL A:332	Vdw, hydrogen bonding, π-anion and alkyl	H-047, ETA_dEpsilon_D and ETA_Beta_ns_d
6	**85** (low pIC_{50})	44.710	THR A:329, GLN A:12, ASP A:228, TYR A:71 and LEU A:30	Vdw, hydrogen bonding, π-anion, π – π stacked and π-alkyl	H-047, ETA_dEpsilon_D and ETA_Beta_ns_d

Molecular docking for the least active compounds from the dataset

Three least active compounds from the dataset (pIC_{50} = −2.70, −2.77 and −2.770, respectively) namely 84, 10 and 85 interact with the active site amino acid residues through different interaction forces like hydrogen bond (carbon hydrogen bonds and conventional hydrogen bonds), pi-interaction (π-anion, π-π-T-stacking π-alkyl and alkyl), salt bridge an attractive charge interaction. The amino acid residues involved in the interaction with these compounds are VAL A:332, GLN A:73, GLY A:230, ASP A:32, TYR A:71, ASP A:228, GLY A:34, ILE A:226, THR A:329, GLN A:12 and LEU A:30 (shown in Figure 11 and Figures S7 and S8 in Supplementary section S3).

Figure S7 (see Figure S7 in Supplementary materials S3) shows that compound 10 (one of the least active compounds from the dataset) interacts with GLN A:73, GLY A:230, ASP A:32, ASP A:228 and GLY A:34 amino acid residues through hydrogen bonding interaction, TYRA:71 and VAL A:332 amino acid residues through π-alkyl, π-π -T-stacking and alkyl, ASP

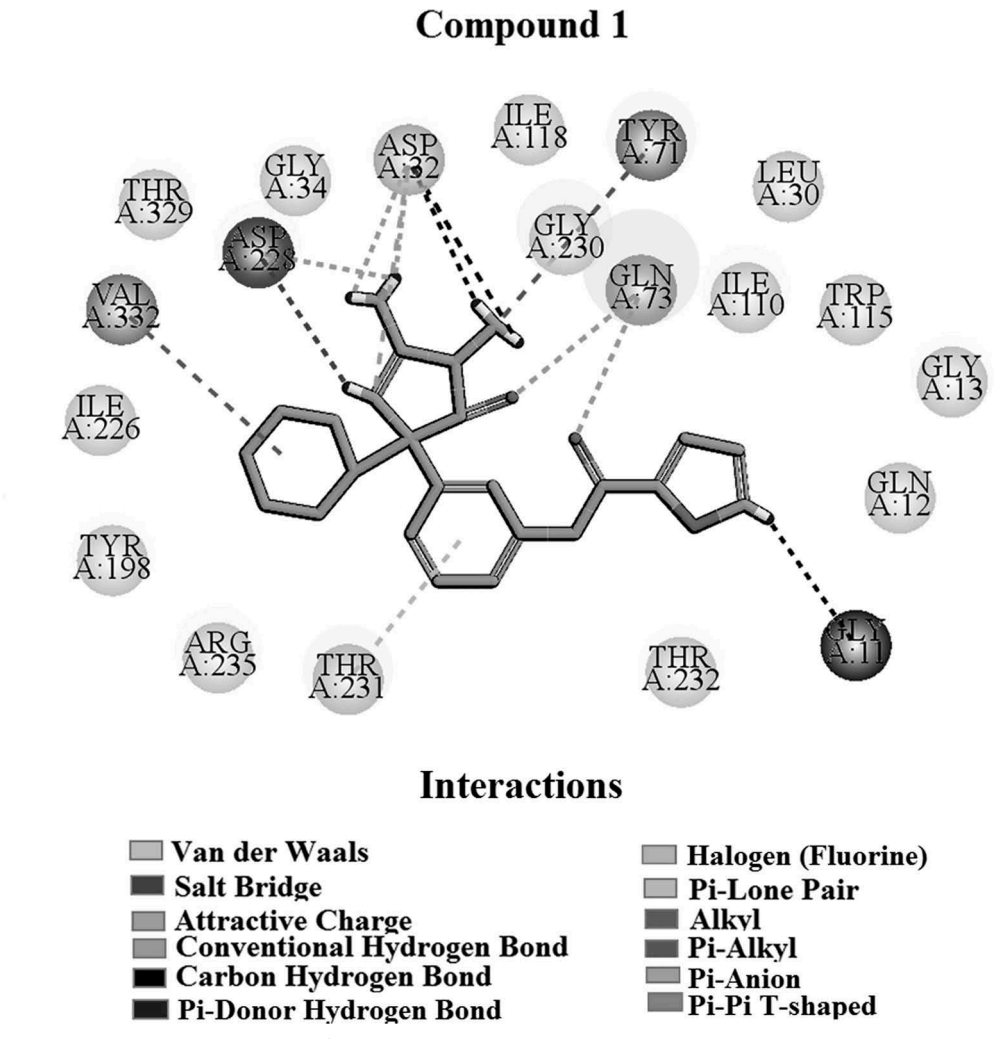


Figure 10. Docking interaction of the most active compound 1.

A:228 amino acid through salt bridge formation and ASP A:32 amino acid through attractive charges interaction. Another least active compound from the dataset, 85, interacts with the amino acid residues through hydrogen bonding (THR A:326 and GLN A:12) and hydrophobic interaction such as π-π-T-stacking, π-alkyl and π anion bonding (TYR A:71, LEU A:30 and ASP A:228) (see Figure S8 in Supplementary materials S3).

Figure 11 shows that compound 84 interacts with GLN A:73 (through hydrogen bonding), ASP A:228 (π-anion) and ILE A:226 and VAL A:332 (π-alkyl bonds) amino acid residues.

As we have observed from docking results that the most active compounds from the dataset such as 1, 45 and 16 (shown in Figure 10 and Figures S5 and S6 in Supplementary section S3) interacted with maximum number of active amino acid residues with higher number of interacting forces (non-covalent forces) in comparison to the least active compounds from the data set like 10, 85 and 84 (shown in Figure 11 and Figures S7

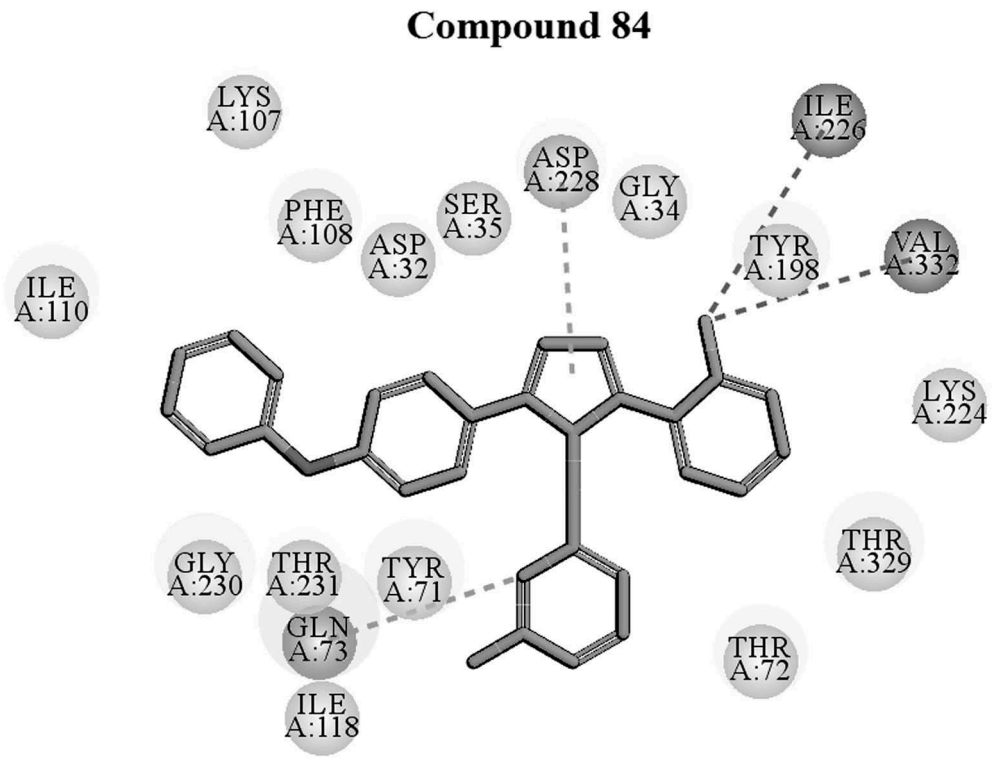


Figure 11. Docking interaction of the least active compound 84.

and S8 in Supplementary section S3). In pharmacophore mapping, we have also observed that the most active compounds from the data set correctly mapped with all features appearing in the model, whereas the least active compounds partially mapped with the model. It is possible that the least active compounds from the data set failed to map in the absence of the features appearing in the developed models, which are most important for inhibitory activity against BACE1 enzyme.

Relation with the 2D-QSAR model

We have observed from docking results that the most active compounds from the dataset such as 1, 45 and 16 (see in Table 4) interacted with a maximum number of amino acid residues with a higher number of interacting forces with the lower range of CDocker interaction energy (−45.340, −39.503 and −51.278, respectively). On the other hand, we have seen from docking results that the least active compounds from the dataset such 84 and 85 interacted with less number of active amino acid residues with a lower number of interacting forces, in comparison to the most active compounds, with CDocker interaction energy of −43.144 and −44.710, respectively. It was noted that the interaction energy (-CDocker interaction energy) depends on the number of interactions and the forces involved in the docking. Actually, it is not obvious that the most active compounds will always show the highest interaction energy and vice versa. Many compounds show

a higher range of interaction energy (as for example, compound 10) due to some insignificant interactions with other elements or amino acids in the active site which have no contribution to the biological activity. We have observed that the formation of hydrogen bonding and π-π stacking between the ligands and receptor plays a crucial role in enzyme inhibitory activity. These observations were also observed from the descriptors ETA_dEpsilon_D (a measure of the contribution of hydrogen bond donor atoms) and ETA_Beta_ns_d (a measure of lone electrons entering into resonance with an aromatic system) in the 2D-QSAR model. The information obtained from the descriptor C-033 (R–CH … X) is correlated with hydrogen bonding, attractive charges and π-donor hydrogen bonding interactions as observed from the docking study (shown in Figure 10 and Figures S5 and S6 in Supplementary section S3). Thus, from above said information, we can conclude that hydrogen bonding effect, hydrophobicity, electrostatic interactions and unsaturation (π-π interaction) features as obtained from both the 2D QSAR model and docking study are essential for the inhibitory activity against the BACE1 enzyme.

Comparisons of the performance of the present model with previously published models

A comparison of the statistical results obtained from the present QSAR model and previously published models is depicted in Table 5. Based on the statistical quality in terms of both internal and external validation criteria, the model reported in this work is statistically significant and robust enough as compared to the previously reported models (Table 5). We have used 2D descriptors only for model development. A number of researchers reported QSAR models for the prediction of bioactivity of BACE1 enzyme inhibitors previously using various techniques such as Multiple Linear Regression (MLR) analysis, Partial least squares (PLS), Comparative Molecular Field Analysis (CoMFA), Comparative molecular similarity index analysis (CoMSIA) and linear heuristic method (LHM). In the present study, prior to the development of the final model, we have performed a variable selection strategy using stepwise regression technique followed by the best subset selection method. The final model was developed by PLS regression technique with five selected descriptors using three LVs. The five selected descriptors reflect the fundamental structural characteristics of molecules which are important in modelling the bioactivity of BACE1 enzyme inhibitors. In comparison with other models, it may be noted the model developed in this study are superior in terms of statistical quality, equation length, LVs, etc. We can see from Table 5 that Ambure et al. [17] developed PLS and MLR models against for BACE1 enzyme inhibitors using only 74 compounds, and the model quality was good. In the present study, we have utilized a wider range of compounds and developed the model with five selected descriptors using three latent variables. We can also see from Table 5 that Jain et al. [18] and Chakraborty et al. [20] developed QSAR models using very narrow group of samples (27 and 30 compounds, respectively) and developed MLR and LHM-based models, respectively. Hossain et al. [19] reported 2D-QSAR and 3D-QSAR models along with molecular docking and pharmacophore mapping utilizing 106 compounds. The details of different internal and external validation parameters obtained from our model and previously reported models are given in Table 5. The docking and pharmacophore mapping results in this study are also well collaborated with descriptors obtained from the developed QSAR model and justify the significance of the developed model. As the present work deals with diverse classes of

Table 5. Comparison of proposed study with previous published studies against BACE1 enzyme.

				Training set			*Test set*	
Sources	*E. L.*	*Compts.*	*Model*	*n*	r^2	Q^2	*n*	r^2_{pred}
Model in this study	**5**	**3**	**PLS**	**76**	**0.826**	**0.795**	**22**	**0.846**
Ambure et al. 2016 [17]	5	4	PLS	52	0.831	0.764	22	0.813
Ambure et al. 2016 [17]	5	-	MLR	51	0.826	0.764	22	0.791
Jain et al. 2013 [18]	2	-	MLR	20	0.895	0.893	7	0.903
Hossain et al. 2013 [19]	-	10	CoMFA	71	0.998	0.765	35	0.772
Hossain et al. 2013 [19]	-	10	CoMSIA	71	0.992	0.730	35	0.713
Hossain et al. 2013 [19]	-	7	PLS	71	0.941	0.792	35	0.713
Chakraborty et al. 2017 [20]	4	-	LHM	20	0.941	0.913	10	0.860

Abbreviations: EL = Equation length, Compts = Number of components, MLR = Multiple linear regression, CoMFA = Comparative Molecular Field Analysis, CoMSIA = Comparative molecular similarity index analysis and LHM = Linear heuristic method.

compounds, the reported model in the present study may be used for screening purpose for discovery and development of leads against BACE1 enzyme.

Overview and conclusions

In this study, we have developed a PLS-regression-based 2D-QSAR model from 98 diverse compounds [21] having defined BACE1 enzyme inhibitory activity to investigate the structural requirements or molecular properties essential for the enzyme inhibitory activity. The QSAR model was developed with simple, meaningful and easily interpretable descriptors. Prior to the development of the final model, we have also performed variable selection strategy using stepwise regression followed by best subset selection. The statistical results of the developed model shows good predictivity based on both internal and external validation parameters. The PLS model was developed by following strict OECD guidelines (a defined endpoint, unambiguous algorithm, acceptable quantitative metrics, applicability domain analysis and mechanistic interpretation). The information obtained from PLS model (as also demonstrated in the regression coefficient plot), variable importance plot and loading plot (see Figure S1 and S4 in Supplementary materials S3), we have concluded that: (i) the presence of hydrogen bond donor groups like –OH, $–NH_2$, –SH, etc., may enhance the inhibitory activity against BACE1 enzyme; (ii) higher number of heteroatoms with a lone pair of electrons capable of resonance with an aromatic nucleus are essential to increase the inhibitory activity against BACE1 enzyme; (iii) the R–CH.X fragment is favourable to enhance the inhibitory activity of β-secretase enzyme inhibitors; (iv) a higher number of H atoms attached to C^1(sp3)/C°(sp2) and presence of N–S fragments at topological distance 3 in the molecules is detrimental to the enzyme inhibitory activity. Alzheimer's disease is one of those circumstances, where genetics is known to play an insightful role but not the sole factor in the disease development and progression. There are intensifying evidences that the environment has a great deal to do with the development of this neurodegenerative disorder. Environment contains a number of chemical elements such as lead, aluminium, arsenic, nitrogen oxides, carbon monoxide, silicon, selenium, etc., which have effects on the disease progression. Relatively few studies [55,56] have examined the influence of these toxic chemical exposures on the risk of dementia/cognitive decline. Furthermore, the

results obtained from pharmacophore mapping and molecular docking studies are well supported with the QSAR analysis. The developed 2D-QSAR model thus may be helpful for prediction of the activity of new analogues even before their synthesis and evaluation.

Acknowledgements

VK thanks Indian Council of Medical Research, New Delhi for a senior research fellowship. PKO thanks University Grant Commission, New Delhi, Government of India for his financial assistance (Letter number and date: F./PDFSS-2015-17-WES-11996; Dated:06/04/2016).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Indian Council of Medical Research.

ORCID

A. Saha http://orcid.org/0000-0002-0205-7719
K. Roy http://orcid.org/0000-0003-4486-8074

References

[1] R. Vassar, D.M. Kovacs, R. Yan, and P.C. Wong, *The β-secretase enzyme BACE in health and Alzheimer's disease: Regulation, cell biology, function, and therapeutic potential*, J. Neurosc. 29 (2009), pp. 12787–12794. doi:10.1523/JNEUROSCI.3657-09.2009.

[2] N. Schaduangrat, V. Prachayasittikul, S. Choomwattana, P. Wongchitrat, K. Phopin, W. Suwanjang, A.A. Malik, B. Vincent, and C. Nantasenamat, *Multidisciplinary approaches for targeting the secretase protein family as a therapeutic route for Alzheimer's disease*, Med. Res. Rev. 39 (2019), pp. 1730–1778. doi:10.1002/med.21563.

[3] R.J. Vassar, *BACE1, The beta-secretase enzyme: A leading therapeutic target for Alzheimer's disease*, Alzheimers Dement. 12 (2016), p. 162. doi:10.1016/j.jalz.2016.06.289.

[4] J. Hardy and D.J. Selkoe, *The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics*, Science 297 (2002), pp. 353–356. doi:10.1126/science.1072994.

[5] K.W. Menting and J.A. Claassen, *β-secretase inhibitor; a promising novel therapeutic drug in Alzheimer's disease*, Front. Aging. Neurosci. 6 (2014), p. 165. doi:10.3389/fnagi.2014.00165.

[6] S. Kitazume, Y. Tachida, R. Oka, K. Shirotani, T.C. Saido, and Y. Hashimoto, *Alzheimer's β-secretase, β-site amyloid precursor protein-cleaving enzyme, is responsible for cleavage secretion of a Golgi-resident sialyltransferase*, PNAS 98 (2001), pp. 13554–13559. doi:10.1073/pnas.241509198.

[7] J.A. Hardy and G.A. Higgins, *Alzheimer's disease: The amyloid cascade hypothesis*, Science 256 (1992), pp. 184–186. doi:10.1126/science.1566067.

[8] R. Yan, M.J. Bienkowski, M.E. Shuck, H. Miao, M.C. Tory, A.M. Pauley, J.R. Brashler, N. C. Stratman, W.R. Mathews, A.E. Buhl, and D.B. Carter, *Membrane-anchored aspartyl protease with Alzheimer's disease β-secretase activity*, Nature 402 (1999), pp. 533–537. doi:10.1038/990107.

[9] A.K. Ghosh, S. Gemma, and J. Tang, *β-Secretase as a therapeutic target for Alzheimer's disease*, Neurotherapeutics 5 (2008), pp. 399–408. doi:10.1016/j.nurt.2008.05.007.

[10] G. Subramanian, B. Ramsundar, V. Pande, and R.A. Denny, *Computational modeling of β-secretase 1 (BACE-1) inhibitors using ligand based approaches*, J. Chem. Inf. Model. 56 (2016), pp. 1936–1949. doi:10.1021/acs.jcim.6b00290.

[11] S.L. Cole and R. Vassar, *The Alzheimer's disease β-secretase enzyme, BACE1*, Mol. Neurodegener. 2 (2007), p. 22. doi:10.1186/1750-1326-2-22.

[12] S.J. Patey, E.A. Edwards, E.A. Yates, and J.E. Turnbull, *Heparin derivatives as inhibitors of BACE-1, the Alzheimer's β-secretase, with reduced activity against factor Xa and other proteases*, J. Med. Chem. 49 (2006), pp. 6129–6132. doi:10.1021/jm051221o.

[13] A. Kumar, S. Roy, S. Tripathi, and A. Sharma, *Molecular docking based virtual screening of natural compounds as potential BACE1 inhibitors: 3D QSAR pharmacophore mapping and molecular dynamics analysis*, J. Biomol. Struct. Dyn. 34 (2016), pp. 239–249. doi:10.1080/07391102.2015.1022603.

[14] K. Roy, P. Chakraborty, I. Mitra, P.K. Ojha, S. Kar, and R.N. Das, *Some case studies on application of "rm2" metrics for judging quality of quantitative structure–Activity relationship predictions: Emphasis on scaling of response data*, J. Comput. Chem. 34 (2013), pp. 1071–1082. doi:10.1002/jcc.23231.

[15] T. Puzyn, J. Leszczynski, and M.T. Cronin (eds.), *Recent Advances in QSAR Studies: Methods and Applications*, Vol. 8, Springer Science & Business Media, 2010. doi:4020-9783-6",1,0,0>10.1007/978-1-4020-9783-6.

[16] M.B. Colovic, D.Z. Krstic, P.T.D. Lazarevic, A.M. Bondzic, and V.M. Vasic, *Acetylcholinesterase inhibitors: Pharmacology and toxicology*, Curr. Neuropharmacol. 11 (2013), pp. 315–335. doi:10.2174/1570159x11311030006.

[17] P. Ambure and K. Roy, *Understanding the structural requirements of cyclic sulfone hydroxyethylamines as hBACE1 inhibitors against Aβ plaques in Alzheimer's disease: A predictive QSAR approach*, RSC Adv. 6 (2016), pp. 28171–28186. doi:10.1039/C6RA04104C.

[18] P. Jain and H.R. Jadhav, *Quantitative structure activity relationship analysis of aminoimidazoles as BACE-I inhibitors*, Med. Chem. Res. 22 (2013), pp. 1740–1746. doi:10.1007/s00044-012-0166-z.

[19] T. Hossain, M.A. Islam, R. Pal, and A. Saha, *Exploring structural requirement and binding interactions of β-amyloid cleavage enzyme inhibitors using molecular modeling techniques*, Med. Chem. Res. 22 (2013), pp. 4766–4774. doi:10.1007/s00044-013-0481-z.

[20] S. Chakraborty and S. Basu, *Multi-functional activities of citrus flavonoid narirutin in Alzheimer's disease therapeutics: An integrated screening approach and in vitro validation*, Int. J. Biol. Macromol. 103 (2017), pp. 733–743. doi:10.1016/j.ijbiomac.2017.05.110.

[21] M.K. Gilson, T. Liu, M.G. Baitaluk, G. Nicola, L. Hwang, and J. Chong, *BindingDB: A public database for medicinal chemistry, computational chemistry and systems pharmacology*, Nucleic Acids Res. 19 (2015), pp. 1045–1053. Available at https://www.bindingdb.org/bind/index.jsp.

[22] S. Das, P.K. Ojha, and K. Roy, *Multilayered variable selection in QSPR: A case study of modeling melting point of bromide ionic liquids*, IJQSPR 2 (2017), pp. 106–124. doi:10.4018/IJQSPR.2017010108.

[23] K. Roy, S. Kar, and P. Ambure, *On a simple approach for determining applicability domain of QSAR models*, Chemom. Intell. Lab. Syst. 145 (2015), pp. 22–29. doi:10.1016/j.chemolab.2015.04.013.

[24] J. Jiaranaikulwanitch, C. Boonyarat, V.V. Fokin, and O. Vajragupta, *Triazolyl tryptoline derivatives as β-secretase inhibitors*, Bioorg. Med. Chem. Lett. 20 (2010), pp. 6572–6576. doi:10.1016/j.bmcl.2010.09.043.

[25] P. Nowak, D.C. Cole, A. Aulabaugh, J. Bard, R. Chopra, R. Cowling, K.Y. Fan, B. Hu, S. Jacobsen, M. Jani, and G. Jin, *Discovery and initial optimization of 5, 5′-disubstituted aminohydantoins as potent β-secretase (BACE1) inhibitors*, Bioorg. Med. Chem. Lett. 20 (2010), pp. 632–635. doi:10.1016/j.bmcl.2009.11.052.

[26] M.S. Malamas, K. Barnes, Y. Hui, M. Johnson, F. Lovering, J. Condon, W. Fobare, W. Solvibile, J. Turner, Y. Hu, and E.S. Manas, *Novel pyrrolyl 2-aminopyridines as potent and selective human β-secretase (BACE1) inhibitors*, Bioorg. Med. Chem. Lett. 20 (2010), pp. 2068–2073. doi:10.1016/j.bmcl.2010.02.075.

[27] G. Larbig and B. Schmidt, *Synthesis of tetramic and tetronic acids as β-secretase inhibitors*, J. Comb. Chem. 8 (2006), pp. 480–490. doi:10.1021/cc0600021.
[28] Y. Niu, C. Ma, H. Jin, F. Xu, H. Gao, P. Liu, Y. Li, C. Wang, G. Yang, and P. Xu, *The discovery of novel β-Secretase inhibitors: Pharmacophore modeling, virtual screening, and docking studies*, Chem. Bio. Drug Des. 79 (2012), pp. 972–980. doi:10.1111/j.1747-0285.2012.01367.x.
[29] D.C. Cole, E.S. Manas, J.R. Stock, J.S. Condon, L.D. Jennings, A. Aulabaugh, R. Chopra, R. Cowling, J.W. Ellingboe, K.Y. Fan, and B.L. Harrison, *Acylguanidines as small-molecule β-secretase inhibitors*, J. Med. Chem. 49 (2006), pp. 6158–6161. doi:10.1021/jm0607451.
[30] T.H. Al-Tel, R.A. Al-Qawasmeh, M.F. Schmidt, A. Al-Aboudi, S.N. Rao, S.S. Sabri, and W. Voelter, *Rational design and synthesis of potent dibenzazepine motifs as β-secretase inhibitors*, J. Med. Chem. 52 (2009), pp. 6484–6488. doi:10.1021/jm9008482.
[31] Marvin ChemAxon Software Solutions and Services for Chemistry, Available at https://chemaxon.com/products/marvin
[32] P. Ambure, J. Bhat, T. Puzyn, and K. Roy, *Identifying natural compounds as multi-target-directed ligands against Alzheimer's disease: An in silico approach*, J. Biomol. Struct. Dyn. 37 (2019), pp. 1282–1306. doi:10.1080/07391102.2018.1456975.
[33] A. Mauri, V. Consonni, M. Pavan, and R. Todeschini, *Dragon software: An easy approach to molecular descriptor calculations*, Match 56 (2006), pp. 237–248. Available at http://www.talete.mi.it/products/dragon_description.htm.
[34] C.W. Yap, *PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints*, J. Comput. Chem. 32 (2011), pp. 1466–1474. doi:10.1002/jcc.21707.
[35] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors, Method and Principles in Medicinal Chemistry*, Vol. 11, John Wiley & Sons, 2008.
[36] H.S. Park and C.H. Jun, *A simple and fast algorithm for K-medoids clustering*, Expert Syst. Appl. 36 (2009), pp. 3336–3341. doi:10.1016/j.eswa.2008.01.039.
[37] K. Roy, R.N. Das, P. Ambure, and R.B. Aher, *Be aware of error measures. Further studies on validation of predictive QSAR models*, Chemom. Intell. Lab. Syst. 152 (2016), pp. 18–33. doi:10.1016/j.chemolab.2016.01.008.
[38] I.N. Minitab, *MINITAB statistical software*, Minitab Release 13 (2000). Available at http://www.minitab.com/en-us/products/minitab/.
[39] K. Roy, *On some aspects of validation of predictive quantitative structure–Activity relationship models*, Expert Opin. Drug .v 2 (2007), pp. 1567–1577. doi:10.1517/17460441.2.12.1567.
[40] A. Umetri, *SIMCA-P for Windows, Graphical Software for Multivariate Process Modeling*, Umea, Sweden, 1996. Available at https://umetrics.com/products/simca.
[41] R.B. Aher and K. Roy, *QSAR and pharmacophore modeling of diverse aminothiazoles and aminopyridines for antimalarial potency against multidrug-resistant* Plasmodium falciparum, Med. Chem. Res. 23 (2014), pp. 4238–4249. doi:10.1007/s00044-014-0997-x.
[42] D. Systemes, *BIOVIA, Discovery Studio Modeling Environment*, Release 4.5, Dassault Systemes, San Diego, CA, 2015. Available at https://www.3dsbiovia.com/.
[43] H.J. Huang, C.C. Lee, and C.Y.C. Chen, *In silico Investigation of traditional chinese medicine compounds to inhibit human histone deacetylase 2 for patients with Alzheimer's disease*, BioMed Res. Int. 2014 (2014), pp. 1–15. doi:10.1155/2014/741703.
[44] S. Pal, V. Kumar, B. Kundu, D. Bhattacharya, N. Preethy, M.P. Reddy, and A. Talukdar, *Ligand-based pharmacophore modeling, virtual screening and molecular docking studies for discovery of potential topoisomerase i inhibitors*, CSBJ 17 (2019), pp. 291–310. doi:10.1016/j.csbj.2019.02.006.
[45] A.G. Murzin, S.E. Brenner, T. Hubbard, and C. Chothia, *SCOP: A structural classification of proteins database for the investigation of sequences and structures*, J. Mol. Biol. 247 (1995), pp. 536–540. doi:(05)80134-2",1,0,0>10.1016/S0022-2836(05)80134-2.
[46] G. Wu, D.H. Robertson, C.L. Brooks, and M. Vieth, *Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm*, J. Comput. Chem. 24 (2003), pp. 1549–1562. doi:10.1002/jcc.10306.
[47] P.L.A. Popelier and P.J. Smith, *QSAR models based on quantum topological molecular similarity*, Eur. J. Med. Chem. 41 (2006), pp. 862–873. doi:10.1016/j.ejmech.2006.03.004.

[48] K. Roy and R.N. Das, *The "ETA" indices in QSAR/QSPR/QSTR research, in pharmaceutical sciences: Breakthroughs in research and practice*, IGI Global (2017), pp. 978–1011. doi:5225-1762-7. ch038",1,0,0>10.4018/978-1-5225-1762-7.ch038.
[49] R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical listing/volume II: Appendices, References*, Vol. 41, John Wiley & Sons, New Jersey, 2009.
[50] P.A. Speck, V.V. Kleandrova, and V.J.A. Rojas, *QSAR model toward the rational design of new agrochemical fungicides with a defined resistance risk using substructural descriptors*, Mol. Divers. 15 (2011), pp. 901–909. doi:10.1007/s11030-011-9320-7.
[51] B. Lei, J. Li, J. Lu, J. Du, H. Liu, and X. Yao, *Rational prediction of the herbicidal activities of novel protoporphyrinogen oxidase inhibitors by quantitative structure activity relationship model based on docking-guided active conformation*, J. Agr. Food. Chem. 57 (2009), pp. 9593–9598. doi:10.1021/jf902010g.
[52] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, and R.K. Agrawal, *Validation of QSAR models-strategies and importance*, RRJoDDD 3 (2011), pp. 511–519.
[53] D. Gadaleta, G.F. Mangiatordi, M. Catto, A. Carotti, and O. Nicolotti, *Applicability domain for QSAR models: Where theory meets reality*, IJQSPR 1 (2016), pp. 45–63. doi:10.4018/IJQSPR.2016010102.
[54] C. Yoo and M. Shahlaei, *The applications of PCA in QSAR studies: A case study on CCR5 antagonists*, Chem. Biol. Drug. Des. 9 (2018), pp. 137–152. doi:10.1111/cbdd.13064.
[55] L.O. Killin, J.M. Starr, I.J. Shiue, and T.C. Russ, *Environmental risk factors for dementia: A systematic review*, BMC Geriatr. 16 (2016), p. 175. doi:10.1186/s12877-016-0342-y.
[56] P.V. Moulton and W. Yang, *Air pollution, oxidative stress, and Alzheimer's disease*, J. Environ. Res. Public Health (2012), p. 9. doi:10.1155/2012/472751.

Computers in Biology and Medicine 118 (2020) 103658

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

Cheminformatic modelling of β-amyloid aggregation inhibitory activity against Alzheimer's disease

Vinay Kumar [a], Probir Kumar Ojha [a], Achintya Saha [b], Kunal Roy [a,*]

[a] Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India
[b] Department of Chemical Technology, University of Calcutta, 92 A P C Road, Kolkata, 700 032, India

ARTICLE INFO

Keywords:
2D-QSAR
Amyloid-β peptide
Validation
Alzheimer's disease
pharmacophore
molecular docking

ABSTRACT

In the current research, we have developed robust two-dimensional quantitative structure-activity relationship (2D-QSAR) and pharmacophore models using a dataset of 314 heterocyclic β-amyloid aggregation inhibitors. The main purpose of this study is to determine the essential structural features which are responsible for the inhibition of β-amyloid aggregation. Prior to the development of the 2D-QSAR model, we applied a multilayered variable selection method to reduce the size of the pool of descriptors, and the final models were built by the partial least squares (PLS) regression technique. The models obtained were thoroughly analysed by applying both internal and external validation parameters. The validation metrics obtained from the analysis suggested that the developed models were significant and sufficient to predict the inhibitory activity of unknown compounds. The structural features obtained from the pharmacophore model, such as the presence of aromatic rings and hydrogen bond acceptor/donor or hydrophobic sites, are well corroborated with those of the 2D-QSAR models. Additionally, we also performed a molecular docking study to understand the molecular interactions involved in binding, and the results were then correlated with the requisite structural features obtained from the 2D-QSAR and 3D-pharmacophore models.

1. Introduction

Alzheimer's disease (AD) is the most common form of dementia. It is characterized by neuronal death with deposition of an abnormal β-amyloid peptide within the brain tissue [1,2]. The amyloid-β peptide is a 39 to 43 amino acid peptide derived from a transmembrane glycoprotein (amyloid precursor protein) [3]. The development of extracellular deposits of amyloid plaques is one of the major indications of AD [1]. According to the previous research, AD is a serious disorder with a long preclinical period and progressive course [4]. Physiologically, amyloid plaques develop in the hippocampus, which is a structure situated deep in the brain that helps to encode memories, and in other areas of the cerebral cortex that are involved in intellectual and occupational functions [4]. It is still unknown whether amyloid plaques themselves cause AD or if they are by-products of the AD process [5]. Still no unifying hypothesis has been proposed to integrate the aggregation and neurotoxic properties of β-amyloid peptide (β-AP) into the framework of disparate biochemical abnormalities observed in AD [6]. There are many drug candidates identified that target Aβ for AD treatment, including beta-secretase 1 (BACE1) inhibitors, γ-secretase inhibitors, Aβ aggregation inhibitors, and Aβ antibodies [2]. But, initially, it was recognized that toxicity is associated with mature fibres. The majority of inhibitors have been directed toward identifying modulators of Aβ fibrillation [2]. In the current area of research, there is an urgent need to develop a novel treatment strategy to block the various key steps in the amyloidosis process as investigated by Kokkoni et al., in 2006 [6]. As per our knowledge, the most effective strategy for designing a molecule to inhibit the initial step of amyloid β-aggregation is one that blocks the production of the amyloidogenic protein.

Development of a novel drug is a time-consuming process, and it requires a huge amount of material and financial resources [7]. In this respect, computational approaches like quantitative structure-activity relationship (QSAR), pharmacophore modelling, and molecular docking etc. are playing important roles for the discovery and development of new compounds with improved therapeutic activity [7]. The QSAR methods are important for the investigation of important structural features and prediction of the biological activity of novel compounds based on mathematical and statistical relations [8,9]. The idea of QSAR is based on the notion that end point values of compounds change systematically with modification of the structural attributes [8,9]. There have been a large number of computational studies performed (Leal

* Corresponding author.
E-mail address: kunal.roy@jadavpuruniversity.in (K. Roy).

https://doi.org/10.1016/j.compbiomed.2020.103658
Received 26 December 2019; Received in revised form 6 February 2020; Accepted 8 February 2020
Available online 13 February 2020

Abbreviations

βAAI	β-amyloid aggregation inhibitors
AD	Alzheimer's disease
AD	Applicability domain
VIP	Variable importance plot
PLS	Partial least squares
SW	Stepwise
GAs	Genetic algorithm
ETA	Extended topochemical atom
LOO	Leave one out
QSAR	Quantitative structure activity relationship
OECD	Organization for economic co-operation and development
MAE	Mean absolute error
CCC	Concordance correlation coefficient
APP	Amyloid precursor protein
KNIME	Konstanz Information Miner
BSS	Best subset selection
LV	Latent variable
CHARMm	Chemistry at Harvard Macromolecular Mechanics
MLR	Multilinear Regression
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative molecular similarity index analysis
Aβ	Amyloid-β
SDF	Structural data format
DTC lab	Drug Theoretics and chemoinformatics lab
HQSAR	Hologram QSAR

et al. [10], Zhao et al. [11], Aswathy et al. [12] and Hossein et al. [13]) so far which have designed new inhibitors against AD, but still we are far from finding a precise treatment strategy for AD [14]. In the current investigation, we have applied a dataset of 314 heterocyclic compounds (downloaded from BindingDB data base) [15] with β-amyloid aggregation inhibitory activity for the purpose of QSAR model development in order to explore the key structural features that are essential for β-amyloid aggregation inhibitory activity. Prior to the development of the final models, we have applied a multilayered variable selection approach to reduce noise in the input, and the final models were developed using the Partial Least Squares (PLS) regression technique. The two-dimensional QSAR (2D-QSAR) models were built with the guidelines of the Organization for Economic Cooperation and Development (OECD) [16]. The developed models were validated by using various strict internal and external validation metrics [17]. Moreover, we also performed pharmacophore modelling to reveal the structural requirements for inhibitory activity against β-amyloid aggregation and also to categorize the compounds into active and less active classes. Furthermore, we performed a molecular docking study with the most active and least active compounds from the dataset and tried to justify the contributions of different descriptors/features obtained from QSAR/pharmacophore models.

2. Materials and methods

2.1. QSAR methodology

2.1.1. The dataset

In this study, we collected 314 heterocyclic compounds (β-amyloid aggregation inhibitors) from the BindingDB database [15] (see S1 sheet 1 in supplementary materials) with β-amyloid aggregation inhibitory activity for the purpose of QSAR model development. The experimental IC_{50} values (nM) of the dataset compounds were converted to pIC_{50} (=-logIC_{50}) values for model development purposes. The dataset compounds utilized in this study followed the same experimental protocol (Thioflavin T (ThT) spectrofluorometric assay method) [18–37]. All the compounds were carefully checked and filtered using different software and tools like KNIME (https://dtclab.webs.com/software-tools), MarvinView and MarvinSketch [38]. The compounds were drawn by using the MarvinView ChemAxon tool [38] and saved in the MDL.mol format.

2.1.2. Preliminary dataset preparation and data curation

Chemical curation is very important when researchers collect data from different sources [7]. In this work, before the development of the regression models, we implemented preliminary dataset preparation and a data curation (chemical and biological) strategy using KNIME work flow (https://dtclab.webs.com/software-tools). The accuracy of KNIME workflow was confirmed by Mariana et al., 2017 [39], Domenico et al., 2018 [40] and Fabian P et al., 2015 [41]. The dataset was downloaded from BindingDB [15] in a structural data format (SDF) containing important information related to the structure and endpoint values against the β-amyloid aggregation. An identifier was given to each and every compound present in the dataset, characterizing the name of a respective protein and a serial number. We extracted the endpoint values from the dataset file to classify the compounds in four orders of magnitude. At the end, we incorporated the dataset to chemical and biological curation.

2.1.2.1. Reading and storing the information obtained from the binding database. In this methodology, we utilized the "SDF reader" using KNIME workflow (https://dtclab.webs.com/software-tools) for the purpose of reading the input file and storing the important structural features of the compounds. The downloaded dataset contained all of the essential information related to the compounds such as the molecule name, coordinates, bond counts, bond order, number of rings, end point, biological assay, etc. The compounds with incorrect information were deleted from the source file and the compounds with correct information were saved for further use.

2.1.2.2. Elimination of salts, mixtures, inorganics and organo-metallics from the dataset. In this study, we removed all of the salts, mixtures, in-organics, and organo-metallic compounds before calculating the descriptors using 'RDKit Salt Stripper', 'Connectivity' node and 'Element Filter' node [42], respectively.

2.1.2.3. Standardization of chemical structures. Normalization is an important step in QSAR modelling to correct the structural pattern before the molecular descriptors calculation. In the current work, we applied the 'RDKit Structure Normalizer' node for correction of the geometry of the chemical structure using the KNIME workflow (https://dtclab.webs.com/software-tools).

2.1.2.4. Biological curation. The molecules obtained from chemical curation were subjected to biological curation. Biological data curation is one of the important steps in the QSAR modelling. In the current work, we implemented biological data curation as described by Ambure et al. [14] using the KNIME workflow (http://dtclab.webs.com/software-tools) for the purpose of duplicate identification and activity cliff determination.

2.1.3. Descriptor calculation and data pre-treatment

A pool of 457 descriptors was computed by applying two software tools, namely, Dragon version 7 [43] and PaDEL-Descriptor software version 2.20 [44]. In this work, we computed only 2D descriptors including constitutional, ring descriptors, connectivity index, functional

group counts, atom centred fragments, 2D atom pairs, atom type E-states, molecular properties and extended topochemical atom (ETA) indices. After descriptor calculation, we performed the data pre-treatment using the tool Pre-treatment V-WSP version 1.2 (http://dtclab.webs.com/software-tools) to discard the descriptors with incomplete information or with nearly constant values.

2.1.4. Dataset division

In this study, our aim was to develop QSAR models having good reliable prediction ability. Therefore, QSAR models were developed by using a training set and validated using new chemical entities, i.e., a test set to check the predictive capacity of the developed models. In this study, the whole data set (n = 314) was divided into a training set (n = 252, 80% of the total number of compounds), and a test set (n = 62, 20% of the total number of compounds) based on an Euclidean distance based algorithm using the "Dataset Division GUI" developed by our group (https://dtclab.webs.com/software-tools).

2.1.5. Multilayered variable selection strategy and model development

Prior to the development of the final model, we tried to extract the important descriptors from the large pool of initial descriptors using various variable selection strategies. For this purpose, we applied a multilayered variable selection strategy before the development of the final model using multistage stepwise regression (using a suitable stepping criterion, e.g., 'F-for-inclusion 'and 'F-for-exclusion' based on partial F-statistic) followed by a genetic algorithm (GA) followed by the best subset selection, and the final models were built by using partial least squares (PLS) regression techniques. The detailed multi-layered variable selection strategy is schematically represented in Fig. S1 (See S3 in supplementary information).

2.1.6. Statistical validation metrics

Validation of the robustness and predictive ability of the developed models is an important step in a QSAR study [17]. In this study, we employed different statistical approaches such as internal and external validation metrics to justify the robustness and predictive quality of the developed models. In the case of internal validation, we determined various statistical metrics such as determination coefficient (r^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), Avg $r^2_{m(LOO)}$, and Δr^2_m [17]. Higher values of the metrics r^2, $Q^2_{(LOO)}$) and Avg $r^2_{m(LOO)}$ indicated a better fit of the model, but all of these parameters are not sufficient to evaluate the robustness and predictivity of significant models [17]. Thus, we determined other statistical validation parameters (external validation parameters) such as Q^2F_1, Q^2F_2, r^2_m parameters like average $r^2_{m(test)}$ and Δr^2_m and concordance correlation coefficient (CCC) to assure the significance of the developed models [17]. The details of validation metrics are depicted in Table S1 (See S4 in supplementary information). Moreover, we also performed a Y-randomization test, checked applicability domain criteria, etc. to investigate the robustness of the developed models. The Y-randomization test was performed using the Simca-P software [45] by randomly reordering (100 permutations) the dependent variable [46]. The details of the methodology are depicted in Fig. S2. (See S3 in supplementary information).

2.2. Development and validation of 3D-pharmacophore model

In the current work, we performed a pharmacophore modelling study to reveal the required features which are essential for β-amyloid aggregation inhibitory activity. The β-amyloid aggregation inhibitory activity stated in terms of IC_{50} (nM) was used as dependent variable for the purpose of the development of pharmacophore models. Previously prepared compound structures were used for this study. The dataset was rationally distributed into training (62 compounds) (for model development) and test sets (252 compounds for validation) based on the biological activity values spanned over four orders of magnitude [47–49]. The BIOVIA Discovery Studio Client 4.1 [50] platform was used to build the pharmacophore models. The details of the methodology for the development of the pharmacophore model are as described by Aher et al. [51]. Validation of the developed models was performed using different parameters like cost analysis, the Fischer randomization test (F-test), and test set prediction in order to judge the robustness and predictive quality of models as described by Aher et al. [51].

2.3. Molecular docking studies

Investigation of important structural features that will be helpful for development of novel inhibitors which control the aggregation of β-amyloid was the goal of this study. Here, we performed a molecular docking study for the purpose of identifying of the interaction pattern between the β-amyloid peptide (PDB ID:1IYT [52]) and selected β-amyloid aggregation inhibitors from the dataset. Molecular docking studies were performed by using BIOVIA Discovery Studio client 4.1 [50] platform using the CDOCKER module of receptor-ligand interactions [53]. After docking, the generated poses were sorted according to CDOCKER interaction energy and the top scoring poses were kept for further analysis.

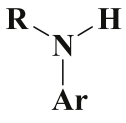
3. Results and discussion

3.1. Mechanistic interpretation of modelled descriptors

In this analysis, we developed statistically significant and robust 2D-QSAR models using the PLS regression based technique; the values of the validation parameters are shown below in Equations 1 and 2 (Box 1). The developed models were validated carefully utilizing the deferent validation metrics (internal and external) for exploration of robust and statistically significant models. The details of their validation metric values are also summarized in Table S1 (see S4 in supplementary materials). Two PLS models were developed with 12 and 13 descriptors using 6 and 7 latent variables (LVs), respectively. The obtained results suggested that the models were acceptable in terms of fitness, stability and classical predictivity measures. Descriptors appearing in the developed models demonstrated the structural and functional requirements which can improve the inhibitory activity of molecules against β-amyloid aggregation. We have presented the scatter plots in Fig. 1 to show the closeness of the observed and predicted values for the β-amyloid aggregation inhibitors (βAAI). The randomization of models assured that the developed models were not found by any chance correlation. The results obtained from the randomized models (for Model 1: R^2intercepts (int) = 0.00833 and Q^2int = −0.296 and for Model 2: R^2int = 0.00756 and Q^2int = −0.392) suggested that the reported models were not obtained by chance (see S3 Fig. S3 in supplementary materials).

The descriptors in the PLS models are arranged according to their importance, and then described separately. The significance level and contribution of the model descriptors towards the β-amyloid aggregation inhibitory activity are determined based on a regression coefficient plot [54] and variable importance plot (VIP) [55] as shown in Figs. S4 and S5 (See S3 in supplementary information). The definition and contributions of all the descriptors obtained from the PLS models are mentioned in Table S2 (see S4 in supplementary materials).

The highest contributing descriptor, nArNHR, a functional group count descriptor, denotes the number of secondary aromatic amines present in the compounds.



The lone pair of electrons on nitrogen is delocalized in the aromatic

Box 1
PLS based 2D QSAR models and their statistical validation metrics obtained from the developed models

Model 1: PLS Model against β-amyloid aggregation

$$pIC_{50}(nM) = -3.017 + 0.843 \times nArNHR + 0.178 \times SaasC - 0.102 \times F07[C-O] - 1.097 \times NNRS - 0.006 \times D/Dtr12 - 1.185 \times B06[C-N] - 0.730 \times B05[N-N] - 0.655 \times B06[N-O] + 1.610 \times B05[C-N] + 0.079 \times F09[C-C] - 1.051 \times nFuranes - 0.031 \times F06[C-C] \quad (1)$$

Internal Validation Parameters: $n_{training} = 252$, $r^2 = 0.664$, $Q^2 = 0.621$, Fitting quality = Moderate.
External Validation Parameters: $n_{test} = 62$, $Q^2F_1 = 0.765$, $Q^2F_2 = 0.763$, Avg $R^2m = 0.601$, $\Delta R^2m = 0.199$, CCC = 0.861, MAE = 0.456, SD = 0.394, Fitting quality = Moderate, LV = 6, No. of descriptors = 12

Model 2: PLS Model against β-amyloid aggregation

$$pIC_{50}(nM) = -2.842 + 0.754 \times nArNHR + 0.191 \times SaasC - 0.114 \times F07[C-O] - 1.089 \times NNRS - 0.005 \times D/Dtr12 - 0.739 \times B05[N-N] - 1.159 \times B06[C-N] - 0.539 \times B06[N-O] + 1.427 \times B05[C-N] + 0.813 \times F05[O-O] - 1.274 \times nFuranes + 0.076 \times F09[C-C] - 0.032 \times F06[C-C] \quad (2)$$

Internal Validation Parameters: $n_{training} = 252$, $r^2 = 0.684$, $Q^2 = 0.638$, Fitting quality = Moderate.
External Validation Parameters: $n_{test} = 62$, $Q^2F_1 = 0.771$, $Q^2F_2 = 0.769$, Avg $R^2m = 0.634$, $\Delta R^2m = 0.186$, CCC = 0.867, MAE = 0.462, SD = 0.375
Fitting quality = Good, LV = 7, No. of descriptors = 13

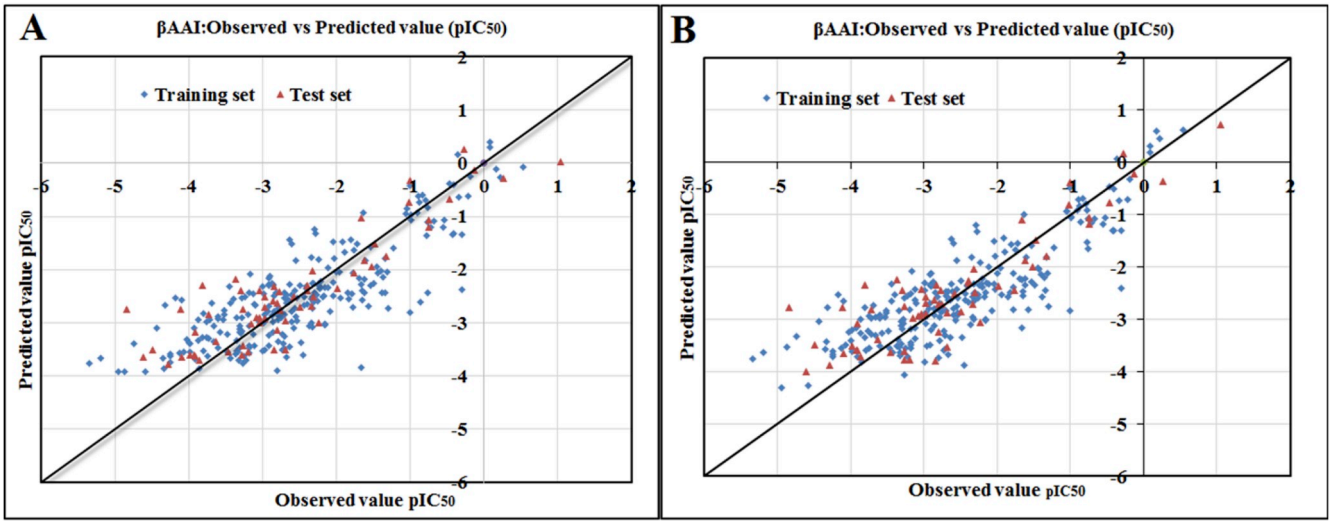


Fig. 1. The scatter plots of observed and predicted values of final PLS models for β-amyloid aggregation inhibitors (**A**: Model 1, **B**: Model 2).

ring thus reducing the electron density on nitrogen [56]. The positive regression coefficient of this descriptor indicated that the activity of inhibitors is directly proportional to the numerical value of nArNHR. Thus, the compounds having a higher number of secondary aromatic amines may enhance the β-amyloid aggregation inhibitory activity as shown in (Fig. 2). Compounds like **88** (pIC_{50}: −0.217), **67** (pIC_{50}: 0.537), and **231** (pIC_{50}: −0.190) and their corresponding descriptor values are 1 each. In contrast, compounds like **294** (pIC_{50}: −5.338), **293** (pIC_{50}: −5.190), and **87** (pIC_{50}: −4.949) have lower inhibitory activity, because these compounds have no such fragment (Fig. 2). From these observations, we have concluded that a secondary aromatic amine is important for β-amyloid aggregation inhibitory activity.

The next significant descriptor, SaasC, an E-state index descriptor, denotes the sum of the atom level E-state values for all non-substituted aromatic carbon atoms (Kier and Hall, 1999) [57]. According to Kier and Hall 1999 [57], each atom or bond in molecules has their intrinsic state, which is altered by every other atom or bond in the same molecule, encoding information related to electronic distribution and topological aspects. SaasC is related to aromatic carbons with an attached substituent atom [58]. The positive regression coefficient of this descriptor indicates that the presence of aromatic carbons with an attached substituent atom is beneficial for the inhibitory activity as shown in compounds (Fig. 3) **71** (pIC_{50}: 0.0861), **70** (pIC_{50}: 0.086), and **74** (pIC_{50}: −0.357), and their corresponding descriptor values are 9.396, 8.807, and 8.779, respectively. On the other hand, the absence of such fragments is detrimental to β-amyloid aggregation inhibitory activity as observed in (Fig. 3) compounds **93** (pIC_{50}: −3.978) and **341** (pIC_{50}: −3.469).

The next significant descriptor, F07 [C–O] is a 2D atom pair descriptor that accounts for the frequency of C and O atoms at topological distance 7. It contributes negatively toward the endpoint value, which suggests that the numerical values of the descriptor are inversely proportional to the inhibitory activity. Thus, the compounds bearing such fragments show lower values of inhibitory activity as evidenced by (Fig. 3) compounds **153** (pIC_{50}: −3.176), **86** (pIC_{50}: −4.869), and **294** (pIC_{50}: −5.338) (their corresponding descriptor values are 14, 13, and 10, respectively). Whereas, compounds having no such fragments show higher β-amyloid aggregation inhibitory activity as shown in compounds **231** (pIC_{50}: −0.190), **106** (pIC_{50}: −0.301) and **201** (pIC_{50}: −0.301) (Fig. 3).

The ring descriptor, NNRS, indicates a normalized number of ring systems. This can be calculated by the following equation [59].

$$NNRS = NRS/Xmax$$

Here, NNRS denotes a normalized number of ring systems, NRS represents the number of ring systems, an X_{max} means the possibility of a

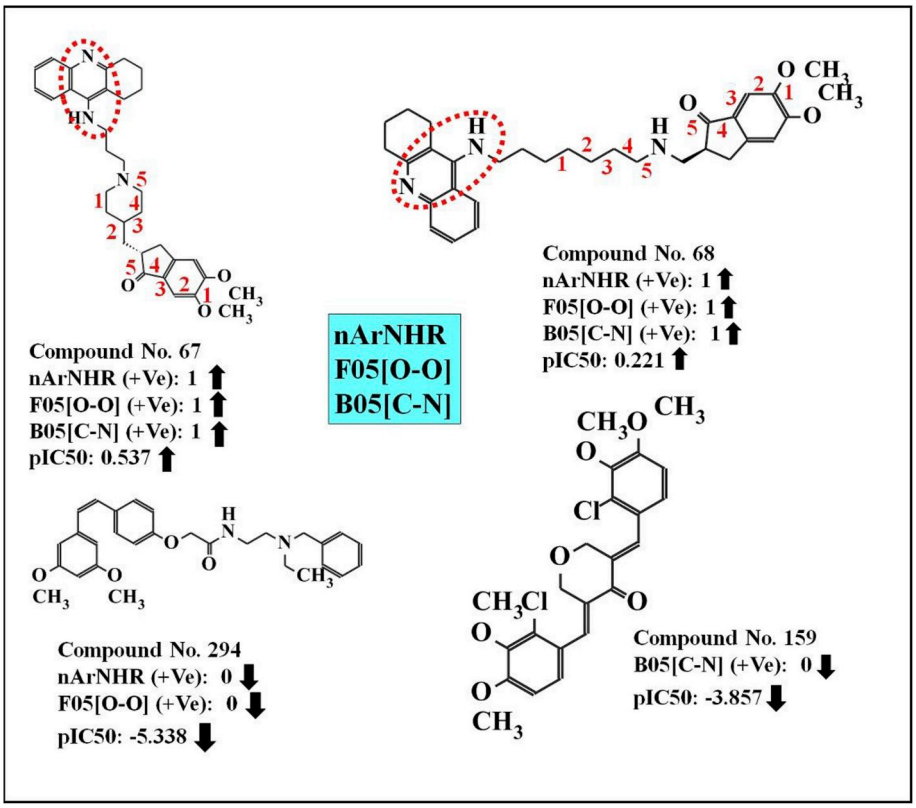


Fig. 2. Contribution of nArNHR, B05 [C–N] and F05 [O–O] descriptors on β-amyloid aggregation inhibition.

maximum number of ring systems [59]. The negative regression coefficient of this descriptor suggests that NNRS negatively influences the inhibitory activity against β-amyloid plaque. The details about this descriptor were described by the Das et al., 2016 [59]. From this, it can be suggested that for the development of inhibitors against beta amyloid aggregation, the normalized number of ring system should be high, as shown in (Fig. 4) compounds **162** (pIC_{50}: −4.243) (one cyclohexane and two benzene rings), **163** (pIC_{50}: −4.350) (one cyclohexane and two benzene rings), and **293** (pIC_{50}: −5.190) (three benzene rings) and vice versa in the case of compounds **88** (pIC_{50}: −0.217) (two fused rings),

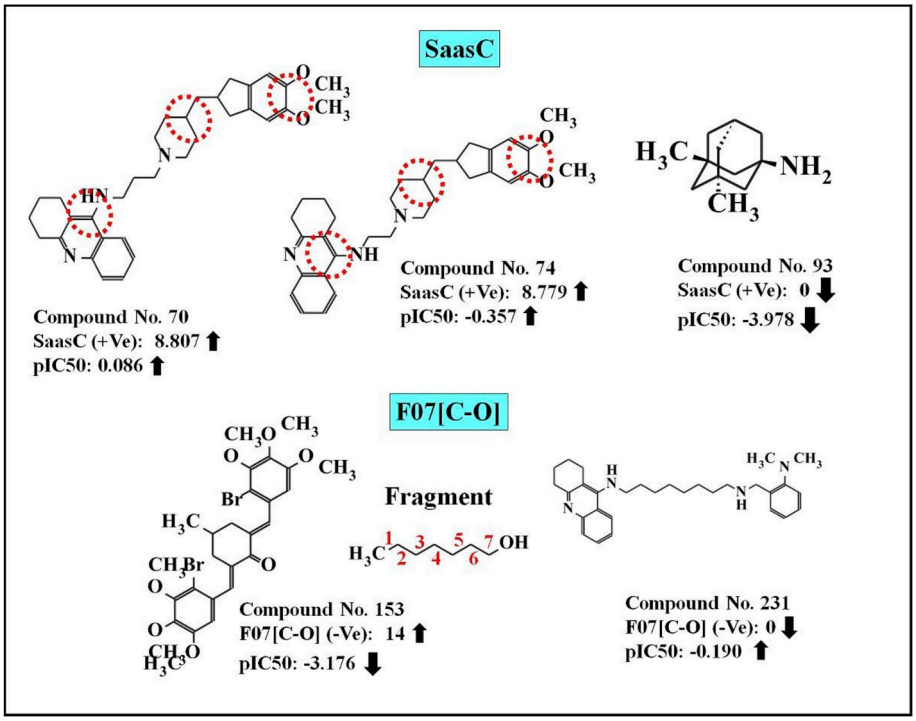


Fig. 3. Contribution of SaasC and F07 [C–O] descriptors on β-amyloid aggregation inhibition.

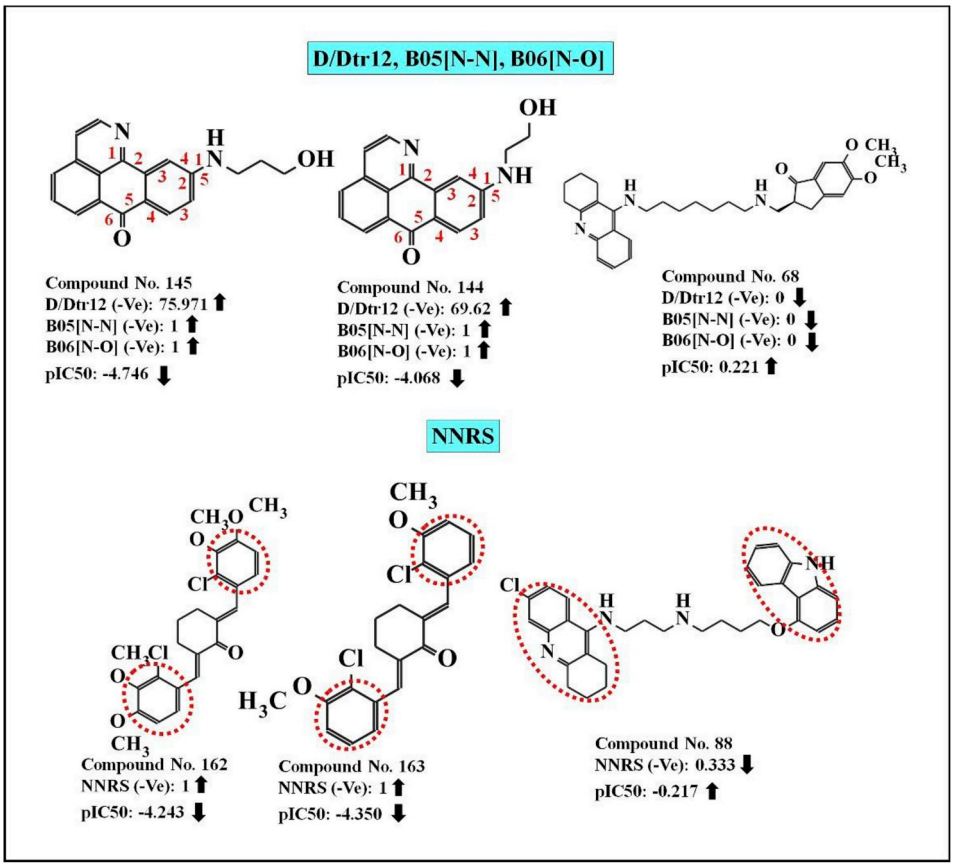


Fig. 4. Contribution of D/Dtr12, B05 [N–N], B06 [N–O], and NNRS descriptors on β-amyloid aggregation inhibition.

106 (pIC_{50}: −0.301) (one fused ring), and **323** (pIC_{50}: −1.38) (one fused ring, one benzene ring, and one pyridine ring) (Fig. 4).

The ring descriptor, D/Dtr12, denotes the distance/detour ring index of order 12 (size of ring) in the compounds. The negative impact of this descriptor recommends that a higher numerical value of this descriptor leads to a lower inhibitory activity as evidenced by the compounds (Fig. 4) **143** (pIC_{50}: −3.760), **145** (pIC_{50}: −4.746), and **144** (pIC_{50}: −4.068) (their corresponding descriptors values are 98.958, 75.971, and 69.627, respectively). On the contrary, the compounds (Fig. 4) having no such fragments show higher β-amyloid aggregation inhibitory activity as observed in case of compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221) and **69** (pIC_{50}: 0.173).

Another 2D atom pair descriptor, B05 [N–N], indicates the presence of two nitrogen atoms at the topological distance of 5. The negative regression coefficient of this descriptor suggests that the presence of an N–N fragment at the topological distance 5 inversely affects the β-amyloid aggregation inhibitory activity. This is observed in compounds (Fig. 4) **50** (pIC_{50}: −4.100), **144** (pIC_{50}: −4.068), and **145** (pIC_{50}: −4.74) (all these compounds have a descriptor value of 1). The opposite is observed in the compounds (Fig. 4) **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173).

Another 2D atom pairs descriptor, B06 [C–N], indicates the presence/absence of C and N atoms at the topological distance 6. This descriptor contributes negatively towards the β-amyloid aggregation as indicated by the negative regression coefficient. For example, compounds (Fig. 5) **50** (pIC_{50}: −4.100), **144** (pIC_{50}: −4.068), and **145** (pIC_{50}: −4.746) (having a descriptor value of 1 each) have lower β-amyloid aggregation inhibitory activity due to the presence of such a fragment at the topological distance 5. On the other hand, the molecules without such a fragment show higher inhibitory activity as shown in compounds (Fig. 5) **106** (pIC_{50}: −0.301), **177** (pIC_{50}: −1.531), and **179** (pIC_{50}: −1.892).

The 2D atom pair descriptors, B06 [N–O], describes the presence/absence of N–O at topological distance 6. The negative regression coefficient of this descriptor suggests that the absence of such a fragment in the molecules showed good β-amyloid aggregation inhibitory activity as shown in (Fig. 4) compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173). While presence of a higher number of this fragment shows lower inhibitory activity as observed in (Fig. 4) compounds **144** (pIC_{50}: −4.068), **145** (pIC_{50}: −4.746), and **293** (pIC_{50}: −5.190) (all of these compounds have a descriptor value of 1).

Another 2D atom pairs descriptor, B05 [C–N], denotes presence/absence of C–N at the topological distance 5. This descriptor positively influences the activity of β-amyloid aggregation inhibitors as suggested by its positive regression coefficient. Thus, the compounds containing a higher number of C–N fragments at topological distance 5 may have high β-amyloid aggregation inhibitory activity as evidenced by (Fig. 2) compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173) (their corresponding descriptors values are 1). On the other hand, the molecules which do not contain such a feature may have lower inhibitory activity as shown in compounds **159** (pIC_{50}: −3.857), **162** (pIC_{50}: −4.243), and **163** (pIC_{50}: −4.350) (Fig. 2).

Another 2D atom pair descriptor, F05 [O–O], stands for the frequency of O–O at the topological distance 5. For the β-amyloid aggregation inhibitors, this fragment indeed plays an important role in the binding process and may influence the inhibitory activity prominently. The positive impact of this descriptor towards the β-amyloid aggregation inhibitors was indicated by its positive regression coefficient. Thus, the information obtained from this descriptor suggested that the molecules containing an O–O fragment at the topological distance 5 show higher β-amyloid aggregation inhibitory activity as shown in compounds (Fig. 2) **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173) (containing descriptor value 1 respectively), while compounds **294** (pIC_{50}: −5.338), **293** (pIC_{50}: −5.190), and **87** (pIC_{50}: −4.949) show

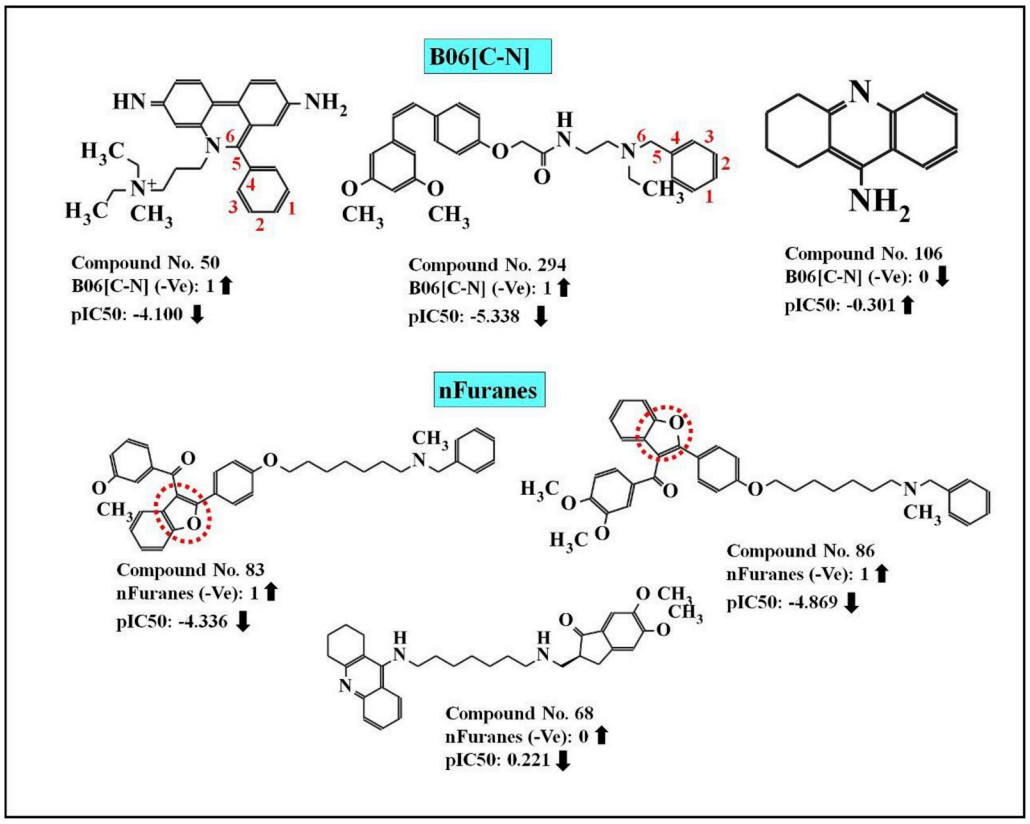


Fig. 5. Contribution of B06 [C–N] and nFuranes descriptors on β-amyloid aggregation inhibition.

lower inhibitory activity due to the absence of this fragment (Fig. 2).

The functional group count descriptor, nFuranes, describes the number of furane rings present in the compounds. The negative regression coefficient of this descriptor suggests that the presence of this ring is inversely proportional to the β-amyloid aggregation inhibitory activity as observed in case of compounds (Fig. 5) **83** (pIC_{50}: −4.33), **84** (pIC_{50}: −4.580), and **86** (pIC_{50}: −4.869) (all of these compounds have descriptor value 1), while the absence of such a ring system in the compounds (Fig. 5) indicated higher inhibitory activity as observed in the compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173).

Another 2D atom pair descriptor, F09 [C–C], stands for the frequency of C–C at the topological distance 9. The positive regression coefficient of this descriptor suggests that an increase in the frequency of a C–C fragment at topological distance 9 may favour β-amyloid aggregation inhibitory activity. Thus, the molecules bearing such a fragment may enhance the β-amyloid aggregation inhibitory activity as shown in (Fig. 6) compounds **67** (pIC_{50}: 0.537), **68** (pIC_{50}: 0.221), and **69** (pIC_{50}: 0.173) (their corresponding descriptors values are 29, 23, and 27, respectively). The opposite is observed in the case of (Fig. 6) compounds **144** (pIC_{50}: −4.068), **145** (pIC_{50}: −4.746), and **93** (pIC_{50}: −3.978) (containing descriptors values 4, 7, and 0, respectively).

Another 2D atom pair descriptor, F06 [C–C], indicates the frequency of C–C at the topological distance 6. The negative regression coefficient of this descriptor suggests that the presence of higher numbers of this fragment is inversely proportional to the β-amyloid aggregation inhibitory activity as observed in case of compounds (Fig. 6) **86** (pIC_{50}: −4.869), **87** (pIC_{50}: −4.949), and **83** (pIC_{50}: −4.336) (corresponding descriptor values are 49, 48, and 48, respectively). In contrast, a lower numerical value of this descriptor may favour the β-amyloid aggregation inhibitory activity as observed in case of compound (Fig. 6) **106** (pIC_{50}: −0.301) (containing descriptor value 6).

3.2. Applicability domain (AD) of PLS models

In this work, the reported PLS models were checked for their applicability domain at a confidence level of 99% according to the DModX (distance to model in the X-space) approach using SIMCA-P 10.0 software [45]. In case of model 1 (see S3 Fig. S6 in supplementary materials), we found that 5 compounds (i.e., compounds number **26**, **81**, **85**, **262,** and **278**) in the test set are located outside the critical DModX value (D-Crit = 1.813). In case of model 2 (see S3 Fig. S6 in supplementary materials), we found that 3 compounds (i.e., compounds number **262**, **274,** and **278**) in the test set are located outside the critical DModX value (D-Crit = 1.814).

3.3. Loading plot

A loading plot of a PLS model (see S3 Fig. S8 in supplementary materials) provides information about the relationship among the X-variables and Y-variables. The amount of the loading for each descriptor to the latent variables can be seen from their corresponding loading plot using SIMCA-P 10.0 software [45]. In the loading plot, a descriptor which is close to zero is not well associated with the trends contained in the related scores [60]. As we have observed from the loading plot (see S3 Fig. S7 in supplementary materials), the X-variables nArNHR, SaasC, F09(C–C), B05(C–N), and F05(O–O) are significant for the Y-variable (pIC_{50}), because they are very close to the Y-variable. On the other hand, the variables NNRS, D/Dtr12, nFuranes, B05 [N–N], B06 [C–N], B06 [N–O], F06 [C–C] F07 [C–O]), which are situated on the opposite side of the plot origin with respect to the activity (Y-variable), contribute negatively.

3.4. 3D-pharmacophore model

In the current work, we developed ten different pharmacophore

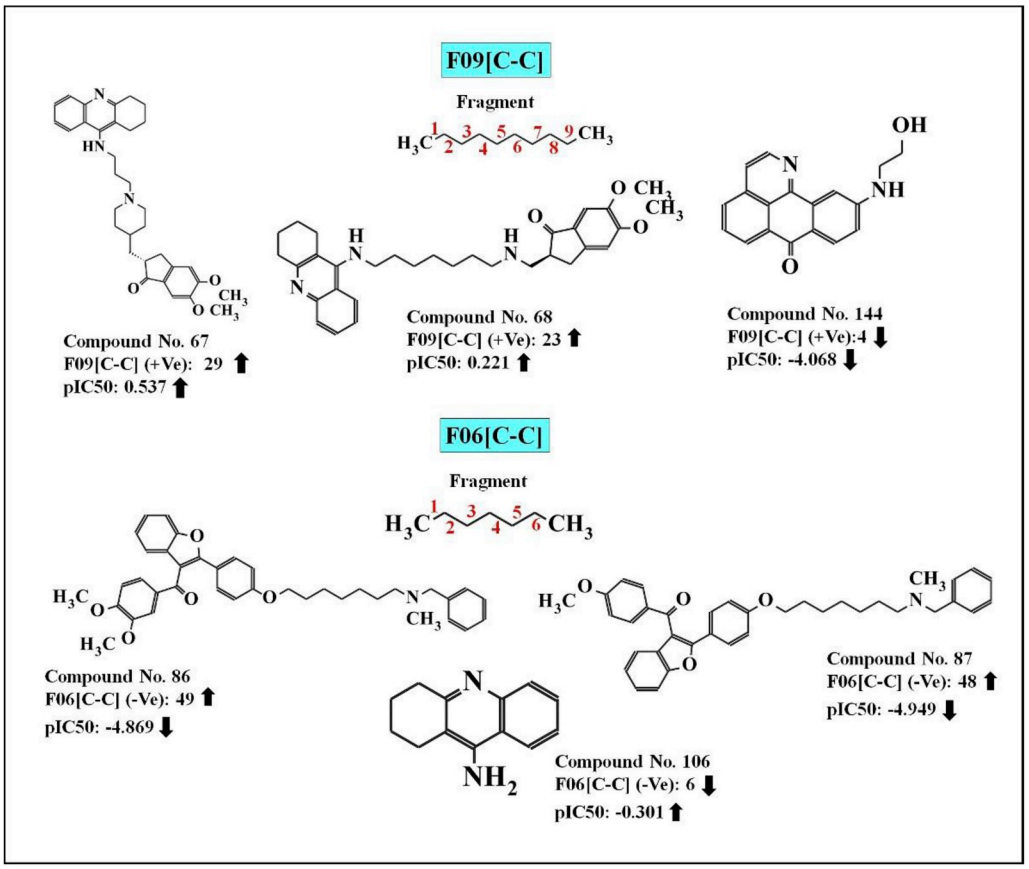


Fig. 6. Contribution of F09 [C–C] and F06 [C–C] descriptors on β-amyloid aggregation inhibition.

models from a training set of 62 compounds. For the development of a pharmacophore model, we used Discovery Studio 2016 Client 4.1 [50]. In terms of internal validation, the best pharmacophore model (**Hypo 1**) was found in the cost analysis with a higher correlation coefficient (r: 0.724), lower root mean square deviation (rmsd: 2.293), total cost (366.258), maximum fit (6.573), configuration cost (21.364), error (343.071), and weight (1.821). These values indicated that the developed model was acceptable. The results of ten pharmacophore hypotheses against β-amyloid peptide are given in Table S3 (See S4 in supplementary information). Based on all reported metrics, Hypo-1 was found to be the best one among the ten hypotheses with one hydrogen bond acceptor (HBA), one hydrophobic (HYD), one hydrophobic aromatic, and one ring aromatic (RA) feature (Fig. 7). External validation of the model has been carried out by mapping the test set molecules (Fig. 7) on Hypo-1 with the same settings as employed for the pharmacophore generation by the FAST method. After mapping, we found that 240 molecules from the data set of 252 compounds were mapped properly. Only 12 compounds failed to map due to the absence of the features found in the developed pharmacophore model. The observed and estimated activity of the training and test set compounds obtained from the analysis using Hypo-1 are given in the supplementary materials section (S2 sheet 1 and 2). The results obtained from the qualitative analysis for the training and test sets using Hypo-1 are given in Table S4 (See S4 in supplementary information) (qualitative validation parameters). The F-test confirms the non-randomness of the developed pharmacophore (Hypo-1) model. The total cost and correlation values obtained from the original and randomized models of the hypothesis for the F-test are given in the supplementary materials section (S2 sheet 3 and 4).

3.4.1. *Relation of the 3D-pharmacophore model with the 2D-QSAR model*

All of the compounds in the dataset have at least one aromatic ring feature. The RA feature is a preliminary requirement for inhibitory activity against β-amyloid aggregation. The RA feature is in accordance with the nArNHR and SaasC descriptors of the 2D-QSAR models. Hydrophobic and hydrophobic aromatic features are in harmony with the F09 [C–C] and SaasC descriptors of the 2D-QSAR models. The hydrogen bond acceptor feature is well corroborated with B05 [C–N] and F05 [O–O] descriptors of the 2D-QSAR models. The most active compound of the training set (**69**, IC_{50}: 0.67 nM) mapped entirely on Hypo-1 with all of the four features (see S3 Fig. S8 in supplementary materials). One benzene ring lies in the RA region, an amino group is in the hydrogen bond acceptor region, and a ring system lies in the hydrophobic region. The least active compound **24** (IC_{50}: 6323 nM) of the training set lacks RA features; thus it does not map completely (see S3 Fig. S8 in supplementary materials). The most active compound of the test set (**66**, IC_{50}: 0.09 nM) mapped completely on Hypo-1 with all of the four features (Fig. 7). The least active compound (**86**, IC_{50}: 74100 nM) of the test set mapped partially with Hypo-1 (Fig. 7). From the above discussion, we have concluded that the absence of any feature among these four features in a molecule decreases the β-amyloid aggregation inhibitory potency.

3.5. *Molecular docking*

In this study, we performed a molecular docking study of the two most active (**67** and **231**), two moderately active (**208** and **276**) and two least active (**87** and **145**) compounds from this dataset. The docking interactions suggest that the molecules interacted with a pocket containing HIS A:13, (hydrophilic nature), LYS A: 16 (Charged), and LEU A:17, VAL A:24, PHE A:20 and ALA A:21 (Hydrophobic nature) amino acid residues.

3.5.1. *Molecular docking of the most active compounds from the dataset*

The two most active compounds (**67** and **231**) from the dataset

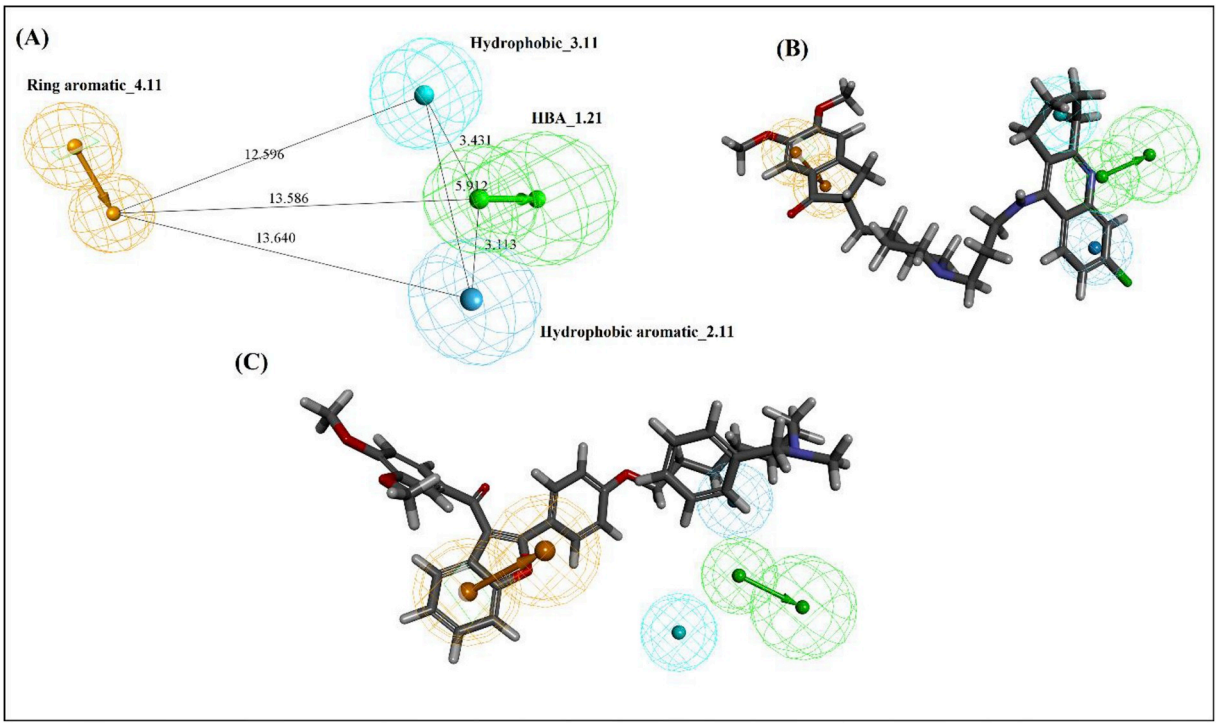


Fig. 7. The best pharmacophore model (Hypo1) of β-amyloid aggregation inhibitors generated by the HypoGen module: (A) the best pharmacophore model Hypo1 represented with distance constraints (Å), (B) **Hypo-1** mapping with one of the most active compounds **66** of test set compounds and (C) **Hypo-1** mapping with one of the least active compounds **86** of test set compounds. Pharmacophoric features are coloured as follows: hydrogen bond acceptor (green), hydrophobic (cyan), hydrophobic aromatic (light blue colour), and ring aromatic (orange).

(pIC_{50} = 0.537 and −0.190, respectively) interacted with the active site amino acid residues through different interaction forces like hydrogen bonding interactions (carbon hydrogen bonds and conventional hydrogen bonds), electrostatic interactions (π-cation), and hydrophobic interactions (π-alkyl bonds, alkyl and π-π-T-shaped). The amino acid residues involved in interactions with these compounds are HIS A:13, LYS A:16, PHE A:20, LEU A:17, VAL A:24 and ALA A:21 (see S3 Figs. S9 and S10 in supplementary materials). The details of docking results are tabulated in Table S5 (see S4 in supplementary materials).

Fig. S9 (see S3 in supplementary materials) shows that compound **67** (one of the most active compounds in the dataset) interacts with HIS A:13, LYS A: 16 and LEU A:17 amino acid residues through a hydrogen bonding interaction, with PHE A:20 through a π-cation, and VAL A:24 and ALA A:21 amino acid residues through alkyl and π-alkyl interactions, respectively.

Another of the most active compounds, compound **231**, interacts with the amino acid residue through hydrogen bonding and π-π T-shaped interaction (HIS A:13) and π-alkyl and alkyl (VAL A:24 and ALA A:21) interactions (See in supplementary section S3, Fig. S10).

3.5.2. *Molecular docking of the moderately active compounds from the dataset*

Two moderately active compounds (**208** and **276**) from the dataset (pIC_{50} = −2.908 and −2.989, respectively) interacted with the active site amino acid residues through various interaction forces like hydrogen bonds (carbon hydrogen bonds and conventional hydrogen bonds) and hydrophobic interactions (π-π-T shaped, π-alkyl and π-sigma). The amino acid residues involved in interactions with these compounds are ALA A:21, LEU A:17, PHE A:20, HIS A:13, VAL A:24 and LYS A:16 (see S3 Figs. S11 and S12 in supplementary materials).

Fig. S11 (see S3 in supplementary materials) shows that compound **208** interacts with LEU A: 17 and LYS A:16 (through hydrogen bonding), PHE A:20 (π-π-T-shaped and pi-sigma), and ALA A:21 (π-alkyl bonds) amino acid residues.

Another moderately active compound (compound **276**) from the dataset interacted with the amino acid residues through hydrogen bonding (LEU A: 17 and LYS A:16) and hydrophobic interactions such as π-π T-shaped and π-alkyl bonding (HIS A:13 ALA A:21 and VAL A:24) (See in supplementary section S3 Fig. S12).

3.5.3. *Molecular docking of the least active molecules from the dataset*

The two least active compounds from the dataset (pIC_{50} = −4.949 and −4.746, respectively) namely **87** and **145** interacted with the active site amino acid residues through different interaction forces like hydrogen bonding (carbon hydrogen bonds and conventional hydrogen bonds), electrostatic (π-cation) and hydrophobic interactions (π-π-stacking and π-alkyl). The amino acid residues involved in interactions with these compounds are ALA A:21, PHE A:24, LYS A:16 and VAL A:24 (see S3 Figs. S13 and S14 in supplementary materials).

Fig. S13 (see S3 in supplementary materials) shows that compound 87 (one of the least active compounds from the dataset) interacted with ALA A:21 and LYS A:16 amino acid residues through hydrogen bonding interaction, LYS A:16 amino acid residue through π-cation and PHE A:20 and ALA A:21 through π-π -T-shaped and π-alkyl interactions.

Another least active compound from the dataset, compound 145, interacts with the amino acid residues through hydrogen bonding (LYS A: 16) and hydrophobic interactions such as π-alkyl bonding (ALA A: 21, VAL A: 24) (See in supplementary section S3 Fig. S14).

We have observed from the docking results that the most active compounds from the dataset (**67** and **231**) (See in supplementary S3 Figs. S9 and S10) interacted with maximum number of active amino acid residues with higher number of interacting forces (non-covalent forces) in comparison with the least active compounds from the data set like **87** and **145** (See in supplementaryS3 Figs. S13 and S14). In pharmacophore mapping, we have also seen that the most active compounds (**66** and **67**) from the data set correctly mapped with all features appeared in the model, whereas the least active compounds (**86** and **87**) partially mapped with the model. The least active compound from the data set

Table 1
Comparisons of proposed study with previous published studies against β-amyloid aggregation.

Sources	E. L.	LV	Model	Training set			Test set	
				n	r^2	Q^2	n	R^2_{pred}
Model-1 in this study	**12**	**6**	**PLS**	**252**	**0.664**	**0.621**	**62**	**0.763**
Model-2 in this study	**13**	**7**	**PLS**	**252**	**0.684**	**0.638**	**62**	**0.769**
Leal et al., 2015 [10]	5	–	HQSAR	36	0.937	0.757	10	0.659
Zhao et al., 2013 [11]	–	5	CoMFA	32	0.877	0.431	7	0.834
Zhao et al., 2013 [11]	–	8	CoMSIA	32	0.836	0.447	7	0.617
Zhao et al., 2013 [11]	–	5	CoMFA	34	0.828	0.522	5	0.915
Zhao et al., 2013 [11]	–	6	CoMSIA	34	0.800	0.493	5	0.902
Aswathy et al., 2018 [12]	4	–	HQSAR	24	0.931	0.615	6	0.956
Aswathy et al., 2018 [12]	–	5	CoMFA	24	0.787	0.687	6	0.731
Aswathy et al., 2018 [12]	–	3	CoMSIA	24	0.972	0.743	6	0.713
Aswathy et al., 2018 [12]	6	–	MLR	24	0.908	0.747	6	0.807
Hossein et al., 2019 [13]	4		MLR	28	0.912	0.915	12	0.836
Xiangji 2006 [61]	4	–	PLS	22	0.857	–	–	–
Yang et al., 2010 [62]	–	6	CoMSIA	21	0.911	0.512	–	–
Najmeh et al., 2014 [63]	5	–	PCA	25	0.631	–	–	–
Sehan et al., 2015 [64]	–	–	3D-QSAR	63	0.93	0.89	26	0.89

Abbreviations: LV = Latent variables, E. L. = Equation length, MLR = Multiple linear regression, CoMFA=Comparative Molecular Field Analysis, CoMSIA= Comparative molecular similarity index analysis, PCA = Principal component analysis and HQSAR= Hologram QSAR.

failed to map in the absence of the features appearing in the developed models, which are most important for inhibitory activity against β-amyloid plaque.

3.5.4. Relation with the QSAR models

From the docking studies, we have observed that the formation of a hydrogen bond and alkyl bond between the ligand and receptor play a vital role in interactions. Hydrogen bonding may correlate with descriptors nArNHR (number of secondary aromatic amines present in the compounds), F05 [O–O] (frequency of O–O at topological distance 5) and B05 [C–N] (presence/absence of C–N at topological distance 5 in the compounds) in the 2D-QSAR models. Descriptors SaasC (aromatic carbons with an attached substituent atom) and nArNHR are well related with interactions formed via π-interactions (π-alkyl and π-π T-shaped) between the protein and ligands and define the importance of this descriptor as we have observed in compound nos. **67**, **231**, **208**, and **276** (See in supplementary S3 Figs. S9, S10, S11 and S12). But in contrast, in case of compounds **87** and **145** (least active), the descriptor NNRS (normalized number of ring systems) contributes negatively to the response and is found to be related with π- π-T-shaped, and π-alkyl bonding interactions with those fragments in the docking experiments (See in supplementary S3 Fig. S13 and S14). Thus, from above said information, we can conclude that hydrogen bonding, hydrophobicity, and alkyl (π interaction) features obtained from both 2D-QSAR model and docking results are essential for inhibitory activity against β-amyloid aggregation.

3.6. Comparisons of the performance of the reported models with previous published models

In this investigation, we have performed a comparison of the best models of this study with previously published models (Leal et al., 2015 [10], Zhao et al., 2013 [11], Aswathy et al., 2018 [12], Hossein et al., 2019 [13], Xiangji 2006 [61], Yang et al., 2010 [62], Najmeh et al., 2014 [63] and Sehan et al., 2015 [64]) for the prediction of the bioactivity against β-amyloid plaques, as depicted in Table 1. The details of different internal and external validation parameters obtained from our models and those obtained from previous models are given in Table 1. Based on the statistical quality in terms of both internal and external validation criteria, the models reported in this work are statistically significant and robust enough as compared to the previously reported models (Table 1). Moreover, the models presented in this study are derived from a larger set of molecules than those reported in the previous studies.

4. Overview and conclusion

In this research, we have applied chemoinformatic tools to examine a set of 314 heterocyclic [15] compounds with defined β-amyloid aggregation inhibitory activity in order to identify the structural requirements essential for the inhibitory activity. The QSAR models were developed with simple, meaningful, and easily interpretable descriptors. Prior to the development of the final models, we applied a multilayered variable selection strategy to investigate the meaningful descriptors, and the final models were built by using the PLS regression based methodology following strict OECD guidelines. The obtained results suggest that the developed model showed good predictivity based on both internal and external validation parameters. The information obtained from the PLS models (as also confirmed by the regression coefficient plot, variable importance plot and loading plot in Fig. S4, S5 and S7 in supplementary information S3), suggested the following: (i) the presence of a secondary aromatic amine in compounds, (ii) the presence of C–N fragments at the topological distance 5, and (iii) aromatic carbons with an attached substituent atom are favourable for enhancing β-amyloid aggregation inhibitory activity; whereas, (iv) the presence of a higher number of the normalized number of ring systems and a furane ring in the molecules is detrimental to the inhibitory activity. The results obtained from pharmacophore mapping and molecular docking studies are well supported with those of the QSAR analysis. The present results may be used for screening purposes for the discovery and development of new β-amyloid aggregation inhibitors.

5. Future perspective

The developed 2D-QSAR models may be useful for predicting the activity of new analogues even before their synthesis and evaluation. Features obtained from the developed models (2D-QSAR and 3D-pharmacophore) and molecular docking can be helpful for the design of novel inhibitors against β-amyloid aggregation, and the overall approach established can be adopted for ligand-based drug-design campaigns.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

VK thanks Indian Council of Medical Research, New Delhi for a senior research fellowship (File No: 5/3/8/27/ITR-F/2018-ITR; dated: 18.05.2018). Financial assistance from the UGC, New Delhi in the form of a fellowship (Letter number and date: F./PDFSS-2015-17-WES-11996; dated: April 06, 2016) to PKO is thankfully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2020.103658.

References

[1] S.W. Snyder, U.S. Ladror, W.S. Wade, G.T. Wang, L.W. Barrett, E.D. Matayoshi, H. J. Huffaker, G.A. Krafft, T.F. Holzman, Amyloid-beta aggregation: selective inhibition of aggregation in mixtures of amyloid with different chain lengths, Biophys. J. 67 (3) (1994) 1216–1228, https://doi.org/10.1016/S0006-3495(94)80591-0.

[2] M. Necula, R. Kayed, S. Milton, C.G. Glabe, Small molecule inhibitors of aggregation indicate that amyloid β oligomerization and fibrillization pathways are independent and distinct, J. Biol. Chem. 282 (14) (2007) 10311–10324, https://doi.org/10.1074/jbc.M608207200.

[3] K. Hensley, J.M. Carney, M.P. Mattson, M. Aksenova, M. Harris, J.F. Wu, R. A. Floyd, D.A. Butterfield, A model for β-amyloid aggregation and neurotoxicity based on free radical generation by the peptide: relevance to Alzheimer disease, Proc. Natl. Acad. Sci. U.S.A. 91 (8) (1994) 3270–3274, https://doi.org/10.1073/pnas.91.8.3270.

[4] T. Yamasaki, H. Muranaka, Y. Kaseda, Y. Mimori, S. Tobimatsu, Understanding the pathophysiology of Alzheimer's disease and mild cognitive impairment: a mini review on fMRI and ERP studies, Neurol. Res. Int. 2012 (2012), https://doi.org/10.1155/2012/719056.

[5] F. Re, C. Airoldi, C. Zona, M. Masserini, B. La Ferla, N. Quattrocchi, F. Nicotra, Beta amyloid aggregation inhibitors: small molecules as candidate drugs for therapy of alzheimers disease, Curr. Med. Chem. 17 (27) (2010) 2990–3006, https://doi.org/10.2174/092986710791959729.

[6] N. Kokkoni, K. Stott, H. Amijee, J.M. Mason, A.J. Doig, N-methylated peptide inhibitors of-amyloid aggregation and toxicity. Optimization of the Inhibitor Structure † 45 (32) (2006) 9906–9918, https://doi.org/10.1021/bi060837s.

[7] D. Fourches, E. Muratov, A. Tropsha, Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research 50 (7) (2010), https://doi.org/10.1021/ci100176x, 1889-204.

[8] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (5) (2007) 694–701, https://doi.org/10.1002/qsar.200610151.

[9] K. Roy, R. Narayan Das, A review on principles, Theory and Practices of 2D-QSAR 15 (4) (2014) 346–349.

[10] F. Leal, C. da Silva Lima, R. de Alencastro, H. Castro, C. Rodrigues, M. Albuquerque, Hologram QSAR models of a series of 6-arylquinazolin-4-amine inhibitors of a new Alzheimer's disease target: dual specificity tyrosine-phosphorylation-regulated kinase-1A enzyme, Int. J. Mol. Sci. 16 (3) (2015) 5235–5253, https://doi.org/10.3390/ijms16035235.

[11] L. Zhao, L. Zhang, M. Lei, 3D-QSAR and docking studies on 2-arylbenzoxazole and linker-Y transthyretin amyloidogenesis inhibitors, Sci. China Chem. 56 (11) (2013) 1550–1563, https://doi.org/10.1007/s11426-013-4894-9.

[12] L. Aswathy, R.S. Jisha, V.H. Masand, J.M. Gajbhiye, I.G. Shibi, Design of novel amyloid β aggregation inhibitors using QSAR, pharmacophore modeling, molecular docking and ADME prediction, Silico Pharmacol 6 (1) (2018) 12, https://doi.org/10.1007/s40203-018-0049-1.

[13] H. Safarizadeh, Z. Garkani-Nejad, Molecular docking, molecular dynamics simulations and QSAR studies on some of 2-arylethenylquinoline derivatives for inhibition of Alzheimer's amyloid-beta aggregation: insight into mechanism of interactions and parameters for design of new inhibitors, J. Mol. Graph. Model. 87 (2019) 129–143, https://doi.org/10.1016/j.jmgm.2018.11.019.

[14] P. Ambure, J. Bhat, T. Puzyn, K. Roy, Identifying natural compounds as multi-target-directed ligands against Alzheimer's disease: an in silico approach, J. Biomol. Struct. Dyn. 37 (5) (2019) 1282–1306, https://doi.org/10.1080/07391102.2018.1456975.

[15] M.K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Res. 44 (2016) D1045–D1053, https://doi.org/10.1093/nar/gkv1072.

[16] K. Roy, S. Kar, P. Ambure, On a simple approach for determining applicability domain of QSAR models, Chemometr. Intell. Lab. Syst. 145 (2015) 22–29, https://doi.org/10.1016/j.chemolab.2015.04.013.

[17] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inform. 29 (6–7) (2010) 476–488, https://doi.org/10.1002/minf.201000061.

[18] H. Tang, H.T. Zhao, S.M. Zhong, Z.Y. Wang, Z.F. Chen, H. Liang, Novel oxoisoaporphine-based inhibitors of acetyl- and butyrylcholinesterase and acetylcholinesterase-induced beta-amyloid aggregation, Bioorg. Med. Chem. Lett 22 (6) (2012) 2257–2261, https://doi.org/10.1016/j.bmcl.2012.01.090.

[19] Y. Li, X. Qiang, Y. Li, X. Yang, L. Luo, G. Xiao, Z. Cao, Z. Tan, Y. Deng, Pterostilbene-O-acetamidoalkylbenzylamines derivatives as novel dual inhibitors of cholinesterase with anti-β-amyloid aggregation and antioxidant properties for the treatment of Alzheimer's disease, Bioorg. Med. Chem. Lett 26 (8) (2016) 2035–2039, https://doi.org/10.1016/j.bmcl.2016.02.079.

[20] D. Panek, A. Więckowska, T. Wichur, M. Bajda, J. Godyń, J. Jończyk, K. Mika, J. Janockova, O. Soukup, D. Knez, J. Korabecny, S. Gobec, B. Malawska, Design, synthesis and biological evaluation of new phthalimide and saccharin derivatives with alicyclic amines targeting cholinesterases, beta-secretase and amyloid beta aggregation, Eur. J. Med. Chem. 125 (2017) 676–695, https://doi.org/10.1016/j.ejmech.2016.09.078.

[21] W. Luo, Y.P. Li, Y. He, S.L. Huang, J.H. Tan, T.M. Ou, D. Li, L.Q. Gu, Z.S. Huang, Design, synthesis and evaluation of novel tacrine-multialkoxybenzene hybrids as dual inhibitors for cholinesterases and amyloid beta aggregation, Bioorg. Med. Chem. 19 (2) (2011) 763–770, https://doi.org/10.1016/j.bmc.2010.12.022.

[22] A. Shi, L. Huang, C. Lu, F. He, X. Li, Synthesis, biological evaluation and molecular modeling of novel triazole-containing berberine derivatives as acetylcholinesterase and β-amyloid aggregation inhibitors, Bioorg. Med. Chem. 19 (7) (2011) 2298–2305, https://doi.org/10.1016/j.bmc.2011.02.025.

[23] J.W. Yan, Y.P. Li, W.J. Ye, S. Bin Chen, J.Q. Hou, J.H. Tan, T.M. Ou, D. Li, L.Q. Gu, Z.S. Huang, Design, synthesis and evaluation of isaindigotone derivatives as dual inhibitors for acetylcholinesterase and amyloid beta aggregation, Bioorg. Med. Chem. 20 (8) (2012) 2527–2534, https://doi.org/10.1016/j.bmc.2012.02.061.

[24] F. Mao, L. Huang, Z. Luo, A. Liu, C. Lu, Z. Xie, X. Li, O-Hydroxyl- or o-amino benzylamine-tacrine hybrids: multifunctional biometals chelators, antioxidants, and inhibitors of cholinesterase activity and amyloid-β aggregation, Bioorg. Med. Chem. 20 (19) (2012) 5884–5892, https://doi.org/10.1016/j.bmc.2012.07.045.

[25] A. Więckowska, K. Więckowski, M. Bajda, B. Brus, K. Sałat, P. Czerwińska, S. Gobec, B. Filipek, B. Malawska, Synthesis of new N-benzylpiperidine derivatives as cholinesterase inhibitors with β-amyloid anti-aggregation properties and beneficial effects on memory in vivo, Bioorg. Med. Chem. 23 (10) (2015) 2445–2457, https://doi.org/10.1016/j.bmc.2015.03.051.

[26] G.F. Zha, C.P. Zhang, H.L. Qin, I. Jantan, M. Sher, M.W. Amjad, M.A. Hussain, Z. Hussain, S.N.A. Bukhari, Biological evaluation of synthetic α,β-unsaturated carbonyl based cyclohexanone derivatives as neuroprotective novel inhibitors of acetylcholinesterase, butyrylcholinesterase and amyloid-β aggregation, Bioorg. Med. Chem. 24 (10) (2016) 2352–2359, https://doi.org/10.1016/j.bmc.2016.04.015.

[27] M.L. Bolognesi, V. Andrisano, M. Bartolini, R. Banzi, C. Melchiorre, Propidium-based polyamine ligands as potent inhibitors of acetylcholinesterase and acetylcholinesterase-induced amyloid-β aggregation, J. Med. Chem. 48 (1) (2005) 24–27, https://doi.org/10.1021/jm049156q.

[28] P. Camps, X. Formosa, C. Galdeano, T. Gómez, D. Muñoz-Torrero, M. Scarpellini, E. Viayna, A. Badia, M.V. Clos, A. Camins, M. Pallàs, M. Bartolini, F. Mancini, V. Andrisano, J. Estelrich, M. Lizondo, A. Bidon-Chanal, F.J. Luque, Novel donepezil-based inhibitors of acetyl- and butyrylcholinesterase and acetylcholinesterase-induced β-amyloid aggregation, J. Med. Chem. 51 (12) (2008) 3588–3598, https://doi.org/10.1021/jm8001313.

[29] S. Rizzo, C. Rivière, L. Piazzi, A. Bisi, S. Gobbi, M. Bartolini, V. Andrisano, F. Morroni, A. Tarozzi, J.-P. Monti, A. Rampa, Benzofuran-based hybrid compounds for the inhibition of cholinesterase activity, β amyloid aggregation, and Aβ neurotoxicity, J. Med. Chem. 51 (10) (2008) 2883–2886, https://doi.org/10.1021/jm8002747.

[30] M. Catto, A.A. Berezin, D. Lo Re, G. Loizou, M. Demetriades, A. De Stradis, F. Campagna, P.A. Koutentis, A. Carotti, Design, synthesis and biological evaluation of benzo[e][1,2,4]triazin-7(1H) -one and [1,2,4]-triazino[5,6,1-jk] carbazol-6-one derivatives as dual inhibitors of beta-amyloid aggregation and acetyl/butyryl cholinesterase, Eur. J. Med. Chem. 58 (2012) 84–97, https://doi.org/10.1016/j.ejmech.2012.10.003.

[31] Y.P. Li, F.X. Ning, M.B. Yang, Y.C. Li, M.H. Nie, T.M. Ou, J.H. Tan, S.L. Huang, D. Li, L.Q. Gu, Z.S. Huang, Syntheses and characterization of novel oxoisoaporphine derivatives as dual inhibitors for cholinesterases and amyloid beta aggregation, Eur. J. Med. Chem. 46 (2011) 1572–1581, https://doi.org/10.1016/j.ejmech.2011.02.005.

[32] F. Belluti, M. Bartolini, G. Bottegoni, A. Bisi, A. Cavalli, V. Andrisano, A. Rampa, Benzophenone-based derivatives: a novel series of potent and selective dual inhibitors of acetylcholinesterase and acetylcholinesterase-induced beta-amyloid aggregation, Eur. J. Med. Chem. 46 (2011) 1682–1693, https://doi.org/10.1016/j.ejmech.2011.02.019.

[33] J. Rouleau, B.I. Iorga, C. Guillou, New potent human acetylcholinesterase inhibitors in the tetracyclic triterpene series with inhibitory potency on amyloid β aggregation, Eur. J. Med. Chem. 46 (2011) 2193–2205, https://doi.org/10.1016/j.ejmech.2011.02.073.

[34] H. Tang, L.Z. Zhao, H.T. Zhao, S.L. Huang, S.M. Zhong, J.K. Qin, Z.F. Chen, Z. S. Huang, H. Liang, Hybrids of oxoisoaporphine-tacrine congeners: novel acetylcholinesterase and acetylcholinesterase-induced β-amyloid aggregation inhibitors, Eur. J. Med. Chem. 46 (2011) 4970–4979, https://doi.org/10.1016/j.ejmech.2011.08.002.

[35] W.J. Shan, L. Huang, Q. Zhou, F.C. Meng, X.S. Li, Synthesis, biological evaluation of 9-N-substituted berberine derivatives as multi-functional agents of antioxidant, inhibitors of acetylcholinesterase, butyrylcholinesterase and amyloid-β aggregation, Eur. J. Med. Chem. 46 (2011) 5885–5893, https://doi.org/10.1016/j.ejmech.2011.09.051.

[36] N. Guzior, M. Bajda, M. Skrok, K. Kurpiewska, K. LewiÅski, B. Brus, A. Pišlar, J. Kos, S. Gobec, B. Malawska, Development of multifunctional, heterodimeric isoindoline-1,3-dione derivatives as cholinesterase and β-amyloid aggregation

inhibitors with neuroprotective properties, Eur. J. Med. Chem. 92 (2015) 738–749, https://doi.org/10.1016/j.ejmech.2015.01.027.
[37] N. García-Font, H. Hayour, A. Belfaitah, J. Pedraz, I. Moraleda, I. Iriepa, A. Bouraiou, M. Chioua, J. Marco-Contelles, M.J. Oset-Gasque, Potent anticholinesterasic and neuroprotective pyranotacrines as inhibitors of beta-amyloid aggregation, oxidative stress and tau-phosphorylation for Alzheimer's disease, Eur. J. Med. Chem. 118 (2016) 178–192, https://doi.org/10.1016/j.ejmech.2016.04.023.
[38] MarvinView 5.9.4 and MarvinSketch 5.9.4 (Internal build id: 5.9.4_b56) year of version release: 11-05-2012, Available at: ChemAxon, http://www.chemaxon.com.
[39] G.M. Mariana, J.J. Naveja, N.S. Cruz, J.L. Medina-Franco, Open chemoinformatic resources to explore the structure, properties and chemical space of molecules, RSC Adv. 7 (85) (2017) 54153–54163, https://doi.org/10.1039/c7ra11831g.
[40] G. Domenico, A. Lombardo, C. Toma, E. Benfenati, A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications, J. cheminformatics 10 (1) (2018) 1–13, https://doi.org/10.1186/s13321-018-0315-6.
[41] S. Fabian P, C.L. Mellor, T. Meinl, M.T.D. Cronin, Screening chemicals for receptor-mediated toxicological and pharmacological endpoints: using public data to build screening tools within a KNIME workflow, Mol. Inform. 34 (2-3) (2015) 171–178, https://doi.org/10.1002/minf.201400188.
[42] G. Landrum, RDKit: Open-Source Cheminformatics, 2006.
[43] A. Mauri, V. Consonni, M. Pavan, R. Todeschini, DRAGON SOFTWARE: AN EASY APPROACH TO MOLECULAR DESCRIPTOR CALCULATIONS, 56, 2006, pp. 237–248.
[44] C.W. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (2011) 1466–1474, https://doi.org/10.1002/jcc.21707.
[45] SIMCA | umetrics, n.d. https://umetrics.com/products/simca. accessed December 19, 2019
[46] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Ravichandran Veerasamy, et al., Validation of QSAR models-strategies and importance validation of QSAR models-strategies and importance 2 (3) (2011) 511–519.
[47] K. Naresh, G. Ramakrishnan, Ligand based pharmacophore modeling and virtual screening studies to design novel HDAC2 inhibitors, Advances bioinform (2014) 1–11, https://doi.org/10.1155/2014/812148.
[48] N.M. miao, J.y. Qin, C.p. Tian, Xia-fei Yan, F. gong Dong, Z. qi Cheng, G. Fida, M. Yang, H. Chen, Y. qing Gu, Tubulin inhibitors: pharmacophore modeling, virtual screening and molecular docking, Acta Pharmacol. Sin. 35 (7) (2014) 967–979, https://doi.org/10.1038/aps.2014.34.
[49] S. Pal, V. Kumar, B. Kundu, D. Bhattacharya, N. Preethy, M.P. Reddy, A. Talukdar, Ligand-based pharmacophore modeling, virtual screening and molecular docking studies for discovery of potential topoisomerase I inhibitors, Comput. Struct. Biotechnol. J. 1 (17) (2019) 291–310, https://doi.org/10.1016/j.csbj.2019.02.006.
[50] Discovery Studio predictive science application | dassault systèmes BIOVIA. https://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/ accessed December 19, 2019.
[51] R.B. Aher, K. Roy, QSAR and pharmacophore modeling of diverse aminothiazoles and aminopyridines for antimalarial potency against multidrug-resistant Plasmodium falciparum, Med. Chem. Res. 23 (9) (2014) 4238–4249, https://doi.org/10.1007/s00044-014-0997-x.
[52] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247 (1995) 536–540, https://doi.org/10.1016/S0022-2836(05)80134-2.
[53] G. Wu, D.H. Robertson, C.L. Brooks, M. Vieth, Detailed analysis of grid-based molecular docking: a case study of CDOCKER?A CHARMm-based MD docking algorithm, J. Comput. Chem. 24 (13) (2003) 1549–1562, https://doi.org/10.1002/jcc.10306.
[54] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs, Environ. Health Perspect. 111 (10) (2003) 1361–1375, https://doi.org/10.1289/ehp.5758.
[55] P.K. Ojha, K. Roy, Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection, Chemometr. Intell. Lab. Syst. 109 (2) (2011) 146–161, https://doi.org/10.1016/j.chemolab.2011.08.007.
[56] D. Castillo-González, M.Á. Cabrera-Pérez, M. Pérez-González, A.M. Helguera, A. Durán-Martínez, Prediction of telomerase inhibitory activity for acridinic derivatives based on chemical structure, Eur. J. Med. Chem. 44 (12) (2009) 4826–4840, https://doi.org/10.1016/j.ejmech.2009.07.029.
[57] L.B. Kier, L.H. Hall, The E-state in database analysis: the PCBs as an example, Il Farmaco 54 (6) (1999) 346–353.
[58] E.B. de Melo, A new quantitative structure-property relationship model to predict bioconcentration factors of polychlorinated biphenyls (PCBs) in fishes using E-state index and topological descriptors, Ecotoxicol. Environ. Saf. 75 (2012) 213–222, https://doi.org/10.1016/j.ecoenv.2011.08.026.
[59] S. Das, P.K. Ojha, K. Roy, Multilayered variable selection in QSPR, Int. J. Quant. Struct. Relationships. 2 (1) (2016) 106–124, https://doi.org/10.4018/ijqspr.2017010108.
[60] L. Eriksson, P.L. Andersson, E. Johansson, M. Tysklind, Megavariate analysis of environmental QSAR data. Part I - a basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD), Mol. Divers. 10 (2) (2006) 169–186, https://doi.org/10.1007/s11030-006-9024-6.
[61] C. Xiangji, QSAR and primary docking studies of trans-stilbene (TSB) series of imaging agents for β-amyloid plaques, J. Mol. Struct. 763 (1–3) (2006) 83–89, https://doi.org/10.1016/j.theochem.2006.01.028.
[62] Y. Yang, L. Zhu, X. Chen, H. Zhang, Binding research on flavones as ligands of β-amyloid aggregates by fluorescence and their 3D-QSAR, docking studies, J. Mol. Graph. Model. 29 (4) (2010) 538–545, https://doi.org/10.1016/j.jmgm.2010.10.006.
[63] M. Najmeh, D. Ajloo, QSAR, docking, and Molecular dynamic studies on the polyphenolic as inhibitors of β-amyloid aggregation, Med. Chem. Res. 25 (10) (2016) 2104–2118, https://doi.org/10.1007/s00044-016-1620-0.
[64] L. Sehan, M.G. Barron, Development of 3D-QSAR model for acetylcholinesterase inhibitors using a combination of fingerprint, molecular docking, and structure-based pharmacophore approaches, Toxicol. Sci. 148 (1) (2015) 60–70, https://doi.org/10.1093/toxsci/kfv160.

International Journal of Quantitative Structure-Property Relationships
Volume 5 • Issue 3 • July-September 2020

Chemometric Modeling of Structurally Diverse Carbamates for the Inhibition of Acetylcholinesterase (AChE) Enzyme in Alzheimer's Disease

Vinay Kumar, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India
Achintya Saha, Department of Chemical Technology, University of Calcutta, Kolkata, India

ABSTRACT

In this research, we have developed two-dimensional quantitative structure-activity relationship (2D-QSAR) and group-based QSAR (GQSAR) models employing a dataset of 78 carbamate derivatives (acetylcholinesterase enzyme inhibitors). The developed models were validated using various stringent validation parameters. From the insights obtained from the developed 2D-QSAR and GQSAR models, we have found that the structural features appearing in the models are responsible for the enhancement of the inhibitory activity against the AChE enzyme. Furthermore, we have performed the pharmacophore modeling to unveil the structural requirements for the inhibitory activity. Additionally, molecular docking studies were performed to understand the molecular interactions involved in binding, and the results are then correlated with the requisite structural features obtained from the QSAR and pharmacophore models.

KEYWORDS

2D-QSAR, AChE, Carbamates, Dataset, Docking, GQSAR, Pharmacophore, Validation

1. INTRODUCTION

Alzheimer's disease (AD) is a neurological disorder, characterized by degenerative changes in a variety of neurotransmitter systems (Wenk, 2003). The initial signs of the disease may be a constant decline in loss of short-term memory and intellectual functions, repeatedly accompanied by abnormal behavior such as aggression and depression (Wenk, 2003; Khachaturian, 1985). Various hypotheses have been proposed to explain AD pathogenesis such as cholinergic hypothesis, tau hypothesis and amyloid hypothesis, etc., that only describe the basic causes of disease progression (Roses, 1996). Among these hypotheses, the most prominent is the cholinergic hypothesis which states that AD is caused by a reduction of the activity of choline acetyltransferase in the cerebral cortex and hippocampus of brain area (Singh and Kaur et al., 2013 and Anand and Singh, 2013). The decreased level of the neurotransmitter causes loss of the cholinergic neurotransmission and large-scale aggregation of Aβ leading to the loss of intellectual abilities (Talesa, 2001). This hypothesis usually suggests that the cholinergic amplification will improve the perception in AD (Anand and Singh, 2013). Thus, the AChE has been proven to be the most promising therapeutic target for the symptomatic treatment of AD (Anand and Singh, 2013; Talesa, 2001). There are now five approved drugs for the treatment of cognitive symptoms of AD, four are AChE inhibitors (Tacrine, Rivastigmine, Galantamine and

DOI: 10.4018/IJQSPR.2020070102

Donepezil) and one is non-competitive glutamate (NMDA) receptors antagonist (Memantine) (Schelterns et al., 2003). The benefit from their use is only symptomatic, and no medicine has been clearly shown to delay or halt the progress of the disease (Schelterns et al., 2003). Therefore, there is an urgent need to develop novel treatment strategies for the proper cure of AD. For this observation, in current study, we have performed 2D-QSAR, GQSAR and 3D pharmacophore modeling along with molecular docking studies to reveal the structural requirements for the AChE enzyme inhibitory activity.

In the current scenario, computational and chemoinformatic methods such as quantitative structure-activity relationships (QSARs) and molecular docking have demonstrated their great potential in designing leads for complex diseases. Among these methodologies, QSARs have been effectively used to identify the important structural features for selective biological activities. Currently, a number of different regression and pattern recognition techniques are available, which can be used for the selection of significant variables and QSAR model development. A number of computational studies have been reported (Brahmachari et al., 2015; Shen et al., 2007; de Souza et al., 2012; Goyal et al., 2014; Solomon et al., 2009; Karmakar et al., 2019; Gupta et al., 2011; Bernd et al., 2003; Saw et al., 2016; Paula et al., 2017; Planche et al., 2013; Planche et al., 2012; Francisco et al., 2012; Planche et al., 2012) so far for the designing of new agents against AD, but still we are far from finding a precise treatment strategy for AD. In the present study, we have employed a dataset of 78 (Sterling et al., 2002) structurally diverse carbamates derivatives with defined AChE enzyme inhibitory activity for the purpose QSAR model development in order to explore the key structural features that are essential for inhibitory activity against AChE enzyme. Prior to the development of final models, we have applied a multilayered variable selection approach to reduce noise in the input, and the final models was developed using the Partial Least Squares (PLS) regression technique. The QSAR models were built with the guidelines of the Organization for Economic cooperation and development (OECD) (Roy et al., 2015). The developed models have been validated taking into consideration various strict internal and external validation metrics. Moreover, we have performed pharmacophore modeling to unveil the structural requirements for the inhibitory activity and to categorize the compounds into more active and less active classes for their inhibitory potential against the AChE enzyme. Furthermore, we have implemented molecular docking studies with most active and least active compounds from the whole dataset and tried to justify the contributions of different descriptors/features as apparent in the QSAR/pharmacophore models.

2. MATERIALS AND METHODS

2.1 QSAR modeling

2.1.1 Dataset

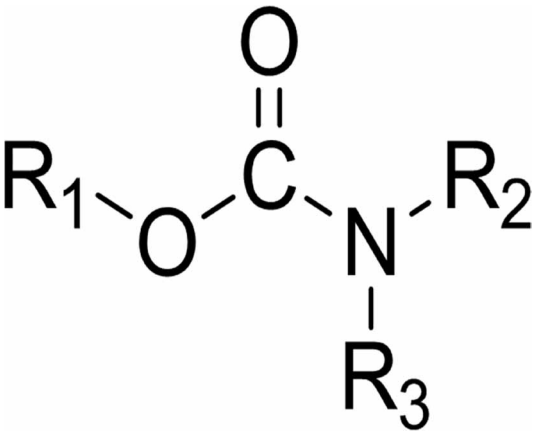
In the current study, we have collected a congeneric series of 78 carbamate derivatives (Sterling et al., 2002) (See Table 5) with inhibitory activity against AChE enzyme from the previous published literature. All 78 structures of the AChEI (carbamate derivatives) present in the dataset with their names, structures and activity (observed and predicted) against the AChE enzyme are depicted in Tables S2-S7. The activity values for all the compounds were measured using the same experimental method (Ellman assay) (Sterling et al., 2002) by the same research group. The activity values of all the dataset compounds expressed as IC_{50} (μM) values were converted to negative logarithm of IC_{50} (i.e., pIC_{50}) values for the purpose of model development. The structures were drawn using Marvin Sketch software version 5.9.4 (Available at https://chemaxon.com/products/marvin) and ChemDraw Ultra software version 12.0 (Available at https://www.perkinelmer.com/category/chemdraw). In GQSAR model development, the designation of common scaffold and substitution sites is a prerequisite step as shown in Figure 1 (Sterling et al., 2002). It can be seen that there are three substitution sites R1, R2 and R3 in the congeneric series used in the GQSAR study. In the GQSAR methodology, every dataset

molecule is considered as a set of fragments, and the fragmentation scheme is either template-based or user defined. Once the common scaffold and substitution sites are defined, various descriptors are calculated for each fragment of the molecule. Now using an appropriate variable selection method, the significant descriptors representing the particular substituent sites are selected.

2.1.2 Descriptor Calculation and Data Pretreatment

In case of 2D-QSAR model development, a pool of 174 descriptors was computed from the congeneric series of 78 carbamate derivatives (Sterling et al., 2002) against AChE enzyme using two software tools, namely, Dragon version 7 (Available at http://www.talete.mi.it/products/dragon_description.htm) and PaDEL-Descriptor version 2.20 software (Available at http://www.yapcwsoft.com/dd/padeldescriptor/). We have calculated only 2D descriptors covering constitutional, ring descriptors, connectivity index, functional group counts, atom centered fragments, 2D atom pairs, atom type E-states and molecular properties (Dragon software version 7) and ETA indices (PaDEL-descriptor 2.20). In case of GQSAR modeling, a pool of 213 descriptors was calculated for every fragment at the substitution sites (R1, R2, and R3) using the tool VLifeMDS version 3 (Available at https://www.vlifesciences.com/support/request_demo.php) software covering physiochemical and atom type count descriptors. We have performed data pretreatment for both the sets of descriptors to remove inter-correlated descriptors from the datasets using the tool Pretreatment V-WSP version 1.2 (available at http://dtclab.webs.com/software-tools).

Figure 1. Common scaffold (carbamate) and the substitution sites assigned for the GQSAR study



2.1.3 Dataset Division

In present study, our aim was to develop QSAR models that are robust enough and capable of making accurate and reliable predictions. Therefore, QSAR models were developed by a training set and validated using new chemical entities, i.e., a test set to check the predictive capacity of the developed models. In this study, the whole data set was divided into a training set and a test set based on activity/property algorithm using the "Dataset Division GUI" developed by our group (available at http://dtclab.webs.com/software-tools). The same strategy was also applied in the case GQSAR study.

2.1.4 Multilayered Variable Selection and QSAR Model Development

Prior to the development of the final models, we tried to extract the important descriptors from the large pool of initial descriptors using various variable selection strategies (Kumar et al. 2019). For

this purpose, we have applied a multilayered variable selection strategy before the development of the final models using stepwise regression (using a suitable stepping criterion, e.g., 'F-for-inclusion 'and 'F-for-exclusion' based on partial F-statistic) (Khan et al., 2019) followed by genetic algorithm (GA) followed by best subset selection, and the final models were build using partial least squares (PLS) regression techniques. The same variable selection strategy was applied in both cases (2D-QSAR and GQSAR). The detail multi-layered variable selection strategy is schematically represented in Figure 5.

2.1.5 Statistical Validation Metrics

To judge the robustness and predictive quality of the developed models is a critical step in the QSAR study (Roy et al., 2007). In this study, we have employed different statistical approaches such as internal and external validation metrics to justify the robustness and predictive quality of developed models. In case of internal validation, we have determined various statistical metrics such as determination coefficient (R^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), Avg $rm^2_{(LOO)}$ and Δrm^2 (Roy et al., 2007). Higher values of the metrics R^2, $Q^2_{(LOO)}$) and Avg $rm^2_{(LOO}$ indicated a better fit of the model, but all these parameters are not sufficient to evaluate the robustness and predictivity of significant models (Roy et al., 2007). Thus, we have determined other statistical validation parameters (external validation parameters) such as Q^2F_1, Q^2_{F2}, r^2m parameters like average rm^2 (test) and Δrm^2 and concordance correlation coefficient (CCC) to assure the significance of developed models (Roy et al., 2007). Moreover, we have also performed Y-randomization test, checked applicability domain criteria, etc., to investigate the robustness of developed models. The Y-randomization test was performed using the Simca-P software (Available at http://www.minitab.com/en-us/products/minitab/) through randomly reordering (100 permutations) the dependent variable (Veerasamy et al., 2011). The details of the methodology are depicted in Figure 6.

2.2 Development and Validation of 3D QSAR Pharmacophore Model

In present study, we have applied pharmacophore modeling study to reveal the required features essential for the inhibitory activity against AChE enzyme. The AChE enzyme inhibitory activity expressed in term of IC_{50} (μM) was used as dependent variable for the purpose of the development of pharmacophore models. Previously prepared compound structures were used for this study. The dataset was rationally distributed into training (23) (for model development) and test set (55) (For validation) compounds based on the span over four order of magnitude (Pal et al., 2019). BIOVIA Discovery Studio client 4.1 (Available at https://www.3dsbiovia.com/) platform was used to build the pharmacophore models. The details of the methodology for the development of pharmacophore model are as described by Aher et al. (Aher et al., 2014). Validation of the developed models was performed using different parameters like cost analysis, Fischer randomization test (F-test) and test set prediction, to judge the robustness and predictive quality of models as described by Aher et al. (Aher et al., 2014).

2.3 Molecular Docking Studies

In this study, we have performed molecular docking studies to identify the interactions of the AChE enzyme (the structure of the protein was retrieved from Protein Data Bank with PDB ID: 4M0E (Murzin et al., 1995)) with the most and least active AChE enzyme inhibitors from the dataset. Molecular docking studies were carried out using BIOVIA Discovery Studio client 4.1 (Available at https://www.3dsbiovia.com/) platform using CDOCKER module of receptor-ligand interaction (Wu et al., 2003). After docking, the generated poses were sorted according to CDOCKER interaction energy and the top scoring (most negative, thus favorable to binding) poses are kept.

3. RESULTS AND DISCUSSION

3.1 Mechanistic Interpretation of Modeled Descriptors

The statistically significant 2D-QSAR and GQSAR models derived using the PLS regression-based technique along with the values of their validation parameters are shown below in equations 1 and 2. The details of their validation metric values are summarized in Table 14. The obtained results suggested that the models were acceptable in terms of fitness, stability and classical predictivity measures. The reported 2D-QSAR model was developed by using 6 descriptors with corresponding latent variables of 3, while the GQSAR model was developed by using 7 descriptors with corresponding latent variables of 6. The descriptors appearing in the models define the structural and functional requirements which can improve the inhibitory activity of molecules against AChE enzyme. The proximity of the observed and predicted values for the AChE enzyme inhibitors in the data set can be further established from the scatter plots as shown in Figure 2. The quantitative contributions of similar/dissimilar descriptors (similar descriptors are placed in close proximity) and the interrelationships between the X-variables and the Y-response are depicted in the loading plots Figure 7. Additionally, we have also performed Y-Randomization test to check whether the models were obtained by any chance or not. The results obtained from the randomized models in case of 2D-QSAR (Model 1: R^2rand = -0.0166 and Q^2rand = -0.379) and GQSAR (Model 2: R^2rand = 0.0545 and Q^2rand = -0.606) suggested that the developed models were not obtained by any chance correlation as given in Figure 8.

3.1.1 2D-QSAR Analysis

Model 1

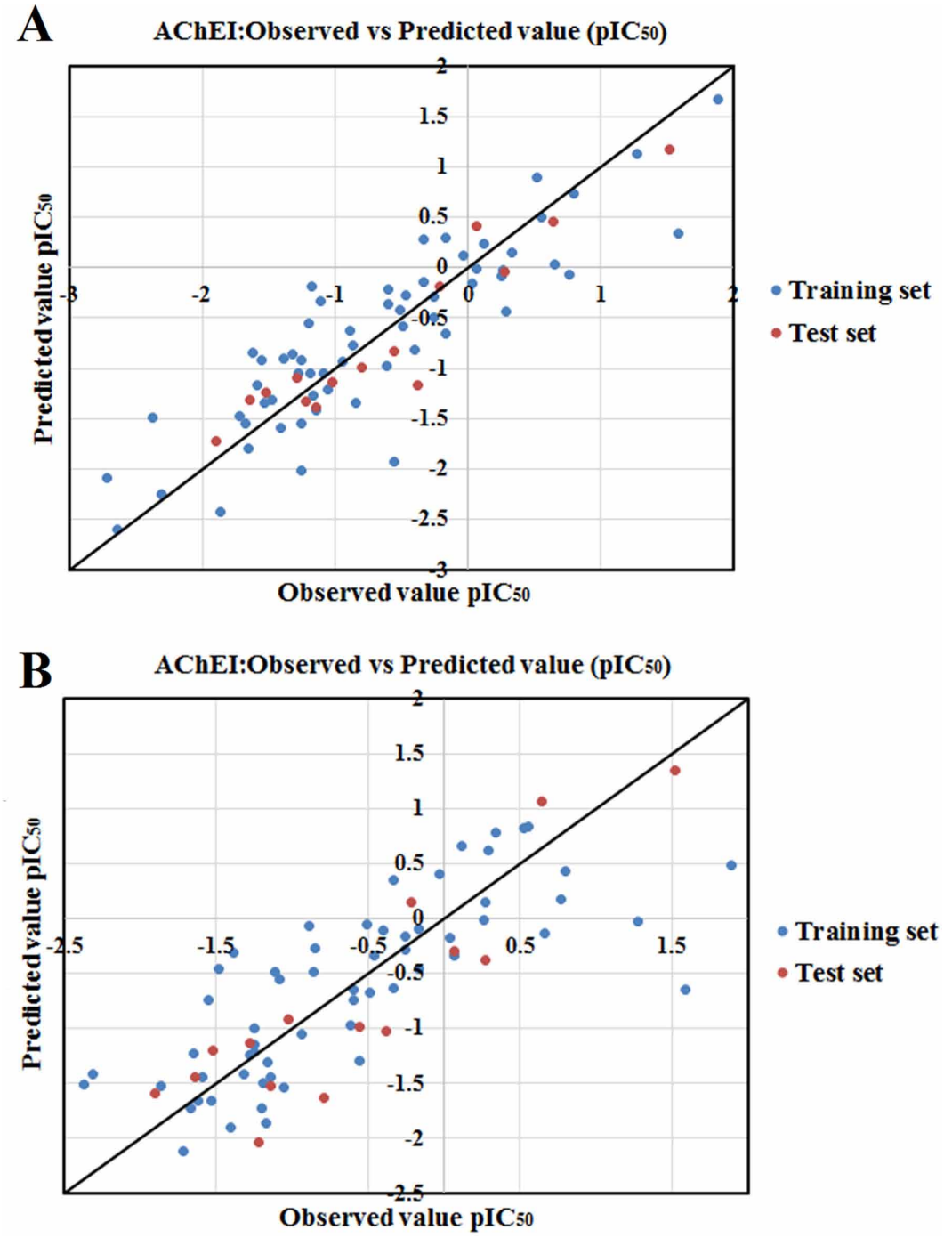
$$pIC_{50}\left(\mu M\right) = -9.011 - 0.541 \times C - 006 - 0.466 \times F07\left[C - N\right] - 0.190 \times H - 052 + 28.395 \times X2A + 0.554 \times F06\left[C - N\right] + 0.481 \times NRS \quad (1)$$

$N_{training} = 63$, $R^2 = 0.789$, $Q^2_{LOO} = 0.732$, $Avgrm^2_{(LOO)} = 0.634$, $\Delta rm^2 = 0.149$, LV = 3, EL = 6, Prediction quality $= MODERATE$; $N_{test} = 15$, $Q^2F_1 = 0.883$, $Q^2F_2 = 0.883$, Avg $rm^2 = 0.852$, $\Delta rm^2 = 0.062$, CCC = 0.935, MAE = 0.261, SD = 0.174, Prediction quality $= GOOD$.

The descriptors in the PLS models are arranged accordingly to their importance, and then described separately. The significance level and contribution of the model descriptors towards the AChE inhibitory activity are determined based on variable importance plot (VIP) and regression coefficient plot as shown in Figure 9 (Popelier et al. 2006).

The descriptor C-006 belongs to the class of atom-centred fragments that encodes information about the topological environment of an atom. This descriptor indicates the number of CH2RX functional groups (X: heteroatom (O, N, S, P, Se, or halogens), R: any group linked through carbon) that describes each atom by its own atom type and the bond types and atom types of its first neighbors (Duchowicz et al., 2007). The neighbors of a carbon atom in this case can be hydrogen (represented as H), carbon (represented as R), or hetero atoms (represented as X). On the other hand, 2D atom pair descriptor, F07[C-N], is simply characterized by the frequency of C-N at topological distance 7. The negative regression coefficient of these descriptors suggests that the absence of such fragments in the compounds may increase the inhibitory activity against AChE enzyme as observed in (See Figure 10) case of compounds 18 (pIC_{50}: 2.045) and 55 (pIC_{50}: 1.886), whereas presence of such fragments correlates to lower inhibitory activity as observed in (See Figure 10) compounds 52 (pIC_{50}: -2.309) and 73 (pIC_{50}: -2.309).

Figure 2. Scatter plots of observed and predicted values of final PLS (2D QSAR and GQSAR) models against AChE enzyme (A: 2D QSAR, B: GQSAR).



Another atom-centred fragments descriptor, H-052, describes H (hydrogen) attached to C(sp3) with 1X (heteroatom) attached to the next C (Todeschini et al. 2009). It is a simple molecular descriptor

defined as the number of specific atom types in a molecule, and it is calculated by knowing the molecular composition and atom connectivity. The negative sign of H-052 indicates that compounds with higher number of H atoms attached to C0(sp3) with 1X (X = O) attached to next C would show lower inhibitory activity as observed in (See Figure 11) compounds 73 (pIC_{50}: -2.369) and 52 (pIC_{50}: -2.309) while lower numerical values of this fragment correlates the higher inhibitory activity as shown in compounds 18 (pIC_{50}: 2.045) and 55 (pIC_{50}: 1.886) (See Figure 11).

Another 2D atom pair descriptor, F06[C-N], indicates the frequency of C-N at the topological distance 6. The positive regression coefficient of this descriptor indicates that the presence of the C-N fragment at the topological distance 6 may favor the inhibitory activity against the AChE enzyme as found in (See Figure 11) compounds 18 (pIC_{50}: 2.045) and 19 (pIC_{50}: 1.858) (containing descriptor value of 4 for all the cases). On the other hand, compounds with lower numerical value of this descriptor show lower inhibitory activity as observed in (See Figure 11 in) compounds 40 (pIC_{50}: -2.720) and 43 (pIC_{50}: -2.642).

The connectivity descriptor, X2A, corresponds to average connectivity index $^2\chi$ and represents the steric feature of the molecule. The positive coefficient of X2A indicates that an increase in the values of the descriptor will result in an increase in inhibitory activity. This is observed in (See Figure 12) case of compound 61 (pIC_{50}: 0.795), 77 (pIC_{50}: 0.769) (descriptor value 0.333 and 0.32, respectively), and the opposite is seen in compound 40 (pIC_{50}: -2.720) and 43 (pIC_{50}: -2.642) as depicted in Figure 12.

The functional group count descriptor, NRS, indicates the number of ring systems present in the compounds, which contributes positively towards the AChE enzyme inhibitory activity. Hydrophobicity plays an important role for better AChE inhibitory activity as we have observed in compounds (See Figure 12) such as 8 (pIC_{50}: 0.522) and 35 (pIC_{50}: -0.250) containing descriptor value 4 and 3, respectively, showing higher inhibitory activity, while compounds 40 (pIC_{50}: -2.720), 43 (pIC_{50}: -2.642), and 73 (pIC_{50}: -2.369) (containing lower descriptor values 1 in all three cases) (See Figure 12) show lower inhibitory activity.

3.1.2 GQSAR Analysis

Model 2

$$\begin{aligned} &pIC_{50}(\mu M) = \\ &1.681 - 0.650 \times R2 - XKMostHydrophobicHydrophilic + 0.071 \times R2 - \\ &+vePotentialSurfaceArea - 0.0001 \times R1 - HosoyaIndex - 0.483 \times R1 - \\ &SssCH2count - \\ \\ &0.015 \times R1 - +vePotentialSurfaceArea - 0.276 \times R3 - MMFF\_1 + \\ &0.0007 \times R3 - MomInertiaX \end{aligned} \quad (2)$$

N_{train} = 63, R^2 = 0.625, Q^2 = 0.538, EL = 7, LV = 6; Ntest = 15, Q^2F_1 = 0.735, Q^2F_2 = 0.734, Avg rm^2 = 0.681, Δrm^2 = 0.151, SD = 0.231, Prediction quality = Moderate.

Equation 2 corresponds to the best GQSAR model that comprises 7 descriptors. The descriptors appearing in the model are arranged accordingly to their importance, and then described separately. The VIP and regression coefficient plot (Popelier et al. 2006) define the importance of each variable obtained from the final PLS models that are responsible to regulate the AChE enzyme inhibitory activity as shown in Figure 13 in the regression coefficient plot.

The physicochemical descriptor, R2-XKMostHydrophobicHydrophilicDistance, belonging to the subclass Hydrophobicity XlogpK, implies the distance between most hydrophobic and hydrophilic point on the VdW surface computed using the Kellog XlogP method (Mannhold et al. 2006). For each fragment, the descriptor value is calculated by generating van der Waals surface of the fragment (R2 position), putting a probe atom at each point on van der Waals surface and calculating the distance between most hydrophobic and hydrophilic points (Available at https://www.vlifesciences.com/support/request_demo.php). It shows a negative contribution to the inhibitory activity which suggests that relatively less distance between most hydrophobic and hydrophilic group at R2 position may favor the inhibitory activity against AChE enzyme. For instance, most active compounds (See Figure 14) 76 (pIC_{50}: -0.938) and 42 (pIC_{50}: -0.845) in the dataset contain methyl group at R2 position with minimum distance from nitrogen group in contrast the least active compounds (See Figure 14) 32 (pIC_{50}: -1.053) and 33 (pIC_{50}: -1.86) in the dataset containing methyl group at R2 position with maximum distance from nitrogen group.

Another physicochemical descriptor, +vePotentialSurfaceArea, belongs to the subclass Electrostatic descriptors, which signifies the total van der Waals surface area with positive electrostatic potential of the compounds having electron accepting or positive centers at substitution sites (Available at https://www.vlifesciences.com/support/request_demo.php). The most contributing descriptor in position R2 with a positive coefficient value is +vePotentialSurfaceArea which suggests that an increase in positive electrostatic potential of fragment R2 may lead to an increase in the inhibitory activity against AChE enzyme. We have observed that the most active compounds (See Figure 15) 75 (pIC_{50}: -0.486) and 58 (pIC_{50}: -0.892) in the dataset show higher inhibitory activity whereas the least active compounds (See Figure 15) 52 (pIC_{50}: -2.31) in the dataset with descriptor values 21.624 show lower inhibitory activity. In contrast, +vePotentialSurfaceArea is detrimental for R1 position in the compounds as we have observed from the least active compounds (See Figure 15) 52 (pIC_{50}: -2.31) and 73 (pIC_{50}: -2.369) in the data set containing higher number electron accepting or positive centers at R1 substitutions site, whereas the most active compounds (See Figure 15) 18 (pIC_{50}: 2.046) and 19 (pIC_{50}: 1.585) from dataset contain lower number of electron accepting group at R1 position.

The distance based topological descriptor, R1-HosoyaIndex, belonging to the family of physicochemical descriptors, signifies the topological index or Z index of a graph which is the total number of matching in it plus 1 (“plus 1” accounts for the number of matchings with 0 edges), and it can be calculated through equation,

$$Z = \sum_{k=0}^{\left\lfloor \frac{n}{2} \right\rfloor} m\left(G, k\right)$$

where n is the number of the vertices of graph G (order of graph G), [n/2] stands for the integer part of n/2 and m (G, k) is the number of k-matchings of graph G (Randić et al., 2004). It shows a negative contribution to the inhibitory activity against AChE enzyme and suggests that relatively lower numerical values of this fragment may contribute to the inhibitory activity. We have observed that the most active compounds (See Figure 16) 58 (pIC_{50}: -0.892) and 19 (pIC_{50}: -1.585) in the data set containing single phenyl ring at R1 position have lower numerical values for this descriptor, while the least active compounds (See Figure 16) 43 (pIC_{50}: -2.642) and 73 (pIC_{50}: -2.369) in the data set contain indene ring on R1 position (which means higher number of edge and vertices) show higher numerical values for this descriptor.

The physicochemical descriptor, R1-SssCH2count, belongs to the sub class E-state numbers. It gives an indication about the total number of –CH2 groups which are connected with the help of two single bonds. It shows a negative contribution at R1 substitution site of the compounds hinting that a reduction in the number of such groups would be better for the inhibitory activity of the compounds.

We have observed that the most active compounds (See Figure 17) 2 (pIC_{50}: 0.036) that absence of such fragment at R1 position shows higher inhibitory activity whereas in case of the least active compounds (See Figure 17) 43 (pIC_{50}: -2.642), 52 (pIC_{50}: -2.31) and 40 (pIC_{50}: -2.72), there are higher number of such fragment at R1 position which is detrimental for inhibitory activity.

R3-MMFF_1 is an atom type count descriptor and is based on MMFF atom types and their count in each molecule (N-CH3 group at R3 position). It shows a negative contribution to the inhibitory activity and suggests that absence of this group at R3 position may favor the inhibitory activity against AChE enzyme. For instance, most active compounds (See Figure 17) 36 (pIC_{50}: 0.26) in the dataset found absence of methyl group at R3 position; in contrast the least active compounds (See Figure 17) 52 (pIC_{50}: -2.31) and 28 (pIC_{50}: -1.615) in the dataset contain N-CH3 group at R3 position, which is detrimental for inhibitory activity.

R3-MomInertiaX, is a distance based topological descriptor. It refers to the moment of inertia at X-axis of the molecules and implies that incorporation of any group at R3 position that increase the resistance or restricts the internal rotation of the molecule will increase the AChE enzyme inhibitory activity. It shows a positive contribution to the inhibitory activity against AChE enzyme. We have observed from the most active compounds (See Figure 17) 18 (pIC_{50}: 2.046) and 55 (pIC_{50}: 1.886) in the data set that an increase in the value of this descriptor adds to the activity profiles of the molecules. On the other hand, the least active compounds (See Figure 17) 73 (pIC_{50}: -2.369) in the data set show lower numerical values of this descriptor.

3.1.3 Applicability Domain of the PLS Models

In current work, we have tested the applicability domain as described by Roy et al (Roy et al., 2019 and Khan et al. 2019) for test set compounds at 99% confidence level using the DModX (distance to model in X-space) approach available in SIMCA-P 10.0 software (Available at https://umetrics.com/products/simca). We can see from Figure 18 that all the test set compounds in both 2D-QSAR and GQSAR models are within the critical DModX values (Model 1: D-Crit =2.412, Model 2: D-Crit = 3.506).

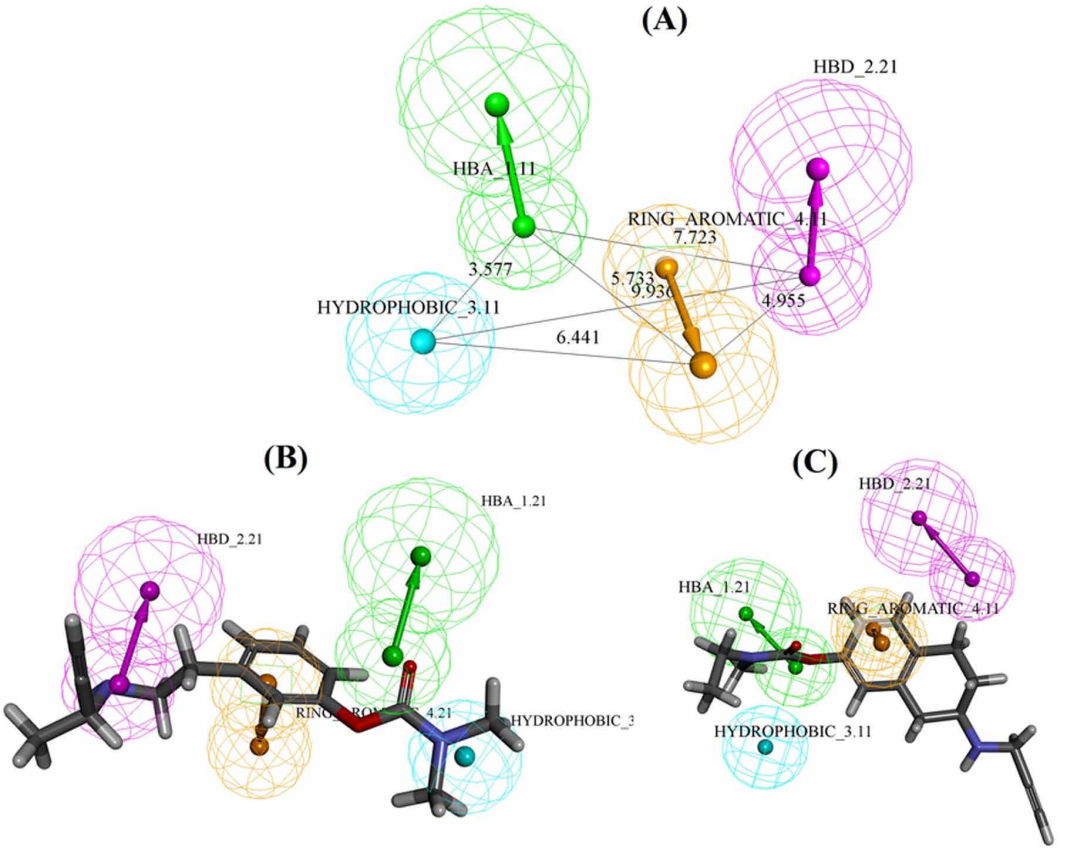
3.2 3D-Pharmacophore Model Analysis

In this study, we have developed ten different pharmacophore models using a training set of 23 compounds. From ten different pharmacophore hypotheses, Hypo-2 with a high correlation coefficient (r: 0.858), lower root mean square deviation (rmsd: 2.061), error 66.77, lower configuration cost (9.909) and weight 1.12 were found to be of acceptable quality. The results of ten pharmacophore hypotheses against AChE enzyme are given in Table 15. Based on all reported metrics, Hypo-2 was found to be the best one among the ten hypotheses with one HBA, one HBD, one RA and one hydrophobic features (Figure 3). External validation of the model has been carried out by mapping the test set molecules (Figure 3) on Hypo-2 with the same settings as employed for the pharmacophore generation by the BEST method. After mapping, we have observed that 50 molecules from the test set of 55 compounds were mapped, only 5 compounds failed in absence of the features found in the developed model. The observed and estimated activity of the training and test sets compounds using Hypo-2 are given in Table 12 and 13. The values of different validation parameters for the training as well as test sets are given in Table 16 (qualitative validation parameters). The F-test confirms the non-randomness of the developed pharmacophore (Hypo-2). The original and randomized total cost values of the hypotheses for F-test are given in Figures 18 and 19. Correlation of 3D pharmacophore model with QSAR (2D-QSAR and GQSAR) models are depicted in Table 1.

3.3. Molecular Docking

We have performed molecular docking study of some most active (19, 18, 44, 55 and 61) (Figure 4 and Figures 21-24) and least active (33, 40, 43, 52 and 73) (Figure 4 and Figures 25-28) compounds of the dataset. The details of docking interactions are depicted in Table 2 and their correlations with

Figure 3. (A) Pharmacophore hypothesis (Hypo-2) with one hydrogen bond acceptor (HBA), one hydrophobic (HYD), one hydrogen bond donor and one ring aromatic (RA) features and interfeature distance (A°); (B) Mapping of the most active compound 77 of the test set (pharmacophore mapping) on the Hypo-2, (C) Mapping of the least active compound 52 of the test set (pharmacophore mapping) on Hypo-2.



the final QSAR (2D-QSAR and GQSAR) models are shown in Table 3. Figure 4 shows the docking interactions of most and least active compounds of dataset (compound 19 and 43).

3.4 Relationship of the Docking Results with QSAR (2D QSAR And GQSAR) Results

In the molecular docking studies, we have observed that the formation of hydrogen bonds (classical and non-classical), hydrophobic bonding (π-π Stacked, π-π-T-Shaped, alkyl and π-alkyl) and some other electrostatic interactions such as π-cation and attractive charges between the ligand and the protein play a vital role in the binding process. Hydrogen and hydrophobic bonding can be correlated with X2A, R2-+vePotentialSurfaceArea, R3-MomInertiaX and NRS descriptors in the QSAR models. X2A, R2-+vePotentialSurfaceArea and R3-MomInertiaX are related to hydrogen bonding, electrostatic interactions between protein and ligand. The descriptor, NRS, gives an evidence of hydrophobic interaction. The descriptor +vePotentialSurfaceArea contributes positively for the substitution site R2; but in contrast, it contributes negatively in case of R1 substitution site as we have observed in the least active compounds from dataset like 33 and 40 (Figures 25 and 26). Furthermore, the R3-MomInertiaX (moment of inertia at X-axis) descriptor supports the evidence of hydrophobic interactions along with hydrogen bonding interactions as we have observed in (Figures 4, 21-24) most active compounds from the dataset (19, 18, 44, 55 and 61), Thus, from above stated information, we

Table 1. Correlation of 3D-QSAR pharmacophore model with QSAR (2D-QSAR and GQSAR) models

Pharmacophoric features	Correlation with QSAR (2D-QSAR and GQSAR) models
Ring aromatic	The compounds present in the dataset have at least one ring aromatic feature, which is either pyrazole/pyridine/thiazole/phenyl or other heterocycle. The RA features are the preliminary requirements for the inhibitory activity against AChE enzyme. The RA features are accordance with the NRS descriptor of the 2D-QSAR model (Eqs. 1). This observation we can see from the most active compound of the test set (**77** IC_{50}: 0.17μM) (Figure 3) one benzene ring lies in the RA region.
Hydrogen bond acceptor	The HBA feature (-CO) of pharmacophore model is also in accordance with the +vePotentialSurfaceArea and MomInertiaX (moment of inertia at X-axis) descriptors of the GQSAR models at R2 and R3 positions respectively in the compounds (Eq. 2). The most contributing descriptor in position R2 is +vePotentialSurfaceArea, which suggests that an increase in positive electrostatic potential of fragment at R2 site may lead to an increase in the inhibitory activity against AChE enzyme. As found the in most active compound of test set (Figure 3) (77 IC_{50}: 0.17μM), the presence of this features in compound shows higher inhibitory activity.
Hydrogen bond donor	Hydrogen bond donor feature from the obtained model accordance with F06 [C-N] descriptor of the 2D-QSAR model.
Hydrophobic	Hydrophobic feature (-CH3) from the developed model is in accordance with the X2A descriptor of the 2D-QSAR model (Eqs. 1). We have observed from most active compound of the test set (77 IC_{50}: 0.17μM) Figure 3, hydrophobic feature on Hypo2 mapped completely with the molecule.

can conclude that hydrogen bonding, hydrophobicity, electrostatic interactions and unsaturation (π-π interaction) features as obtained from both QSAR and docking study are essential for the inhibitory activity against the AChE enzyme.

3.5. Comparison of the Performance of the Present QSAR Models with Previously Published Models

In the present work, we have performed a comparison of the best models of this study with previously published models (Brahmachari et al., 2015, Shen et al., 2007, de Souza et al., 2012, Goyal et al., 2014, Solomon et al., 2009, Karmakar et al., 2019, Gupta et al. 2011 and Bernd., et al 2003) for the prediction of the bioactivity against AChE enzyme, as depicted in the Table 4. The details of different internal and external validation parameters obtained from our models and those obtained from previous models are given in Table 4. It is important to note that the models developed in this study show better quality models of low equation length and less number of variables of LVs and consider more diverse compounds as compared to previously reported models.

4. OVERVIEW AND CONCLUSION

In the current area of drug research and development, computational approaches have their own importance, and they are widely used for the prediction of numerous activity/property/toxicity of different organic materials against different targets. In the current study, we have developed 2D-QSAR and GQSAR models for the prediction of inhibitory activity against AChE enzyme using dataset containing 78 carbamate derivatives (Sterling et al., 2002). The statistical results of the developed models show good predictivity based on both internal and external validation parameters. From the evidences obtained from developed models (2D QSAR and GQSAR) we have concluded that: (1) higher number of steric features in the molecule may enhance the inhibitory activity against AChE enzyme, (2) higher number of ring systems present in the molecules are essential to increase the inhibitory activity against AChE enzyme as also corroborated with the developed pharmacophore model, (3) the CH2RX fragment and H attached to C0(sp3) with 1X attached to next “C” are

Table 2. Details of docking interactions (Most and Least active compounds from dataset)

Sr. No.	Name of compound	Docking interactions
1	19	Most active compound from the dataset; it interacted with the amino acid residues through hydrogen bonding (SER A: 293, TYR A: 341 and TYR A: 124), alkyl, π-alkyl, π-π stacked and π-cation interactions (TRP A: 286 and TYR A: 341 respectively) and salt bridge formation (LEU A: 289 and ASP A: 74) (Figure 4).
2	18	Figure 21 showed that compound 18 (most active compound in dataset) interacted with SER A: 293, TYR A: 341 and TYR A: 124 amino acid residues through hydrogen bonding interaction, TYR A: 341 and TRP A: 286 through π-cation, π-alkyl and π-π stacked) and ASP A: 74 amino acid through salt bridge formation.
3	44	Figure 22 showed that compound 44 interacted with SER A: 293 and TYR A: 124 (through hydrogen bonding), LEU A: 289, VAL A: 294, TYR A: 124, PHE A: 297, TRP A: 286 and TYR A: 341 (alkyl, π-alkyl and π-π-stacked bond respectively) amino acid residues.
4	55	Compound 55 (See Figure 23) interacted with the amino acid residues through hydrogen bonding (TYR A: 341 and SER A: 293) and π-alkyl and π-π-stacked (TYR A: 337, PHE A: 297, PHE A: 338, TYR A: 341 and TRP A: 286 respectively) amino acid residues.
5	61	Most active compound from the dataset, compound 61 (See Figure 24) interacted with TYR A: 337 and TYR A: 72 (through hydrogen bonding), TYR A: 337, PHE A: 297 TRP A: 286 (through π-alkyl and π-cation respectively), TYR A: 341 and ASP A: 74 (via π-π stacked and pi-cation) amino acid residues.
6	33	The least active compound from the dataset, compound 33 (Figure 25), interacted with TYR A: 124 and TYR A: 73 (through hydrogen bonding), PHE A: 338, TRY A: 337, PHE A: 297, TYR A: 124, TYR A: 341, TYR A: 286 (through π-alkyl interaction), TRP A: 286 (via π-π stacked interaction) and LEU A: 76 (through alkyl bond) amino acid residues.
7	40	Another least active compound from the dataset, compound 40, interacted with the amino acid residues through hydrogen bonding (SER A: 293, TRP A: 286, TYR A: 341), π-alkyl (TYR A: 72, TRP A: 286 and TYR A: 341) alky (LEU A: 289 and VAL A: 294) and π-π stacked interactions (TRP A: 286 and TYR A: 341) (See Figure 26).
8	43	Figure 4 showed that compound 43 (one of the least active compound from dataset) interacted with SER A: 293 and TYR A: 73 amino acid residues through hydrogen bonding interaction, TYR A: 72, TRP A: 286, TYR A: 341 amino acid residues through π-alkyl interaction, LEU A: 289 via alkyl bond and TYR A: 341 and TRP A: 286 amino acids via π-π stacked interaction.
9	52	Compound 52 (See Figure 27 interacted with the amino acid residues through hydrogen bonding (TYR A: 124), π-alkyl (TYR A: 72, TYR A: 337, TRP A: 286 and TYR A: 341) and π-π-stacked interactions (TRP A: 286 and TYR A: 341) amino acid residues.
10	73	Figure 28 showed that compound 73 interacted with TYR A: 124 (through hydrogen bonding), TYR A: 337, PHE A: 338, TYR A: 72 and TYR A: 341 through π-alkyl, TYR A: 341 and VAL A: 294 via alkyl bond and TRP A: 286 via π-π-stacked bond) amino acid residues.

detrimental for the AChE enzyme inhibitory activity, (4) relatively less distance between most hydrophobic and hydrophilic group at R2 position may favor the inhibitory activity against AChE enzyme, (5) an increase in positive electrostatic potential of fragment R2 may lead to an increase in the inhibitory activity against AChE enzyme, (6) reduction in the number of –CH2 groups which are connected with the help of two single bonds at R1 position would be better for the inhibitory activity of the compounds and (7) absence of N-CH3 group at R3 position may favor the inhibitory activity against AChE enzyme. Moreover, the results obtained from the QSAR analysis are well supported by pharmacophore mapping and molecular docking studies. The developed QSAR models thus may be helpful for prediction of the activity of new analogues even before their synthesis and evaluation.

Table 3. Docking results and their correlation with the final QSAR (2D QSAR and GQSAR) models

S. No.	Compound Number	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
1	18 (high pIC_{50})	40.306	SER A: 293, TYR A: 341, ASP A: 74, TYR A: 124 and TRP A: 286	Hydrogen bonding (Classical and non-classical), electrostatics (Salt bridge, attractive charges and π-cation) and hydrophobic (π-π stacked and π-alkyl)	NRS, R2-+vePotentialSurfaceArea and R3-MomInertiaX
2	19 (high pIC_{50})	42.26	LEU A: 289, SER A: 293, TRP A: 286, TYR A: 341, ASP A: 74 and TYR A: 124	Hydrogen bonding (Classical, non-classical and Salt bridge), electrostatics (Salt bridge, attractive charges and π-cation) and hydrophobic (π-π stacked, π-alkyl and alkyl)	NRS and R2-+vePotentialSurfaceArea
3	44 (high pIC_{50})	36.76	LEU A: 289, SER A: 293, TRP A: 286, TYR A: 341, VAL A: 249, PHE A: 297 and TYR A: 124	Hydrogen bonding (Classical, non-classical and Salt bridge) and hydrophobic (π-π stacked, π-alkyl and alkyl)	NRS, X2A, R2-+vePotentialSurfaceArea and R3-MomInertiaX
4	55 (high pIC_{50})	40.88	TRP A: 286, SER A: 293, TYR A: 341, PHE A: 297, PHE A: 338 and TYR A: 337	Hydrogen bonding (Classical and non-classical) and hydrophobic (π-π stacked and pi-alkyl)	NRS, R2-+vePotentialSurfaceArea and R3-MomInertiaX
5	61 (high pIC_{50})	42.625	PHE A: 297, TYR A: 341, TRP A: 286, ASP A: 74, TYR A: 72 and TYR A: 337	Hydrogen bonding (Non-classical), Electrostatics (attractive charges and π-cation), hydrophobic (π-π stacked and π-alkyl)	NRS, R2-+vePotentialSurfaceArea and R3-MomInertiaX
6	33 (low pIC_{50})	34.295	TYR A: 124, TYR A: 73, PHE A: 338, TRY A: 337, PHE A: 297, TYR A: 341, TRP A: 286 and LEU A: 76	Hydrogen bonding (Classical and non-classical), hydrophobic (π-π stacked, alkyl and π-alkyl)	NRS, X2A, R1-+vePotentialSurfaceArea
7	40 (low pIC_{50})	31.686	SER A: 293, TRP A: 286, TYR A: 341, TYR A: 72, LEU A: 289, VAL A: 294 and TYR A: 341	Hydrogen bonding (non-classical), hydrophobic (π-π stacked and π-alkyl)	NRS, X2A, R1-+vePotentialSurfaceArea
8	43 (low pIC_{50})	31.129	SER A: 293, TYR A: 72, TYR A: 341, LEU A: 289 and TRP A: 286	Hydrogen bonding (non-classical), hydrophobic (π-π stacked, alkyl and π-alkyl)	NRS, X2A, R2-+vePotentialSurfaceArea
9	52 (low pIC_{50})	34.114	TYR A: 124, TYR A: 72, TYR A: 337, TRP A: 286 and TYR A: 341	Hydrogen bonding (Classical and non-classical), hydrophobic (π-π stacked and π-alkyl)	X2A, R2-+vePotentialSurfaceArea
10	73 (low pIC_{50})	32.807	TYR A: 124, TYR A: 337, PHE A: 338, TYR A: 72, TYR A: 341, VAL A: 294, TRP A: 286	Hydrogen bonding (non-classical), hydrophobic (π-π stacked, alkyl and π-alkyl)	NRS, X2A, R2-+vePotentialSurfaceArea

Figure 4. Docking interactions of most and least active compounds of dataset (compound 19 and 43 respectively)

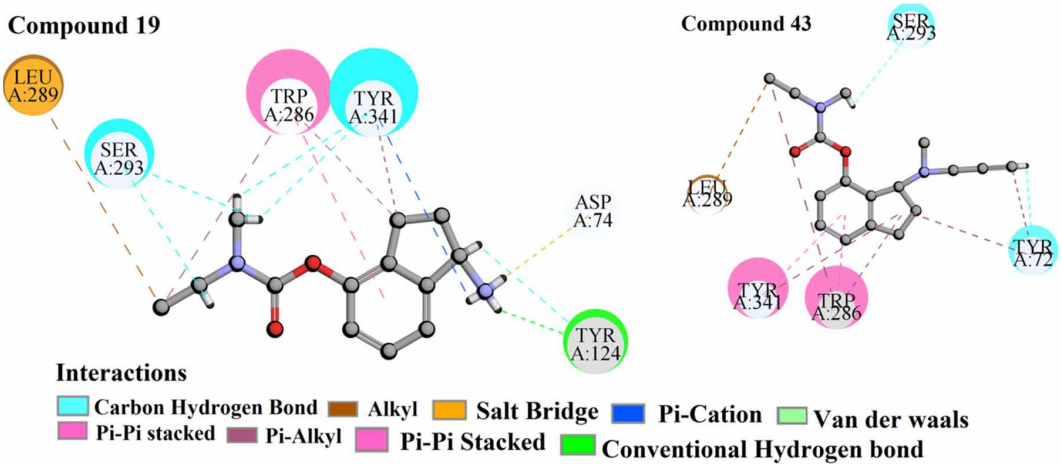


Table 4. Comparisons of proposed study with previous published studies against AChE enzyme

Sources	EL	LV	Model	Training set			Test set	
				n	R^2	Q^2	n	R^2_{pred}
Model 1 (2D QSAR)	**6**	**3**	**PLS**	**63**	**0.789**	**0.732**	**15**	**0.88**
Model 2 (GQSAR)	**7**	**6**	**PLS**	**63**	**0.625**	**0.538**	**15**	**0.735**
Brahmachari et al. 2015	8	5	PLS	325	0.647	0.625	105	0.675
Shen et al. 2007	-	5	CoMFA	36	0.974	0.784	9	0.968
Shen et al. 2007	-	4	CoMSIA	36	0.947	0.736	9	0.927
de Souza et al. 2012	-	8	HQSAR	29	0.965	0.787	7	-
Goyal M et al. 2014	4	4	PLS	19	0.822	0.683	5	0.789
Solomon KA et al. 2009	5	-	GFA	62	0.857	0.803	26	0.882
Karmakar A et al. 2019	3	-	GFA	28	0:683	0:589	-	0:641
Gupta et al 2011	6	-	GFA	31	0.88	0.838	11	0.75
Gupta et al 2011	6	-	GPLS	31	0.889	0.739	11	0.706
Gupta et al 2011	-	-	SVM	31	0.798	-	11	0.762
Gupta et al 2011	-	-	ANN	31	0.753	-	11	0.694
Bernd., et al 2003	-	-	CoMFA	28	0.974	0.671	4	-

Abbreviations: LV= Latent variables, MLR= Multiple linear regression, CoMFA= Comparative Molecular Field Analysis, CoMSIA= Comparative molecular similarity index analysis, ANN = Artificial neural network, SVM = Support vector machine, PLS= Partial least square, GFA= genetic function approximation and HQSAR= Hologram QSAR.

ACKNOWLEDGMENT

Financial assistance from the Indian Council of Medical Research (ICMR), New Delhi in the form of a senior research fellowship (File No: 5/3/8/27/ITR-F/2018-ITR; dated: 18.05.2018) to VK is thankfully acknowledged. CONFLICT OF INTEREST: None.

REFERENCES

Umetrics. (1996). SIMCA-P for windows, Graphical Software for Multivariate Process Modeling. Retrieved from https://umetrics.com/products/simca

Gupta, S., Fallarero, A., Vainio, M. J., Saravanan, P., Santeri Puranen, J., Järvinen, P., ... Mohan, C. G. (2011). Molecular docking guided comparative GFA, G/PLS, SVM and ANN models of structurally diverse dual binding site acetylcholinesterase inhibitors. Molecular informatics, 30(8), 689-706. PMID:27467261

Aher, R. B., & Roy, K. (2014). QSAR and pharmacophore modeling of diverse aminothiazoles and aminopyridines for antimalarial potency against multidrug-resistant Plasmodium falciparum. *Medicinal Chemistry Research*, *23*(9), 4238–4249. doi:10.1007/s00044-014-0997-x

Anand, P., & Singh, B. (2013). A review on cholinesterase inhibitors for Alzheimer's disease. *Archives of Pharmacal Research*, *36*(4), 375–399. doi:10.1007/s12272-013-0036-3 PMID:23435942

Brahmachari, G., Choo, C., Ambure, P., & Roy, K. (2015). In vitro evaluation and in silico screening of synthetic acetylcholinesterase inhibitors bearing functionalized piperidine pharmacophores. *Bioorganic & Medicinal Chemistry*, *23*(15), 4567–4575. doi:10.1016/j.bmc.2015.06.005 PMID:26105711

Wellenzohn, B., Tonmunphean, S., Khalid, A., Choudhary, M. I., & Rode, B. M. (2003). 3D-QSAR Studies on natural acetylcholinesterase inhibitors of Sarcococca saligna by comparative molecular field analysis (CoMFA). *Bioorganic & Medicinal Chemistry*, *13*(24), 4375–4380. doi:10.1016/j.bmcl.2003.09.034 PMID:14643329

ChemDraw software for chemistry. (n.d.). Perkinelmer. Retrieved from https://www.perkinelmer.com/category/chemdraw

de Souza, S. D., de Souza, A. M., de Sousa, A. C., Sodero, A. C., Cabral, L. M., Albuquerque, M. G., & Rodrigues, C. R. et al. (2012). Hologram QSAR models of 4-[(diethylamino) methyl]-phenol inhibitors of acetyl/butyrylcholinesterase enzymes as potential anti-Alzheimer agents. *Molecules (Basel, Switzerland)*, *17*(8), 9529–9539. doi:10.3390/molecules17089529 PMID:22878227

Duchowicz, P. R., Talevi, A., Bellera, C., Bruno-Blanch, L. E., & Castro, E. A. (2007). Application of descriptors based on Lipinski's rules in the QSPR study of aqueous solubilities. *Bioorganic & Medicinal Chemistry*, *15*(11), 3711–3719. doi:10.1016/j.bmc.2007.03.044 PMID:17418580

Goyal, M., Grover, S., Dhanjal, J. K., Goyal, S., Tyagi, C., & Grover, A. (2014). Molecular modelling studies on flavonoid derivatives as dual site inhibitors of human acetyl cholinesterase using 3D-QSAR, pharmacophore and high throughput screening approaches. *Medicinal Chemistry Research*, *23*(4), 2122–2132. doi:10.1007/s00044-013-0810-2

Khachaturian, Z. S. (1985). Diagnosis of Alzheimer's disease. *Archives of Neurology*, *42*(11), 1097–1105. doi:10.1001/archneur.1985.04060100083029 PMID:2864910

Karmakar, A., Ambure, P., Mallick, T., Das, S., Roy, K., & Begum, N. A. (2019). Exploration of synthetic antioxidant flavonoid analogs as acetylcholinesterase inhibitors: An approach towards finding their quantitative structure–activity relationship. *Medicinal Chemistry Research*, *28*(5), 723–741. doi:10.1007/s00044-019-02330-8

Khan, K., Baderna, D., Cappelli, C., Toma, C., Lombardo, A., Roy, K., & Benfenati, E. (2019). Ecotoxicological QSAR modeling of organic compounds against fish: Application of fragment based descriptors in feature analysis. *Aquatic Toxicology (Amsterdam, Netherlands)*, *212*, 162–174. doi:10.1016/j.aquatox.2019.05.011 PMID:31128417

Kumar, V., Ojha, P. K., Saha, A., & Roy, K. (2019). Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease. *SAR and QSAR in Environmental Research*. doi:10.1080/1062936X.2019.1695226 PMID:31865778

Khan, K., & Roy, K. (2019). Ecotoxicological QSAR modelling of organic chemicals against Pseudokirchneriella subcapitata using consensus predictions approach. *SAR and QSAR in Environmental Research*, *30*(9), 665–681. doi:10.1080/1062936X.2019.1648315 PMID:31474156

Mannhold, R. (2006). Calculation of lipophilicity: A classification of methods. In *Pharmacokinetic profiling in drug research* (pp. 333–352). Zurich: Wiley-VCH.

Marvin ChemAxon Software Solutions and Services for Chemistry. (n.d.). Chemaxon. Retrieved from https://chemaxon.com/products/marvin

Minitab. (2000). *MINITAB statistical software Minitab Release 13*. Retrieved from http://www.minitab.com/en-us/products/minitab/

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, *247*(4), 536–540. doi:10.1016/S0022-2836(05)80134-2 PMID:7723011

Mauri, A., Consonni, V., Pavan, M., & Todeschini, R. (2006). Dragon software: An easy approach to molecular descriptor calculations. *Match (Mülheim an der Ruhr, Germany)*, *56*(2), 237–248. Retrieved from http://www.talete.mi.it/products/dragon_description.htm

Pal, S., Kumar, V., Kundu, B., Bhattacharya, D., Preethy, N., Reddy, M. P., & Talukdar, A. (2019). Ligand-based pharmacophore modeling, virtual screening and molecular docking studies for discovery of potential topoisomerase I inhibitors. *Computational and Structural Biotechnology Journal*, *17*, 291–310. doi:10.1016/j.csbj.2019.02.006

Abeijon, P., Garcia, M. X., Caamano, X. O., Yanez, M., Lopez, C. E., Romero-Duran, F. J., & Gonzalez, D. H. (2017). Multi-Target Mining of Alzheimer Disease Proteome with Hansch's QSBR-Perturbation Theory and Experimental-Theoretic Study of New Thiophene Isosters of Rasagiline. *Current Drug Targets*, *18*(5), 511–521. doi:10.2174/1389450116666151102095243 PMID:26521774

Prado, P. F., Garcia, M. X., Escobar, M., Alonso, N., Caamano, O., Yanez, M., & Gonzalez, D. H. (2012). 3D MI-DRAGON: New model for the reconstruction of US FDA drug- target network and theoretical-experimental studies of inhibitors of rasagiline derivatives for AChE. *Current Topics in Medicinal Chemistry*, *12*(16), 1843–1865. doi:10.2174/156802661120906184 3 PMID:23030618

Popelier, P. L., & Smith, P. J. (2006). QSAR models based on quantum topological molecular similarity. *European Journal of Medicinal Chemistry*, *41*(7), 862–873. doi:10.1016/j.ejmech.2006.03.004 PMID:16697489

Speck-Planche, A., V Kleandrova, V., Luan, F., & Cordeiro, N. D. (2013). Multi-target inhibitors for proteins associated with Alzheimer: In silico discovery using fragment-based descriptors. *Current Alzheimer Research*, *10*(2), 117–124. doi:10.2174/1567205011310020001 PMID:22515494

Speck-Planche, A., Luan, F., & Cordeiro, M. (2012). Discovery of anti-Alzheimer agents: Current ligand-based approaches toward the design of acetylcholinesterase inhibitors. *Mini-Reviews in Medicinal Chemistry*, *12*(6), 583–591. doi:10.2174/138955712800493744 PMID:22587771

Speck-Planche, A., Luan, F., & NDS Cordeiro, M. (2012). Role of ligand-based drug design methodologies toward the discovery of new anti-Alzheimer agents: Futures perspectives in Fragment-Based Ligand Design. *Current Medicinal Chemistry*, *19*(11), 1635–1645. doi:10.2174/092986712799945058 PMID:22376033

Randić, M. (2004). Wiener Hosoya Index A Novel Graph Theoretical Molecular Descriptor. *Journal of Chemical Information and Computer Sciences*, *44*(2), 373–377. doi:10.1021/ci030425f PMID:15032514

Roy, K. (2007). On some aspects of validation of predictive quantitative structure–activity relationship models. *Expert Opinion on Drug Discovery*, *2*(12), 1567–1577. doi:10.1517/17460441.2.12.1567 PMID:23488901

Roses, A. D. (1996). The Alzheimer diseases. *Current Opinion in Neurology*, *6*(5), 644–650. doi:10.1016/S0959-4388(96)80098-5 PMID:8937829

Roy, J., Ghosh, S., Ojha, P. K., & Roy, K. (2019). Predictive quantitative structure–property relationship (QSPR) modeling for adsorption of organic pollutants by carbon nanotubes (CNTs). *Environmental Science. Nano*, *6*(1), 224–247. doi:10.1039/C8EN01059E

Roy, K., Kar, S., & Ambure, P. (2015). On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems*, *145*, 22–29. doi:10.1016/j.chemolab.2015.04.013

BIOVIA, Discovery Studio Modeling Environment, Release 4.5. (2015). Dassault Systemes. Retrieved from https://www.3dsbiovia.com/

Schelterns, P., & Feldman, H. (2003). Treatment of Alzheimer's disease; current status and new perspectives. *Lancet Neurology*, *2*(9), 539–547. doi:10.1016/S1474-4422(03)00502-7 PMID:12941576

Sterling, J., Herzig, Y., Goren, T., Finkelstein, N., Lerner, D., Goldenberg, W., & Toth, G. et al. (2002). Novel dual inhibitors of AChE and MAO derived from hydroxy aminoindan and phenethylamine as potential treatment for Alzheimer's disease. *Journal of Medicinal Chemistry*, *45*(24), 5260–5279. doi:10.1021/jm020120c PMID:12431053

Singh, M., Kaur, M., Kukreja, H., Chugh, R., Silakari, O., & Singh, D. (2013). Acetylcholinesterase inhibitors as Alzheimer therapy: From nerve toxins to neuroprotection. *European Journal of Medicinal Chemistry*, *70*, 165–188. doi:10.1016/j.ejmech.2013.09.050 PMID:24148993

Shen, L. L., Liu, G. X., & Tang, Y. (2007). Molecular docking and 3D-QSAR studies of 2-substituted 1-indanone derivatives as acetylcholinesterase inhibitors. *Acta Pharmacologica Sinica*, *28*(12), 2053–7254. doi:10.1111/j.1745-7254.2007.00664.x PMID:18031622

Simeon, S., Anuwongcharoen, N., Shoombuatong, W., Malik, A. A., Prachayasittikul, V., Wikberg, J. E., & Nantasenamat, C. (2016). Probing the origins of human acetylcholinesterase inhibition via QSAR modeling and molecular docking. *PeerJ*, *4*, 2322. doi:10.7717/peerj.2322 PMID:27602288

Solomon, K. A., Sundararajan, S., & Abirami, V. (2009). QSAR studies on N-aryl derivative activity towards Alzheimer's disease. *Molecules*, *14*(4), 1448–1455. doi:10.3390/molecules14041448 PMID:19384276

Talesa, V. N. (2001). Acetylcholinesterase in Alzheimer's disease. *Mechanisms of Ageing and Development*, *122*(16), 1961–1969. doi:10.1016/S0047-6374(01)00309-8 PMID:11589914

Todeschini, R., & Consonni, V. (2009). Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references (Vol. 41). John Wiley & Sons.

VLife MDS software for descriptor calculation. (n.d.). Retrieved from https://www.vlifesciences.com/support/request_demo.php

Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., & Agrawal, R. K. (2001). Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov*, *3*(2), 511–519.

Wenk, G. L. (2003). Neuropathologic changes in Alzheimer's disease. *J. Clin. Psychiat*, *64*(9), 7–10. PMID:12934968

Wu, G., Robertson, D. H., Brooks, C. L., & Vieth, M. (2003). Detailed analysis of grid-based molecular docking: A case study of CDOCKER A CHARMm-based MD docking algorithm. *Journal of Computational Chemistry*, *24*(13), 1549–1562. doi:10.1002/jcc.10306 PMID:12925999

Yap, C. W. (2001). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, *32*(7), 1466–1474. doi:10.1002/jcc.21707 PMID:21425294

APPENDIX 1

Abbreviations

AChE: Acetylcholinesterase enzyme
AD: Alzheimer's disease,
AD: Applicability domain,
VIP: Variable importance plot,
PLS: Partial least square,
SW: Stepwise,
ETA: Extended topochemical atom,
LOO: Leave one out,
QSAR: Quantitative structure activity relationship,
GQSAR: Group based QSAR,
OECD: Organization for economic co-operation and development,
MAE: Mean absolute error
CCC: Concordance correlation coefficient
APP: Amyloid precursor protein
BSS: Best subset selection
LV: Latent variable
CHARMm:Chemistry at Harvard Macromolecular Mechanics
MLR:Multilinear Regression
CoMFA:Comparative Molecular Field Analysis
CoMSIA:Comparative molecular similarity index analysis

APPENDIX 2

Supplementary Tables

Table 5. Dataset employed in present study

NAME	Smiles	$pIC_{50}\mu M$
1	c1ccc(cc1[C@H](N(C)C)C)OC(=O)N(C)C	1.522879
2	c1ccc(cc1[C@H](N(C)C)C)OC(=O)N(CC)C	0.036212
3	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)N	0.119186
4	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C)N	-1.27875
5	c12CC[C@H](c1cc(cc2)OC(=O)N(CCC)C)N	-0.86332
6	c12CC[C@H](c1cc(cc2)OC(=O)N(C)CCCCCC)N	0.275724
7	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C1CCCCC1)N	-0.5977
8	c12CC[C@H](c1cc(cc2)OC(=O)N(C)c1ccc(cc1)OC)N	0.522879
9	c12CC[C@H](c1cc(cc2)OC(=O)NCC)N	-1.24797
10	c12CC[C@H](c1cc(cc2)OC(=O)NCCC)N	-0.17026
11	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)NC	-0.02938
12	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C)NC	-1.58659
13	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)NCC	-1.38917
14	c12CC[C@H](c1cc(cc2)OC(=O)N(C)CC)NCCC	-0.5563
15	c1c2c([C@@H](CC2)N)c(cc1)OC(=O)N(C)C	0.337242
16	c1c2c(c(cc1)OC(=O)N(CC)C)[C@@H](CC2)N	-1.02119
17	c1c2c(c(cc1)OC(=O)N(C)C)[C@@H](CC2)NC	0.29243
18	c1ccc(c2CC[C@H](c12)N)OC(=O)N(C)C	2.045757
19	c1ccc(c2CC[C@H](c12)N)OC(=O)N(CC)C	1.585027
20	c12CCC[C@H](c1cc(cc2)OC(=O)N(C)C)N	-0.17026
21	c12CCC[C@H](c1cc(cc2)OC(=O)N(CC)C)N	-0.79379
22	c12CC[C@H](Cc1cc(cc2)OC(=O)N(C)C)N	-0.51055
23	c12CC[C@H](Cc1cc(cc2)OC(=O)N(CC)C)N	-1.90091
24	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)NCC#C	-0.4624
25	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C)NCC#C	-1.6721
26	c12CC[C@H](c1cc(cc2)OC(=O)N(CCC)C)NCC#C	-1.16435
27	c12CC[C@H](c1cc(cc2)OC(=O)N(C)CCCCCC)NCC#C	-1.1959
28	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C1CCCCC1)NCC#C	-1.6149
29	c12CC[C@H](c1cc(cc2)OC(=O)N(C)c1ccc(cc1)OC)NCC#C	0.065502
30	c12CC[C@H](c1cc(cc2)OC(=O)NCC)NCC#C	-1.13672
31	c12CC[C@H](c1cc(cc2)OC(=O)NCCC)NCC#C	-0.37658
32	c12CC[C@H](c1cc(cc2)OC(=O)N(CCCC)C)NCC#C	-1.05308
33	c12c(cc(cc2)OC(=O)N(CCCC)CC)[C@@H](CC1)NCC#C	-1.85974
34	c12CC[C@H](c1cc(cc2)OC(=O)N(CC)C1CCCCC1)NCC#C	-1.25285
35	c12CC[C@H](c1cc(cc2)OC(=O)N(C)Cc1ccccc1)NCC#C	-0.25042
36	c12CC[C@H](c1cc(cc2)OC(=O)N(C)c1ccccc1)NCC#C	0.259637
37	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)N(C)CC#C	-1.11059
38	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)N(CC)CC#C	-1.25285
39	c1c2c([C@@H](CC2)NCC#C)c(cc1)OC(=O)N(C)C	-0.39794
40	c1c2c([C@@H](CC2)NCC#C)c(cc1)OC(=O)N(CC)C	-2.72016
41	c1c2c(c(cc1)OC(=O)N(CCC)C)[C@@H](CC2)NCC#C	-1.65321
42	c1c2c([C@@H](CC2)N(C)CC#C)c(cc1)OC(=O)N(C)C	-0.8451
43	c1c2c([C@@H](CC2)N(C)CC#C)c(cc1)OC(=O)N(CC)C	-2.64246
44	c1ccc(c2CC[C@H](c12)NCC#C)OC(=O)N(C)C	1.275724
45	c1ccc(c2CC[C@H](c12)NCC#C)OC(=O)N(C)CC	-0.33244

continued on following page

Table 5.Continued

NAME	Smiles	$pIC_{50}\mu M$
46	c12CC[C@H](c1cc(c(Cl)c2)OC(=O)N(CC)C)NCC#C	-1.40654
47	c12CC[C@H](c1cc(c(Cl)c2)OC(=O)N(CCC)C)NCC#C	-1.64246
48	c12CC[C@H](c1cc(cc2)OC(=O)N(C)C)N[C@H](C#C)C	-0.25527
49	c12CCC[C@H](c2cc(cc1)OC(=O)N(C)C)NCC#C	-0.5955
50	c12CCC[C@H](c2cc(cc1)OC(=O)N(CC)C)NCC#C	-1.71933
51	c12CC[C@H](Cc2cc(cc1)OC(=O)N(C)C)NCC#C	-0.61278
52	c12ccc(cc2C[C@@H](CC1)NCC#C)OC(=O)N(C)CC	-2.30963
53	c1ccc(c2CC[C@H](c12)N(C)CC#C)OC(=O)N(CC)C	-1.17319
54	c12CC[C@H](c2cc(cc1)OC(=O)N(C)C)N(C)C	-0.33244
55	c1ccc(c2CC[C@H](c12)N(C)C)OC(=O)N(C)C	1.886057
56	c1ccc(cc1CCN)OC(=O)N(C)C	0.638272
57	c1ccc(cc1CCN)OC(=O)N(CC)C	-1.55023
58	c1ccc(cc1CCN)OC(=O)N(CCC)C	-0.89209
59	c1ccc(cc1CCNC)OC(=O)N(C)C	0.552842
60	c1ccc(cc1CCNC)OC(=O)N(CC)C	-1.31597
61	c1ccc(cc1CCN(C)C)OC(=O)N(C)C	0.79588
62	c1ccc(cc1CCNCC#C)OC(=O)N(C)C	0.657577
63	c1ccc(cc1CCNCC#C)OC(=O)N(C)CC	-1.48001
64	c1ccc(cc1CCNCC#C)OC(=O)N(CCC)C	-1.18752
65	c1ccc(cc1CCN(CC#C)C)OC(=O)N(C)C	0.070581
66	c1ccc(cc1CCN(CC#C)C)OC(=O)N(CC)C	-1.22011
67	c1ccc(cc1C[C@H](C)NCC#C)OC(=O)N(C)C	0.267606
68	c1ccc(cc1C[C@@H](C)NCC#C)OC(=O)N(CC)C	-1.5302
69	c1ccc(cc1C[C@@H](C)NCC#C)OC(=O)N(CCC)C	-1.28103
70	c1ccc(cc1C[C@@H](NCC#C)C)OC(=O)N(C)C1CCCCC1	-0.5563
71	c1ccc(cc1C[C@@H](C)NCC#C)OC(=O)N(C)CCCC	-1.08636
72	c1ccc(cc1C[C@H](C)N(CC#C)C)OC(=O)N(C)C	-0.21484
73	c1ccc(cc1C[C@@H](C)N(C)CC#C)OC(=O)N(CC)C	-2.36922
74	c1ccc(cc1C[C@@H](N(C)CC#C)C)OC(=O)N(CCC)C	-1.51983
75	c1ccc(cc1C[C@@H](N(CC#C)C)C)OC(=O)N(CCCCCC)C	-0.48572
76	c1ccc(cc1C[C@H](C)N(CC#C)C)OC(=O)N(C)C1CCCCC1	-0.93802
77	c1c(CCN[C@@H](C#C)C)cc(cc1)OC(=O)N(C)C	0.769551
78	c1ccc(cc1CCN[C@@H](C#C)C)OC(=O)N(CC)C	-1.14301

Table 6. The list of Carbamate derivatives present in the dataset with their name, structure, and activity (observed and predicted) against AChE enzyme

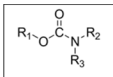
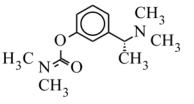
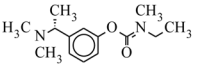
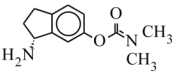
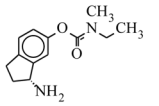
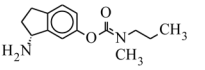
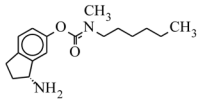
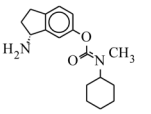
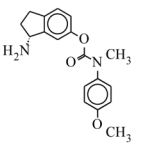
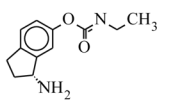
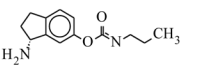
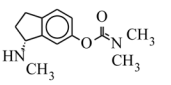
 Common scaffold and the substitution sites				
Cpd.name	Molecular structures	Observed value (pIC_{50} µM)	Predicted value (pIC_{50} µM)	
			2D QSAR	GQSAR
			Model 1	Model 2
1		1.522	1.171	1
2		0.036	-0.153	-0.175
3		0.119	0.235	0.660
4		-1.278	-1.05	-1.237
5		-0.863	-0.765	-0.481
6		0.275	-0.035	-0.375
7		-0.597	-0.364	-0.649
8		0.522	0.900	0.828
9		-1.247	-0.916	-1.002
10		-0.170	-0.656	-0.458
11		-0.029	0.115	0.409

Table 7. The list of Carbamate derivatives present in the dataset with their name, structure, and activity (observed and predicted) against AChE enzyme

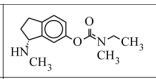
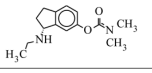
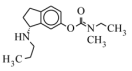
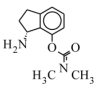
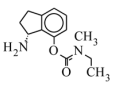
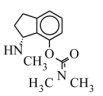
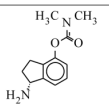
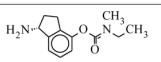
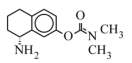
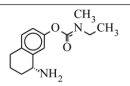
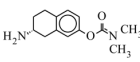
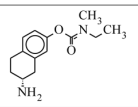
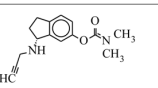
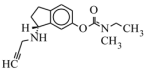
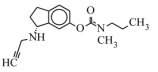
12		-1.586	-1.162	-1.439
13		-1.389	-1	-0.308
14		-0.556	-2	-1.300
15		0.337	0.147	0.788
16		-1.021	-1	-0.915
17		0.292	-0.435	0.626
18		2.045	1.622	1
19		1.585	0.336	-1
20		-0.170	0.300	-0.091
21		-0.793	-1	-2
22		-0.510	-0.426	-0.056
23		-1.900	-2	-2
24		-0.462	-0.268	-0.340
25		-1.672	-2	-2
26		-1.164	-1	-1.303

Table 8. The list of Carbamate derivatives present in the dataset with their name, structure, and activity (observed and predicted) against AChE enzyme

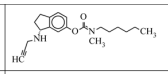
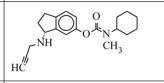
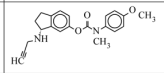
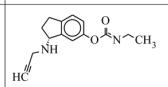
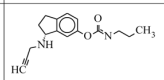
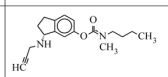
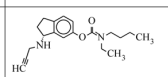
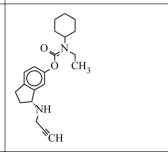
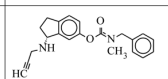
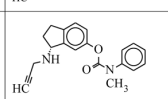
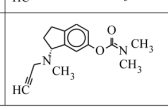
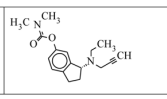
27		-1.195	-1	-1.720
28		-1.614	-1	-1.654
29		0.065	0.414	-0.293
30		-1.136	-1	-1
31		-0.376	-1	-1.026
32		-1.053	-1	-1.542
33		-1.859	-2	-2
34		-1.252	-2	-1.207
35		-0.250	-0.492	-0.282
36		0.259	-0.086	-0.282
37		-1.110	-0.335	-0.487
38		-1.252	-2	-1

Table 9. The list of Carbamate derivatives present in the dataset with their name, structure, and activity (observed and predicted) against AChE enzyme

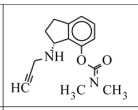
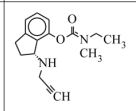
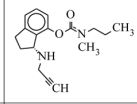
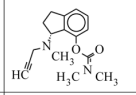
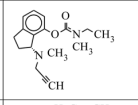
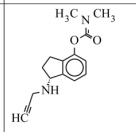
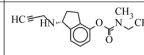
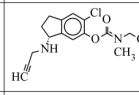
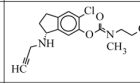
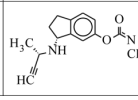
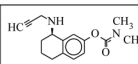
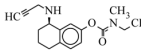
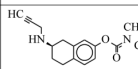
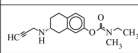
39		-0.397	-1	-0.100
40		-2.720	-2	-2
41		-1.653	-1.803	-1
42		-0.845	-1.340	-0.272
43		-2.642	-2.59	-2
44		1.275	1	-0.023
45		-0.332	-0.139	-1
46		-1.406	-2	-2
47		-1.642	-1	-1
48		-0.255	-0.287	-0.161
49		-0.595	-0.212	-1
50		-1.719	-1	-2
51		-0.612	-1	-1
52		-2.309	-2	-1

Table 10. The list of Carbamate derivatives present in the dataset with their name, structure, and activity (observed and predicted) against AChE enzyme

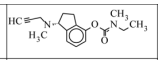
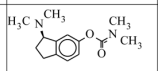
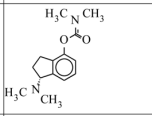
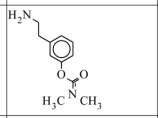
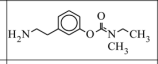
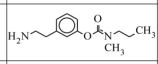
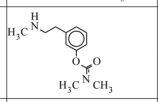
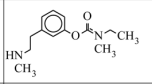
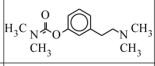
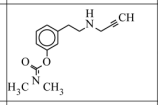
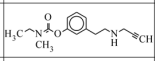
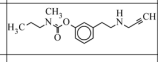
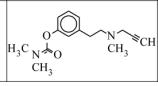
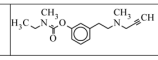
53		-1.173	-0.188	-2
54		-0.332	0.279	0.353
55		1.886	2	0.482
56		0.638	0.458	1.071
57		-1.550	-1	-1
58		-0.892	-1	-0.067
59		0.552	1	0.832
60		-1.315	-1	-1.414
61		0.795	1	-0.128
62		0.657	0.033	-0.128
63		-1.480	-1.309	-0.450
64		-1.187	-1.047	-1
65		0.070	-0.005	-0.335
66		-1.220	-1	-2

Table 11. The list of Carbamate derivatives present in the dataset with their name, structure, and activity (observed and predicted) against AChE enzyme

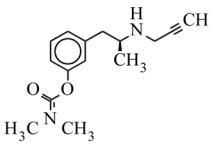
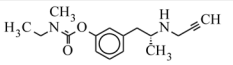
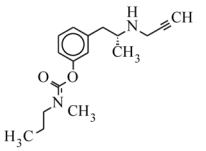
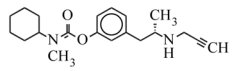
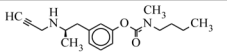
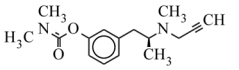
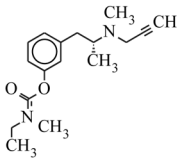
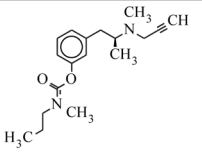
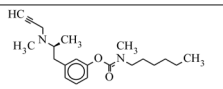
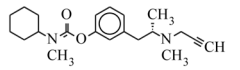
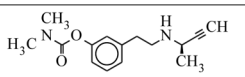
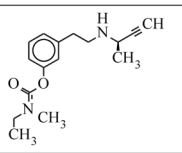
67		0.267	-0.021	0.150
68		-1.530	-1.345	-2
69		-1.281	-1	-1.126
70		-0.556	-1	-1
71		-1.086	-1.054	-1
72		-0.214	-0.193	0.148
73		-2.369	-1.495	-2
74		-1.519	-1.237	-1.197
75		-0.485	-0.581	-1
76		-0.938	-0.935	-1
77		0.769	-0.067	0.179
78		-1.143	-1.389	-1.521

Table 12. Training set implemented in 2D QSAR modeling

NAME	C-006	F07[C-N]	H-052	X2A	F06[C-N]	NRS	pIC50(μm)
2	1	1	6	0.312265	3	1	0.036212
3	0	2	2	0.296444	3	1	0.119186
4	1	2	5	0.290247	3	1	-1.27875
5	1	2	4	0.293715	3	1	-0.86332
7	0	2	6	0.285245	3	2	-0.5977
8	0	2	2	0.283394	4	2	0.522879
9	1	2	5	0.295095	3	1	-1.24797
10	1	2	4	0.297531	3	1	-0.17026
11	0	2	2	0.292228	3	1	-0.02938
12	1	2	5	0.286431	3	1	-1.58659
13	1	2	5	0.29562	3	1	-1.38917
14	2	2	7	0.292185	3	1	-0.5563
15	0	2	2	0.293335	3	1	0.337242
17	0	3	2	0.289244	3	1	0.29243
18	0	0	2	0.29293	4	1	2.045757
19	1	0	5	0.286874	4	1	1.585027
20	0	2	2	0.298728	3	1	-0.17026
22	0	3	4	0.302979	3	1	-0.51055
24	1	2	2	0.297766	3	1	-0.4624
25	2	2	5	0.292185	3	1	-1.6721
26	2	2	4	0.295228	3	1	-1.16435
27	2	2	4	0.300696	4	1	-1.1959
28	1	2	6	0.287225	3	2	-1.6149
30	2	2	5	0.296624	3	1	-1.13672
32	2	2	4	0.297172	3	1	-1.05308
33	3	2	7	0.293749	3	1	-1.85974
34	2	2	9	0.285162	3	2	-1.25285
35	2	2	2	0.292019	3	2	-0.25042
36	1	2	2	0.287225	3	2	0.259637
37	1	2	2	0.29542	3	1	-1.11059
38	2	2	5	0.291941	3	1	-1.25285
39	1	3	2	0.295003	3	1	-0.39794
40	2	3	5	0.289521	3	1	-2.72016
41	2	3	4	0.292656	3	1	-1.65321
42	1	4	2	0.292848	3	1	-0.8451
43	2	4	5	0.287803	3	1	-2.64246
44	1	0	2	0.294643	4	1	1.275724
45	2	0	5	0.289174	4	1	-0.33244
46	2	2	5	0.290363	3	1	-1.40654
48	0	2	5	0.298157	3	1	-0.25527
49	1	2	2	0.299758	3	1	-0.5955
50	2	2	5	0.294301	3	1	-1.71933
51	1	3	4	0.30277	3	1	-0.61278
52	2	3	7	0.297209	3	1	-2.30963
53	2	0	5	0.287479	4	1	-1.17319
54	0	2	2	0.297972	3	1	-0.33244
55	0	0	2	0.294849	4	1	1.886057
57	2	2	5	0.314293	3	1	-1.55023
58	2	2	4	0.317442	3	1	-0.89209
59	1	2	2	0.324894	3	1	0.552842
60	2	2	5	0.316162	3	1	-1.31597

continued on following page

Table 12. Continued

NAME	C-006	F07[C-N]	H-052	X2A	F06[C-N]	NRS	pIC50(μm)
61	1	2	2	0.333039	3	1	0.79588
62	2	2	2	0.3275	3	1	0.657577
63	3	2	5	0.319413	3	1	-1.48001
64	3	2	4	0.321956	3	1	-1.18752
65	2	2	2	0.326119	3	1	0.070581
67	1	2	5	0.326615	3	1	0.267606
68	2	2	8	0.319211	3	1	-1.5302
71	2	2	7	0.32275	3	1	-1.08636
73	2	2	8	0.31393	3	1	-2.36922
75	2	2	7	0.31991	4	1	-0.48572
76	1	2	9	0.304347	3	2	-0.93802
77	1	2	5	0.324999	3	1	0.769551

Table 13. Training set implemented in GQSAR modeling

Name	R2-+vePotentialSurfaceArea	R1-HosoyaIndex	R3-MomInertiaX	R3-MMFF_1	R1-+vePotentialSurfaceArea	R1-SssCH2count	R2-XKMostHydrophobicHydrophilicDistance	pIC50 (μM)
2	34.66	141	3.213	1	99.541	0	3.782	0.036
3	23.841	126	3.295	1	92.561	2	0	0.119
4	32.941	126	3.203	1	91.927	2	3.929	-1.279
5	41.084	126	3.212	1	92.199	2	3.652	-0.863
7	22.574	126	162.121	6	90.048	2	0	-0.598
8	22.303	126	334.82	1	91.158	2	0	0.523
9	32.669	126	0.225	0	92.426	2	3.947	-1.248
10	40.223	126	0.228	0	92.335	2	3.941	-0.17
11	22.665	216	3.207	1	102.298	2	0	-0.029
12	32.352	216	3.218	1	102.027	2	3.909	-1.587
13	22.303	348	3.208	1	114.3	3	0	-1.389
14	21.669	575	12.987	2	124.447	4	0	-0.556
15	22.031	124	304.169	1	90.965	2	0	0.337
17	22.077	216	260.871	1	98.454	2	0	0.292
18	22.167	126	172.521	1	90.74	2	0	2.046
19	34.614	126	3.209	1	89.154	2	3.27	1.585
20	21.986	208	3.205	1	100.417	3	0	-0.17
22	21.941	203	3.205	1	98.046	3	0	-0.511
24	22.031	575	3.211	1	112.95	3	0	-0.462
25	35.384	575	3.209	1	112.814	3	3.587	-1.672
26	44.567	575	3.207	1	113.313	3	3.938	-1.164
27	21.715	575	37.393	6	113.041	3	0	-1.196
28	21.715	575	118.106	6	112.814	3	0	-1.615
30	35.384	575	0.064	0	112.95	3	3.575	-1.137
32	54.703	575	3.204	1	113.268	3	5.417	-1.053
33	53.797	575	447.913	2	113.815	3	5.37	-1.86
34	34.342	575	2753.977	6	113.9	3	3.788	-1.253
35	21.896	575	97.237	1	113.226	3	0	-0.25
36	21.624	575	87.021	0	111.454	3	0	0.26
37	22.077	847	3.213	1	119.879	3	0	-1.111
38	22.167	1418	3.216	1	126.027	4	0	-1.253
39	22.258	575	268.74	1	111.682	3	0	-0.398
40	35.384	575	3.212	1	110.277	3	3.574	-2.72
41	44.522	575	3.208	1	111.86	3	3.86	-1.653
42	22.258	847	231.153	1	118.156	3	0	-0.845
43	35.384	847	3.213	1	117.522	3	3.573	-2.642
44	22.258	575	418.186	1	114.129	3	0	1.276
45	21.896	575	15.803	2	113.537	3	0	-0.332
46	35.384	830	3.209	1	122.079	3	3.581	-1.407
48	22.031	842	3.212	1	129.672	2	0	-0.255
49	22.031	937	3.214	1	103.993	4	0	-0.595
50	35.429	937	3.214	1	104.175	4	3.589	-1.719
51	22.077	922	3.215	1	119.013	4	0	-0.613
52	21.624	922	9.9	2	127.848	4	0	-2.31

continued on following page

Table 13. Continued

Name	R2-+vePotentialSurfaceArea	R1-HosoyaIndex	R3-MomInertiaX	R3-MMFF_1	R1-+vePotentialSurfaceArea	R1-SssCH2count	R2-XKMostHydrophobicHydrophilicDistance	pIC50 (μM)
53	35.565	847	3.214	1	119.516	3	3.604	-1.173
54	21.986	306	3.215	1	101.894	2	0	-0.332
55	22.077	306	274.513	1	107.373	2	0	1.886
57	34.66	63	159.949	1	82.522	2	3.782	-1.55
58	43.209	63	124.779	1	78.917	2	3.729	-0.892
59	22.258	106	370.681	1	92.645	2	0	0.553
60	33.393	106	3.252	1	108.198	2	3.86	-1.316
61	22.348	149	3.211	1	99.857	2	0	0.796
62	22.348	288	3.211	1	103.656	3	0	0.658
63	22.212	288	13.891	2	106.464	3	0	-1.48
64	39.047	288	3.202	1	108.639	3	3.815	-1.188
65	22.574	419	3.214	1	116.523	3	0	0.071
67	22.529	417	3.213	1	116.2	2	0	0.268
68	34.344	417	3.223	1	121.047	2	3.95	-1.53
71	23.705	417	140.011	4	120.281	2	0	-1.086
73	33.032	626	3.212	1	124.264	2	3.449	-2.369
75	59.546	626	3.215	1	116.331	2	5.259	-0.486
76	23.796	626	198.639	6	118.096	2	0	-0.938
77	23.66	427	3.297	1	119.329	2	0	0.77

Table 14. Statistical validation parameters obtained from the developed PLS (2D QSAR and GQAR) models

Models	EL	LV	Internal matrices		External matrices					
			R^2	Q^2	Q^2F_1	Q^2F_2	$\overline{rm2}$	Δrm^2	MAE	Quality
Model 1 (2D QSAR)	6	3	0.789	0.732	0.88	0.88	0.85	0.06	0.261	Good
Model 2 (GQSAR)	7	6	0.625	0.538	0.735	0.734	0.68	0.151	0.412	Moderate

Note. EL: Equation length, LV: Latent variables, MAE: Mean average error.

Table 15. Results of 3D-QSAR pharmacophore development against AChE enzyme

Hypo.	Total cost	ΔCost[a]	ΔCost[b]	Correlation (R)	RMS	Features
1	136.016	110.71	58.212	0.849	2.104	HBA, HBD, RA
2[C]	**137.453**	**109.27**	**59.649**	**0.858**	**2.061**	**HBA, HBD, HYD, RA**
3	140.601	106.12	62.797	0.823	2.253	HBA, HBD, HYD, RA
4	147.077	99.648	69.273	0.799	2.382	HBA,HBD,HYD,HYD_A
5	152.184	94.541	74.38	0.786	2.454	HBA, HBD, HYD, HYD
6	153.808	92.917	76.004	0.770	2.526	HBA, HBD, HYD, RA
7	155.333	91.392	77.529	0.759	2.574	HBA, HBD, HYD, RA
8	157.052	89.673	79.248	0.757	2.585	HBD, HBA, RA
9	161.523	85.202	83.719	0.742815	2.653	HBA, HBD, RA
10	161.814	84.911	84.01	0.750	2.626	HBA, HBD, RA

Cost difference[a]= Null cost - total cost, Cost difference[b]= Total cost - fixed cost, Null cost = 246.725, Fixed Cost = 77.804, Best records in pass: 3, Config. Cost: 9.909, C= Best Hypothesis, Note: RA: Ring aromatic, HYD: Hydrophobic, HYD_A: Hydrophobic aromatic, HBA: Hydrogen bond acceptor, HBD: Hydrogen bond donor

Table 16. Different qualitative validation parameters (%) of Hypo-2 model obtained by classification of more active and less active compounds for the training and test sets of AChEI

Dataset	No. of compounds	Qualitative validation parameters (%)					
		Sensitivity	Specificity	Accuracy	Precision	F-measure	G-Means
Train	**23**	100	95	97.36	94.73	97.29	97.46
Test	**55**	80	95.91	94.44	66.66	72.72	87.59

*Compounds with IC_{50} ≤0.20 µM: more active (H) and IC_{50} > 0.20 µM: least active (L)

Table 17. Training set predicted values from pharmacophore model

Comp no.	EXP. (IC_{50}µM)	Estimated (IC_{50}µM)	Error	Activity Scale (exp.)	Activity Scale (est.)	Fit value
18	0.009	0.0101925	-0.00119	H	H	12.7369
19	0.026	0.09681	-0.07081	H	H	11.7593
56	0.23	0.103692	0.126308	L	H	11.7295
44	0.053	0.157293	-0.10429	H	H	11.5485
36	0.55	2.3706	-1.8206	L	L	10.3703
58	7.8	2.73119	5.06881	L	L	10.3088
12	38.6	5.34238	33.25762	L	L	10.0175
20	1.48	6.4641	-4.9841	L	L	9.93469
22	3.24	8.21193	-4.97193	L	L	9.83075
54	2.15	8.25217	-6.10217	L	L	9.82863
17	0.51	8.37215	-7.86215	L	L	9.82236
51	4.1	8.41832	-4.31832	L	L	9.81997
71	12.2	9.44707	2.75293	L	L	9.7699
78	13.9	11.4896	2.4104	L	L	9.68489
4	19	12.8935	6.1065	L	L	9.63483
50	52.4	17.4649	34.9351	L	L	9.50303
66	16.6	17.5582	-0.9582	L	L	9.50072
53	14.9	18.0564	-3.1564	L	L	9.48857
41	45	19.5647	25.4353	L	L	9.45373
40	525	19.6188	505.3812	L	L	9.45253
43	439	19.8227	419.1773	L	L	9.44804
21	6.22	20.044	-13.824	L	L	9.44322
16	10.5	20.8033	-10.3033	L	L	9.42707

Table 18. Test set predicted values from pharmacophore model

Comp no.	EXP. (IC_{50}µM)	Estimated (IC_{50}µM)	Error	Activity Scale (exp.)	Activity Scale (est.)	Fit value
63	30	0.112044	30	L	H	11.6958
45	2.15	0.14687	2	L	H	11.5783
57	35.5	0.341721	35	L	L	11.2115
67	1	0.387513	0	L	L	11.1569
11	1.07	1.68274	-1	L	L	10.5192
60	21	1.72756	19	L	L	10.5078
59	0	1.76714	-1	H	L	10.4979
77	0	1.95491	-2	H	L	10.4541
64	15.4	2.04088	13	L	L	10.4354
5	7.3	4.52902	3	L	L	10.0892
7	3.96	4.6176	-1	L	L	10.0808
62	0.22	4.95031	-5	L	L	10.0506
24	2.9	5.01809	-2	L	L	10.0447
76	8.67	7.35166	1	L	L	9.87881
28	41.2	7.91492	33	L	L	9.84675
6	0.53	7.9489	-7	L	L	9.84489
10	1	11.5544	-10	L	L	9.68245
34	18	12.9399	5	L	L	9.63327
46	26	15.3451	10	L	L	9.55923
33	72.4	17.9202	54	L	L	9.49186
8	0.3	18.7173	-18	L	L	9.47296
48	1.8	22.9839	-21	L	L	9.38378
52	204	28.7447	175	L	L	9.28664
68	34	35.1554	-1	L	L	9.19921
32	11.3	35.5421	-24	L	L	9.19446
23	79.6	46.1456	33	L	L	9.08107
30	13.7	76.78	-63	L	L	8.85997
69	19	79.09	-60	L	L	8.8471
31	2.38	109.02	-107	L	L	8.7077
25	47	113.22	-66	L	L	8.69129
27	16	129.38	-114	L	L	8.63333
29	1	266.81	-266	L	L	8.319
3	1	284.75	-284	L	L	8.29074
70	4	299.49	-296	L	L	8.26881
9	17.7	366.92	-349	L	L	8.18063
14	4	441.16	-438	L	L	8.10061
35	2	446.26	-444	L	L	8.09561
47	44	555.91	-512	L	L	8.0002
73	234	639.43	-405	L	L	7.93941
75	3	684.17	-681	L	L	7.91004
26	14.6	1,135.01	-1,120	L	L	7.6902
49	4	1,454.94	-1,451	L	L	7.58235
13	25	4,368	-4,343	L	L	7.10495
65	1	8,403.49	-8,403	L	L	6.82074
74	33.1	8,935.01	-8,902	L	L	6.79411
37	13	11,212	-11,200	L	L	6.6955
72	1.64	118,701.00	-118,699	L	L	5.67074
39	2.5	2,228,060.00	-2,228,058	L	L	4.39727
15	0	3,994,510.00	-3,994,510	H	L	4.14374
42	7	5,035,180,000.00	-5,035,179	L	L	1.04318

APPENDIX 3

Supplementary Figures

Figure 5. Schematic representation of multi-layered variable selection strategy for the development of model

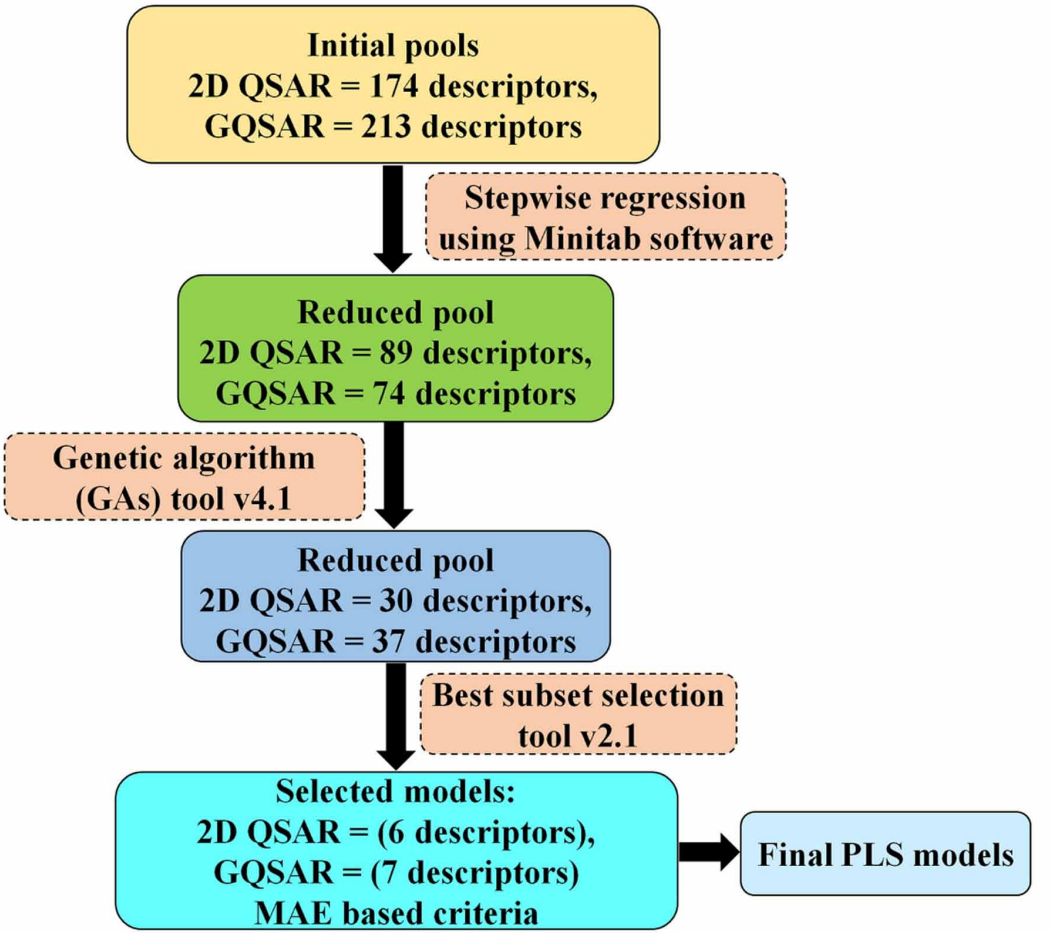


Figure 6. Schematic work flow of QSAR model development against AChE enzyme [PLS = Partial least squares, SR = Stepwise regression, BSS = Best subset selection, GAs = Genetic algorithms]

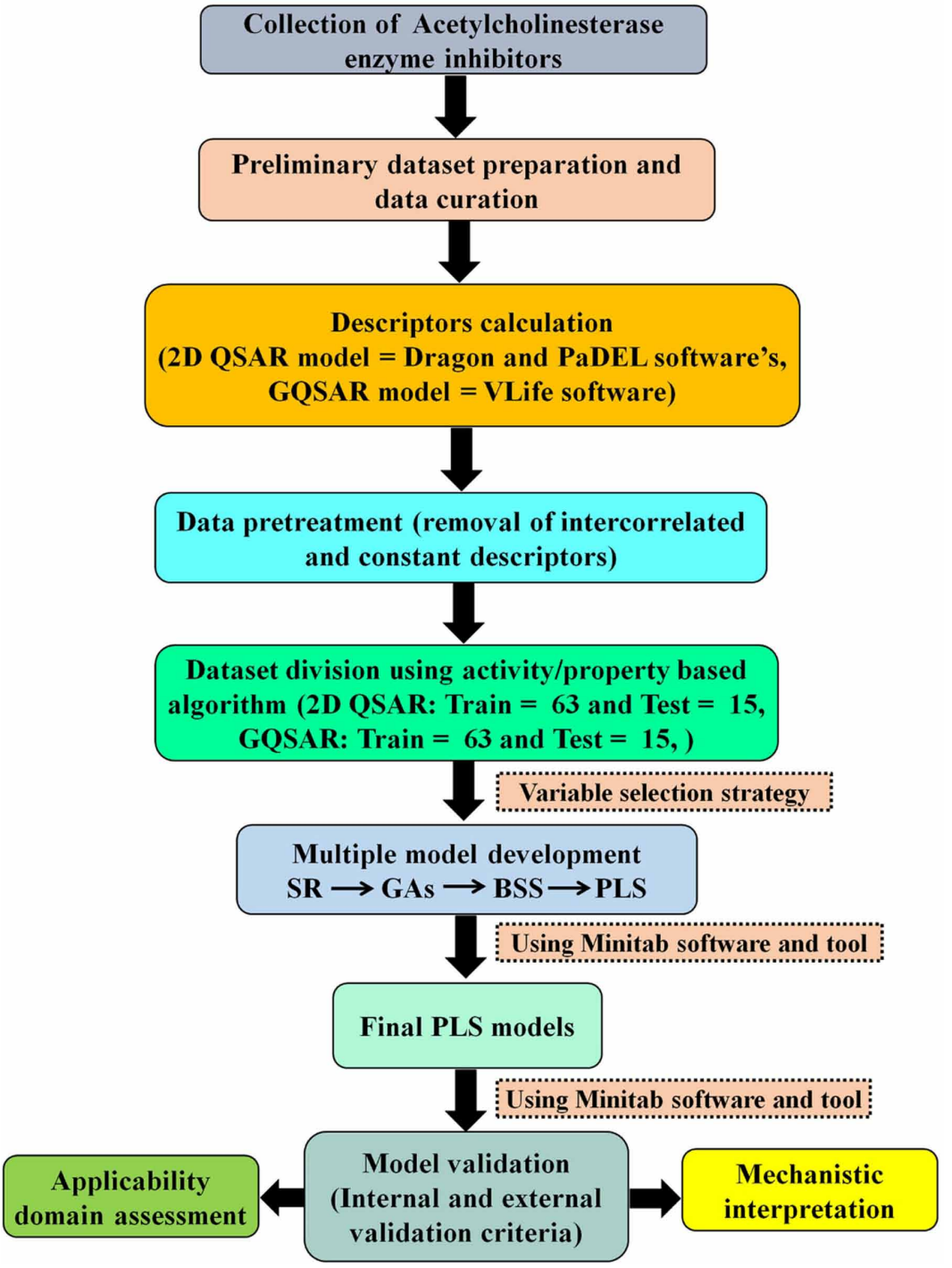


Figure 7. Loading plot for final PLS (2D QSAR and GQSAR) models against AChE enzyme (A: model 1 and B: model 2)

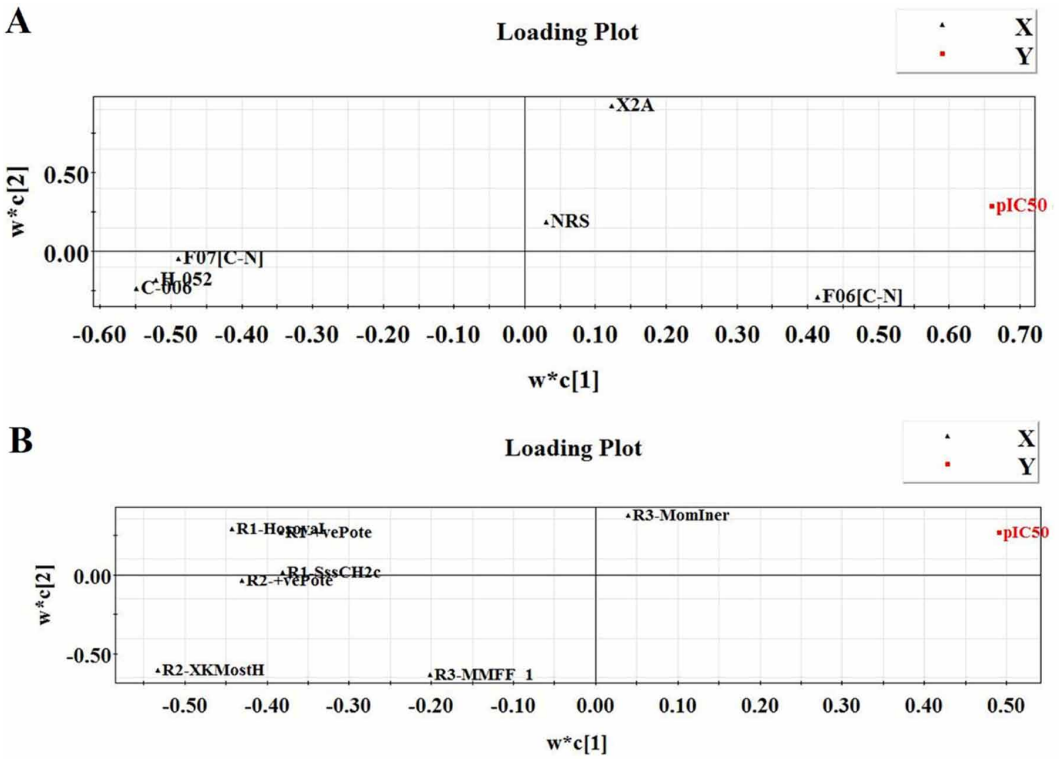


Figure 8. Model Randomization plots for final PLS (2D QSAR and GQSAR) models against AChE enzyme (A: model 1 and B: model 2)

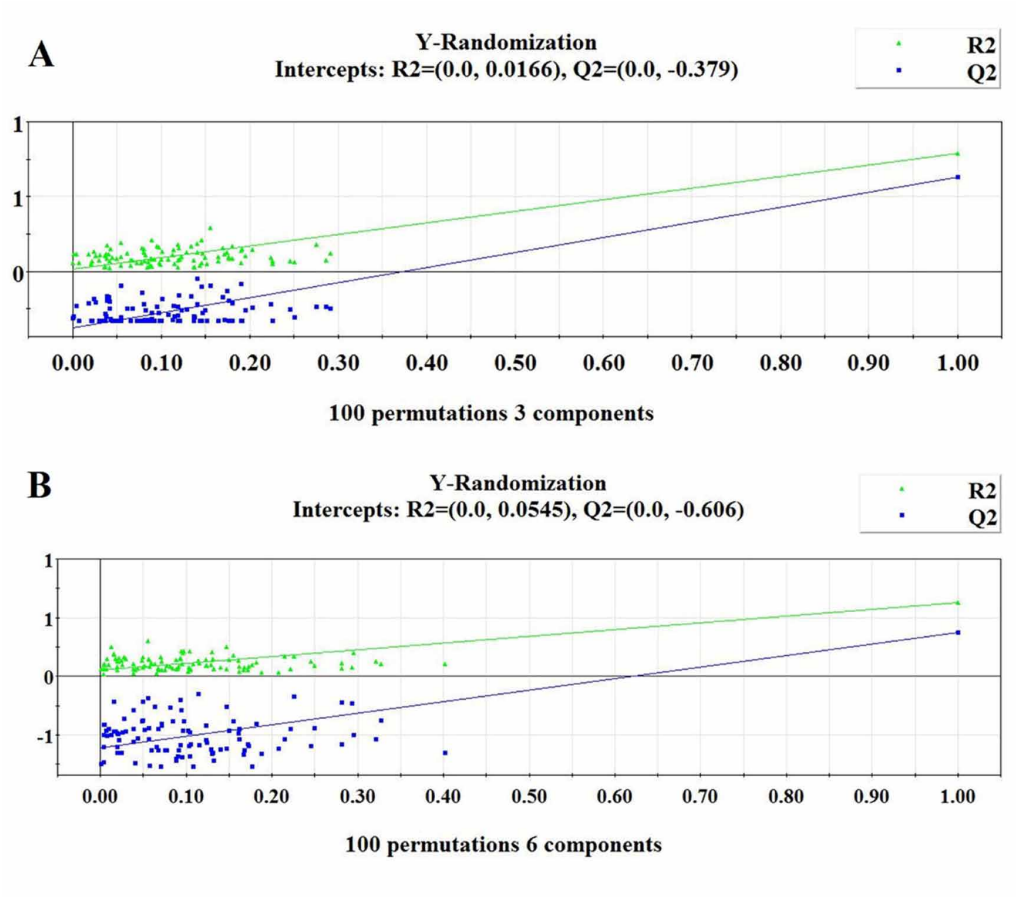


Figure 9. Variable importance plot (VIP) and regression coefficient plot of final PLS (2D QSAR) models against AChE enzyme (A: VIP plot and B: Regression coefficient plot)

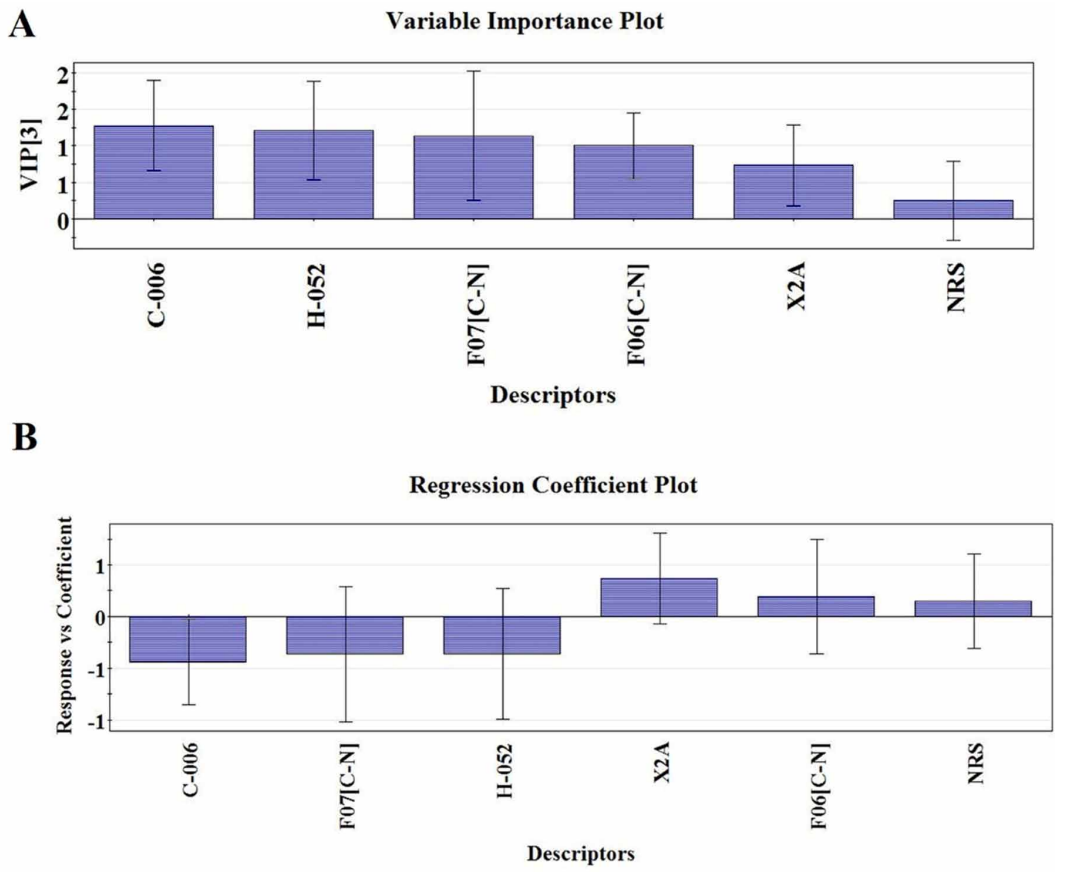


Figure 10. Contributions of C-006 and F07[C-N] descriptors on AChE enzyme inhibitory activity

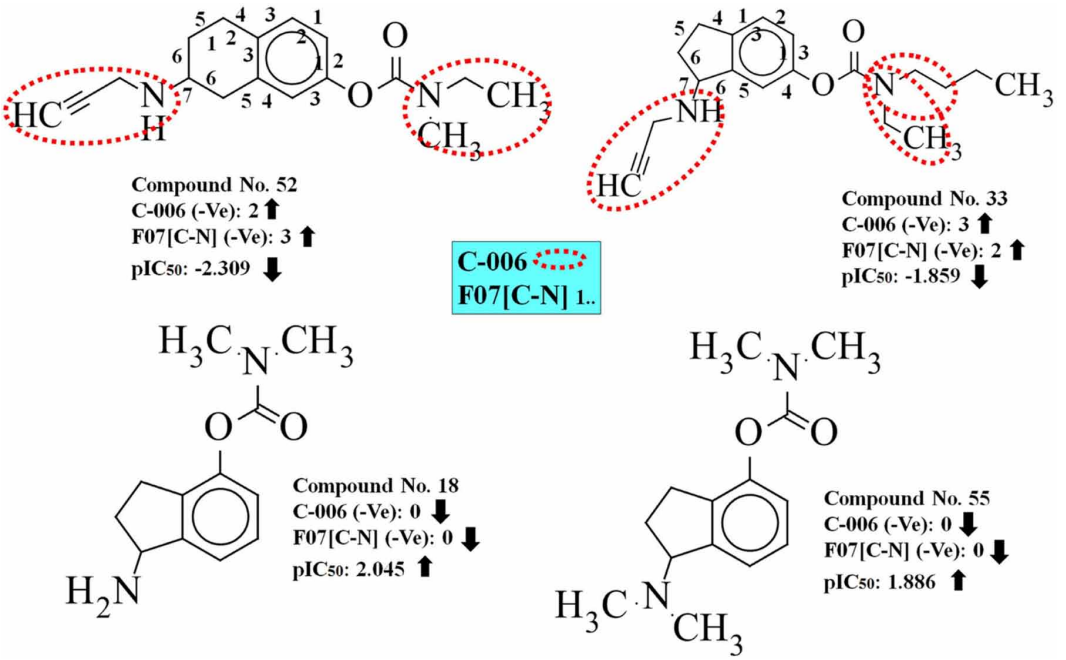


Figure 11. Contributions of H-052 and F06[C-N] descriptors on AChE enzyme inhibitory activity

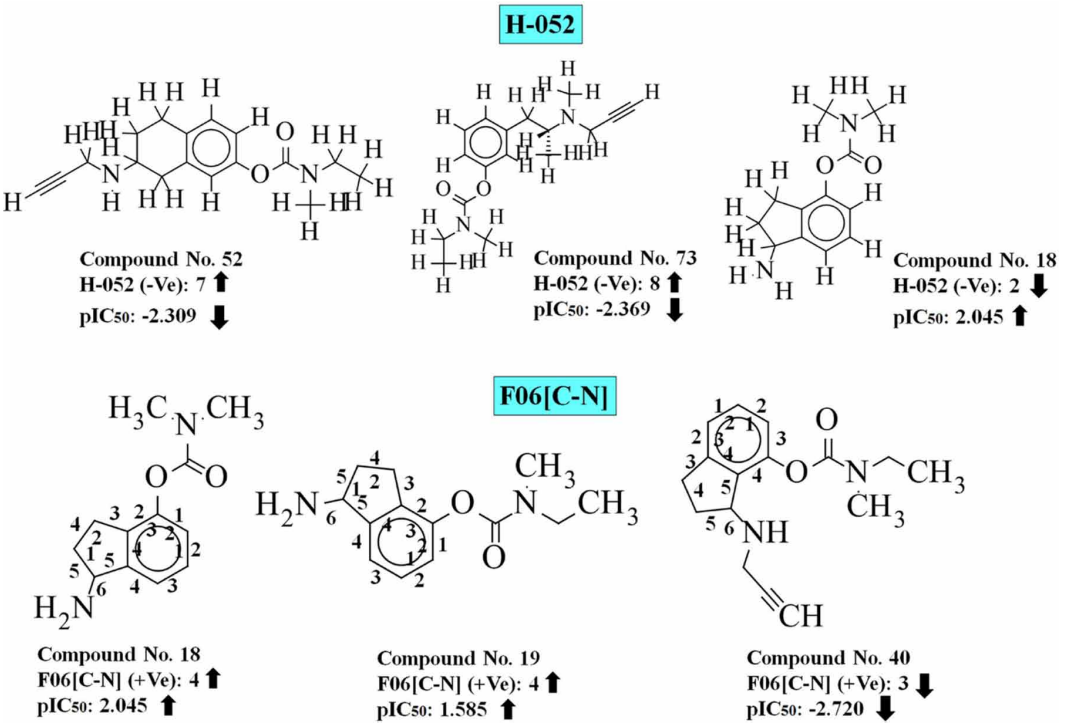


Figure 12. Contributions of X2A and NRS descriptor on AChE enzyme inhibitory activity

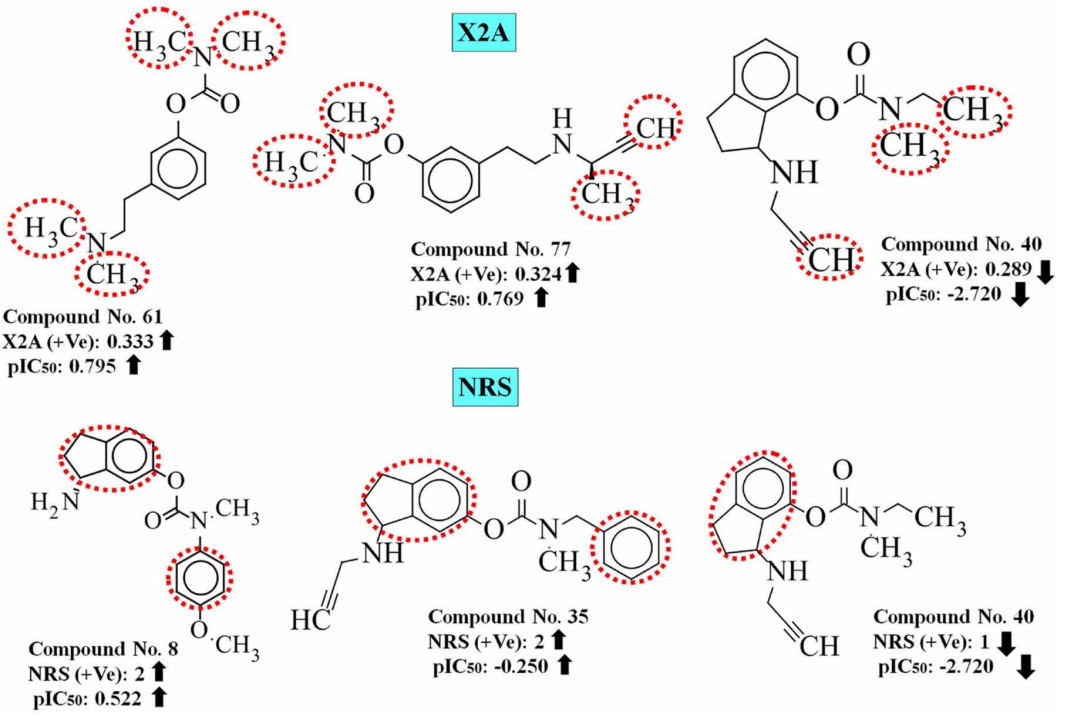


Figure 13. Variable importance plot (VIP) and Regression coefficient plot of final PLS (GQSAR) models against AChE enzyme (A: VIP and B: Regression coefficient plot)

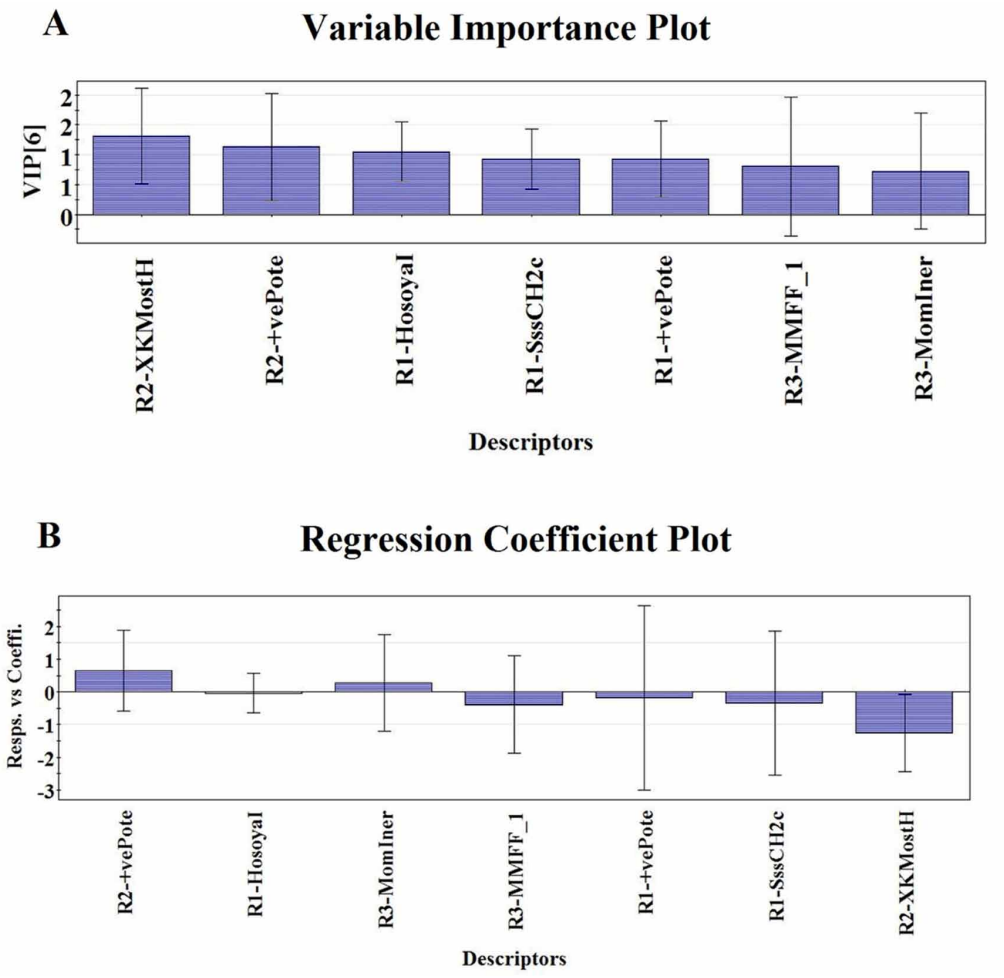


Figure 14. Contribution of XKMostHydrophobicHydrophilicDistance descriptor at R2 position on AChE enzyme inhibitory activity

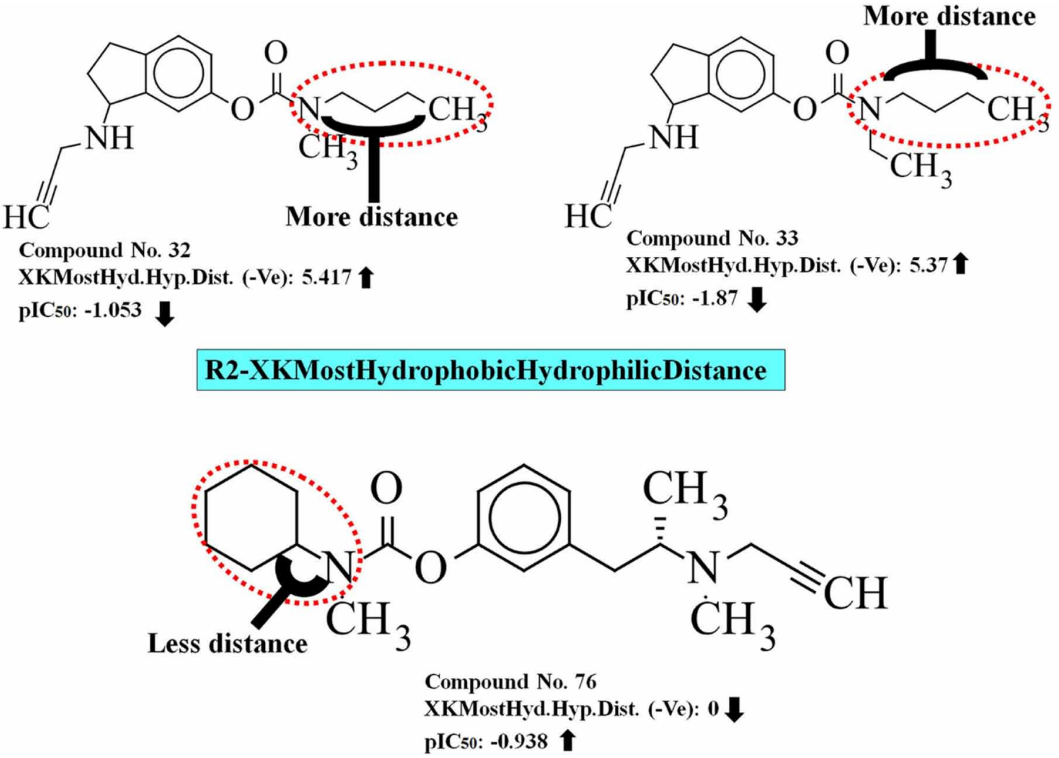


Figure 15. Contributions of +vePotentialSurfaceArea descriptor at R2 position and +vePotentialSurfaceArea descriptor at R1 position on AChE enzyme inhibitory activity

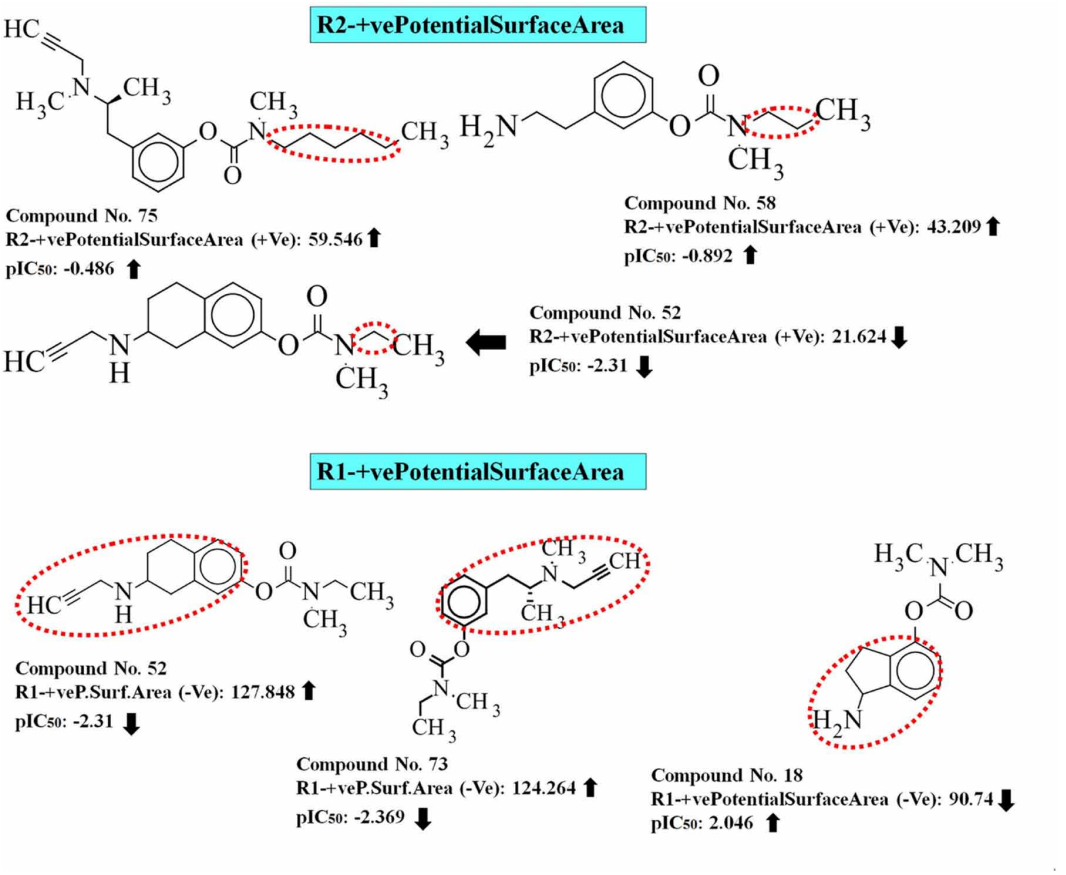


Figure 16. Contributions of HosoyaIndex descriptor at R1 position on AChE enzyme inhibitory activity

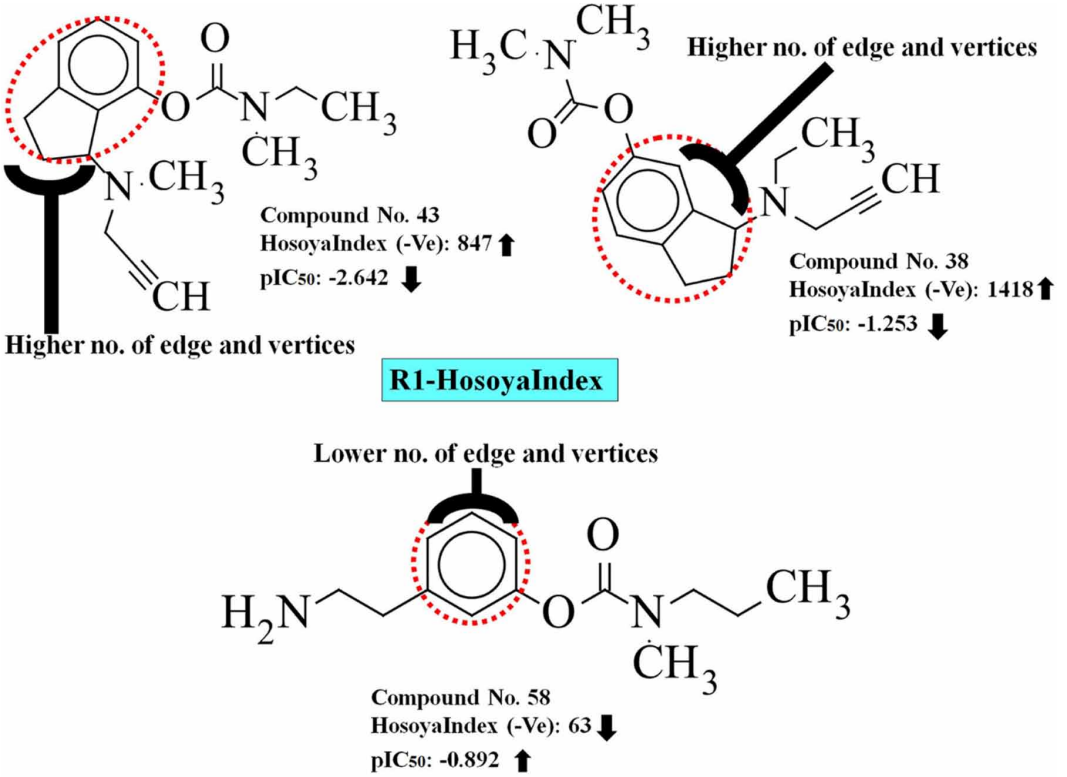


Figure 17. Contributions of SssCH2count descriptor at R1 position, MMFF_1 descriptor at R3 position and MomInertiaX descriptor at R3 position on AChE enzyme inhibitory activity

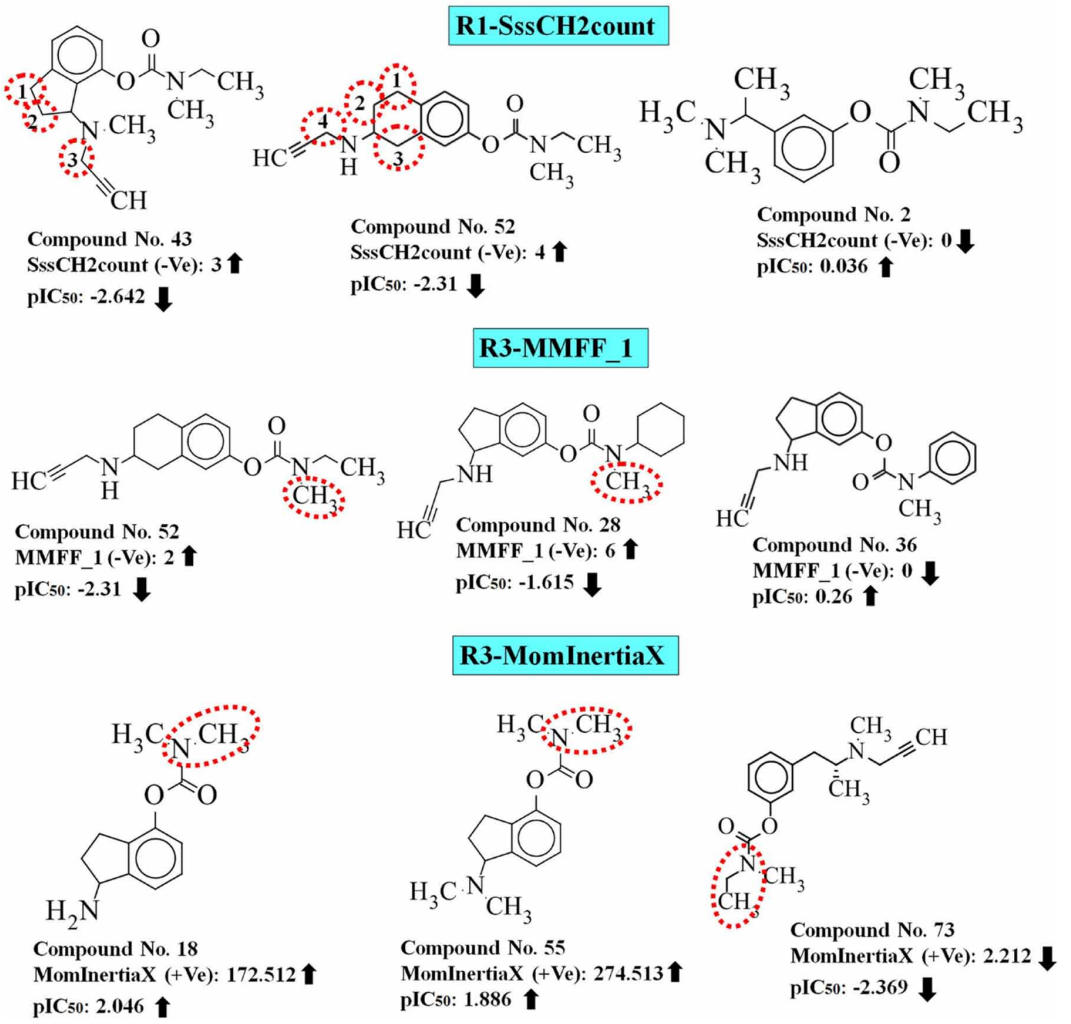


Figure 18. Applicability domain DModX values of the test set compounds at 99% confidence level of the developed PLS (2D QSAR and GQSAR) model against AChE enzyme (A: 2D QSAR and B: GQSAR)

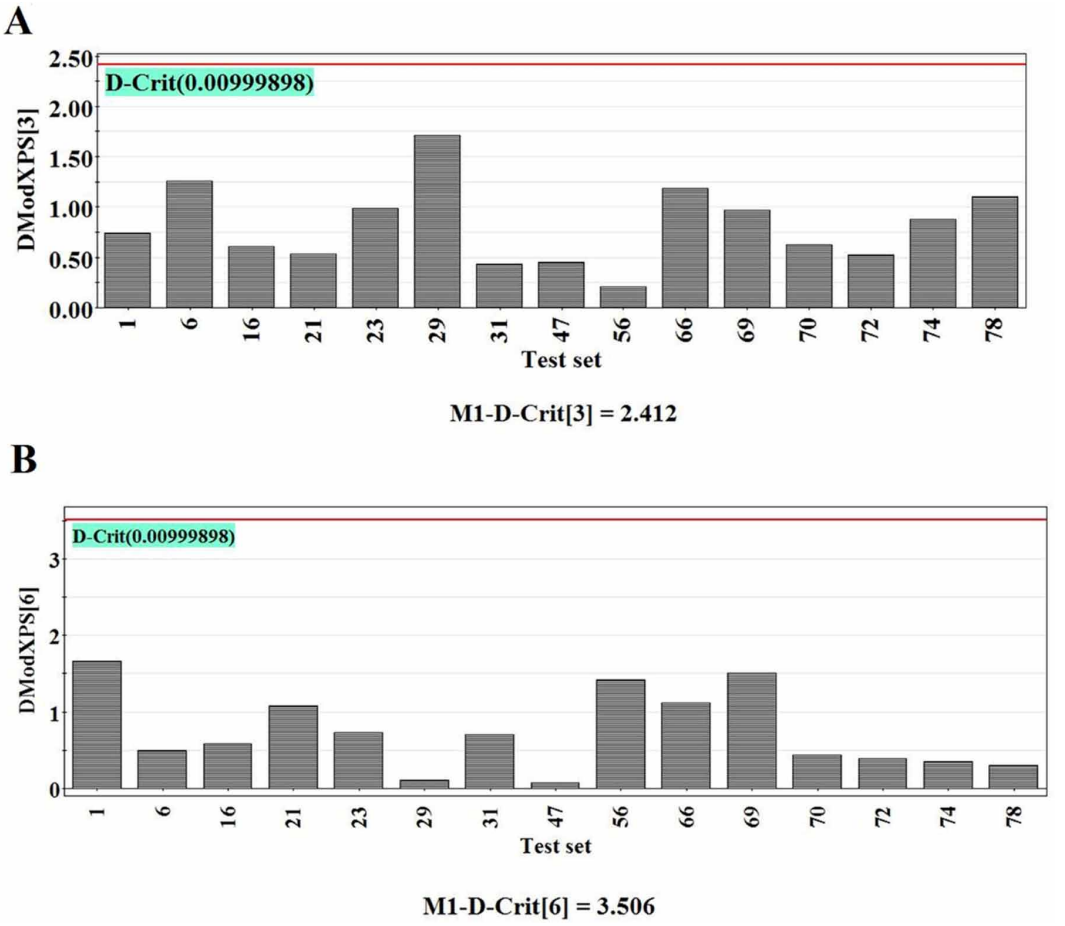


Figure 19. The original and randomized total cost values of the hypotheses for F-test

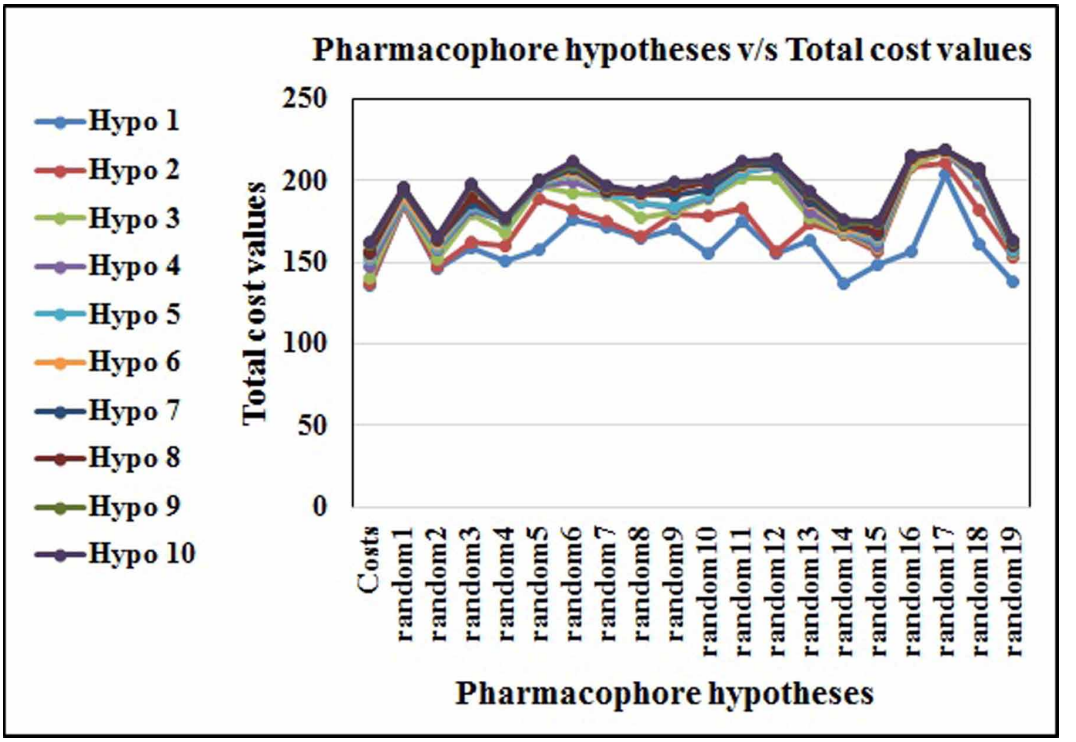


Figure 20. The original and randomized correlation values of the hypotheses for F-test

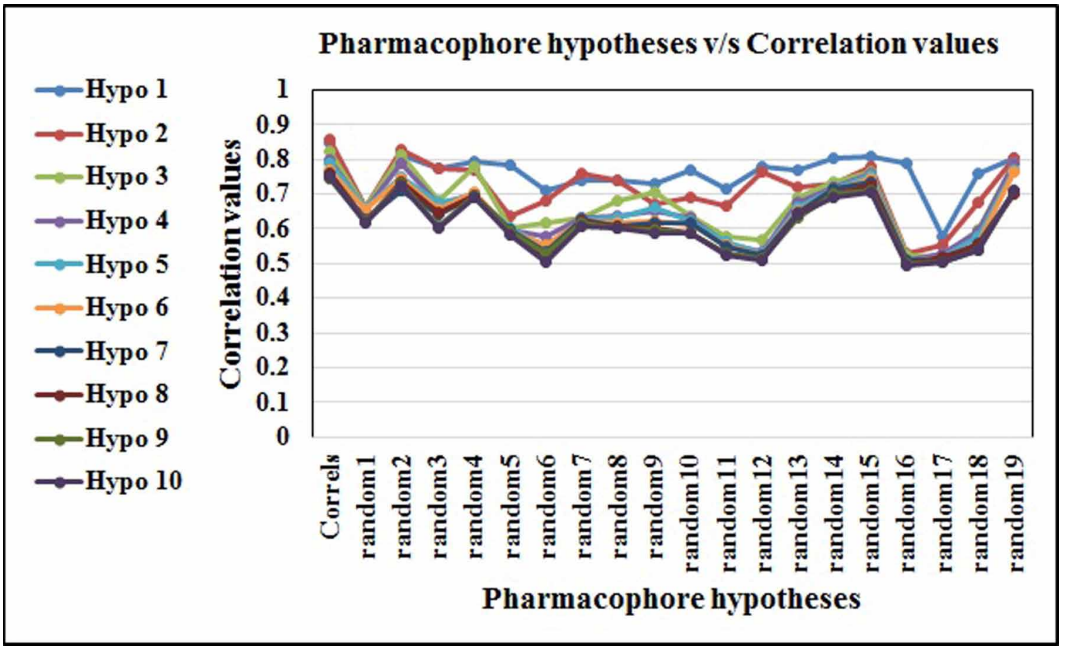


Figure 21. Docking interactions of most active compound of dataset (compound 18)

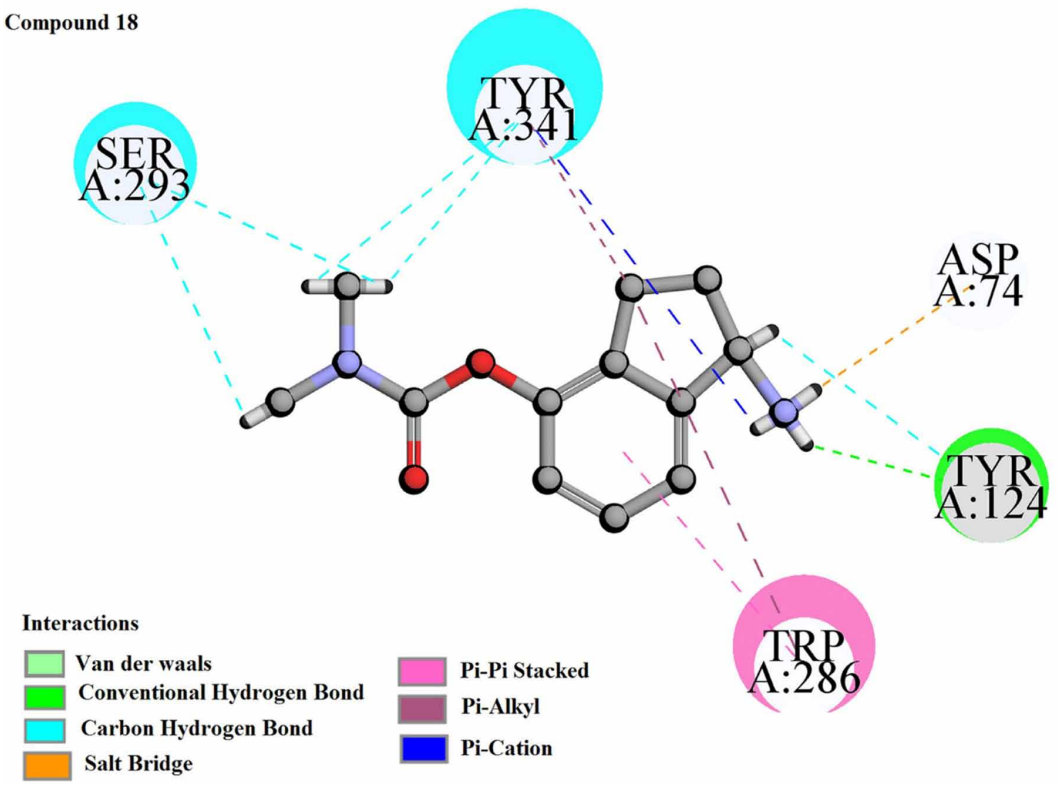


Figure 22. Docking interactions of most active compound of dataset (compound 44)

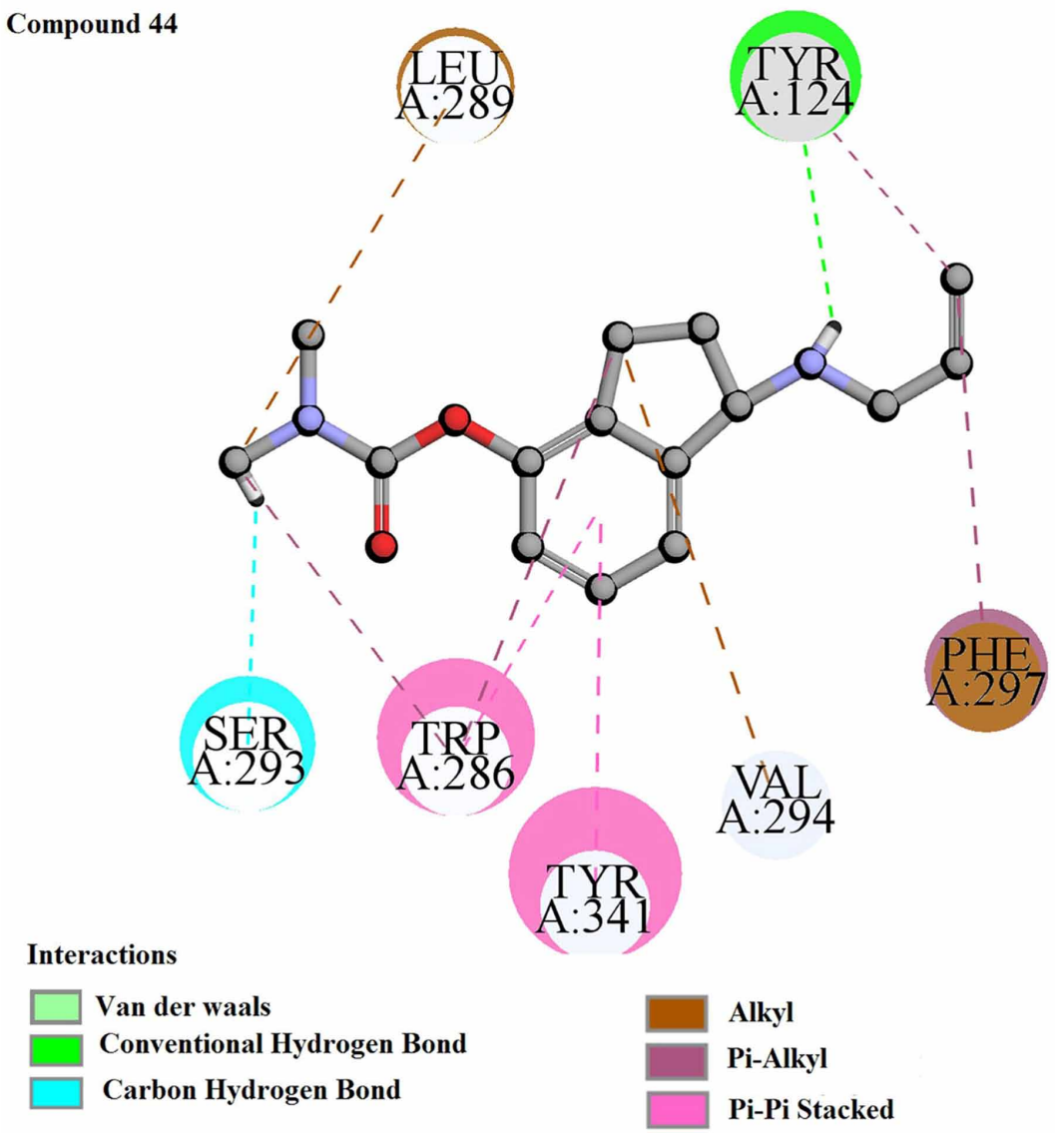


Figure 23. Docking interactions of most active compound of dataset (compound 55)

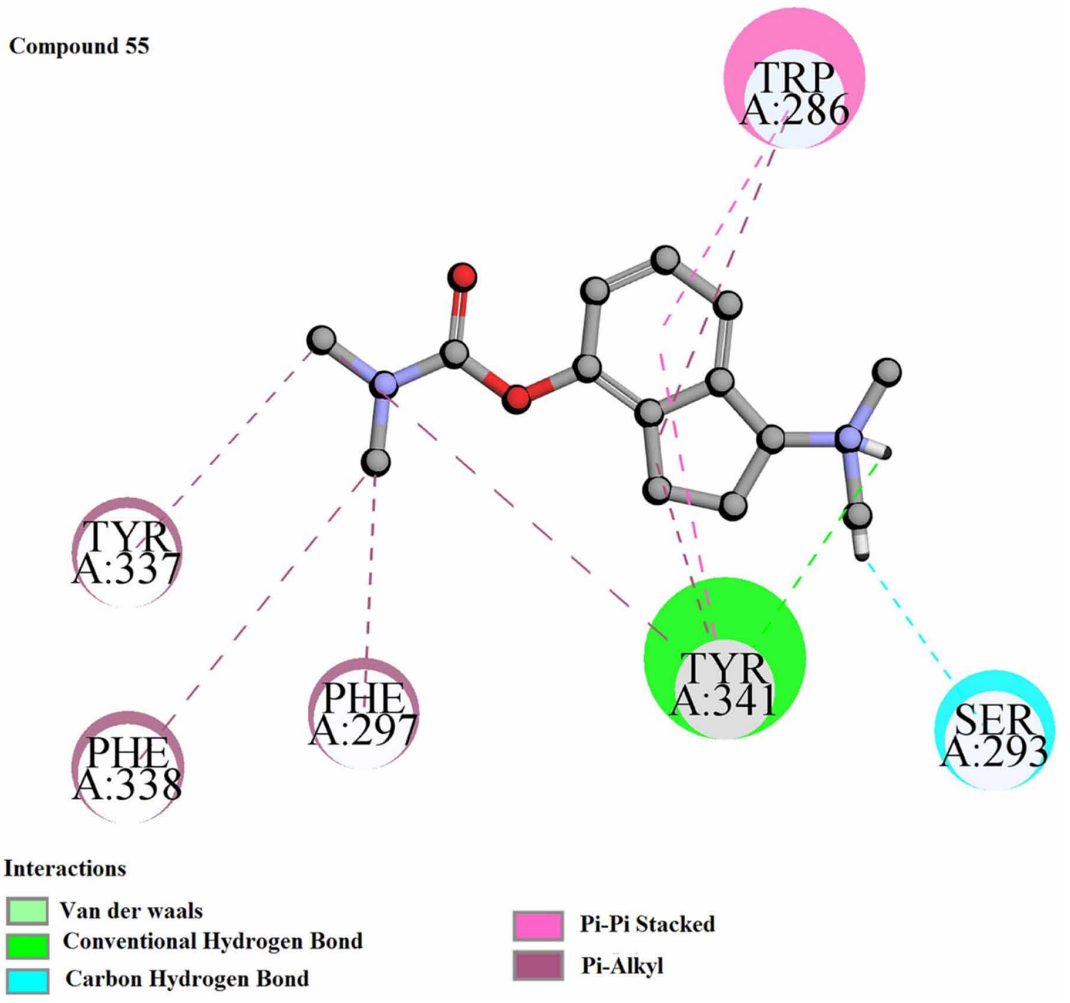


Figure 24. Docking interactions of most active compound of dataset (compound 61)

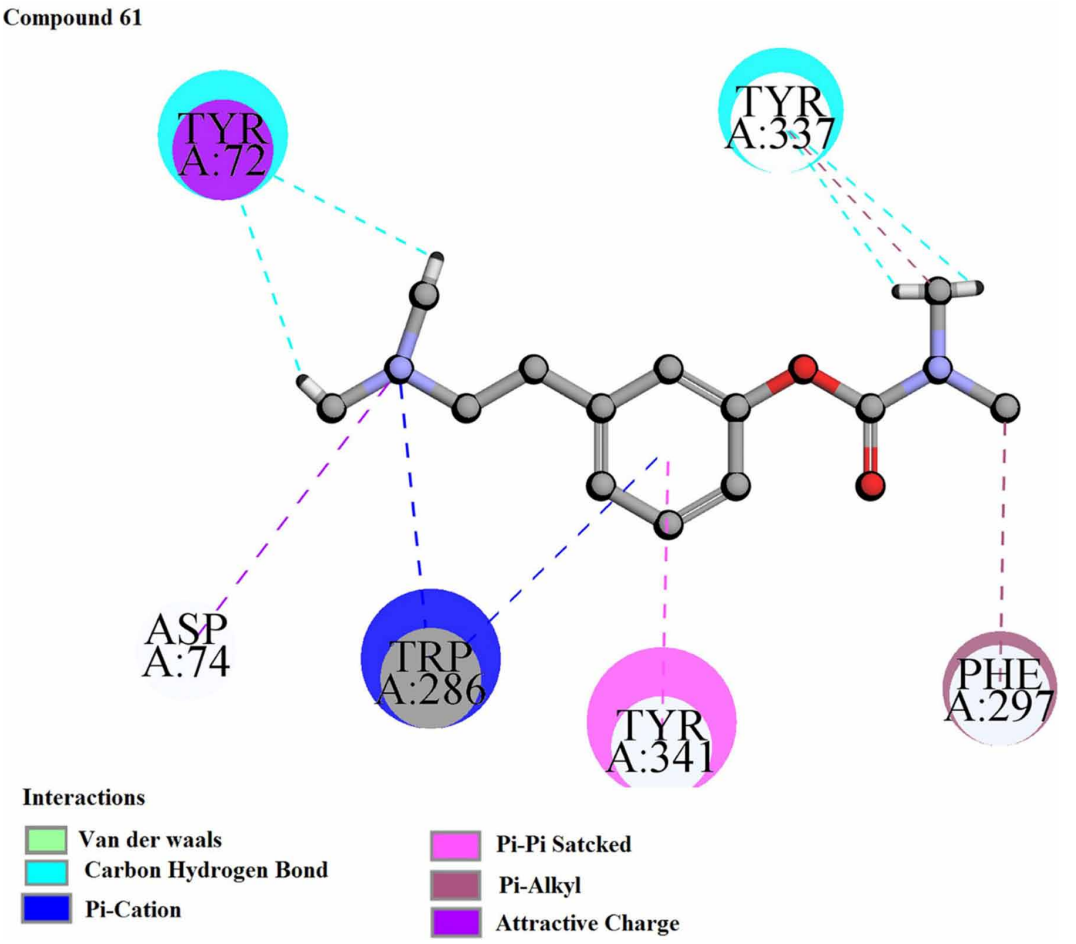


Figure 25. Docking interactions of least active compound of dataset (compound 33)

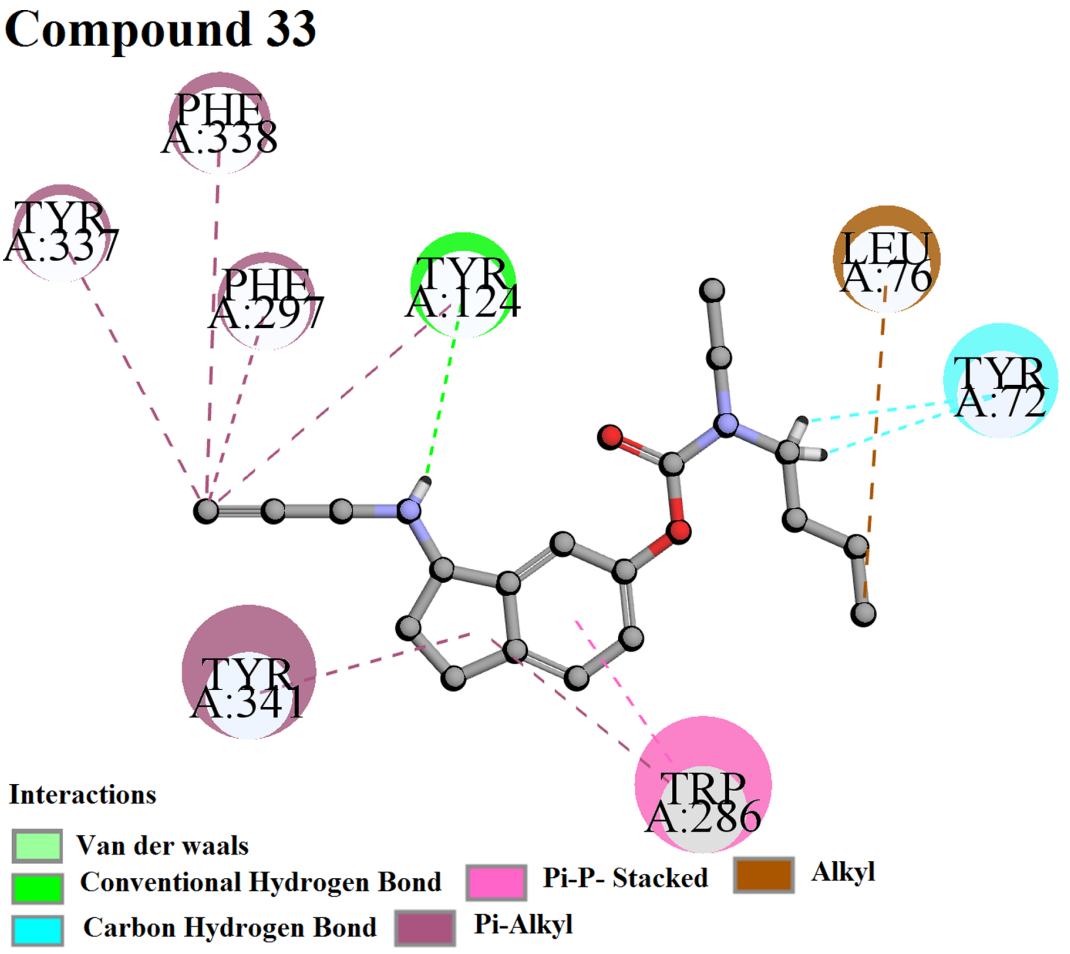


Figure 26. Docking interactions of least active compound of dataset (compound 40)

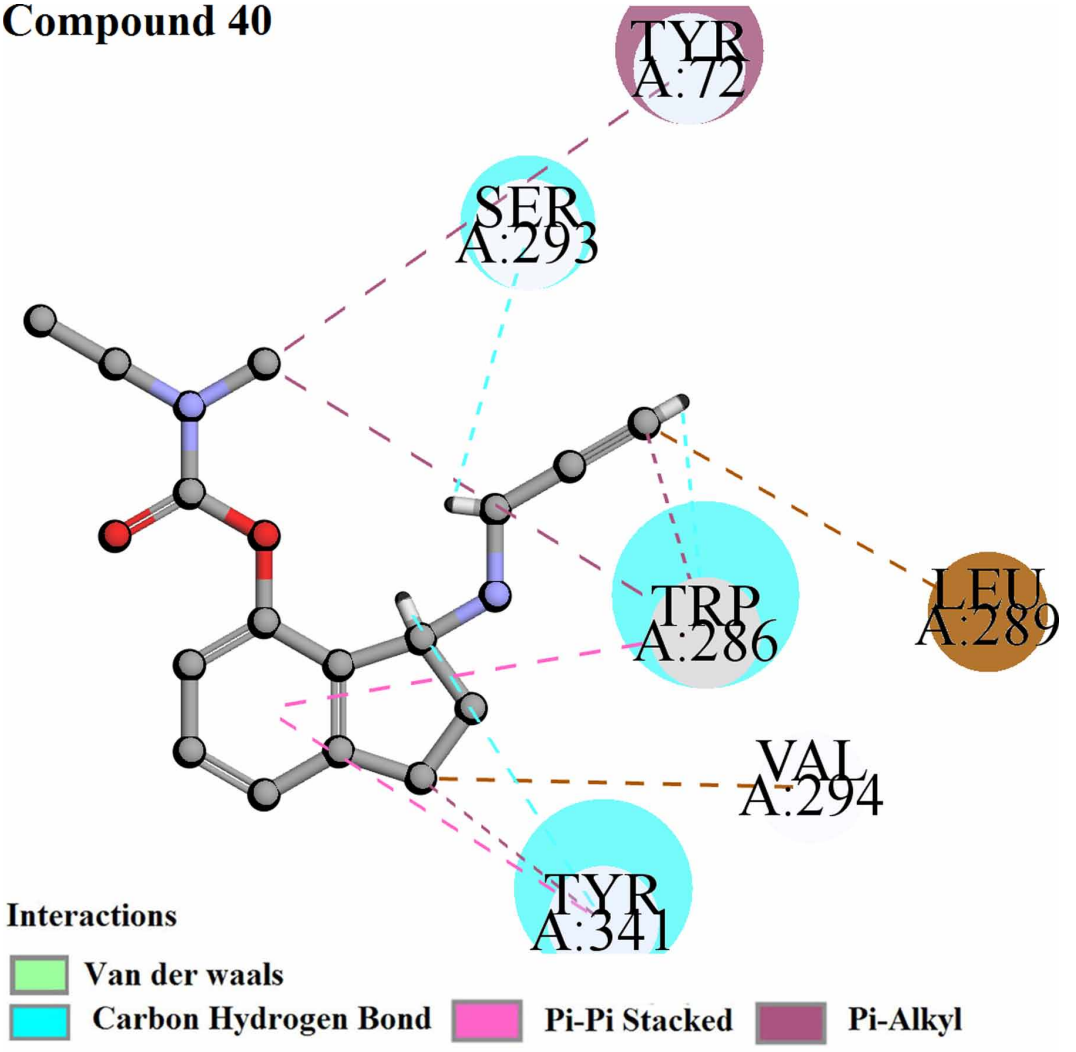


Figure 27. Docking interactions of least active compound of dataset (compound 52)

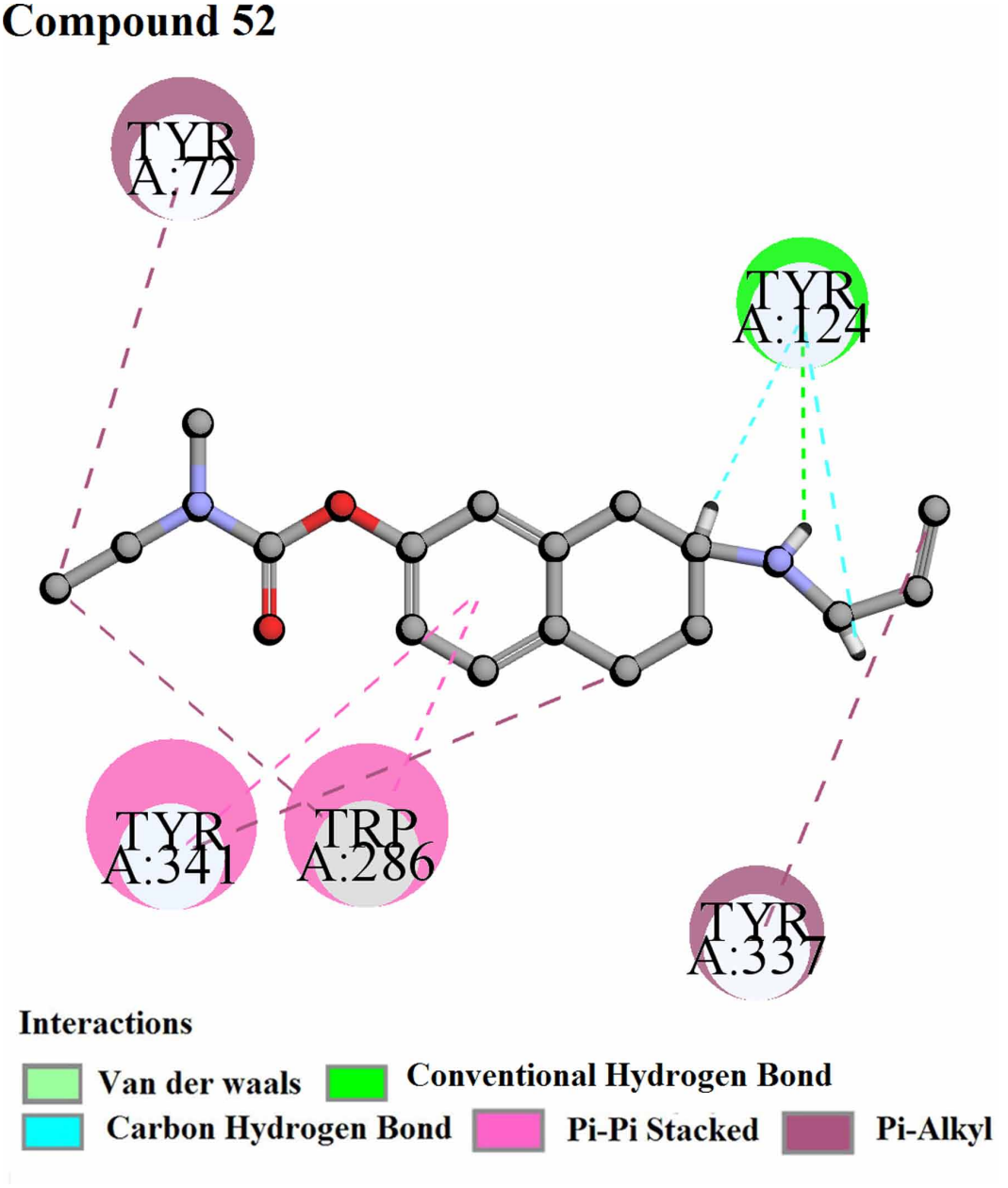
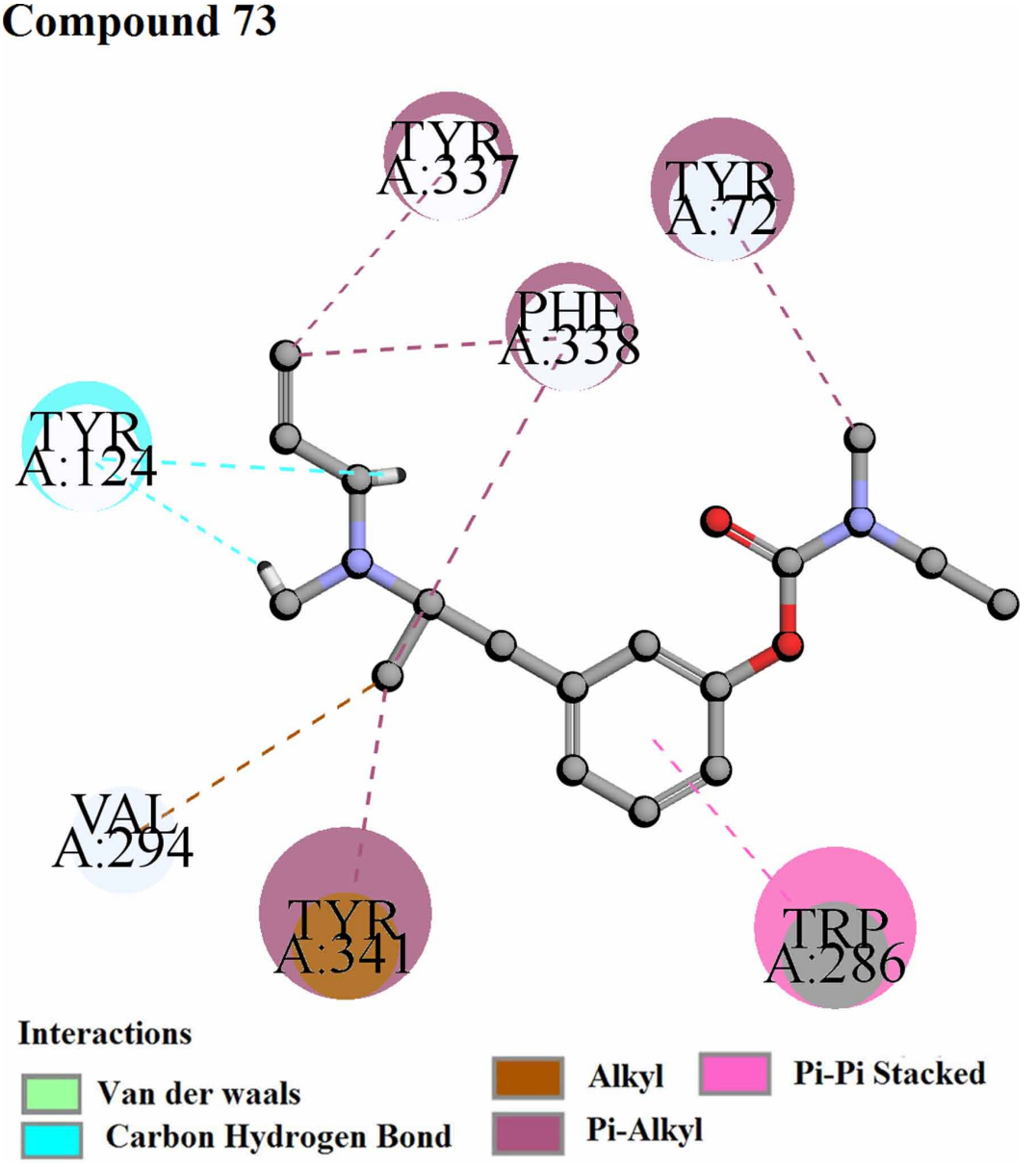


Figure 28. Docking interactions of least active compound of dataset (compound 73)



Vinay Kumar is working on the "Molecular Modelling of Potential anti-Alzheimer Agents Using Chemoinformatics Tools" as a registered Ph.D. scholar in the Department of Pharmaceutical Technology, Jadavpur University, Kolkata (India). He completed his B. Pharmacy (2015) degree from Goel group of institutions, Lucknow, Uttar Pradesh (UP) and M.S. in Pharmacy (2017) degree from National Institute of Pharmaceutical Education and Research (NIPER), Kolkata (WB). Mr. Vinay's research interests include QSAR, Chemoinformatics and molecular modeling with special reference to development of inhibitors for the treatment of Alzheimer disease.

Achintya Saha is a faculty member of Dept. of Chemical Technology, University of Calcutta. He has published more than 135 papers in standard journals.

Computational Biology and Chemistry 88 (2020) 107355

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac

Research Article

In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease

Vinay Kumar[a], Achintya Saha[b], Kunal Roy[a,*,1]

[a] Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India
[b] Department of Chemical Technology, University of Calcutta, 92 A P C Road, Kolkata 700 009, India

ARTICLE INFO

Keywords:
2D-QSAR
PLS
AChE
BuChE
Selectivity
Docking

ABSTRACT

In this research, we have implemented two-dimensional quantitative structure-activity relationship (2D-QSAR) modeling using two different datasets, namely, acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzyme inhibitors. A third dataset has been derived based on their selectivity and used for the development of partial least squares (PLS) based regression models. The developed models were extensively validated using various internal and external validation parameters. The features appearing in the model against AChE enzyme suggest that a small ring size, higher number of −CH2- groups, higher number of secondary aromatic amines and higher number of aromatic ketone groups may contribute to the inhibitory activity. The features obtained from the model against BuChE enzyme suggest that the sum of topological distances between two nitrogen atoms, higher number of fragments X-C(=X)-X, higher number of secondary aromatic amides, fragment R–CR-X may be more favorable for inhibition. The features obtained from selectivity based model suggest that the number of aromatic ethers, unsaturation content relative to the molecular size and molecular shape may be more specific for the inhibition of the AChE enzyme in comparison to the BuChE enzyme. Moreover, we have implemented the molecular docking studies using the most and least active molecules from the datasets in order to identify the binding pattern between ligand and target enzyme. The obtained information is then correlated with the essential structural features associated with the 2D-QSAR models.

1. Introduction

Alzheimer's disease (AD) is a neuropathological disorder, found in the most common type of dementia, which mostly affects elderly persons (Orhan et al., 2004). The word 'common type of dementia' describes a set of symptoms that may comprise memory loss, difficulties with thinking, problem-solving or language. These changes are often small to start with, but for someone with dementia, they become severe enough to affect the daily life. A person with dementia may also experience changes in their mood or behavior (Schelterns and Feldman, 2003). AD is characterized by the multiple cortical disturbances, such loss of memory, learning skill, unable to perform daily life activity, intellectual functions, and repeatedly accompanied by abnormal behavior, such as aggression and depression (Orhan et al., 2004; Kumar, 2015). There are several hypotheses reported to explain the AD, but none of them is able to provide the exact cause of AD (Du et al., 2018). Among them, the cholinergic hypothesis (covering acetylcholinesterase or AChE and butyrylcholinesterase or BuChE enzymes) is the most prominent hypothesis for the treatment of AD (Liston et al., 2004). Acetylcholine is an important neurotransmitter for healthy brain, playing an important role to perform the cognitive functions (Liston et al., 2004; Nordberg and Svensson, 1998; Ganeshpurkar et al., 2019; Shi et al., 2017). BuChE is a sister enzyme of AChE (Nordberg and Svensson, 1998; Awasthi et al., 2018). In AD, the activity level of AChE enzyme declines and the level of BuChE activity increases and the normal ratio of BuChE and AChE in the brain can alter from 0.6–11 (Kumar, 2015). Based on these facts, the dual inhibition strategy for these enzymes has been proposed to increase the effectiveness of the treatment strategy and expand the indications (Orhan et al., 2004; Kumar, 2015). Therefore,

Abbreviations: AD, Alzheimer's disease; AChE, Acetylcholinesterase; BuChE, Butyrylcholinesterase; 2D-QSAR, Two-dimensional quantitative structure-activity relationship; CADD, Computer-Aided Drug Design; QSAR, Quantitative structure-activity relationship; OECD, Organization for Economic Co-operation and Development; GA, Genetic algorithm; BSS, Best subset selection; PLS, Partial least squares; DCV, Double cross-validation; VIP, Variable importance plot; ETA, Extended topochemical atom; DModX, Distance to model in X-space; HQSAR, Hologram QSAR; 3D-QSAR, Three dimensional quantitative structure activity relationship; GA-MLR, Genetic algorithm multiple linear regression; ICMR, Indian Council of Medical Research

* Corresponding author.
E-mail address: kunal.roy@jadavpuruniversity.in (K. Roy).
[1] URL: http://sites.google.com/site/kunalroyindia/.

https://doi.org/10.1016/j.compbiolchem.2020.107355
Received 21 April 2020; Received in revised form 29 July 2020; Accepted 30 July 2020
Available online 05 August 2020
1476-9271/

AChE and BuChE inhibitors have become significant tools in the treatment of AD. Though the effects of the enzymes appear mainly to be symptomatic, but have possible neuroprotective effects. Currently four cholinesterase inhibitors, like tacrine, donepezil, galantamine and rivastigmine (FDA approved) are used for the symptomatic treatment of AD (Liston et al., 2004). In the present study, we have examined the essential features, which are responsible for dual inhibition of these enzymes.

The development of inhibitors against AD is a challenging and difficult procedure due to the complication of the molecular pathways involved in the progression of the disease (Mouchlis et al., 2020). Computer-Aided Drug Design (CADD) uses computer power, three-dimensional graphics, mathematics, and statistics to understand and predict the binding mode and energy of small molecule inhibitors with potential targets (Mouchlis et al., 2020). The most common in-silico techniques employed by medicinal chemists to help them rationalize the selection of hit compounds and to perform hit-to-lead optimization include structure-based design like molecular docking and dynamics and ligand-based design like quantitative structure-activity relationships (QSAR), and pharmacophore mapping (Mouchlis et al., 2020). The main molecular targets employed in the development of drugs against AD include β-secretase, γ-secretase, acetylcholinesterase, glycogen synthase kinase, muscarinic acetylcholine receptor and Tau-protein (Mouchlis et al., 2020). Among different in silico methodologies, the quantitative structure-activity relationship (QSAR) and molecular docking have their great applications in the area of in-silico search. The QSAR methods are essential for the exploration of important structural features and prediction of the biological activity of novel compounds based on mathematical and statistical relations (Ekins et al., 2007). The idea of QSAR is based on the concept that end point values of compounds change systematically with modification of the structural attributes (Ekins et al., 2007). There have been a large number of computational studies performed (Shrivastava et al. 2019 (Shrivastava et al., 2019), Bukhari et al. 2014 (Bukhari et al., 2014), De Souza et al. 2012 (De Souza et al., 2012), Pang et al. 2017 et al. (Pang et al., 2017), F. Türkan 2019 (Türkan, 2019), K Buldurun et al. 2019 (Buldurun et al., 2020) and Taslimi P. et al. 2020 (Taslimi et al., 2020)) so far for designing new inhibitors against AD, but still we are far from finding a precise treatment strategy for AD. In the current research, we have developed the PLS-regression based 2D-QSAR models using two different datasets namely, AChE and BuChE enzymes' inhibitors. We have also developed another PLS-regression based QSAR model for specificity against both these enzymes (198 compounds containing inhibitory activity against both these enzymes) to identify which structural fragments/properties are specifically essential to inhibit a specific type of enzyme. Multilayered variable section strategy has been applied for the selection of important descriptors for the bioactivity before the development of final models. The validated models showed acceptable values for various internal and external validation metrics. The models were also developed based on the Organization for Economic Co-operation and Development (OECD) guidelines (https://www.oecd.org/env/ehs/risk-assessment/validationofqsarmodels.htm). The obtained models highlight the structural requirements or molecular properties necessary to design safer dual inhibitors. The study has identified the structural attributes in small molecules that confer human BuChE and AChE enzymes inhibitory activities and their correlation to ligand-receptor interactions for further development of new generation drugs for AD. Additionally, we have also implemented the molecular docking studies using the most and least active compounds from the datasets and tried to rationalize the influences of different descriptors/features as apparent from the 2D-QSAR models.

2. Materials and methods

2.1. 2D-QSAR analysis

2.1.1. Dataset selection

In the current work, we have used two different datasets against two important targets, namely AChE (number of compounds = 997) and BuChE (number of compounds = 761) enzymes (see supplementary information S1 Sheets 1 and 2) collected from previously published literatures (Jaén et al., 1996; Högenauer et al., 2001; Andreani et al., 2008; Carlier et al., 1999a; Ceschi et al., 2016; Contreras et al., 1999; da Costa et al., 2013; DeBernardis et al., 1988; Anand and Singh, 2012a; Barreiro et al., 2003; Cardoso et al., 2004; Li et al., 2016; Erlanson et al., 2004; Feng et al., 2005a; Fink et al., 1995; Girisha et al., 2009; Gray et al., 1985; Han et al., 1991; He et al., 2007; Hu et al., 2002, 2013; Huang et al., 2011; Ishihara et al., 1994; Jia et al., 2009; Kapples et al., 1993; Kavitha et al., 2007; Li et al., 2014, 2017a; Liu et al., 2014; McKenna et al., 1997; Morini et al., 2008; Pool et al., 1996; Anand and Singh, 2012b; Sadashiva et al., 2006; Sang et al., 2015a, b; Sang et al., 2015c; dos Santos et al., 2010; Shao et al., 2004; Shen et al., 2008; Sheng et al., 2005, 2009a; Sheng et al., 2009b; Shi et al., 2013; Shinada et al., 2012; Shutske et al., 1989; Simoni et al., 2012; Sugimoto et al., 1992, 1995; Valenti et al., 1997; Vidaluc et al., 1994, 1995; Wong et al., 2003; Yang et al., 2017; Zeng et al., 1999; Zhan et al., 2010; Zheng et al., 2010; Zhu et al., 2009; Szymański et al., 2011; Krátký et al., 2015, 2017; Rodríguez-Franco et al., 2006; Conejo-García et al., 2011; Rodríguez et al., 2016; Yurttaş et al., 2013; Zelík et al., 2010; Ahmad and Fatima, 2008; Ahmad et al., 2016; Ahmed et al., 2006; Bacalhau et al., 2016; Bagheri et al., 2015; Bolognesi et al., 2005; Camps et al., 2008; Carlier et al., 1999b; Cho et al., 2017; Czarnecka et al., 2017; Decker, 2006; Decker et al., 2008; Fang et al., 2008a, b; Feng et al., 2005b; Lin et al., 1998; Gregor et al., 1992; Hameed et al., 2015, 2016; Hasan et al., 2005; Huang et al., 2010; Jiang et al., 2011; Kanhed et al., 2015; Kurt et al., 2015; Leader et al., 2002; Leng et al., 2016; Li et al., 2013, 2017b; Luo et al., 2011; Mohamed et al., 2011; Mohammadi et al., 2015; Najafi et al., 2016; Pouramiri et al., 2017; Rydberg et al., 2006; Saeed et al., 2015; Saeedi et al., 2017; Samadi et al., 2010, 2012; Sarfraz et al., 2017; Shi et al., 2011; Skrzypek et al., 2013; Sterling et al., 2002; Tang et al., 2007; Villalobos et al., 1994; Wen et al., 2007; Yanovsky et al., 2012; Zakhari et al., 2011). In both the datasets, some compounds have inhibitory activities both against AChE and BuChE enzymes (number of compounds 198) (Högenauer et al., 2001; Andreani et al., 2008; Carlier et al., 1999a; Anand and Singh, 2012a; Barreiro et al., 2003; Feng et al., 2005a; Fink et al., 1995; He et al., 2007; Jia et al., 2009; Li et al., 2014, 2017a; Sang et al., 2015c; Shao et al., 2004; Shen et al., 2008; Sheng et al., 2005, 2009a; Sheng et al., 2009b; Shinada et al., 2012; Sugimoto et al., 1992, 1995; Valenti et al., 1997; Yang et al., 2017; Zhan et al., 2010; Zheng et al., 2010; Camps et al., 2008; Carlier et al., 1999b; Feng et al., 2005b; Li et al., 2017b; Rydberg et al., 2006), Therefore, we have used the difference of activities against AChE and BuChE enzymes as the dependent variable (AChE-BuChE) for modeling selectivity (the third dataset) (See supplementary information S1 Sheet 3). The datasets comprise diverse classes of heterocyclic compounds, and the experimental activity of each compound is expressed in IC_{50} (nM) values. Although the original authors adopted different bioassay protocols in different publications to check the inhibitory activity on different species (rat brain, *Electrophorous electricus*, human or horse plasma) for the comparison of inhibition rate, in the present work, we have selected only the compounds which were examined for the inhibitory activity using modified colorimetric Ellman assay on rat brain showing good inhibitory activity (modified colorimetric Ellman assay (Komersova et al., 2007)). For the purpose of the models development, we have converted the IC_{50} values to pIC_{50} (pIC_{50} = -logIC_{50}) values as customary in QSAR analysis, and all the compounds from each dataset set were drawn using MarvienSketch (ChemAxon - Software Solutions and, 2020), followed by cleaning of molecules. Then, hydrogens were added and the file was saved as sdf format. Prior to descriptor calculation, we have carefully checked all structures in the datasets for development of significant 2D-QSAR models.

2.1.2. Descriptor calculation and pretreatment

In this section, we have calculated only 2D descriptors using two software, namely Dragon 7 (Mauri et al., 2006) (covering functional

group counts, constitutional, ring descriptors, connectivity index, atom centered fragments, 2D atom pairs, atom type E-states and molecular properties) and PaDEL-Descriptor 2.20 (Yap, 2011) (for extended topochemical atom indices). After descriptor calculation, we have performed data curation utilizing the tool Pretreatment V-WSP version 1.2 (available at http://dtclab.webs.com/software-tools) in order to eliminate the descriptors with missing or near constant values.

2.1.3. Dataset division

In this method, we have split the whole dataset in to training and test sets based on the sorted activity based algorithm using the "Dataset Division GUI" developed by our group (available at http://dtclab.webs.com/software-tools). Initially the dataset was divided into training and test sets in a random manner (andom, 2020) for 30 trials. Then the whole range of activities was sorted through ascending order, and every fourth compound was assigned to the test set. Finally an attempt was made to rationalize the division process, in which the division was performed so that points representing both training and test sets were distributed within the whole descriptor space occupied by the entire dataset, and each point of the test set was close to at least one point of the training set (Leonard and Roy, 2000). The training set was used for the development of models, and the test set compounds for the validation of the developed models.

2.1.4. Multi-layered variable selection strategy and 2D-QSAR model development

Selection of important descriptors from the large pool of initial descriptors using different variable selection methodologies is an important task in QSAR modeling. Based on this concept, we have adopted a multi-layered variable selection strategy before the development of the final models. In the multi-layered variable selection, first we have applied stepwise regression in successive iterations using the Minitab software (Minitab IN, 2000) followed by genetic algorithm (GA) using the GA software (available at http://dtclab.webs.com/software-tools), and afterwards we applied double cross-validation (DCV) for the AChE and BuChE enzyme inhibition models. On the other hand, for the selectivity based models, we have applied Best Subset Selection (BSS), while partial least squares (PLS) regression technique was used for the development of final models in all cases. Additionally, we have applied additional selection strategies for the purpose of checking the statistical quality of the developed models. Thus, to address the above said situation, we have performed (1) PLS without stepwise regression (SR) + GA, (2) PLS + SR, (3) PLS + GA, (4) PLS + SR + GA to find out the optimal combinations of predictors (see supplementary information S3 (in case of model against the AChE enzyme), S4 (in case of model against the BuChE enzyme) and S5 (in case of selectivity based model)). From the developed models, we have found that models reported in this study are more robust than models obtained from the above strategies. The details of the steps used in the multi-layered variable selection strategy are schematically represented in **Figure S1** (see supplementary materials S6).

2.1.5. Statistical validation of the generated2D-QSAR models

In present work, we have applied various statistical methodologies like internal (determination coefficient (r^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), Avg $r^2_{m(LOO)}$ and Δrm^2) and external (Q^2F_1, Q^2_{F2}, r_m^2 parameters like average r_m^2 (test) and Δr_m^2 and concordance correlation coefficient (CCC)) validation methods in order to assure the significant level of the generated models (Tropsha, 2010; Veerasamy et al., 2011). Additionally, we have also implemented the Y-randomization test (Kumar et al., 2020a), checked applicability domain criteria etc using Simca-P 10.0 software (SIMCA, 2020). The details of various statistical validation parameters are discussed in the Box S1 (see supplementary materials S4). The detailed methodologies are depicted in **Figure S2** (see supplementary materials S6).

2.2. Molecular docking studies

In this analysis, we have applied the molecular docking studies in order to investigate the binding pattern of molecules (most and least actives from the dataset) with the respective enzymes, such as AChE and BuChE. The enzyme crystal structures were extracted from the protein databank with the PDB id: 4M0E (structure of human acetylcholinesterase bound to Dihydrotanshinone I) and 4BDS (crystal structure of human butyrylcholinesterase in Complex with Tacrine) (Murzin et al., 1995) (AChE and BuChE enzymes, respectively). The molecular docking study was performed by using BIOVIA discovery studio client 4.1 (Discovery Studio Predictive Science Application, 2020) platform with the CDOCKER module of receptor-ligand interaction as discussed by Pal S et al, Singh N et al, Ercan S et al. and Kumar V et al. (Pal et al., 2019; Singh et al., 2020; Ercan and Şenses, 2020; Kumar et al., 2020b). Prior to the docking analysis, we have prepared the target enzyme and selected inhibitors using the protein and ligand preparation protocol available in BIOVIA discovery studio client 4.1 (Discovery Studio Predictive Science Application, 2020). The active site in the enzyme was defined by the "define and edit binding site" protocol available in BIOVIA discovery studio client 4.1 (Discovery Studio Predictive Science Application, 2020). After docking analysis, we have sorted the generated poses as per the CDOCKER interaction energy, and the poses with top scoring values were used for further analysis. The obtained poses were validated using the bound ligand present in the crystal structure of the enzyme. On the basis of number of interactions and the active residues interacting with the bound ligand, we have selected the final pose for the further study. From the ligplot (**Figure S3** See supplementary information S6), we can see the number of interactions and active residues responsible for the significant interactions in crystal structure of AChE and BuChE enzyme with their bound ligand.

Box 1
2D-QSAR model and statistical validation parameters obtained from the developed model against AChE enzyme.

pIC_{50}-

$$= -2.806 + 0.148 \times SssCH2 - 0.905 \times totalcharge - 0.027 \times F09[C-C] + 0.504 \times nArNHR + 0.004 \times D/Dtr05 + 0.588 \times B08[N-N] + 6.32076 \times ETA\_Shape\_P - 0.007 \times D/Dtr12 + 0.009 \times D/Dtr08 + 1.125 \times Ui - 0.284 \times B06[C-N] - 0.337 \times B08[C-N] + 0.361 \times F10[N-O] + 0.669 \times nArCO - 0.209 \times F04[O-O] \quad (1)$$

Internal Validation Parameters: n_{train}= 798, r^2= 0.662, Q^2= 0.645, EL = 15, LV = 6, Prediction quality = Good.
External Validation Parameters: n_{test}= 199, Q^2F_1= 661, Q^2F_2= 660, Prediction quality = Good.

Box 2
2D QSAR model and statistical validation parameters obtained from the developed model against BuChE enzyme.

$$pIC_{50}\text{-} = -4.13393 + 0.617 \times C-028 - 0.059 \times F05[C-O] + 0.018 \times T(N..N) - 0.219 \times B06[O-O] - 0.931 \times B03[N-N] + 0.041 \times F09[C-O] - 2.192 \times nThiazoles + 0.276 \times B09[C-O] + 0.727 \times C-041 - 0.981 \times C-019 + 0.374 \times nArCONHR + 0.194 \times B07[N-N] - 0.097 \times H-053 \quad (2)$$

Internal Validation Parameters: n_{train}= 603, r^2= 0.674, Q^2= 0.656, Average rm^2= 0.54, EL = 13, LV = 5.
External Validation Parameters: n_{test}= 158, Q^2F_1= 0.663, Q^2F_2= 0.660, Average rm^2= 0.499,

3. Results and discussion

In this research, we have generated the PLS regression based 2D-QSAR models for AChE and BuChE enzyme inhibitory activities. In the datasets, some compounds have both AChE and BuChE inhibitory values. A PLS regression-based 2D-QSAR model (equation 3) has been developed based on their selectivity (AChEI-BuChEI). The details of the different validation metrics values for the models are summarized in equations 1, 2 and 3 (Boxs , and) and also in Table S1 (See S7 in supplementary materials). The statistical results obtained from the models suggested that the developed models are acceptable in terms of stability, predictive and fitness criteria. The nearness of the observed and predicted values for the AChE and BuChE enzyme inhibitors and also the selectivity values can be further established from the scatter plots as shown in Fig. 1 and 2. The quantitative contributions of similar/dissimilar descriptors and the interrelationships between the X-variables and the Y-response are depicted in the loading plots of **Figure S4** (See S6 in supplementary materials). Moreover, we have also employed Y-Randomization test in order to cross verify whether the models were obtained by any chance or not using Simca-P 10.0 software (SIMCA, 2020). The results obtained from the randomized models, in case of AChE inhibitors (Model 1: R^2int (intercept values) = -0.0011 and Q^2int = -0.157), whereas in case of BuChE inhibitors (Model 2: R^2int = 0.0017 and Q^2int = -0.184) and in case of selectivity based model (Model 3: R^2int = -0.0013 and Q^2int = -0.26), suggested that the developed models were not obtained by any chance correlation as depicted in **Figure S5** (See S6 in supplementary materials).

3.1. Mechanistic interpretation of modeled descriptors

3.1.1. 2D-QSAR analysis against AChE enzyme: Model 1

The descriptors appearing in the model are ranked accordingly to their significance, and then described individually. The importance and contribution of the obtained descriptors in the models towards the AChE inhibitory activity are identified based on variable importance plot (VIP) and regression coefficient plot as shown in Fig. 3 (Kumar et al., 2020c). The importance of each descriptor against AChE enzyme has been analyzed with their appropriate examples.

Box 3
2D QSAR model and statistical validation parameters obtained from the developed model based on selectivity **(AChEI-BuChEI)**.

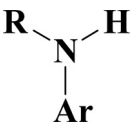
$$\text{Selectivity } [pIC_{50}]\text{-} = -0.891 - 0.720 \times F05[N-O] + 79.839 \times ETA\_Shape\_X - 2.607 \times nCq + 0.505 \times nArOR + 5.599 \times ETA\_dBetaP + 2.962 \times B10[C-C] \quad (3)$$

Internal Validation Parameters: n_{train}= 159, r^2= 0.679, Q^2= 0.650, Average rm^2= 0.524, EL = 6, LV = 5.
External Validation Parameters: n_{test}= 39, Q^2F_1= 0.787, Q^2F_2= 0.785, Average rm^2= 0.686, Δrm^2= 0.176,

According to the VIP plot, SssCH2 is the most important descriptor contributing positively to the response. It is an atom-type E-state index encoding information about sum of ssCH2 count in the compounds. Sharma et al. (Sharma et al., 2020) suggested that the descriptor represents the electro-topological state for the number of $-CH_2$- groups. The positive contribution suggests that the AChE enzyme inhibitory activity may be increased by an increase in the number of such $-CH_2$- groups in the molecules. A higher number of −CH2– groups (corresponding to the descriptor values of SssCH2 not more than 40) in compounds leads to better inhibitory activity against the AChE enzyme as longer chain compounds with higher number −CH2 groups would be more lipophilic resulting in improved brain permeability. Compounds like **597** and **841** (Fig. 4) have higher number of $-CH_2$- groups in their structure, showing higher descriptor values (32.05 and 24.37), leading to their higher range of AChE enzyme inhibitory activity (-1.418 and -0.365 respectively). On the other hand, compounds 884 and **896** showed AChE enzyme inhibitory activity in very a lower range (-5.505 and -5.361 respectively), due to the absence of such group in the compounds.

The extended topochemical atom (ETA) descriptor, ETA_Shape_P, represents the effect of branching in the cationic structure (one central atom is attached to three other non-hydrogen atoms) on the inhibitory activity against AChE enzyme. This descriptor particularly denotes the branching where one central atom is attached to three other non-hydrogen atoms, making a Y-shaped structural fragment (Das et al., 2017a). The positive contribution of this descriptor indicates that the inhibitory activity of compounds is directly proportional to the numerical value of ETA_Shape_P. Therefore, the compounds with higher numerical value of this descriptor may enhance the AChE enzyme inhibitory activity as shown in (Fig. 4) compounds like **735** (pIC_{50}: -1.929) and **740** (pIC_{50}: -1.875) and their corresponding descriptor values are 0.367 and 0.314, respectively. In contrast, compounds like **379** (pIC_{50}: -5.414) and **723** (pIC_{50}: -5) have no such fragment show lower inhibitory activity (Fig. 4).

The functional group count descriptor, nArNHR, simply represents the higher number of secondary aromatic amines (corresponding to descriptor values not more than 4) in the compounds. The fragment contributes significantly to increase the intermolecular interactions by strong H-bonds (discussed later in the docking part). Here, we have represented the general structure of secondary aromatic amine (**1**) to allow the beginners for proper understanding of the important structural fragments responsible for the inhibition.



(1)

As per the regression coefficient plot (**See** Fig. 3), this descriptor contributed positively towards AChE enzyme inhibitory activity. Thus, the compounds with the higher number this fragment (secondary aromatic amines) may enhance the AChE inhibitory activity as shown in compounds **3** (pIC_{50}: 0.698) and **18** (pIC_{50}: 1.154) (containing

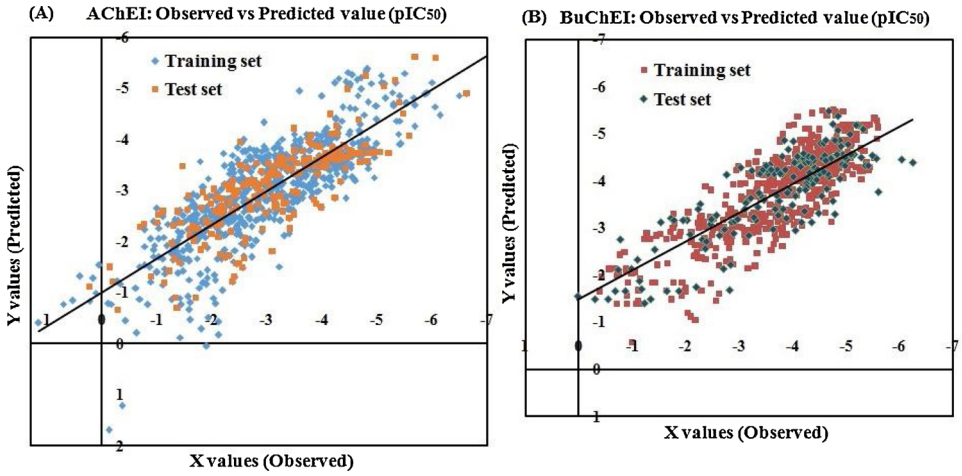


Fig. 1. Scatter plots of developed PLS models (observed and predicted values; the best fit lines are shown) against (A) AChE and (B) BuChE enzymes.

descriptor values 2) (Fig. 5). On the other hand, the compounds containing no such fragments have lower inhibitory activity as observed in compounds **675** (pIC_{50}: -7.683) and **681** (pIC_{50}: -7.152).

The descriptor Ui belongs to the class of molecular properties of an unsaturation index. The positive contribution (as per the regression coefficient plot) of this descriptor implies that it has a positive impact towards the AChE enzyme inhibitory activity. So, the compounds bearing higher unsaturation index may have enhanced inhibitory activity as presented in compounds (Fig. 5) **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698) (containing descriptor value 3.584), while the compounds with a lower unsaturation index may have lower AChE enzyme inhibitory activity as displayed in compounds (Fig. 5) **848** (pIC_{50}: -5.414).

The 2D atom pair descriptor, B08[N-N], signifies the presence/absence of N-N at the topological distance 8. As per the regression coefficient plot (**see** Fig. 3), this descriptor positively influences the inhibitory activity against AChE enzyme. So, the compounds containing higher number of N-N fragments at the topological distance 8 may have higher AChE enzyme inhibitory activity as shown by compounds **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698) (their corresponding descriptors values are 1) (Fig. 5). Again, the compounds with no such fragments shows lower inhibitory activity as shown in compounds **675** (pIC_{50}: -7.683), **681** (pIC_{50}: -7.152).

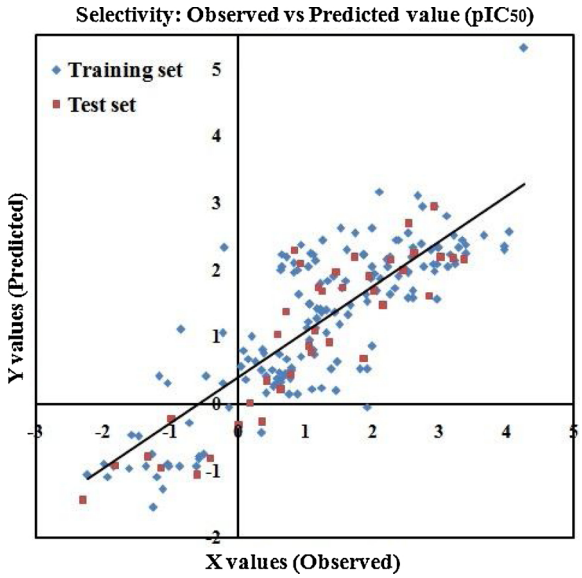


Fig. 2. Scatter plot of the model based on selectivity (AChE-BuChE enzyme) (observed and predicted values; the best fit line is shown).

2D atom pair descriptor, F09[C-C], represents the frequency of C - C at the topological distance 9. This descriptor negatively influences the inhibitory activity against the AChE enzyme as per the negative regression coefficient value. So, the higher number of this fragment correlates to lower AChE enzyme inhibitory activity as observed in compounds numbers **675** (pIC_{50}:-7.683), **681** (pIC_{50}: -7.152) (Fig. 6), while absence of this feature correlates to higher potency of AChE enzyme inhibitory activity as observed in compounds **747** (pIC_{50}: -1.785) and **740** (pIC_{50}: -1.875).

The descriptor, B06[C-N] belonging to the family of 2D atom pair descriptor defines the presence/absence of C-N at the topological distance 6. The negative contribution (as per the regression coefficient plot) of this descriptor suggests that the descriptor is inversely related to the AChE inhibitory activity. The same has been observed in compounds (Fig. 6) **675** (pIC_{50}: -7.683) and **681** (pIC_{50}: -7.152) (lower enzyme inhibitory activity as their corresponding numerical descriptor values are in higher range), whereas the inverse phenomena has been observed in compounds (Fig. 6) **747** (pIC_{50}: -1.785), **740** (pIC_{50}: -1.875) (increases in enzyme inhibitory activity as their corresponding numerical descriptor value is in the lower range).

Another 2D atom pair descriptor, B08[C-N], signifies the presence/absence of C-N at the topological distance 8. This descriptor negatively affects the activity of AChE enzyme inhibitors as indicated by its negative regression coefficient. So, the compounds with higher number of C-N fragments at the topological distance 8 may have lower AChE enzyme inhibitory activity as evidenced by compounds **675**(pIC_{50}: -7.683), **681** (pIC_{50}: -7.152) (their corresponding descriptors values are 1). Again, the compounds with no such fragments show improved inhibitory activity as supported by compounds **747** (pIC_{50}: -1.785), **740** (pIC_{50}: -1.875) (Fig. 6).

The ring descriptors, D/Dtr05, D/Dtr08 and D/Dtr12, denote the distance/detour ring index of order 5, 8 and 12 respectively (size of the ring system) in the compounds. Among these descriptors, D/Dtr05 and D/Dtr08 contribute positively to the inhibitory activity, whereas the descriptor D/Dtr12 influences the inhibitory activity negatively (**See** Fig. 3) against the AChE enzyme. As per the above information, it can be concluded that lower the size of ring (size of the ring corresponding to ring index of not more than order 8) may be more favorable for the inhibitory activity against AChE enzyme instead of a larger the size of the ring. The positive regression coefficients of the descriptors (D/Dtr05 and D/Dtr08) suggest that a higher numerical value of these descriptors lead to improved inhibitory activity as verified by the compounds (Fig. 6) **947** (pIC_{50}: -1.633) and **717** (pIC_{50}: -1.278) in case of D/Dtr05 and **841** (pIC_{50}: -0.365) and **837** (pIC_{50}: -1.907) in case of D/Dtr08. Again, the compounds with no such fragments show lower AChE

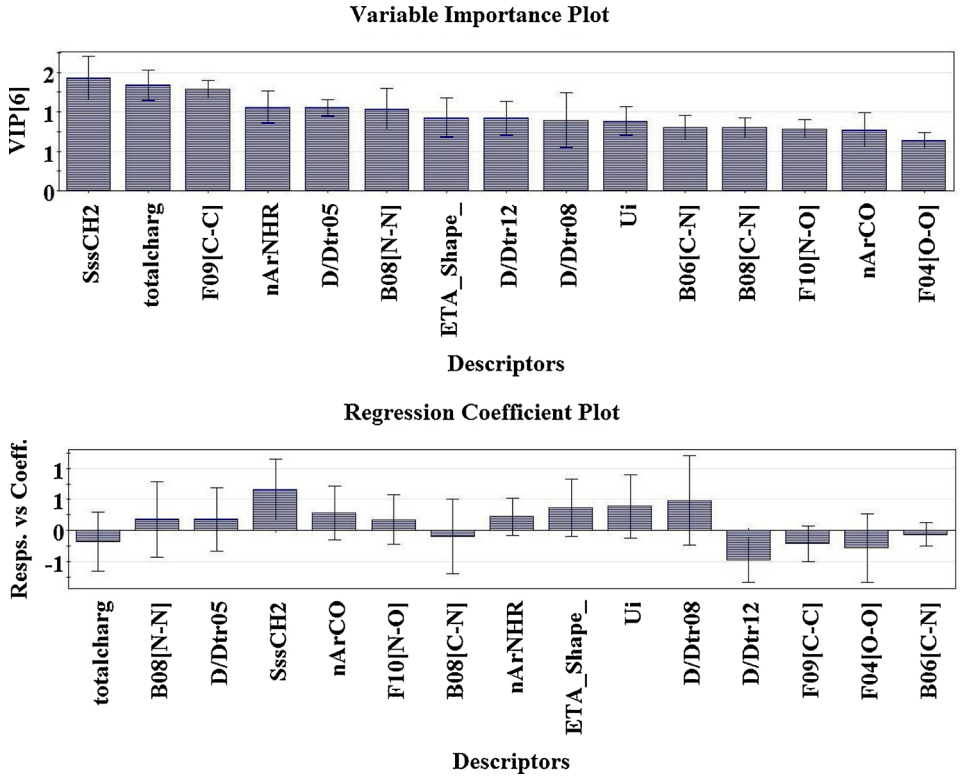


Fig. 3. Variable importance plot (VIP) and regression coefficient plot of final PLS model against AChE enzyme.

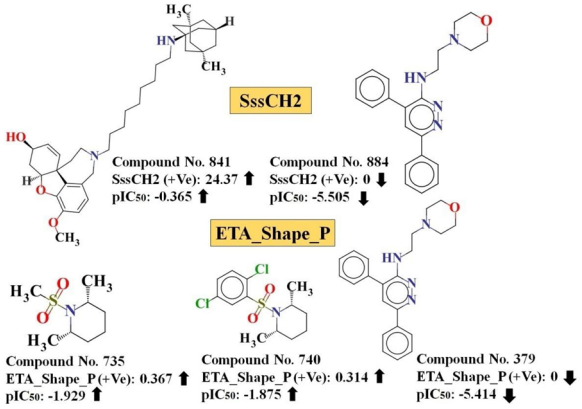


Fig. 4. Impact of SssCH2 and ETA_Shape_P descriptors on AChE enzyme inhibitory activity.

enzyme inhibitory activity as found in the compounds, such as **123** (pIC_{50}: -6.143) and **871** (pIC_{50}: -6) (D/Dtr05 and D/Dtr08). The descriptor D/Dtr12 negatively influences the inhibitory activity as suggested by the regression coefficient plot (**see** Fig. 3).The negative contribution indicates the higher number of this fragment in the compounds shows lower inhibitory activity as supported by compounds (Fig. 7) **675** (pIC_{50}: -7.683) and **681** (pIC_{50}: -7.152). On the other hand, absence of this feature in the compounds shows higher AChE enzyme inhibitory activity as observed in case of compounds (Fig. 7) **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698).

Totalcharge is a constitutional indices type descriptor which negatively influences the inhibitory activity against AChE enzyme. In the dataset, there are only 44 compounds with non-zero values for this descriptor, with a descriptor value of 2 in case of compound number **872,** and 1 in rest of all 43 compounds (compound numbers **685–687, 847–849, 851–858, 860, 862–866, 869–871, 873–877, 879, 880, 882–884, 888, 890, 892, 893, 897, 898, 900, 934, 937** and **938**) (see supplementary materials S1 Sheet 1). The descriptor encoded the information about the total charges present in the compounds. The descriptor negatively influences the inhibitory activity and suggested that the presence of this feature in the compounds correlates to low inhibitory activity as shown in compounds **872** (pIC_{50}: -5.079) and **871** (pIC_{50}: -6) (containing descriptor value 2 and 1 respectively) (Fig. 7). On the other hand, the uncharged compounds have higher inhibitory activity as shown in compounds **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698).

The next descriptor in this model, F04[O-O] belongs to the family of 2D atom pair descriptors that defines the frequency of two oxygen atoms at the topological distance 4. As per the regression coefficient plot (**see** Fig. 3), the descriptor contributes negatively to the inhibitory activity against AChE enzyme. So, the compounds with such fragment express the lower inhibitory activity as proved by the (Fig. 7) compounds **902** (pIC_{50}: -4.799) and **133** (pIC_{50}: -5.423) (their corresponding descriptor values are 6 and 5 respectively), whereas compounds with no such fragments show higher AChE enzyme inhibitory activity as shown in compounds **18** (pIC_{50}: 1.154) and **3** (pIC_{50}: 0.698) (Fig. 7).

2D atom pair descriptor, F10[N–O], defines the frequency of N–O at the topological distance 10. As per the regression coefficient plot (**see** Fig. 3), the descriptor contributed positively towards the inhibitory activity against the AChE enzyme. Thus, the evidence obtained from this descriptor suggests that the molecules containing the N–O fragment at the topological distance 10 show higher AChE enzyme inhibitory activity as shown in compounds (Fig. 7) **107** (pIC_{50}: -0.692) and **719** (pIC_{50}: -0.342) (containing descriptor value 2 respectively), while compounds **675** (pIC_{50}: -7.683), **681** (pIC_{50}: -7.152) show lower inhibitory activity due to the absence of this fragment (Fig. 7).

The functional group count descriptor, nArCO, denotes the number of aromatic ketone groups present in molecules. The positive regression coefficient of this descriptor suggests that the ketone group attached with an aromatic ring (not less than 2 fragments) is favorable for AChE

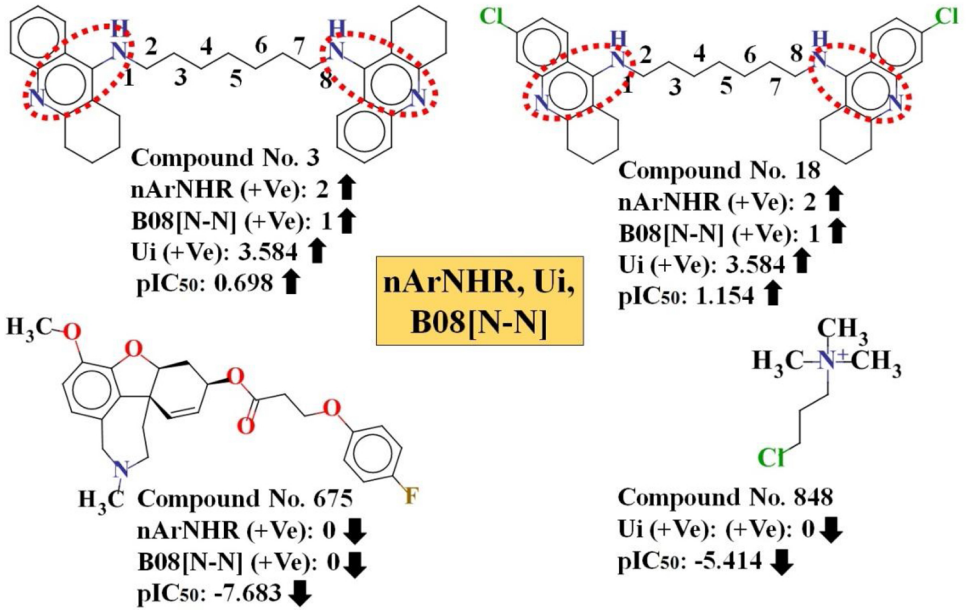


Fig. 5. Impact of nArNHR, Ui and B08[N-N] descriptors on AChE enzyme inhibitory activity.

enzyme inhibitory activity, as found in case of compounds (Fig. 7) **653** (pIC_{50}: -0.619) and **1131** (pIC_{50}: -0.619) and vice versa founds in case of compounds **675** (pIC_{50}: -7.683) and **681** (pIC_{50}: -7.152).

3.1.2. 2D-QSAR analysis against BuChE enzyme: Model 2

Equation 2 corresponds to the best PLS regression based 2D-QSAR model that comprises 13 descriptors with corresponding latent variables 5 against BuChE enzyme. The descriptors appearing in the model are organized accordingly to their significance, and then defined individually. The VIP and regression coefficient plot (Kumar et al., 2020c) define the significance level of each variable found from the final PLS model that are responsible to regulate the BuChE enzyme inhibitory activity as presented in Fig. 8 in the regression coefficient plot.

The next 2D atom pair descriptor, F09[C—O], describes the frequency of C—O at the topological distance 9. The positive contribution (as per regression coefficient plot) of this descriptor points out that the number of the C—O group at the topological distance 9 may favor the inhibitory activity against BuChE enzyme as found in compounds (Fig. 9) **513** (pIC_{50}: -0.863) and **536** (pIC_{50}: -1.778) (containing descriptor values 6 and 7, respectively) and opposite found in case of compounds (Fig. 9) **624** (pIC_{50}: -5.605) and **224** (pIC_{50}: -5.430) (absence of such fragment).

The next atom centered fragment descriptor, C-041, exemplifies the fragment (X-C(=X)-X) indicating the number of fragments containing C(sp2) atoms that are attached with two electronegative atoms (O, N, S, Se and halogens), i.e., one by a single bond and another by a double bond (Todeschini and Consonni, 2020). The positive regression coefficient suggests the influential effect of the feature containing C(sp2) atoms directly attached with two electronegative atoms towards the BuChE inhibitory activity. This is witnessed by the compounds (Fig. 9) **748** (pIC_{50}: -1.462) and **750** (pIC_{50}: -1.531) (descriptor value 1 in both cases), and the opposite is seen in compounds **624** (pIC_{50}: -5.605) and **772** (pIC_{50}: -5.594) as depicted in Fig. 9.

The functional group count descriptor, nArCONHR, stands for the number of secondary amides (aromatic) in the compounds. The fragment contributes significantly to increase the intermolecular

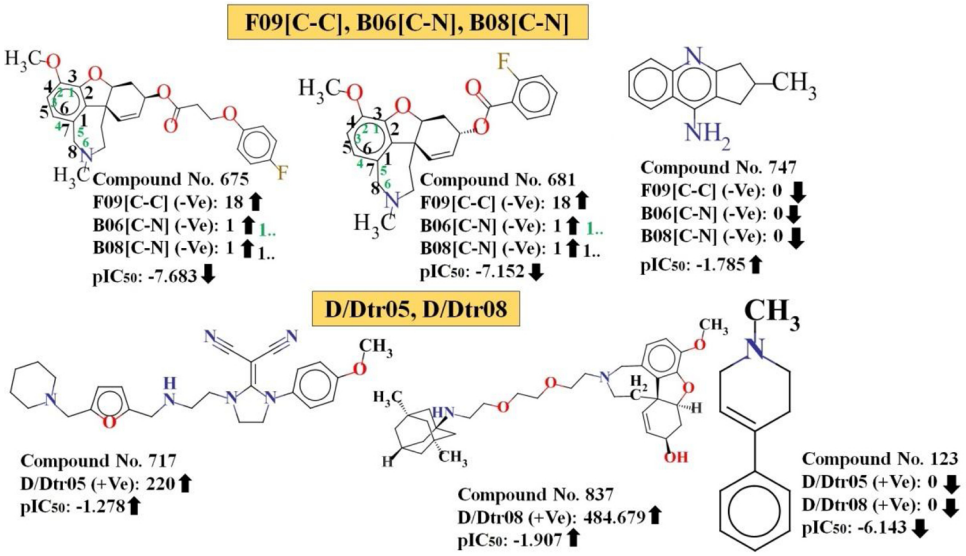


Fig. 6. Impact of F09[C-C], B06[C-N], B08[C-N], D/Dtr05 and D/Dtr08 descriptors on AChE enzyme inhibitory activity.

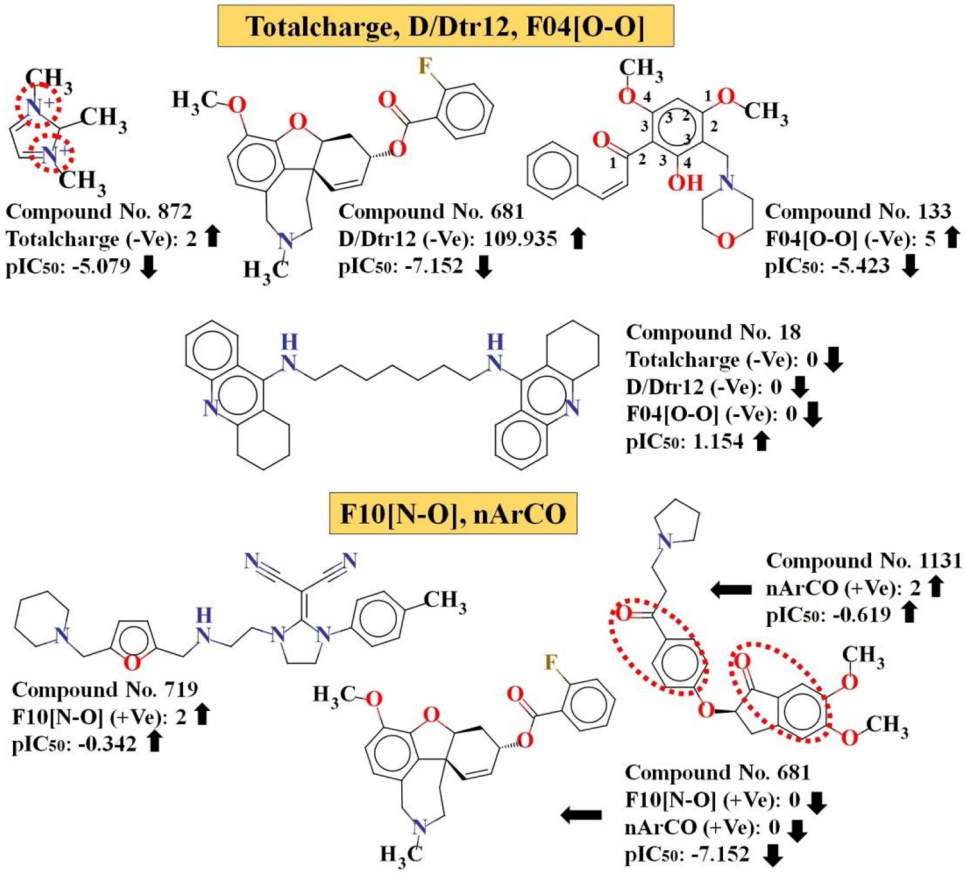


Fig. 7. Impact of Totalcharge, D/Dtr12, F04[O-O], F10[N-O] and nArCO descriptors on AChE enzyme inhibitory activity.

interactions by strong H-bonds (discussed later in the docking part). The descriptor contributes positively to the BuChE inhibitory activity as specified by the positive regression coefficient. Thus, the molecules bearing such fragment may have enhanced BuChE inhibitory activity as presented in compounds (Fig. 9) **391** (pIC_{50}: -0.578) and **388** (pIC_{50}: -0.892) (containing descriptor values 1). Although the compounds containing no such fragments have lower inhibitory activity as shown in compounds (Fig. 9) **624** (pIC_{50}: -5.605) and **772** (pIC_{50}: -5.594).

The 2D atom pair descriptor, B06[O-O] represents the presence/absence of two oxygen atoms at the topological distance 6. The negative contribution of this descriptor indicates that the presence of two oxygen atoms at the topological distance 6 may be detrimental for BuChE enzyme inhibitory activity. This is evidenced by compounds (Fig. 9) such as **772** (pIC_{50}: -5.594) and **621** (pIC_{50}: -5.593) (with descriptor value 1 in each case), while an absence of this fragment in the compounds leads to a higher inhibitory activity as observed in compounds (Fig. 9) **13** (pIC_{50}: -0.397) and **547** (pIC_{50}: -0.428).

The functional group count descriptor, nThiazoles, designates the number of thiazole ring present in the molecules. As per the regression coefficient plot (**see** Fig. 8) this descriptor contributed negatively to the BuChE enzyme inhibitory activity. The information obtained from the regression coefficient plot suggests that presence of such fragments in the compounds is inversely proportional to the BuChE enzyme inhibitory activity as witnessed by the compounds (Fig. 9) **135** (pIC_{50}: -5.202) and **137** (pIC_{50}: -5.073) (both compounds have descriptor value 1), even though the absence of such ring system in the compounds (Fig. 9) leads to an improved inhibitory activity as detected in the compounds **13** (pIC_{50}: -0.397) and **547** (pIC_{50}:-0.428).

The next atom-centred fragment descriptor, C-019, simply refers to CRX, where R represents any group connected through carbon atom; X signifies any heteroatom (O, N, S, P, Se, and halogens) (Todeschini and Consonni, 2020). This descriptor negatively influences the inhibitory activity of BuChE enzyme inhibitors as specified by its negative regression coefficient, which indicates that this feature does not enhance the BuChE enzyme inhibitory activity of molecules as found in compounds (Fig. 9) **66** (pIC_{50}: -4.100) and **793** (pIC_{50}: -5.096) (containing descriptor values 2 in both cases). But, the compounds with no such fragment have higher inhibitory activity as shown in compounds (Fig. 9) **13** (pIC_{50}: -0.397) and **547** (pIC_{50}: -0.428).

The most contributing descriptor, C-028, an atom-centred fragment descriptor, describes each atom by its own atom type and the bond types and atom types of its first neighbors (R–CR-X) (Das et al., 2017a). In this situation, R–CR-X can be defined as a central carbon atom (C) on an aromatic ring that has one carbon neighbor (R) and one heteroatom neighbor (X) on the same aromatic ring and the third neighbor outside this ring is a carbon (R) (Todeschini and Consonni, 2020; Nekoei et al., 2011). The positive contribution of this descriptor to the BuChE inhibitors indicates that by increasing the number of heteroatoms (with R–CR-X format) in compounds, the value of this descriptor increases, resulting in an increase of its activity values. This has been noticed in compounds (Fig. 10) **13** (pIC_{50}:-0.397) and **547** (pIC_{50}: -0.428) having corresponding descriptor value 2 in each case showing higher inhibitory activity, whereas in case of compounds (Fig. 10) like **624** (pIC_{50}: -5.594) and **772** (pIC_{50}: -5.605), the absence of such fragments in the compounds shows lower inhibitory activity.

The 2D atom pair descriptor, B09[C-O], stands for the presence/absence of C-O at the topological distance 9. According to the regression coefficient plot, this feature positively affects the activity of the BuChE enzyme inhibitors. So, the compounds with the higher number of C-O fragments at the topological distance 9 may have improved BuChE enzyme inhibitory activity as verified by (Fig. 10) compounds **13** (pIC_{50}: -0.397) and **547** (pIC_{50}: -0.428) (their corresponding

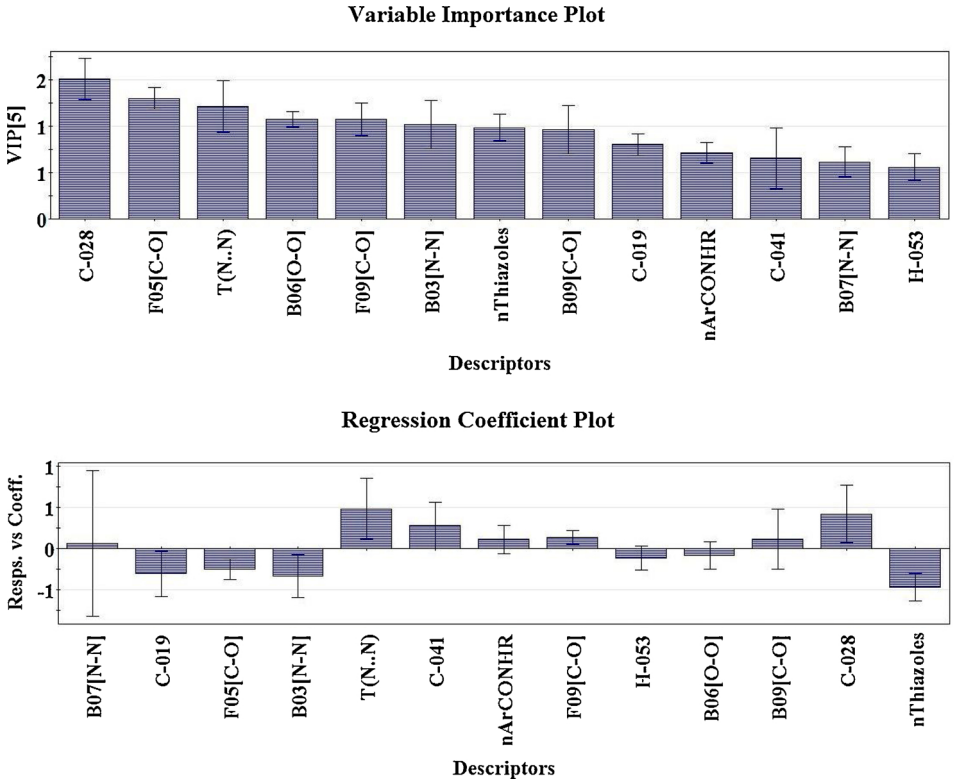


Fig. 8. Variable importance plot (VIP) and regression coefficient plot of final PLS model against BuChE enzyme.

descriptors values are 1). But, the compounds with no such fragment show lower inhibitory activity as revealed in compounds **624** (pIC_{50}: -5.605) and **224** (pIC_{50}: -5.430).

The 2D atom pair descriptor in this model, B07[N-N], describes the presence/absence of two nitrogen atoms at the topological distance 7. As per the regression coefficient plot (**see** Fig. 8), the descriptor positively influences the inhibitory activity against BuChE enzyme. This phenomenon is well observed in compounds (Fig. 10) **547** (pIC_{50}: -0.428) and **555** (pIC_{50}: -0.715), and the reverse is seen in case of compounds (Fig. 10) **624** (pIC_{50}: -5.605) and **772** (pIC_{50}: -5.594) (no such fragment at topological distance 7).

The descriptor, B03[N-N] belongs to the family of 2D atom pair, describes the presence/absence of two nitrogen atoms at the topological distance 3. As per the regression coefficient plot (**see** Fig. 8), the descriptor contributed negatively to the inhibitory activity against BuChE enzyme, suggesting that compounds containing lower number or absence of such fragments have good inhibitory activity against BuChE enzyme as presented in compounds (Fig. 10) **13** (pIC_{50}: -0.397) and **547** (pIC_{50}: -0.428) (absence of such fragment), while higher number of this fragment shows lower inhibitory activity as detected in compounds (Fig. 10) **624** (pIC_{50}: -5.605) and **772** (pIC_{50}: -5.594).

The next important descriptor, F05[C-O], belongs to the family of 2D atom pair descriptor that defines the frequency of C and O atoms at the topological distance 5. As per the regression coefficient plot (**see**

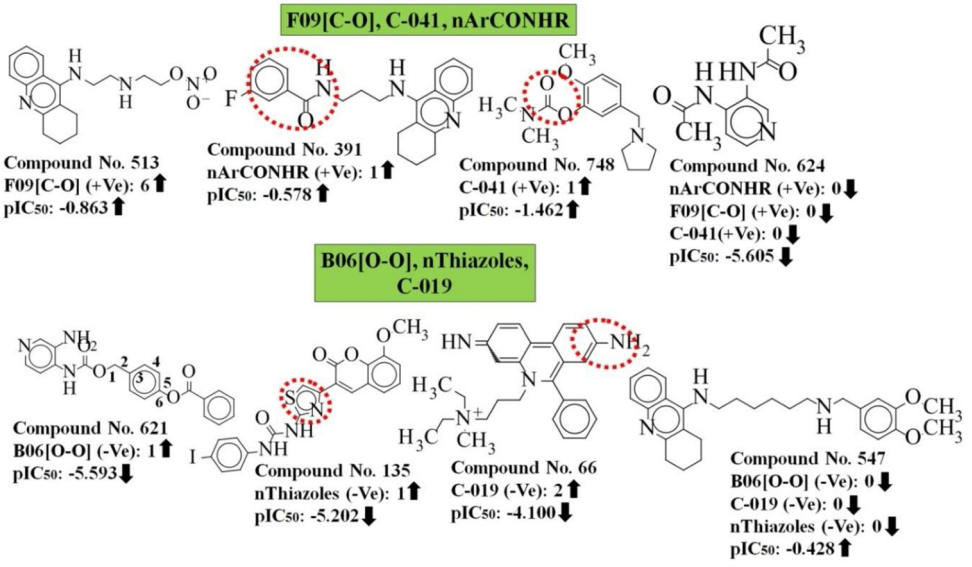


Fig. 9. Contribution of F09[C-O], C-041, nArCONHR, B06[O-O], nThiazoles and C-019 descriptors for the inhibition of BuChE enzyme.

Fig. 8), the descriptor negatively influenced the inhibitory activity against the BuChE enzyme. So, the compounds having such fragment show lower values of inhibitory activity as proved by (Fig. 11) compounds **74** (pIC_{50}: -4.212) and **75** (pIC_{50}: -4.394) (their corresponding descriptor values are 21 and 19 respectively), whereas the absence of such fragment in the compounds show higher BuChE enzyme inhibitory activity as presented in compounds **466** (pIC_{50}: -0.929) and **2** (pIC_{50}: -1) (Fig. 11).

The descriptor, T(N..N), stands for the 2D atom pair descriptor, simply characterizes the sum of topological distances between two nitrogen atoms. This descriptor contributes positively to the BuChE inhibitory activity as suggested by positive regression coefficient. Thus, the molecules bearing higher topological distance between two nitrogen atoms may have higher BuChE inhibitory activity as presented in (Fig. 11) compounds **8** (pIC_{50}: -0.908) and **466** (pIC_{50}: -0.929) (containing descriptor values 59 and 124 respectively), whereas in contrary, compounds **770** (pIC_{50}: -5.539) and **779** (pIC_{50}: -5.459) which do not contain any such fragment showed less BuChE inhibitory activity. From this observation, it can be concluded that the topological distances between two nitrogen atoms should be higher for improved inhibitory activity against BuChE enzyme.

The last atom centered fragment descriptor in this model, H-053, simply refers to H atoms attached to C°(sp^3) with 2X connected to next C, where X signifies any heteroatom (O, N, S, P, Se, and halogens) and the superscript characterizes the formal oxidation number (Todeschini and Consonni, 2020). The formal oxidation number of a carbon atom equals to the sum of the conventional bond orders with electronegative atoms (Todeschini and Consonni, 2020). This descriptor is defined as the number of definite atom types in a compound and can be calculated by knowing only molecular composition and atom connectivity (Todeschini and Consonni, 2020). The negative regression coefficient of this descriptor suggested that compounds containing larger number of such hydrogen atoms have lower inhibitory activity against BuChE enzyme as shown in compounds (Fig. 11) **280** (pIC_{50}: -4.989) and **296** (pIC_{50}: -4.935), and their corresponding descriptor values are 6 in both cases. On the other hand, compounds **13** (pIC_{50}: -0.397) and **597** (pIC_{50}: -0.428) show higher inhibitory activity because of the absence of such H atom.

3.1.3. 2D-QSAR analysis based on selectivity of AChE and BuChE enzyme inhibitors: Model 3

In brain, two varieties of cholinesterase enzymes (AChE and BuChE) are capable of hydrolyzing acetylcholine neurotransmitter (Du et al., 2018). Throughout the development of AD, BuChE activity increases by 40–90 % in the temporal cortex and hippocampus, while at the same time AChE activity decreases (Du et al., 2018). Therefore, the concurrent inhibition of both AChE and BuChE should provide extra benefits in the treatment of AD. In this research, we have developed PLS regression-based 2D-QSAR model based on selectivity for both AChE and BuChE (difference of pIC_{50} values between them). The descriptors found in the developed model are organized as per the VIP plot (**see** Fig. 12) and then defined individually. The regression coefficient plot (Kumar et al., 2020c) describes the contribution of each descriptor in the model for the inhibition of both cholinesterase enzymes (**see** Fig. 12).

The most significant descriptor, ETA_Shape_X, simply refers the shape index X (Das et al., 2017b; Roy and Das, 2017). We can find out the effect of molecular shape on the inhibitory activity against cholinesterase enzymes with the help of this descriptor. The positive regression coefficient of the descriptor suggests that the above feature of the compounds is more specific for AChE enzyme inhibitory activity as compared to the BuChE enzyme inhibitory activity. The higher numerical value of this descriptor correlates to higher AChE enzyme inhibitory activity as observed in case of compounds (Fig. 13) **31** (pIC_{50}: 2.986) and **104** (pIC_{50}: 2.98) and opposite in case of compounds (Fig. 13) **57** (pIC_{50}: -2.238) and **58** (pIC_{50}: -1.995).

The extended topochemical atom (ETA) descriptor, ETA_dBetaP, demonstrates the measure of the unsaturation content relative to molecular size (Roy and Das, 2017). The positive regression coefficient of this descriptor proposes that the highest unsaturation content related to their molecular size in the compounds is more specific for AChE enzyme inhibitory activity. It has been found that with an increase in the numerical value of this descriptor, the inhibitory activity of compound also increases against AChE enzyme as presented in case of compounds (Fig. 13) **203** (pIC_{50}: 3.212) and **11** (pIC_{50}: 4.325), while the lower numerical value indicated the less specificity against the AChE enzyme as found in case of compounds (Fig. 13) **160** (pIC_{50}: -1.122) and **158** (pIC_{50}: -1.477). Thus, we can conclude from above information that unsaturation content related to their molecular size is more specific for AChE enzyme inhibitory activity compared to the BuChE enzyme inhibitory activity.

The descriptor, F05 [N–O], belonging to the class of 2D atom pairs, indicates the frequency of N–O at the topological distance 5. The negative contribution (as per the regression coefficient plot) of this

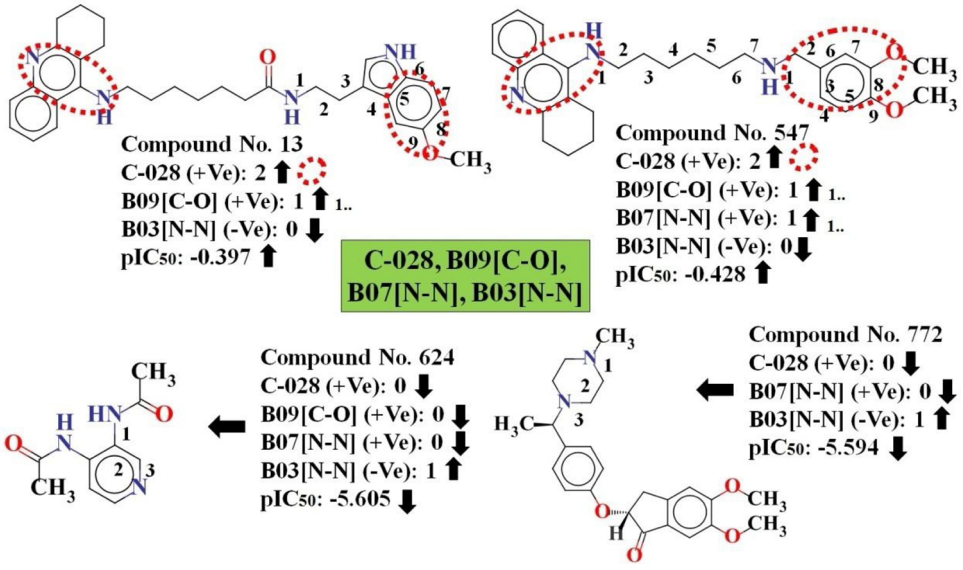


Fig. 10. Contribution of C-028, B09[C-O], B07[N-N] and B03[N-N] descriptors for the inhibition of BuChE enzyme.

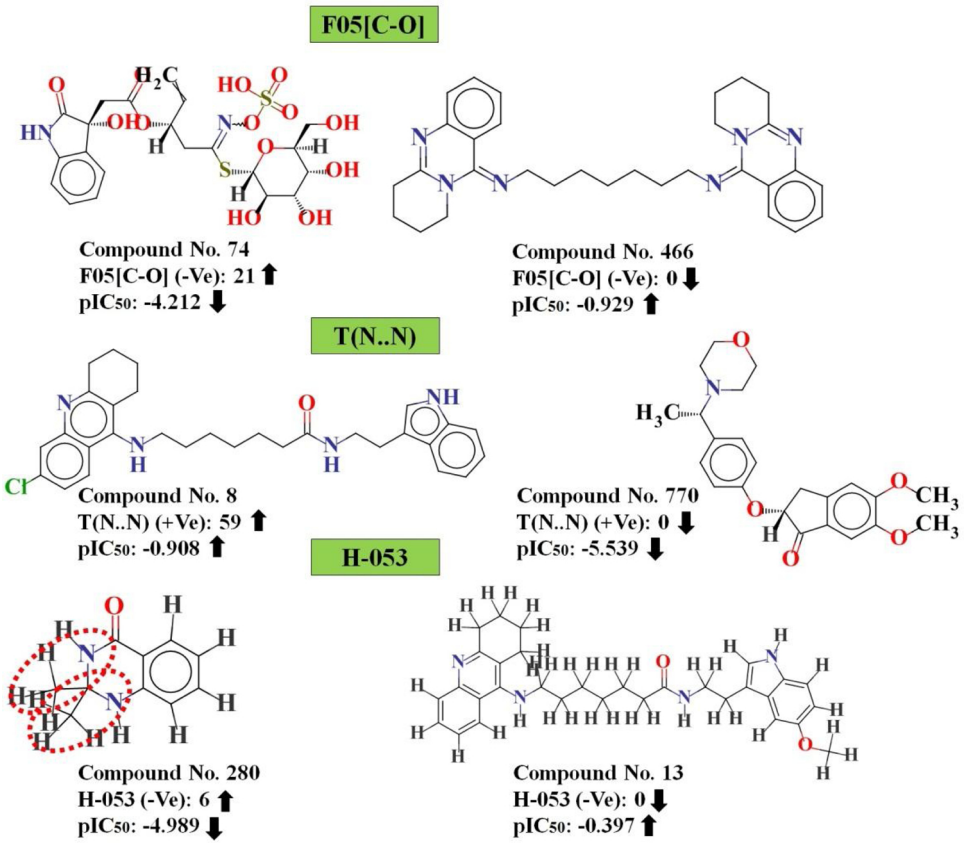


Fig. 11. Contribution of F05[C-O], T(N..N) and H-053 descriptors for the inhibition of BuChE enzyme.

descriptor suggested that frequency of N–O at the topological distance 5 is more specific to BuChE enzyme inhibitory activity than AChE enzyme inhibitory activity. Thus, higher number of this fragment correlates with lower AChE enzyme inhibitory activity as noticed in case of compounds (Fig. 14) **160** (pIC_{50}: -1.122) and **57** (pIC_{50}: -2.238). On the other hand, compounds having no such fragments show better AChE enzyme inhibitory activity values as observed in the compounds (Fig. 14) **11** (pIC_{50}: 4.256) and **44** (pIC_{50}: 4.041).

The functional group count descriptor, nCq, denotes the number of total quaternary carbons (sp3) present in the molecules. As per the regression coefficient plot (**see** Fig. 12), the descriptor negatively influences the inhibitory activity and suggests that the presence of the number of quaternary carbon in the compounds is more specific to BuChE enzyme inhibitory activity compared to the AChE enzyme inhibitory activity. From the descriptor contribution, it can be suggested that molecules containing this fragment may not be favorable for the AChE enzyme inhibitory activity as presented in compound numbers (Fig. 14) **164** (pIC_{50}: -1.852) and **158** (pIC_{50}: -1.477). The reverse is seen in compounds (Fig. 14) **11** (pIC_{50}: 4.256) and **44** (pIC_{50}: 4.041).

The descriptor, nArOR, belongs to the family of functional group count descriptor, designates the number of aromatic ether groups present in a compounds. The positive regression coefficient of this descriptor advises that the ether group attached with an aromatic ring is more specific and favorable for the AChE enzyme inhibitory activity, as found in case of compound numbers (Fig. 14) **97** (pIC_{50}: 3.974) and **88** (pIC_{50}: 3.960), whereas, the lack of such fragment in the compounds leads to a decrease in the AChE enzyme inhibitory activity as shown in compound number (Fig. 14) **2** (pIC_{50}: -1.249) and **204** (pIC_{50}: -1.175). Thus, it can be concluded that the fragment is more specific for AChE enzyme inhibitory activity compared to the BuChE enzyme inhibitory activity.

The last 2D atom pair descriptor in this model, B10[C-C], describes the presence/absence of two carbon atoms at the topological distance 10. The positive contribution (as per the regression coefficient plot) of this descriptor specifies that presence of two carbon fragments at the topological distance 10 is more specific for AChE enzyme inhibitory activity than the BuChE enzyme inhibitory activity. This phenomenon is well noticed in compounds (Fig. 14) **11** (pIC_{50}: 4.256) and **44** (pIC_{50}: 4.041) and the reverse is seen in case of compounds (Fig. 14) **195** (pIC_{50}: -1.570) and **2** (pIC_{50}: -1.249) (no such fragment at the topological distance 10).

3.1.4. Applicability domain criteria

The applicability domain for the test set compounds was checked at 99% confidence level applying the DModX (distance to model in X-

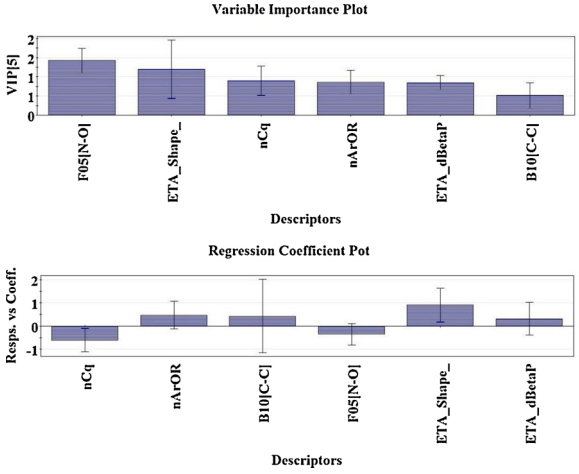


Fig. 12. Variable importance plot (VIP) and regression coefficient plot of final PLS selectivity based model.

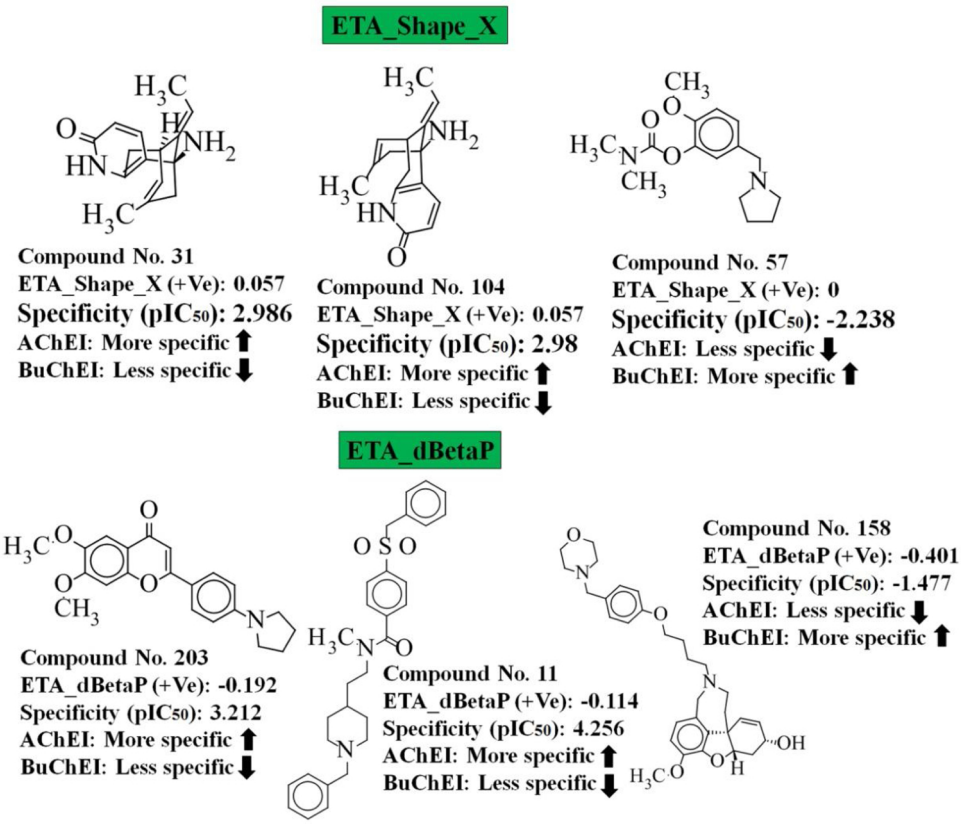


Fig. 13. Contribution of ETA_Shape_X and ETA_dBetaP descriptors for the dual inhibition of AChE and BuChE enzyme.

space) approach available in SIMCA-P 10.0 software (SIMCA, 2020) as described by the Roy et al. and Khan et al.(Roy et al., 2019; Khan and Roy, 2019). We have noticed from the **Figure S6** (see S6 in supplementary materials), there is only one compound (compound **666**) in the test set found outside the DModX value (D-Crit = 1.623) in case of **Model 1** (AChEI model). In case of the reported **Model 2** (see **Figure S6** in supplementary materials S6), we found that one compound (compound **138**) in the test set is traced outside the critical DModX value (D-Crit = 1.67), whereas in case of **Model 3** (selectivity based model) (see **Figure S6** in supplementary materials S6), we have also found the only one compound (Compound **30**) in the test set is outside the DModX value (D-Crit = 3.066).

3.2. Molecular docking

In the present work, we have applied molecular docking studies using most and least active compounds from the two datasets, i.e., AChE and BuChE enzyme inhibitors. In case of AChE enzyme inhibitors, we have selected two most active compounds, i.e., **15** and **19** and two least active compounds **123** and **674**, from the dataset; in case of BuChE inhibitors dataset, we have selected two most active compounds, such as **13** and **547** and least active compounds **621** and **624** to understand the docking interactions with the active site of enzymes. The details of docking interactions and their relation with 2D-QSAR results are depicted in Table 1. Here, we have discussed the details of docking interaction as well as their analysis below.

3.2.1. Molecular docking analysis of the selected compounds from AChE enzyme inhibitors dataset

In this analysis, two most active compounds (**15** and **19**) from the dataset (pIC_{50} = 0.221 and0.522 respectively) linked with the active site amino acid residues (LEU A:289, TRP A:286, SER A: 293, TYR A: 124, TRP A: 86, HIS A:447, GLU A:292, GLN A: 291, ASN A: 87, TYR A: 72, TYR A:337 and TYR A: 341) through interacting forces, such as hydrogen bonding (conventional and carbon hydrogen bonds), π-bonding (π-donor hydrogen bond, π-π stacked, π-alkyl, π-cation, π-π T-shaped), alkyl, and halogen.

One of the most active compounds form the dataset, compound **15** (see Fig. 15), interacts with the active site cavity through Hydrogen bonding (TYR A:337, ASN A:87, SER A:293, GLU A:292 and GLN A:291), π-donor hydrogen bonding (TYR A:124), Halogen (fluorine) (TRP A: 86 and TRP A: 87), π-π stacked (TRP A:286, TRP A:86), π-π T-shaped (TYR A:124), Alkyl (LEU A:289) and π-Alkyl (TRP A: 86, HIS A:447, TYR A:72, TRP A:286 and LEU A:289).

Another most active compound **19**of the dataset interacted with amino acid residues, such as TYR A: 341, SER A: 293 and TYR A: 337 (hydrogen bond), TYR A: 124 (π-donor hydrogen bond), TRP A: 286 (π-cation), TRP A: 286 and TRP A: 86 (π- π stacked), PRO A: 88 and LEU A: 289 (alkyl bond), TRP A: 86 and HISA: 447 (π-alkyl bonding) and ASN A: 87 (halogen bond) (see **Figure S7** in supplementary materials S6).

In this investigation, two least active compounds (**123** and **674**) from the dataset (pIC_{50} = -6.143 and -6.495 respectively) interacted with the active site amino acid residues (ARG A:296, TYR A:341, SER A:293, TRP A: 286, HIS A:287, PHE A: 338, TYR A: 337) through interacting forces, such as Hydrogen bonding (conventional and carbon hydrogen bonds) and π-bonding (π-π stacked, π-alkyl).

One of the least active compounds of the dataset, compound **123** (Fig. 15) interacts with amino acid residues like ARG A: 296 and SER A: 293 through hydrogen bonding and TYR A: 341 via π- π stacked bonding.

Figure S8 (see S6 in supplementary materials) displays that compound **674** interacts with amino acid residues, such as HIS A: 287, SER A: 293 (through hydrogen bonding) and TRP A: 286, TYR A: 337 (via π-alkyl bond).

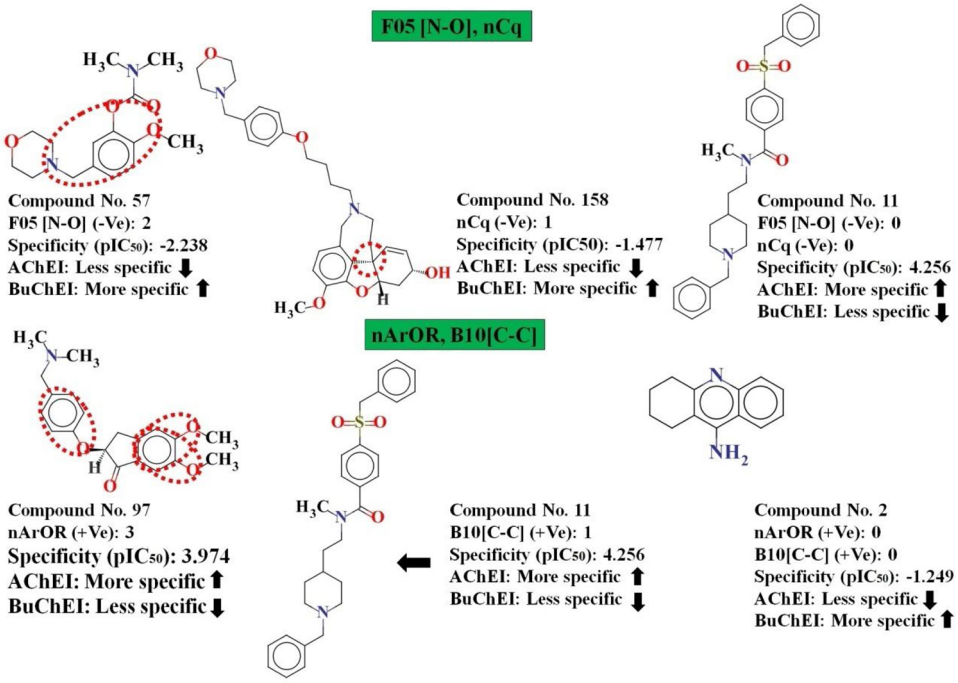


Fig. 14. Contribution of F05 [N-O], nCq, nArOR and B10[C-C] descriptors for the inhibition of both AChE and BuChE enzyme.

3.2.2. Molecular docking analysis of the selected compounds from BuChE enzyme inhibitors dataset

In this work, two most active compounds (**13** and **547**) from the dataset (pIC_{50} = -0.397 and -0.428 respectively) are linked with the active site amino acid residues, like THR A: 120, GLU A: 197, TRP A: 82, HIS A: 438, TYR A: 440, GLY A: 116, PHE A: 329, SER A: 287, SER A: 79, ASP A: 70, TYR A: 332, SER A: 198 and ALA A: 328, through interacting forces such as hydrogen bonding (conventional and carbon hydrogen bonds), π-bonding (π-π stacked, π-alkyl, amide π-stacked, π-cation, π-π T-shaped), alkyl, and attractive charges.

Fig. 16 shows that compound **547** interacts with amino acid residues, such as SER A: 79, ASP A: 70, TYR A: 332, HIS A: 438, SER A: 198 (hydrogen bonding), PHE A: 329, TRP A; 82 (π-π T-shaped), ASP A: 70, GLU A: 197 (attractive charges), TYR A: 332 (π-π stacked, π-cation), TRP A; 82, ALA A: 328 (π-alkyl), GLY A: 116 (Amide π-stacked).

One of the most active compounds from dataset, compound **13** (**see Figure S9 in supplementary information S6**), interacts with amino acid residues, THR A: 120, GLU A: 197 TRP A: 82, HIS A: 438, TYR A: 440, SER A: 287 via hydrogen bonds, PHE A: 329 through π-π T-shaped, TRP A: 82 via π-π stacked, GLY A: 116 Amide π-stacked.

In this analysis, we have selected two least active compounds (**621** and **624**) from the dataset (pIC_{50} = -5.593 and -5.605 respectively) which interacted with the active site amino acid residues, such as SER A: 79, ASP A: 70, GLN A: 71, TYR A: 332, GLU A: 197 and TRP A: 82, through interacting forces, such as hydrogen bonding (conventional, carbon hydrogen bonds and π-donor hydrogen bond), π-bonding (π-π stacked, π-cation) and salt bridge.

One of the least active compounds from dataset, compound **624** (Fig. 16), interacts with amino acid residues such as SER A: 79, ASP A: 70 and TRP A: 82 through hydrogen bonding and TRP A: 82 via π-π stacked bond.

Another least active compound of the dataset, compound **621** (see **Figure S10** in supplementary materials S6), interacts with amino acid residues such as GLU A: 197, GLN A: 71 (via hydrogen bonding), TYR A: 332, TRP A: 82 (π-π stacked), TRP A: 82 (through π-cation) and GLU A: 197 (via salt bridge).

3.2.3. Correlation with the developed 2D-QSAR models

From the above investigation, we have concluded that hydrogen bonding and π-interaction among the ligand and receptor play important roles in the interactions. Hydrogen bonding may associate through the descriptor nArNHR (against AChE enzyme inhibitor model), C-028, F09[C—O], B09[C—O], B07[N—N] and T(N..N) (against BuChE enzyme inhibitors model) of the developed 2D-QSAR models. Descriptors nArNHR, ETA_Shape_P and Ui (against AChE enzyme), C-028, B07[N—N] and T(N..N) (against BuChE enzyme) are well corroborated with interactions made via π- interactions (π-π stacked, π-cation, π-alkyl and π-π T-shaped) between the protein and ligand. The above mentioned features are observed in compounds in case of AChE enzyme inhibitors such as **15** and **19** (most active) (Fig. 15 and **Figure S7** in supplementary materials S6) and in case of BuChE enzyme inhibitors, such as **547** and **13** (most active) (Fig. 16 and **Figure S9** in supplementary materials S6). But in contrast, the descriptors B06[C—N], D/Dtr12, F09[C—C], F04[O—O] and B06[C—N] (against AChE enzyme), F05[C—O], B03[N—N], B06[O—O] (against BuChE enzyme) contributed negatively in the 2D-QSAR model and this has been observed in case of AChE enzyme inhibitors such as **123**, **674**, **621** and **624** (least active) respectively (see Fig. 15**, Figure S8, S10** in supplementary materials S6 and Fig. 16). Thus, from above investigation, we can conclude that features obtained from molecular docking studies and 2D-QSAR models are in agreement and essential for the inhibitory activity against the both AChE and BuChE enzymes.

4. Comparisons of the performance of the reported models with previous published models

In this investigation, we have performed a comparison of the best models of this study with previously published models (Shrivastava et al. 2019 (Shrivastava et al., 2019), Bukhari et al. 2014 (Bukhari et al., 2014), De Souza et al. 2012 (De Souza et al., 2012) and Pang et al. 2017 et al. (Pang et al., 2017)) for the prediction of the bioactivity against AChE and BuChE enzymes, as depicted in the Table 2. The details of different internal and external validation parameters obtained

Table 1
Docking results and correlation with 2D-QSAR models in this study.

S. No.	Compound Number	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
AChE enzyme inhibitors					
1	15 (high pIC_{50})	70.674	TYR A:337, ASN A:87, SER A:293, GLU A:292, GLN A:291, TYR A:124, TRP A: 87, TRP A:286, TRP A:86, LEU A:289, HIS A:447, TYR A:72 and LEU A:289	Hydrogen bonding, π-donor hydrogen bonding, Halogen (fluorine), π-π stacked, π-π T-shaped, Alkyl and π-Alkyl	SssCH2, nArNHR, ETA_Shape_P and Ui,
2	19 (high pIC_{50})	80	TYR A: 341, SER A: 293, TYR A: 337, TYR A: 124, TRP A: 286, TRP A: 86, PRO A: 88, LEU A: 289, HISA: 447 and ASN A: 87	Hydrogen bond, π-donor hydrogen bond, π-cation, π- π stacked, alkyl, π-alkyl and Halogen (fluorine)	SssCH2, nArNHR, ETA_Shape_P and Ui
3	123 (low pIC_{50})	31.26	ARG A: 296, SER A: 293 and TYR A: 341	Hydrogen bonding and π- π stacked	B06[C-N]
4	674 (low pIC_{50})	43	HIS A: 287, SER A: 293, TRP A: 286 and TYR A: 337	Hydrogen bonding and π-alkyl	D/Dtr12, F09[C-C], F04[O-O] and B06[C-N]
BuChE enzyme inhibitors					
1	13 (high pIC_{50})	63.977	THR A: 120, GLU A: 197, TRP A: 82, HIS A: 438, TYR A: 440, GLY A: 116, PHE A: 329, SER A: 287	Hydrogen bonding, π-π stacked, Amide π-stacked and π-π T-shaped	F09[C-O], B09[C-O] and C-028
2	547 (high pIC_{50})	72.494	SER A: 79, ASP A: 70, TYR A: 332, HIS A: 438, TRP A: 82, SER A: 198, GLY A: 116, GLU A: 197, PHE A: 329, ALA A: 328	Hydrogen bonding, attractive charges, π-cation, π-π stacked, π-π T shaped, Amide π-stacked, alkyl and π-alkyl	B07[N-N], T(N..N), F09[C-O], B09[C-O] and C-028
3	624 (low pIC_{50})	20.491	SER A: 79, ASP A: 70 and TRP A: 82	Hydrogen bonding, π-π stacked and π-donor hydrogen bond	F05[C-O], B03[N-N]
4	621 (low pIC_{50})	38.333	GLN A: 71, TYR A: 332, GLU A: 197 and TRP A: 82	Hydrogen bonding, π-cation, π-donor hydrogen bond and π-π stacked bond	F05[C-O], B03[N-N], B06[O-O]

from our models and those obtained from previous models are given in Table 2. Based on the statistical quality in terms of both internal and external validation criteria, the models reported in this work are statistically significant and robust enough as compared to the previously reported models (Table 2). Moreover, the models presented in this study are derived from a larger set of molecules than those reported in the previous studies.

5. Conclusion and future prospective

In present investigation, we have employed chemoinformatic tools to investigate the datasets of 997 and 761 heterocyclic compounds (Jaén et al., 1996; Högenauer et al., 2001; Andreani et al., 2008; Carlier et al., 1999a; Ceschi et al., 2016; Contreras et al., 1999; da Costa et al., 2013; DeBernardis et al., 1988; Anand and Singh, 2012a; Barreiro et al., 2003; Cardoso et al., 2004; Li et al., 2016; Erlanson et al., 2004; Feng et al., 2005a; Fink et al., 1995; Girisha et al., 2009; Gray et al., 1985; Han et al., 1991; He et al., 2007; Hu et al., 2002, 2013; Huang et al., 2011; Ishihara et al., 1994; Jia et al., 2009; Kapples et al., 1993; Kavitha et al., 2007; Li et al., 2014, 2017a; Liu et al., 2014; McKenna et al., 1997; Morini et al., 2008; Pool et al., 1996; Anand and Singh, 2012b; Sadashiva et al., 2006; Sang et al., 2015a, b; Sang et al., 2015c; dos Santos et al., 2010; Shao et al., 2004; Shen et al., 2008; Sheng et al., 2005, 2009a; Sheng et al., 2009b; Shi et al., 2013; Shinada et al., 2012; Shutske et al., 1989; Simoni et al., 2012; Sugimoto et al., 1992, 1995; Valenti et al., 1997; Vidaluc et al., 1994, 1995; Wong et al., 2003; Yang et al., 2017; Zeng et al., 1999; Zhan et al., 2010; Zheng et al., 2010; Zhu et al., 2009; Szymański et al., 2011; Krátký et al., 2015, 2017; Rodríguez-Franco et al., 2006; Conejo-García et al., 2011; Rodríguez et al., 2016; Yurttaş et al., 2013; Zelík et al., 2010; Ahmad and Fatima, 2008; Ahmad et al., 2016; Ahmed et al., 2006; Bacalhau et al., 2016; Bagheri et al., 2015; Bolognesi et al., 2005; Camps et al., 2008; Carlier et al., 1999b; Cho et al., 2017; Czarnecka et al., 2017; Decker, 2006; Decker et al., 2008; Fang et al., 2008a, b; Feng et al., 2005b; Lin et al., 1998; Gregor et al., 1992; Hameed et al., 2015, 2016; Hasan et al., 2005; Huang et al., 2010; Jiang et al., 2011; Kanhed et al., 2015; Kurt et al., 2015; Leader et al., 2002; Leng et al., 2016; Li et al., 2013, 2017b; Luo et al., 2011; Mohamed et al., 2011; Mohammadi et al., 2015; Najafi et al., 2016; Pouramiri et al., 2017; Rydberg et al., 2006; Saeed et al., 2015; Saeedi et al., 2017; Samadi et al., 2010, 2012; Sarfraz et al., 2017; Shi et al., 2011; Skrzypek et al., 2013; Sterling et al., 2002; Tang et al., 2007; Villalobos et al., 1994; Wen et al., 2007; Yanovsky et al., 2012; Zakhari et al., 2011) with defined AChE and BuChE enzyme inhibitory activities, respectively, to investigate the important structural features for enzyme inhibition. Additionally, 198 heterocyclic compounds from the same datasets having dual inhibitory activity against AChE and BuChE enzymes have been considered for exploring selectivity pattern. Significant and easily interpretable 2D descriptors are calculated for the purpose of model development. Prior the development of the final models, we have employed a multilayered variable selection strategy for the selection of significant descriptors. The PLS regression based methodology was used for the developed final models following the OECD guidelines. The statistical results obtained from the developed models exhibited acceptable quality in terms of both internal and external validation matrices. From the insights obtained from generated PLS models (as also revealed in the variable importance plot, regression coefficient plot, and loading plot as shown in Figs. 3, 8, 12 and S4 in supplementary materials S6), we have concluded that: higher number of $-CH_2-$ groups (corresponding SssCH2 descriptor values not more than 40), number of secondary aromatic amines (corresponding descriptor values not more than 4), smaller ring size (size of the ring corresponding to ring index not more than order 8), branching in the cationic structure (one central atom is attached to three other non-hydrogen atoms) and number of aromatic ketone groups (not less than 2 fragments) may be more favorable for the inhibitory activity against AChE enzyme. In case of the BuChE inhibitor model, a central carbon

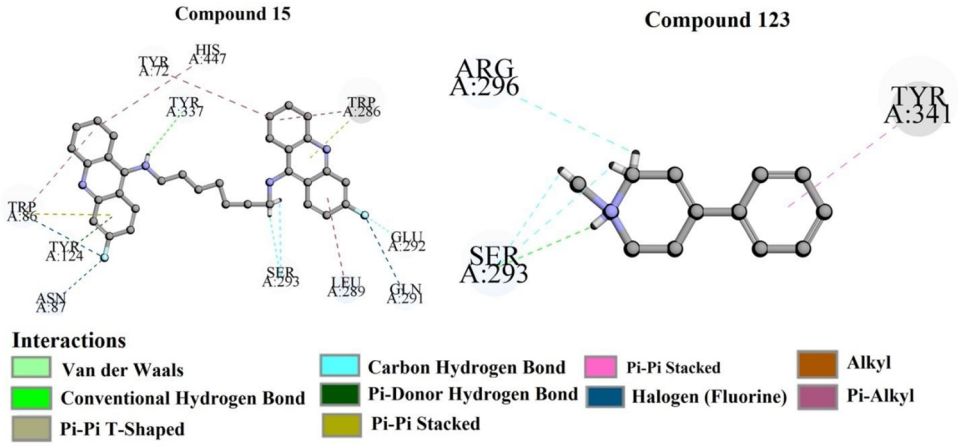


Fig. 15. Docking interactions of most active (Compound **15**) and least active (Compound **123**) compounds from dataset of AChE enzyme inhibitors.

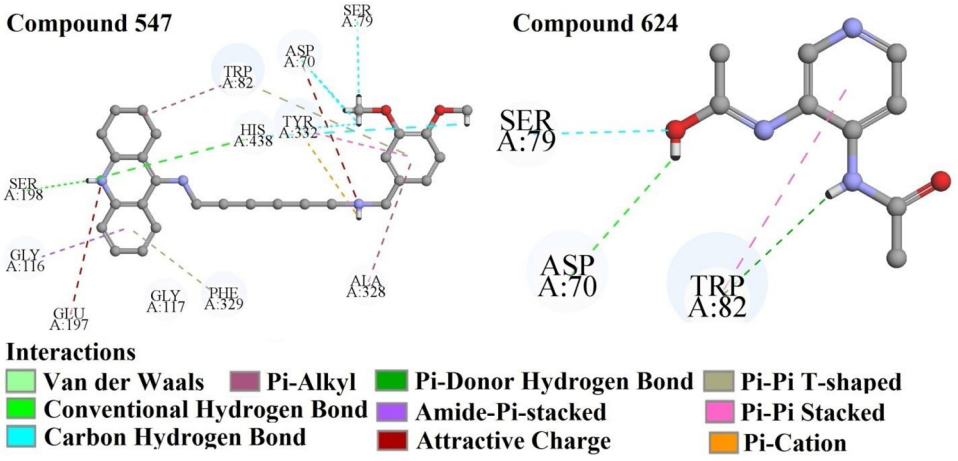


Fig. 16. Docking interactions of most active (Compound **547**) and least active (Compound **624**) compounds from dataset of BuChE enzyme inhibitors.

atom (C) on an aromatic ring that has one carbon neighbor (R) and one heteroatom neighbor (X) on the same aromatic ring and the third neighbor outside this ring is a carbon (R), sum of topological distances between two nitrogen atoms, number of fragments containing C(sp2) atoms that are attached with two electronegative atoms (O, N, S, Se and halogens), i.e., one by a single bond and another by a double bond and the number of secondary aromatic amides may influence the inhibitory activity of BuChE enzyme inhibitors, whereas higher number of thiazole rings, CRX fragments and H atoms connected to C°(sp^3) with 2X attached to next 'C' are detrimental to the BuChE enzyme inhibitory activity. The features obtained from selectivity based model suggests that the number of aromatic ethers, unsaturation content related to their molecular size and molecular shape may be more specific for the inhibition AChE enzyme in comparison to the BuChE enzyme, whereas the number of total quaternary carbons (sp3) may be more specific for BuChE inhibitory activity. The identified features are responsible for increasing brain permeability and an entropically more favorable binding to the receptor and intermolecular interactions by strong H-bonds for the improvement of the inhibitory activity of both enzymes (AChE and BuChE). Finally, molecular docking analysis has performed to identify the interactions between target proteins (AChE and BuChE enzyme) and inhibitors in this datasets., and the results showed the

Table 2
Comparisons of proposed study with previous published studies against AChE and BuChE enzymes.

Sources	E. L.	LV	Model	Training set			Test set	
				n	r^2	Q^2	n	Q^2F_1
Against AChE (Present work)	**15**	**6**	**PLS**	**798**	**0.662**	**0.645**	**199**	**0.661**
Against BuChE (Present work)	**13**	**5**	**PLS**	**603**	**0.674**	**0.656**	**158**	**0.663**
Selectivity (Present work)	**6**	**5**	**PLS**	**159**	**0.679**	**0.650**	**39**	**0.787**
Shrivastava et al. 2019a (Shrivastava et al., 2019)	–	–	PLS	26	0.792	0.713	6	0.542
Bukhari et al 2014 (Bukhari et al., 2014)	2	–	GA-MLR	14	0.855	0.792	3	0.771
De Souza et al. 2012 (De Souza et al., 2012)	2	–	HQSAR	29	0.965	0.787	7	–
De Souza et al. 2012 (De Souza et al., 2012)	2	–	HQSAR	29	0.952	0.904	7	–
Pang et al 2017 et al. (Pang et al., 2017)	–	–	3D-QSAR	35	0.925	–	10	0.850
Pang et al 2017 et al. (Pang et al., 2017)	–	–	3D-QSAR	35	0.883	–	10	0.881

Abbreviations: LV = Latent variables, E.L. = Equation length, PLS = Partial least square, GA-MLR = Genetic algorithm multiple linear regression, 3D-QSAR = Three dimensional quantitative structure activity relationship and HQSAR= Hologram QSAR.

active compounds (compounds **15** and **19** in case of AChE inhibitors and **13** and **547** in case of BuChE inhibitors), formed hydrogen bond and hydrophobic π interactions with amino acid residues that lead to identification of active binding site of target protein. Moreover, the information obtained from molecular docking analysis well supported the features obtained from the 2D-QSAR analysis results. The validated models might be supportive for estimation of the inhibitory activity of novel compounds against the AChE and BuChE enzymes, and the information obtained from the 2D-QSAR analysis and molecular docking studies can be useful for the development of new analogues.

CRediT authorship contribution statement

Vinay Kumar: Methodology, Validation, Investigation, Writing - original draft. **Achintya Saha:** Data curation, Supervision. **Kunal Roy:** Conceptualization, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Financial assistance from the Indian Council of Medical Research (ICMR), New Delhi in the form of a senior research fellowship (File No: 5/3/8/27/ITR-F/2018-ITR; dated: 18.05.2018) to VK is thankfully acknowledged. KR thanks Science and Engineering Research Board (SERB), New Delhi for financial assistance under the MATRICS scheme (File number MTR/2019/000008).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.compbiolchem.2020.107355.

References

Ahmad, I., Fatima, I., 2008. Butyrylcholinesterase, lipoxygenase inhibiting and antifungal alkaloids from Isatistinctoria. J. Enzyme Inhib. Med. Chem. 23 (3), 313–316.
Ahmad, S., Iftikhar, F., Ullah, F., et al., 2016. Rational design and synthesis of dihydropyrimidine based dual binding site acetylcholinesterase inhibitors. Bioorg. Chem. 69, 91–101.
Ahmed, E., Nawaz, S.A., Malik, A., et al., 2006. Isolation and cholinesterase-inhibition studies of sterols from Haloxylonrecurvum. Bioorg. Med. Chem. Lett. 16 (3), 573–580.
Anand, P., Singh, B., 2012a. Synthesis and evaluation of novel 4-[(3H, 3aH, 6aH)-3-phenyl)-4, 6-dioxo-2-phenyldihydro-2H-pyrrolo [3, 4-d] isoxazol-5 (3H, 6H, 6aH)-yl] benzoic acid derivatives as potent acetylcholinesterase inhibitors and anti-amnestic agents. Bioorg. Med. Chem. 20 (1), 521–530.
Anand, P., Singh, B., Singh, N., 2012b. A review on coumarins as acetylcholinesterase inhibitors for Alzheimer's disease. Bioorg. Med. Chem. 20 (3), 1175–1180.
https://www.random.org/.
Andreani, A., Burnelli, S., Granaiola, M., et al., 2008. Chemiluminescent high-throughput microassay applied to imidazo [2, 1-b] thiazole derivatives as potential acetylcholinesterase and butyrylcholinesterase inhibitors. Eur. J. Med. Chem. 43 (3), 657–661.
Awasthi, M., Upadhyay, A.K., Singh, S., et al., 2018. Terpenoids as promising therapeutic molecules against Alzheimer's disease: amyloid beta-and acetylcholinesterase-directed pharmacokinetic and molecular docking analyses. Mol. Simul. 44 (1), 1–11.
Bacalhau, P., San Juan, A.A., Marques, C.S., et al., 2016. New cholinesterase inhibitors for Alzheimer's disease: structure activity studies (SARs) and molecular docking of isoquinolone and azepanone derivatives. Bioorg. Chem. 67, 1–8.
Bagheri, S.M., Khoobi, M., Nadri, H., et al., 2015. Synthesis and anticholinergic activity of 4-hydroxycoumarin derivatives containing substituted benzyl-1, 2, 3-triazole moiety. Chem. Bio. Drug Des. 86 (5), 1215–1220.
Barreiro, E.J., Camara, C.A., Verli, H., 2003. Design, synthesis, and pharmacological profile of novel fused pyrazolo [4, 3-d] pyridine and pyrazolo [3, 4-b][1, 8] naphthyridineisosteres: a new class of potent and selective acetylcholinesterase inhibitors. J. Med. Chem. 46 (7), 1144–1152.
Bolognesi, M.L., Andrisano, V., Bartolini, M., et al., 2005. Propidium-based polyamine ligands as potent inhibitors of acetylcholinesterase and acetylcholinesterase-induced amyloid-β aggregation. J. Med. Chem. 48 (1), 24–27.
Bukhari, S.N., Jantan, I., Masand, V.H., et al., 2014.) 'Synthesis of α, β-unsaturated carbonyl based compounds as acetylcholinesterase and butyrylcholinesterase inhibitors: characterization, molecular modeling, QSAR studies and effect against amyloid β-induced cytotoxicity'. Eur. J. Med. Chem. 83 (18), 355–365.
Buldurun, K., Turan, N., Bursal, E., et al., 2020. Synthesis, spectroscopic properties, crystal structures, antioxidant activities and enzyme inhibition determination of Co (II) and Fe (II) complexes of Schiff base. Res. Chem. Intermed. 46 (1), 283–297.
Camps, P., Formosa, X., Galdeano, C., et al., 2008. Novel donepezil-based inhibitors of acetyl-and butyrylcholinesterase and acetylcholinesterase-induced β-amyloid aggregation. J. Med. Chem. 51 (12), 3588–3598.
Cardoso, C.L., Castro-Gamboa, I., Silva, D.H., 2004. Indoleglucoalkaloids from Chimarrhisturbinata and their evaluation as antioxidant agents and acetylcholinesterase inhibitors. J. Nat. Prod. 67 (11), 1882–1885.
Carlier, P.R., Chow, E.S., Han, Y., 1999a. Heterodimerictacrine-based acetylcholinesterase inhibitors: investigating ligand− peripheral site interactions'. J. Med. Chem. 42 (20), 4225–4231.
Carlier, P.R., Du, D.M., Han, Y., et al., 1999b. Potent, easily synthesized huperzine A-tacrine hybrid acetylcholinesterase inhibitors. Bioorg. Med. Chem. Letts 9 (16), 2335–2338.
Ceschi, M.A., da Costa, J.S., Lopes, J.P., et al., 2016. Novel series of tacrine-tianeptine hybrids: synthesis, cholinesterase inhibitory activity, S100B secretion and a molecular modeling approach. Eur. J. Med. Chem. 121, 758–772.
ChemAxon - Software Solutions and Services for Chemistry & Biology. Available at https://chemaxon.com/products/marvin (accessed December 19, 2019).
Cho, J.K., Ryu, Y.B., Curtis-Long, M.J., 2017. Cholinestrase inhibitory effects of geranylated flavonoids from Paulownia tomentosa fruits. Bioorg. Med. Chem. 20 (8), 2595–2602.
Conejo-García, A., Pisani, L., del Carmen Núnez, M., et al., 2011. Homodimericbis-quaternary heterocyclic ammonium salts as potent acetyl-and butyrylcholinesterase inhibitors: a systematic investigation of the influence of linker and cationic heads over affinity and selectivity. J. Med. Chem. 54 (8), 2627–2645.
Contreras, J.M., Rival, Y.M., Chayer, S., et al., 1999. Aminopyridazines as acetylcholinesterase inhibitors'. J. Med. Chem. 42 (4), 730–741.
Czarnecka, K., Szymański, P., Girek, M., et al., 2017. Tetrahydroacridine derivatives with fluorobenzoic acid moiety as multifunctional agents for alzheimer's disease treatment. Bioorg. Chem. 72, 315–322.
da Costa, J.S., Lopes, J.P., Russowsky, D., et al., 2013. Synthesis of tacrine-lophine hybrids via one-pot four component reaction and biological evaluation as acetyl-and butyrylcholinesterase inhibitors. Eur. J. Med. Chem. 62, 556–563.
Das, S., Ojha, P.K., Roy, K., 2017a. Multilayered variable selection in QSPR: a case study of modeling melting point of bromide ionic liquids. Int. J. Quant. Struct. Prop. Rel. 2 (1), 106–124.
Das, S., Ojha, P.K., Roy, K., 2017b. Development of a temperature dependent 2D-QSPR model for viscosity of diverse functional ionic liquids. J. Mol 40, 454–467.
De Souza, S.D., De Souza, A.M., De Sousa, A.C., et al., 2012. Hologram QSAR models of 4-[(diethylamino) methyl]-phenol inhibitors of acetyl/butyrylcholinesterase enzymes as potential anti-Alzheimer agents. Molecules 17 (8), 9529–9539.
DeBernardis, J.F., Gifford, P., Rizk, M., 1988. Evaluation of the side arm of (naphthylvinyl) pyridinium inhibitors of choline acetyltransferase'. J. Med. Chem. 31 (1), 117–121.
Decker, M., 2006. Homobivalentquinazolinimines as novel nanomolar inhibitors of cholinesterases with dirigible selectivity toward butyrylcholinesterase. J. Med. Chem. 49 (18), 5411–5413.
Decker, M., Kraus, B., Heilmann, J., 2008. Design, synthesis and pharmacological evaluation of hybrid molecules out of quinazolinimines and lipoic acid lead to highly potent and selective butyrylcholinesterase inhibitors with antioxidant properties. Bioorg. Med. Chem. 16 (8), 4252–4261.
Discovery Studio Predictive Science Application | Dassault Systèmes BIOVIA. https://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/ (accessed December 19, 2019).
dos Santos, P.D., da Costa, J.S., Gamba, D., et al., 2010. Synthesis and AChE inhibitory activity of new chiral tetrahydroacridine analogues from terpeniccyclanones. Eur. J. Med. Chem. 45 (2), 526–535.
Du, X., Wang, X., Geng, M., 2018. Alzheimer's disease hypothesis and related therapies. Transl. Neurodegener. 7 (1), 2.
Ekins, S., Mestres, J., Testa, B., 2007. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. Br. J. Pharmacol. 152 (1), 9–20.
Ercan, S., Şenses, Y., 2020. Design and molecular docking studies of new inhibitor candidates for EBNA1 DNA binding site: a computational study. Mol. Simul. 7, 1–8.
Erlanson, D.A., McDowell, R.S., O'Brien, T., 2004. Fragment-based drug discovery. J. Med. Chem. 47 (14), 3463–3482.
Fang, L., Appenroth, D., Decker, M., et al., 2008a. NO-donating tacrine hybrid compounds improve scopolamine-induced cognition impairment and show less hepatotoxicity. J. Med. Chem. 51 (24), 7666–7669.
Fang, L., Kraus, B., Lehmann, J., et al., 2008b. Design and synthesis of tacrine–ferulic acid hybrids as multi-potent anti-Alzheimer drug candidates. Bioorg. Med. Chem. Lett. 18 (9), 2905–2909.
Feng, S., Xia, Y., Han, D., 2005a. Synthesis and acetylcholinesterase inhibition of derivatives of huperzine B. Bioorg. Med. Chem. Lett. 15 (3), 523–526.
Feng, S., Wang, Z., He, X., et al., 2005b. Bis-huperzine B: highly potent and selective acetylcholinesterase inhibitors. J. Med. Chem. 48 (3), 655–657.
Fink, D.M., Bores, G.M., Effland, R.C., et al., 1995. Synthesis and Evaluation of 5-Amino-5, 6, 7, 8-tetrahydroquinolinones as Potential Agents for the Treatment of Alzheimer's Disease. J. Med. Chem. 38 (18), 3645–3651.

Ganeshpurkar, A., Singh, R., Gore, P.G., et al., 2019. Structure-based screening and molecular dynamics simulation studies for the identification of potential acetylcholinesterase inhibitors. Mol. Simul. 30, 1–7.
Girisha, H.R., Chandra, J.N., Boppana, S., et al., 2009. Active site directed docking studies: synthesis and pharmacological evaluation of cis-2, 6-dimethyl piperidine sulfonamides as inhibitors of acetylcholinesterase. Eur. J. Med. Chem. 44 (10), 4057–4062.
Gray, A.P., Platz, R.D., Chang, T.C., et al., 1985. Synthesis of some quaternary ammonium alkylating agents and their effects on soman-inhibited acetylcholinesterase. J. Med. Chem. 28 (1), 111–116.
Gregor, V.E., Emmerling, M.R., Lee, C., et al., 1992. The synthesis and in vitro acetylcholinesterase and butyrylcholinesterase inhibitory activity of tacrine (Cognex®) derivaties. Bioorg. Med. Chem. Lett. 2 (8), 861–864.
Hameed, A., Zehra, S.T., Shah, S.J., et al., 2015. Syntheses, cholinesterases inhibition, and molecular docking studies of pyrido [2, 3-b] pyrazine derivatives. Chem. Bio. Drug Des. 86 (5), 1115–1120.
Hameed, A., Zehra, S.T., Abbas, S., et al., 2016. One-pot synthesis of tetrazole-1, 2, 5, 6-tetrahydronicotinonitriles and cholinesterase inhibition: probing the plausible reaction mechanism via computational studies. Bioorg. Chem. 65, 38–47.
Han, S.Y., Mayer, S.C., Schweiger, E.J., 1991. Synthesis and biological activity of galanthamine derivatives as acetylcholinesterase (AChE) inhibitors. Bioorg. Med. Chem. Lett. 1 (11), 579–580.
Hasan, A., Khan, K.M., Sher, M., et al., 2005. Synthesis and inhibitory potential towards acetylcholinesterase, butyrylcholinesterase and lipoxygenase of some variably substituted chalcones. J. Enzyme Inhib. Med. Chem. 20 (1), 41–47.
He, X.C., Feng, S., Wang, Z.F., 2007. Study on dual-site inhibitors of acetylcholinesterase: highly potent derivatives of bis-and bifunctionalhuperzine B. Bioorg. Med. Chem. 15 (3), 1394–1408.
Högenauer, K., Baumann, K., Enz, A., et al., 2001. Synthesis and acetylcholinesterase inhibition of 5-desamino huperzineA derivatives. Bioorg. Med. Chem. Lett. 11 (19), 2627–2630.
Hu, M.K., Wu, L.J., Hsiao, G., et al., 2002. Homodimerictacrine congeners as acetylcholinesterase inhibitors. J. Med. Chem. 45 (11), 2277–2282.
Hu, Y., Zhang, J., Chandrashankra, O., 2013. Design, synthesis and evaluation of novel heterodimers of donepezil and huperzine fragments as acetylcholinesterase inhibitors. Bioorg. Med. Chem. 21 (3), 676–683.
Huang, L., Luo, Z., He, F., et al., 2010. Berberine derivatives, with substituted amino groups linked at the 9-position, as inhibitors of acetylcholinesterase/butyrylcholinesterase. Bioorg. Med. Chem. Lett. 20 (22), 6649–6652.
Huang, W., Tang, L., Shi, Y., 2011. Searching for the multi-target-directed ligands against Alzheimer' s disease: discovery of quinoxaline-based hybrid compounds with AChE, H3R and BACE 1 inhibitory activities. Bioorg. Med. Chem. 19 (23), 7158–7167.
Ishihara, Y., Hirai, K., Miyamoto, M., 1994. Central cholinergic agents. 6. Synthesis and evaluation of 3-[1-(phenylmethyl)-4-piperidinyl]-1-(2, 3, 4, 5-tetrahydro-1H-1-benzazepin-8-yl)-1-propanones and their analogs as central selective acetylcholinesterase inhibitors. J. Med. Chem. 37 (15), 2292–2299.
Jaén, J.C., Gregor, V.E., Lee, C., et al., 1996. Acetylcholinesterase inhibition by fused dihydroquinazoline compounds. Bioorg. Med. Chem. Lett. 6 (6), 737–742.
Jia, P., Sheng, R., Zhang, J., 2009. Design, synthesis and evaluation of galanthamine derivatives as acetylcholinesterase inhibitors. Eur. J. Med. Chem. 44 (2), 772–784.
Jiang, H., Wang, X., Huang, L., et al., 2011. Benzenediol-berberine hybrids: multifunctional agents for alzheimer's disease. Bioorg. Med. Chem. 19 (23), 7228–7235.
Kanhed, A.M., Sinha, A., Machhi, J., et al., 2015. Discovery of isoalloxazine derivatives as a new class of potential anti-Alzheimer agents and their synthesis. Bioorg. Chem. 61, 7–12.
Kapples, K.J., Shutske, G.M., Bores, G.M., et al., 1993. Synthesis and in vitro acetylcholinesterase inhibitory activity of some 1-substituted analogues of velnacrine. Bioorg. Med. Chem. Lett. 3 (12), 2789–2792.
Kavitha, C.V., Gaonkar, S.L., Chandra, J.N., 2007. Synthesis and screening for acetylcholinesterase inhibitor activity of some novel 2-butyl-1, 3-diaza-spiro [4, 4] non-1-en-4-ones: derivatives of irbesartan key intermediate. Bioorg. Med. Chem. 5 (23), 7391–7398.
Khan, K., Roy, K., 2019. Ecotoxicological QSAR modelling of organic chemicals against Pseudokirchneriellasubcapitata using consensus predictions approach. SQER 30 (9), 665–681.
Komersova, A., Komers, K., Čegan, A., 2007. New findings about Ellman's method to determine cholinesterase activity. Z. Naturforsch 62 (1-2), 150–154.
Krátký, M., Štěpánková, Š., Vorčáková, K., et al., 2015. Salicylanilide diethyl phosphates as cholinesterases inhibitors. Bioorg. Chem. 58, 48–52.
Krátký, M., Štěpánková, Š., Vorčáková, K., et al., 2017. Synthesis of readily available fluorophenylalanine derivatives and investigation of their biological activity. Bioorg. Chem. 71, 244–256.
Kumar, S., 2015. Dual inhibition of acetylcholinesterase and butyrylcholinesterase enzymes by allicin. Indian J. Pharmacol. 47 (4), 444–446.
Kumar, V., De, P., Ojha, P.K., Saha, A., Roy, K., 2020a. A multi-layered variable selection strategy for QSAR modeling of butyrylcholinesterase inhibitors. Curr. Top. Med. Chem. 20 (18). https://doi.org/10.2174/1568026620666200616142753.
Kumar, V.; Ojha, P.K.; Saha, A.; Roy, K. Cheminformaticmodelling of β-amyloid aggregation inhibitory activity against Alzheimer's disease', Comput. Biol. Med. 13, pp. 103658.
Kumar, V., Ojha, P.K., Saha, A., et al., 2020c. Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease'. SQER 31 (2), 87–133.
Kurt, B.Z., Gazioglu, I., Sonmez, F., et al., 2015. Synthesis, antioxidant and anticholinesterase activities of novel coumarylthiazole derivatives. Bioorg. Chem. 59, 80–90.
Leader, H., Wolfe, A.D., Chiang, P.K., et al., 2002. Pyridophens: binary pyridostigmine – aprophenprodrugs with differential inhibition of acetylcholinesterase, butyrylcholinesterase, and muscarinic receptors. J. Med. Chem. 45 (4), 902–910.
Leng, J., Qin, H.L., Zhu, K., et al., 2016. Evaluation of multifunctional synthetic tetralone derivatives for treatment of alzheimer's disease. Chem. Bio. Drug Des 88 (6), 889–898.
Leonard, J.T., Roy, K., 2000. On selection of training and test sets for the development of predictive QSAR models. Qsar & Combinat. Sci. Mar. 25 (3), 235–251.
Li, Z., Wang, B., Hou, J.Q., et al., 2013. 2-(2-indolyl-)-4 (3 H)-quinazolinesderivates as new inhibitors of AChE: design, synthesis, biological evaluation and molecular modelling. J. Enzyme Inhib. Med. Chem. 28 (3), 583–592.
Li, Y., Peng, P., Tang, L., 2014. Design, synthesis and evaluation of rivastigmine and curcumin hybrids as site-activated multitarget-directed ligands for Alzheimer's disease therapy. Bioorg. Med. Chem. 22 (17), 4717–4725.
Li, X., Wang, H., Lu, Z., 2016. Development of multifunctional pyrimidinylthiourea derivatives as potential anti-Alzheimer agents'. J. Med. Chem. 59 (18), 8326–8344.
Li, Y., Qiang, X., Luo, L., et al., 2017a. Multitarget drug design strategy against Alzheimer's disease: HomoisoflavonoidMannich base derivatives serve as acetylcholinesterase and monoamine oxidase B dual inhibitors with multifunctional properties. Bioorg. Med. Chem. 25 (2), 714–726 2017.
Li, Y., Qiang, X., Luo, L., et al., 2017b. AuroneMannich base derivatives as promising multifunctional agents with acetylcholinesterase inhibition, anti-β-amyloid aggregation and neuroprotective properties for the treatment of alzheimer's disease. Eur. J. Med. Chem. 126, 762–775.
Lin, G., Chen, G.H., Ho, H.C., 1998. Conformationally restricted carbamate inhibitors of horse serum butyrylcholinesterase. Bioorg. Med. Chem. Lett. 8 (19), 2747–2750.
Liston, D.R., Nielsen, J.A., Villalobos, A., et al., 2004. Pharmacology of selective acetylcholinesterase inhibitors: implications for use in Alzheimer's disease'. Eur. J. Pharmacol. 486 (1), 9–17.
Liu, H.R., Huang, X.Q., Lou, D.H., et al., 2014. Synthesis and acetylcholinesterase inhibitory activity of Mannich base derivatives flavokawain B. Bioorg. Med. Chem. Lett. 24 (19), 4749–4753.
Luo, W., Li, Y.P., He, Y., et al., 2011. Design, synthesis and evaluation of novel tacrine-multialkoxybenzene hybrids as dual inhibitors for cholinesterases and amyloid beta aggregation. Bioorg. Med. Chem. 19 (2), 763–770.
Mauri, A., Consonni, V., Pavan, M., et al., 2006. Dragon software: an easy approach to molecular descriptor calculations. Match 56 (2), 237–248.
McKenna, M., Proctor, G.R., Young, L.C., et al., 1997. Novel tacrine analogues for potential use against Alzheimer's disease: potent and selective acetylcholinesterase inhibitors and 5-HT uptake inhibitors. J. Med. Chem. 40 (22), 3516–3523.
Minitab IN, 2000. MINITAB Statistical Software, Minitab Release 13. Available at. http://www.minitab.com/en-us/products/minitab/.
Mohamed, T., Zhao, X., Habib, L.K., et al., 2011. Design, synthesis and structure–activity relationship (SAR) studies of 2, 4-disubstituted pyrimidine derivatives: dual activity as cholinesterase and Aβ-aggregation inhibitors. Bioorg. Med. Chem. 19 (7), 2269–2281.
Mohammadi, K.M., Mahdavi, M., Saeedi, M., et al., 2015. Design, Synthesis, Biological Evaluation, and Docking Study of Acetylcholinesterase Inhibitors: New Acridone-1, 2, 4-oxadiazole-1, 2, 3-triazole Hybrids. Chem. Bio. Drug Des. 86 (6), 1425–1432.
Morini, G., Comini, M., Rivara, M., et al., 2008. Synthesis and structure–activity relationships for biphenyl H3 receptor antagonists with moderate anti-cholinesterase activity. Bioorg. Med. Chem. 16 (23), 9911–9924.
Mouchlis, V.D., Melagraki, G., Zacharia, L.C., et al., 2020. Computer-aided drug design of β-Secretase, γ-Secretase and anti-tau inhibitors for the discovery of novel alzheimer's therapeutics. Int. J. Mol. Sci. 21 (3), 703.
Murzin, A.G., Conte, L., Andreeva, A., et al., 1995. Scop: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247 (536), 40.
Najafi, Z., Saeedi, M., Mahdavi, M., et al., 2016. Design and synthesis of novel anti-alzheimer's agents: acridine-chromenone and quinoline-chromenone hybrids. Bioorg. Chem. 67, 84–94.
Nekoei, M., Salimi, M., Dolatabadi, M., et al., 2011. A quantitative structure-activity relationship study of tetrabutylphosphonium bromide analogs as muscarinic acetylcholine receptors. J. Serb. Chem. Soc. 76 (8), 1117–1127.
Nordberg, A., Svensson, A.L., 1998. Cholinesterase inhibitors in the treatment of Alzheimer's disease. Drug Saf. 19 (6), 465–480.
Orhan, I., Şener, B., Choudhary, M.I., et al., 2004. Acetylcholinesterase and butyrylcholinesterase inhibitory activity of some Turkish medicinal plants. J. Ethnopharmacol. 91 (1), 57–60.
Pal, S., Kumar, V., Kundu, B., et al., 2019. Ligand-based pharmacophore modeling, virtual screening and molecular docking studies for discovery of potential topoisomerase I inhibitors. CSBJ. 17, 291–310.
Pang, X., Fu, H., Yang, S., et al., 2017. Evaluation of novel dual acetyl-and butyrylcholinesterase inhibitors as potential Anti-Alzheimer's disease agents using pharmacophore, 3D-QSAR, and molecular docking approaches. Molecules 22 (8), 1254.
Pool, W.F., Woolf, T.F., Reily, M.D., et al., 1996. Identification of a 3-hydroxylated tacrine metabolite in rat and man: metabolic profiling implications and pharmacology. J. Med. Chem. 39 (15), 3014–3018.
Pouramiri, B., Moghimi, S., Mahdavi, M., et al., 2017. Synthesis and anticholinesterase activity of new substituted benzooxazole-based derivatives. Chem. Bio. Drug Des. 89 (5), 783–789.
Rodríguez, Y.A., Gutiérrez, M., Ramírez, D., et al., 2016. Novel N-allyl/propargyltetrahydroquinolines: Synthesis via three-component cationic iminodiels–alder reaction, binding prediction, and evaluation as cholinesterase inhibitors. Chem. Bio. Drug Des. 88 (4), 498–510.
Rodríguez-Franco, M.I., Fernández-Bachiller, M.I., Pérez, C., et al., 2006. Novel tacrine –

melatonin hybrids as dual-acting drugs for Alzheimer disease, with improved acetylcholinesterase inhibitory and antioxidant properties. J. Med. Chem. 49 (2), 459–462.
Roy, K., Das, R.N., 2017. The “ETA” indices in QSAR/QSPR/QSTR research. Pharmaceutical Sciences: Breakthroughs in Research and Practice (978-1011). IGI Global.
Roy, J., Ghosh, S., Ojha, P.K., et al., 2019. Predictive quantitative structure–property relationship (QSPR) modeling for adsorption of organic pollutants by carbon nanotubes (CNTs). Environ. Sci. (Ruse) 6 (1), 224–247.
Rydberg, E.H., Brumshtein, B., Greenblatt, H.M., et al., 2006. Complexes of Alkylene-linked Tacrine dimers with Torpedo c alifornicaacetylcholinesterase: binding of Bis (5)-tacrine produces a dramatic rearrangement in the active-site gorge. J. Med. Chem. 49 (18), 5491–5500.
Sadashiva, C.T., Chandra, J.N., Ponnappa, K.C., 2006. Synthesis and efficacy of 1-[bis (4-fluorophenyl)-methyl] piperazine derivatives for acetylcholinesterase inhibition, as a stimulant of central cholinergic neurotransmission in Alzheimer’s disease. Bioorg. Med. Chem. Lett. 16 (15), 3932–3936.
Saeed, A., Zaib, S., Ashraf, S., et al., 2015. Synthesis, cholinesterase inhibition and molecular modelling studies of coumarin linked thiourea derivatives. Bioorg. Chem. 63, 58–63.
Saeedi, M., Safavi, M., Karimpour, R.E., et al., 2017. Synthesis of novel chromenones linked to 1, 2, 3-triazole ring system: investigation of biological activities against alzheimer’s disease. Bioorg. Chem. 70, 86–93.
Samadi, A.’, Marco, C.J., Soriano, E., et al., 2010. Multipotent drugs with cholinergic and neuroprotective properties for the treatment of alzheimer and neuronal vascular diseases. I. Synthesis, biological assessment, and molecular modeling of simple and readily available 2-aminopyridine-, and 2-chloropyridine-3, 5-dicarbonitriles. Bioorg. Med. Chem. 18 (16), 861–872.
Samadi, A., de los Ríos, C., Bolea, I., et al., 2012. Multipotent MAO and cholinesterase inhibitors for the treatment of alzheimer’s disease: synthesis, pharmacological analysis and molecular modeling of heterocyclic substituted alkyl and cycloalkylpropargyl amine. Eur. J. Med. Chem. 52, 251–262.
Sang, Z., Qiang, X.’, Li, Y., et al., 2015a. Design, synthesis and evaluation of scutellarein-O-alkylamines as multifunctional agents for the treatment of Alzheimer’s disease. Eur. J. Med. Chem. 94, 348–366.
Sang, Z., Li, Y., Qiang, X., et al., 2015b. Multifunctional scutellarin–rivastigmine hybrids with cholinergic, antioxidant, biometal chelating and neuroprotective properties for the treatment of Alzheimer’s disease. Bioorg. Med. Chem. 23 (4), 668–680.
Sang, Z.P., Qiang, X.M., Li, Y., et al., 2015c. Design, synthesis, and biological evaluation of scutellareincarbamate derivatives as potential multifunctional agents for the treatment of alzheimer’s disease. Chem. Bio. Drug Des. 86 (5), 1168–1177.
Sarfraz, M., Sultana, N., Rashid, U., et al., 2017. Synthesis, biological evaluation and docking studies of 2, 3-dihydroquinazolin-4 (1H)-one derivatives as inhibitors of cholinesterases. Bioorg. Chem. 70, 237–244.
Schelterns, P., Feldman, H., 2003. Treatment of Alzheimer’s disease; current status and new perspectives. Lancet Neurol. 2 (9), 539–547.
Shao, D., Zou, C., Luo, C., et al., 2004. Synthesis and evaluation of tacrine–E2020 hybrids as acetylcholinesterase inhibitors for the treatment of alzheimer’s disease. Bioorg. Med. Chem. Lett. 14 (18), 4639–4642.
Sharma, M.C.; Sharma, S.; Sahu, N.K. et al. ‘QSAR studies of some substituted imidazolinones angiotensin II receptor antagonists using Partial Least Squares Regression (PLSR) method based feature selection’, J. Saudi Chem. Soc. 17(2), pp. 219-225.
Shen, Y., Sheng, R., Zhang, J., 2008. 2-Phenoxy-indan-1-one derivatives as acetylcholinesterase inhibitors: a study on the importance of modifications at the side chain on the activity. Bioorg. Med. Chem. 16 (16), 7646–7653.
Sheng, R., Lin, X., Li, J., 2005. Design, synthesis, and evaluation of 2-phenoxy-indan-1-one derivatives as acetylcholinesterase inhibitors. Bioorg. Med. Chem. Lett. 15 (17), 3834–3837.
Sheng, R., Xu, Y., Hu, C., 2009a. Design, synthesis and AChE inhibitory activity of indanone and aurone derivatives. Eur. J. Med. Chem. 44 (1), 7–17.
Sheng, R., Lin, X., Zhang, J., et al., 2009b. Design, synthesis and evaluation of flavonoid derivatives as potent AChE inhibitors. Bioorg. Med. Chem. 17 (18), 6692–6698.
Shi, A., Huang, L., Lu, C., et al., 2011. Synthesis, biological evaluation and molecular modeling of novel triazole-containing berberine derivatives as acetylcholinesterase and β-amyloid aggregation inhibitors. Bioorg. Med. Chem. 19 (7), 2298–2305.
Shi, D.H., Huang, W., Li, C., et al., 2013. Synthesis, biological evaluation and molecular modeling of aloe-emodin derivatives as new acetylcholinesterase inhibitors. Bioorg. Med. Chem. 21 (5), 1064–1073.
Shi, J., Tu, W., Luo, M., et al., 2017. Molecular docking and molecular dynamics simulation approaches for identifying new lead compounds as potential AChE inhibitors. Mol. Simul. 43 (2), 102–109.
Shinada, M., Narumi, F., Osada, Y., 2012. Synthesis of phenserine analogues and evaluation of their cholinesterase inhibitory activities. Bioorg. Med. Chem. 20 (16), 4901–4914.
Shrivastava, S.K., Sinha, S.K., Srivastava, P., et al., 2019. Design and development of novel p-aminobenzoic acid derivatives as potential cholinesterase inhibitors for the treatment of Alzheimer’s disease. Bioorg. Chem. 82 (1), 11–23.
Shutske, G.M., Pierrat, F.A., Kapples, K.J., 1989. 9-Amino-1, 2, 3, 4-tetrahydroacridin-1-ols. Synthesis and evaluation as potential Alzheimer’s disease therapeutics’. J. Med. Chem. 32 (8), 1805–1813.
SIMCA | Umetrics. https://umetrics.com/products/simca (accessed December 19, 2019).
Simoni, E.’, Daniele, S., Bottegoni, G., et al., 2012. Combining galantamine and memantine in multitargeted, new chemical entities potentially useful in Alzheimer’s disease. J. Med. Chem. 55 (22), 9708–9721.
Singh, N.; Dalal, V.; Kumar, P. ‘Molecular docking and simulation analysis for elucidation of toxic effects of dicyclohexyl phthalate (DCHP) in glucocorticoid receptor-mediated adipogenesis’, Mol. Simul. 46(1), pp. 9-21.
Skrzypek, A., Matysiak, J., Karpińska, M.M., et al., 2013. Synthesis and anticholinesterase activities of novel 1, 3, 4-thiadiazole based compounds. J. Enzyme Inhib. Med. Chem. 28 (4), 816–823.
Sterling, J., Herzig, Y., Goren, T., et al., 2002. Novel dual inhibitors of AChE and MAO derived from hydroxyaminoindan and phenethylamine as potential treatment for alzheimer’s disease. J. Med. Chem. 45 (24), 5260–5279.
Sugimoto, H., Iimura, Y., Yamanishi, Y., et al., 1992. Synthesis and anti-acetylcholinesterase activity of 1-benzyl-4-[(5, 6-dimethoxy-1-indanon-2-yl) methyl] piperidine hydrochloride (E2020) and related compounds. Bioorg. Med. Chem. Lett. 2 (8), 871–876.
Sugimoto, H., Iimura, Y., Yamanishi, Y., et al., 1995. Synthesis and structure-activity relationships of acetylcholinesterase inhibitors: 1-benzyl-4-[(5, 6-dimethoxy-1-oxoindan-2-yl) methyl] piperidine hydrochloride and related compounds. J. Med. Chem. 38 (24), 4821–4829.
Szymański, P., Markowicz, M., Mikiciuk-Olasik, E., 2011. Synthesis and biological activity of derivatives of tetrahydroacridine as acetylcholinesterase inhibitors. Bioorg. Chem. 39 (4), 138–142.
Tang, H., Ning, F.X., Wei, Y.B., et al., 2007. Derivatives of oxoisoaporphine alkaloids: a novel class of selective acetylcholinesterase inhibitors. Bioorg. Med. Chem. Lett. 17 (13), 3765–3768.
Taslimi, P., Turhan, K., Türkan, F., et al., 2020. Cholinesterases, α-glycosidase, and carbonic anhydrase inhibition properties of 1H-pyrazolo [1, 2-b] phthalazine-5, 10-dione derivatives: synthetic analogues for the treatment of Alzheimer’s disease and diabetes mellitus. Bioorg. Chem. 97, 103647.
Todeschini, R. and Consonni, V. Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices. 2009; references, Vol. 41, John Wiley & Sons, Weinheim (Germany).
Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. Mol. Inform. 29 (6-7), 476–488.
Türkan, F., 2019. Investigation of the toxicological and inhibitory effects of some benzimidazole agents on acetylcholinesterase and butyrylcholinesterase enzymes. Arch. Physiol. Biochem. 27, 1–5.
Valenti, P., Rampa, A., Bisi, A., et al., 1997. Acetylcholinesterase inhibition by tacrine analogues. Bioorg. Med. Chem. Lett. 7 (20), 2599–2602.
Veerasamy, R., Rajak, H., Jain, A., et al., 2011. Validation of QSAR models-strategies and importance. Int. J. Drug Des. Discov. 2 (3), 511–519.
Vidaluc, J.L., Calmel, F., Bigg, D., et al., 1994. Novel [2-(4-piperidinyl) ethyl](thio) ureas: synthesis and antiacetylcholinesterase activity. J. Med. Chem. 37 (5), 689–695.
Vidaluc, J.L., Calmel, F., Bigg, D.C., et al., 1995. Flexible 1-[(2-aminoethoxy) alkyl]-3-ar (o) yl (thio) ureas as novel acetylcholinesterase inhibitors. Synthesis and biochemical evaluation. J. Med. Chem. 38 (15), 2969–2973.
Villalobos, A., Blake, J.F., Biggers, C.K., et al., 1994. Novel benzisoxazole derivatives as potent and selective inhibitors of acetylcholinesterase. J. Med. Chem. 37 (17), 2721–2734.
Wen, H., Zhou, Y., Lin, C., et al., 2007. Methyl 2-(2-(4-formylphenoxy) acetamido)-2-substituted acetate derivatives: a new class of acetylcholinesterase inhibitors. Bioorg. Med. Chem. Lett. 17 (8), 2123–2125.
Wong, D.M., Greenblatt, H.M., Dvir, H., 2003. Acetylcholinesterasecomplexed with bivalent ligands related to huperzine A: experimental evidence for species-dependent protein− ligand complementarity. ACS 125 (2), 363–373.
Yang, X., Qiang, X., Li, Y., 2017. Pyridoxine-resveratrol hybrids Mannich base derivatives as novel dual inhibitors of AChE and MAO-B with antioxidant and metal-chelating properties for the treatment of alzheimer’s disease. Bioorg. Chem. 71, 305–314.
Yanovsky, I., Finkin, G.E., Zaikin, A., et al., 2012. Carbamate derivatives of indolines as cholinesterase inhibitors and antioxidants for the treatment of alzheimer’s disease. J. Med. Chem. 55 (23), 10700–10715.
Yap, C.W., 2011. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 32 (7), 1466–1474.
Yurttaş, L., Kaplancıklı, Z.A., Özkay, Y., 2013. Design, synthesis and evaluation of new thiazole-piperazines as acetylcholinesterase inhibitors. J. Enzyme Inhib. Med. Chem. 28 (5), 1040–1047.
Zakhari, J.S., Kinoyama, I., Hixon, M.S., et al., 2011. Formulating a new basis for the treatment against botulinum neurotoxin intoxication: 3, 4-Diaminopyridine prodrug design and characterization. Bioorg. Med. Chem. 19 (21), 6203–6209.
Zelík, P., Lukešová, A., Čejka, J., et al., 2010. Nostotrebin 6, a bis (cyclopentenedione) with cholinesterase inhibitory activity isolated from Nostoc sp. Str. Lukešová 27/97. J. Enzyme Inhib. 25 (3), 414–420.
Zeng, F., Jiang, H., Zhai, Y., et al., 1999. Synthesis and acetylcholinesterase inhibitory activity of huperzine A—E2020 combined compound. Bioorg. Med. Chem. Lett. 9 (23), 3279–3284.
Zhan, Z.J., Bian, H.L., Wang, J.W., et al., 2010. Synthesis of physostigmine analogues and evaluation of their anticholinesterase activities. Bioorg. Med. Chem. Lett. 20 (5), 1532–1534.
Zheng, H., Youdim, M.B., Fridkin, M., 2010. Site-activated chelators targeting acetylcholinesterase and monoamine oxidase for alzheimer’s therapy. ACS Chem. Bio. 5 (6), 603–610.
Zhu, Y., Xiao, K., Ma, L., et al., 2009. Design, synthesis and biological evaluation of novel dual inhibitors of acetylcholinesterase and β-secretase. Bioorg. Med. Chem. 17 (4), 1600–1613.

Chemometrics and Intelligent Laboratory Systems 233 (2023) 104734

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease

Vinay Kumar [a], Achintya Saha [b], Kunal Roy [a,*]

[a] Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, 700032, India
[b] Department of Chemical Technology, University of Calcutta, 92 A P C Road, Kolkata, 700 032, India

ARTICLE INFO

Keywords:
Alzheimer's disease
2D-QSAR
Read-across
QSAAR
AChE
BuChE
BACE1

ABSTRACT

Alzheimer's disease (AD) is an age-related neurodegenerative disorder, which is the most common cause of dementia in elderly individuals. It is characterized by selective neuronal cell death that affects the brain area related to memory and learning. So far, various computational research targeting AD have been reported, but we are still far from finding a precise treatment strategy for AD. It appeared of interest to us to carry out a two-dimensional quantitative structure-activity relationship (2D-QSAR) analysis against multiple targets of AD using large datasets to determine the essential structural features which are responsible for the inhibition of the enzymes/targets. In the present research, we have implemented 2D-QSAR modeling against twelve major targets (AChE, BuChE, BACE1, β-amyloid, 5-HT6, CDK-5, Gamma-secretase, Glutaminyl Cyclase, GSK-3β, MAO-B, NMDA and Phosphodiester (PDE10A) enzymes) of AD for the identifications of novel multitarget inhibitors. The models were used to check the applicability domain of a pool of ~19 million compounds obtained from the four chemical drug-like databases (ZINC[12], Asinex, NCI, and InterBioscreen databases) and provided prioritized compounds for experimental detection of their performance as anti-Alzheimer's drug. Additionally, we have also developed the quantitative structure activity-activity relationship (QSAAR) and selectivity-based models to explore the most important features contributing to the dual inhibition against the respective targets. Furthermore, we have also performed chemical Read-Across predictions using the Read-Across-v3.1 tool (https://dtclab.webs.com/software-tools), the results for the external validation metrics were found to be better than the 2D-QSAR-derived predictions. Furthermore, molecular docking experiments have been performed to understand the molecular interactions between ligands and enzymes at the atomic level, and the observations are compared with the structural features acquired from QSAR models that justified the mechanistic aspect of binding phenomena. The proposed models and read-across hypotheses could be used as potential tools to identify essential molecular features for designing suitable drug(s) for Alzheimer's therapy using rational design of multi-target inhibitors.

1. Introduction

Alzheimer's disease (AD) is a progressive neuropathological disorder, found in the most common form of dementia, which causes severe brain deterioration and cognitive function loss [1,2]. AD is a degenerative ailment that is thought to begin decades before symptoms appear. Clinicians are only able to identify even the first signs of AD after significant damage has already been done to crucial biological components [2,3]. Despite the challenge of researchers to definitively identify the initial trigger that leads apart a series of harmful processes, a lot of studies have identified essential components in AD pathogenesis [1–3]. According to evidences from autosomal dominant and sporadic types of

Abbreviations: 2d-QSAR, Two-dimensional quantitative structure-activity relationship; QSAAR, Quantitative structure activity-activity relationship; OECD, Organization for Economic Co-operation and Development; DCV, Double cross-validation; MLR, Multiple linear regression; PLS, Partial least squares; DModX, Distance to model in the X-space; AD, Alzheimer's disease; AChE, Acetylcholinesterase; BuChE, Butyrylcholinesterase; BACE1, Beta-secretase 1; 5HT6, 5-hydroxytryptamine 6; CDK-5, Cyclin-dependent kinase 5; GSK-3β, Glycogen synthase kinase-3β; QC, Glutaminyl Cyclase; MAO-B, Monoamine oxidase B; NMDA, N-methyl-D-aspartate; PDE 10A, Phosphodiesterase10A; NME, Novel molecular entity; FDA, Food and Drug Administration; IBS, InterBioscreen; LK, Laplacian kernel; GK, Gaussian kernel; ED, Euclidean distance.

* Corresponding author.
E-mail address: kunal.roy@jadavpuruniversity.in (K. Roy).

https://doi.org/10.1016/j.chemolab.2022.104734
Received 22 August 2022; Received in revised form 6 December 2022; Accepted 17 December 2022
Available online 22 December 2022

AD, amyloid plaques and tau protein-based neurofibrillary tangles can develop for up to 20 years before the onset of clinical dementia [2,4]. The staging of AD pathological abnormalities during the preclinical stage of the disease is facilitated by the recent growth of imaging and fluid biomarkers for AD pathogenesis [2,4]. AD is a developing healthcare concern, with increased life expectancy as the primary risk factor [2,5]. Disease prevalence is expected to more than double over the next several decades in the absence of adequate prevention and treatment alternatives [3]. According to the World Alzheimer Report 2021, there were around 55 million people worldwide living with dementia and are expected to exceed 78 million in 2030 and 139 million in 2050 worldwide [2]. Long-term care for affected individuals involves a significant economic burden in addition to its direct impact on human health and welfare [2,4]. About 200 clinical studies have been conducted to date to identify disease-modifying treatments for AD, but these efforts have generally failed, with many failures being attributable to ineffectiveness or excessive toxicity [6,7]. Every failed clinical study of a novel molecular entity (NME) takes a significant amount of time and money. Repurposing medications that have already been approved by the Food and Drug Administration (FDA) for a different indication, however, is less expensive, involves known potential toxicities, and has a greater success rate (30%) than developing an NME [6,7]. Significant effort has been devoted in recent years [4,8–12] to identify therapies that halt neurodegeneration in AD, but we are still far from finding exact treatment techniques [7]. The early diagnosis and treatment of AD is now a fast-developing field of both scientific and clinical research because current treatments only help with the symptoms of the disease. There are now only five approved drugs for the treatment of cognitive symptoms of Alzheimer's disease. Among them, four drugs are acetylcholinesterase enzyme (AChE) inhibitors (Tacrine, Rivastigmine, Galantamine, and Donepezil), and the remaining one drug is non-competitive glutamate (NMDA) receptor antagonist (Memantine "FDA approved") [2,4,6]. Their use is only symptomatic, and no treatment has been proven to slow or stop the progression of the disease [6]. The long-term effects of AChE inhibitors have recently been postulated to be due to these medications interfering with the metabolism of amyloid precursor protein (APP) [5]. The cause and progression of AD are still not well understood. The search for treatments in the field of neurodegenerative disorders is extremely active, yet there is still no cure for AD. Therefore, there is an urgent need for improving current treatment strategies, for example by increasing the selectivity toward targets (e.g., dual binding site AChE inhibitors), exploring other targets for finding novel leads, and finding potential scaffolds acting on more than one target (multitarget strategy). In the past few decades, other than AChE and NMDA receptor, many more potential targets have been identified and investigated such as butyrylcholinesterase (BuChE) enzyme, beta-secretase 1 (BACE1) enzyme, β-amyloid aggregation, 5-hydroxytryptamine 6 (5-HT6), Cyclin-Dependent Kinase 5 (CDK-5), gamma-secretase enzyme, Glutaminyl Cyclase enzyme (QC), glycogen synthase kinase-3β (GSK-3β) enzyme, monoamine oxidase B (MAO-B) enzyme and phosphodiesterase enzyme 10A (PDE 10A); therefore, the multi-target-directed (MTDs) is still the promising treatment method of AD [13,14]. In this paradigm, drug design and discovery have shifted from the molecular and cellular to the systems-biology level to reflect minor processes in biological networks that lead to disease. MTDs provide various advantages over single-target medications, such as increased efficacy due to synergistic or additive effects, improved target tissue distribution, accelerated therapeutic efficacy in terms of clinical onset and maximal effect, predictable pharmacokinetic profile and less drug-drug interactions, reduced risk of toxicity, improved bioavailability, and acceptance, and lesser likelihood of target-based drug resistance due to modulation of a few targets. However, designing potent MTDs is not easy, with issues ranging from proper target selection to affinity balancing to avoiding affinity to related off-targets.

In this regard, computational approaches such as quantitative structure-activity relationship (QSAR), chemical Read-Across, pharmacophore modeling, molecular docking, Molecular Dynamics (MD) Simulations, etc. are playing imperative roles in the design and discovery of new compounds with enhanced therapeutic activity [8–12]. For decades, chemoinformatics and molecular modeling approaches have been utilized to identify and optimize novel compounds with improved therapeutic potential in various fields [8–12]. Currently, *in silico* modeling is a part of the conventional drug discovery process, and such methods are usually employed in the search for novel drugs or the optimization of the therapeutic activity of a chemical series at the initial phases of drug development [8–12]. In the current research, we have developed 2D-QSAR models against twelve major targets namely, 5-HT6, AChE enzyme, BuChE enzyme, BACE1 enzyme, β-amyloid aggregation, CDK-5 protein, Gamma-secretase enzyme, Glutaminyl Cyclase (QCs) enzyme, GSK-3β enzyme, MAO-B enzyme, NMDA receptor, and PDE 10A enzyme. Subsequently, the validated models were used to predict the multiple drugs like databases such as the Asinex database (338604 compounds) (available from https://www.asinex.com/screening-libraries-(all-libraries), InterBioscreen (IBS) database (552793 compounds) (available from https://www.ibscreen.com/), NCI Open Database (265242 compounds) (available from https://cactus.nci.nih.gov/download/nci/), and Zinc[12] Database (17900742 compounds) (available from https://zinc12.docking.org/subsets/drug-like) for the search of novel inhibitors with multitarget inhibitory activity. Furthermore, we have developed the quantitative structure activity-activity relationship (QSAAR) and selectivity-based models to determine structural fragments/properties that are precisely necessary to inhibit a specific type of enzyme. Before developing the final models, a multilayered variable section technique was used to choose relevant descriptors for bioactivity. For several internal and external validation metrics, the validated models showed acceptable results. Additionally, the models were developed following OECD (Organization for Economic Co-operation and Development) standards (https://www.oecd.org/env/ehs/risk-assessment/validationofqsarmodels.htm).The developed models emphasize the structural requirements or molecular properties essential for the invention of safer dual inhibitors. Moreover, we have also performed chemical Read-Across predictions using the Read-Across-v3.1 tool (https://dtclab.webs.com/software-tools). In addition, we also executed molecular docking analyses with the most and least active molecules from the datasets, attempting to explain the influences of different properties as seen in the 2D-QSAR models. The study identified structural features in small molecules, which provide dual enzyme inhibitory effects, as well as their relationship to ligand-receptor interactions, which will assist in the design of next-generation Alzheimer's treatments.

2. Materials and methods

2.1. 2D-QSAR modeling

2.1.1. Data collection, curation, and dataset preparation

The activity data against twelve major targets of AD were collected from the BindingDB database [15] (available from www.bindingdb.org). Initially, 80 inhibitors against 5-hydroxytryptamine receptor 6 (5-HT6) following cell-based Radio ligand binding assay, 1733 compounds against acetylcholinesterase (AChE) enzyme following modified colorimetric Ellman assay, 2507 compounds against butyrylcholinesterase (BuChE) enzyme following modified colorimetric Ellman assay, 905 inhibitors against beta-secretase 1 (BACE1) enzyme following FRET (fluorescence resonance energy transfer) assay, 262 β-amyloid aggregation inhibitors following Thioflavin T-based fluorometric assay, 225 compounds against Cyclin Dependent Kinase 5 (CDK-5) protein following Scintillation proximity assay, 217 inhibitors against gamma-secretase enzyme following cell-based sandwich ELISA assay, 132 compounds against Glutaminyl Cyclase (QCs) enzyme following Continuous Spectrometric Assay, 159 inhibitors against glycogen synthase kinase-3 beta (GSK-3β) enzyme following Kinase-Glo

reagent based luminescence assay, 170 compounds against Monoamine oxidase B (MAO-B) enzyme following Fluorometric method, 356 compounds against N-methyl-D-aspartate (NMDA) receptor following Fluorescence-based assay, 289 compounds against Phosphodiesterase 10A (PDE 10A) enzyme following TR-FRET assay were collected from the BindingDB database [15] (available from www.bindingdb.org) (see supplementary information SI-1, sheet 1–12 sheet). The datasets comprise diverse classes of heterocyclic compounds, and the experimental activity values are quantified in IC_{50} (nM). Before proceeding with the development of the regression models, we executed preliminary dataset preparation and data curation (chemical and biological) strategy using a KNIME workflow (available from https://dtclab.webs.com/software-tools) following the protocol as discussed by Kumar et al. [16]. The precision of the KNIME workflow was confirmed by Mariana et al., 2017 [17], Domenico et al., 2018 [18], and Fabian P et al., 2015 [19]. After dataset curation, screening of the activity datasets was performed to find the common compounds having dual inhibitory activity against the listed targets. Accordingly, we have found that the 43 compounds with dual inhibitory activities both against AChE and BACE1 enzymes, 83 compounds against AChE and β-amyloid, 113 compounds against AChE and BuChE enzymes, 52 compounds against AChE and MAO-B enzymes, 20 compounds against BACE1 and GSK-3β enzymes, 51 compounds against BuChE and BACE1 enzymes, 23 compounds against BuChE and β-amyloid, 48 compounds against BuChE and MAO-B enzymes, 21 compounds against AChE and GSK-3β enzymes and 21 compounds against BuChE and GSK-3β enzymes were retained and used for the development of the respective QSAAR and selectively based models. Marvin Sketch software version 5.5.0.1 (available from https://chemaxon.com) was used to draw the chemical structures of all compounds, followed by the addition of explicit hydrogens in the structures. The activity end point values (IC_{50}) were converted to the negative logarithmic scale, pIC_{50}, as customary in QSAR modeling.

2.1.2. Computation of the molecular descriptors and data pretreatment

In this section, we have calculated only 2D descriptors using software, namely the alvaDesc (v2.0.12) tool (available from https://www.alvascience.com/alvaDesc/) covering atom-type E-state indices, 2D Atom Pairs, 2D autocorrelations, 2D matrix-based descriptors, atom-centred fragments, 2D Autocorrelation, connectivity indices, constitutional indices, ETA indices, functional group counts, information indices, MDE descriptors, molecular properties, P_VSA-like descriptors Rotatable Bonds Count Descriptor, Pharmacophore descriptors, Ring descriptors, Rule of Five Descriptor and Topological indices. After descriptor calculation, we have executed data pretreatment employing the tool Pretreatment V-WSP version 1.2 (available from http://dtclab.webs.com/software-tools) to remove the descriptors with missing or near constant values.

2.1.3. Dataset division

After data pretreatment, all the datasets were divided into training and test sets. In this work, the division of the data sets was implemented following three different dataset division methods, namely, activity-property, Euclidean distance based, and modified k-medoid clustering techniques using "Dataset Division GUI" version 1.2 and "Modified k-Medoid" version 1.3 software tools, respectively (Available at: http://teqip.jdvu.ac.in/QSAR_Tools/). The training set was used for the development of models, and the test set compounds for the validation of the obtained models. The datasets containing less than 20 molecules in the whole dataset have been modeled by the application of the "small dataset modeler_beta version" (available from http://teqip.jdvu.ac.in/QSAR_Tools/) without dividing the dataset into training and test sets. Instead of dividing the small dataset into training and test sets, the double cross-validation (DCV) technique is utilized here to model for small data sets [20–25]. Therefore, all potential combinations (k) of the validation set, which contains r compounds, and the calibration set, which contains n-r compounds, are calculated [20–25]. This is because the inner loop does not produce the "modeling set" (containing *n* compounds) [20–25]. The software enables the user to define the number of compounds to be kept in the validation set (*r*) depending on how the calibration and validation sets are made [20–25]. Genetic algorithm-multiple linear regression (GA-MLR) models are generated using calibration set chemicals [20–25]. Several internal and external validation metrics are generated for each of the chosen models during the thorough double cross-validation process. Furthermore, for each MLR model, the software generates partial least squares (PLS) regression models.

2.1.4. Multi-layered variable selection strategy and model development

In the current investigation, we have adopted a multi-layered variable selection strategy to extract the meaningful and important descriptors before developing the final model. In this approach, initially, we have applied stepwise regression in successive iterations using the Minitab software [26] using with the whole pool of descriptors, followed by a genetic algorithm using the GeneticAlgorithm_v4.1 software (available from http://dtclab.webs.com/software-tools), with a reduced pool of descriptors. Finally, we have implemented the best subset selection (available from http://dtclab.webs.com/software-tools) on the reduced pool of descriptors obtained from the genetic algorithm step. Finally, the acquired pool of descriptors was used to develop the final model. All of the final QSAR, selectivity and QSAAR models were developed using the partial least squares (PLS) regression method, except the QSAAR models (between β-amyloid and BuChE enzyme inhibitory activity, and BACE1 enzyme and BuChE enzyme inhibitory activity), and selectivity based models (between BACE1 and GSK-3β enzyme, BuChE and BACE1 enzyme, AChE and GSK-3β enzyme and BuChE and GSK-3β enzyme inhibitors), which were developed by using the multiple linear regression (MLR) technique.

2.1.5. Statistical validation of the developed 2D-QSAR models

To establish a model's significance and reliability in terms of robustness and prediction accuracy, statistical validation is one of the most important steps in the model development process. In the current work, we have calculated different internal and external validation metrics to establish that the developed models are robust and predictive enough to satisfy the acceptability criteria. For statistical quality and internal validation, we have calculated metrics like determination coefficient (R^2), leave-one-out cross-validated correlation coefficient ($Q^2_{(LOO)}$), leave-many-out cross-validated correlation coefficient ($Q^2_{(LMO)}$), $Avg\ rm^2_{(LOO)}$, and $\Delta rm^2_{(LOO)}$) [27] using only training set compounds. Higher values of the metrics R^2, $Q^2_{(LOO)}$), $Avg\ rm^2_{(LOO)}$, and lower $\Delta rm^2_{(LOO)}$) indicate a better fit and robustness of the model [28–34]. Since the internal validation metrics are insufficient to assess the predictive accuracy and robustness of the developed model, comprehensive validation of test set compounds using various external validation metrics like Q^2F_1, Q^2F_2, r^2_m parameters, and concordance correlation coefficient (*CCC*) are required to guarantee the predictive nature of the developed models [28–34]. Furthermore, we have also performed the Y-randomization test [35], applicability domain criteria (DModX (distance to model) in the X-space), etc using Simca-P 10.0 software [36].

2.1.6. Database preparation and activity prediction using developed 2D-QSAR models

To predict the inhibitory activity using developed models, we have used four chemical drug-like databases, namely, Asinex database (338604 compounds) (available from https://www.asinex.com/screening-libraries-(all-libraries), InterBioscreen (IBS) database (552793 compounds) (available from https://www.ibscreen.com/), NCI Open Database (265242 compounds) (available from https://cactus.nci.nih.gov/download/nci/), and Zinc[12] Database (17900742 compounds) (available from https://zinc12.docking.org/subsets/drug-like). Before the prediction, we developed the alvaModel by establishing 2D QSAR model descriptors against each listed target and then converted the

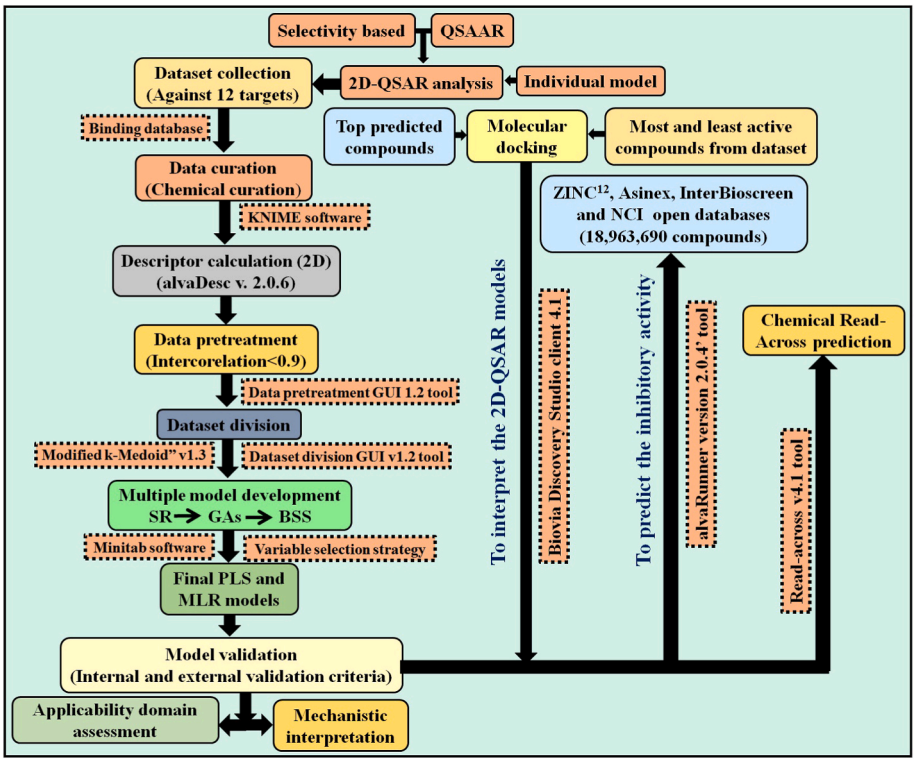


Fig. 1. The framework of the methodologies implemented in this investigation.

alvaModel into the alvaRunner project file using software, namely the alvaModel v2.0.4′ tool (available from https://www.alvascience.com/alvaModel/). The established alvaRunner project files were individually used to compute the predicted values of the above databases' compounds; the validated models were capable of precisely predicting the inhibitory activity of the majority of the compounds, as suggested by 'alvaRunner version 2.0.4' tool (https://www.alvascience.com/alvarunner/) product of Alvascience solution. We have predicted the inhibitory activity of these compounds considering the applicability domain of our PLS-based 2D QSAR models against the respective targets.

2.2. Similarity-based read-across prediction

Read-across prediction is a similarity-based *in silico* technique that predicts the biological response of unknown compounds based on known activity values [37–40]. In this study, we employed the chemical read-across approach based on machine learning to estimate the activity of the test set chemicals using the modeled descriptors. For a successful prediction with Laplacian kernel similarity-based (LK), Gaussian kernel similarity-based (GK), and Euclidean distance-based (ED) estimations, we have optimized the hyperparameter sigma (σ) and gamma (γ) respectively using validation sets. For the optimization, the initial training set is randomly divided into sub-training and sub-test sets in 3:1 proportion. The sub-training and sub-test sets were then subjected to 'Read-across v3.1' (Available from https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home) with different σ and γ values. During optimization, the other tool parameters, including the number of nearby training compounds, the distance threshold, and the similarity threshold were kept constant. The optimized setting has been selected by checking the external validation metrics (Q^2F_1 and Q^2F_2). Finally, the optimized setting was combined with the original training and test sets to get the final prediction. To obtain the best predictions, we gradually reduced the number of similar training compounds from 10 to 2.

2.3. Molecular docking study

In this investigation, we have performed the molecular docking study using the most and least active compounds from the initial datasets and also the top predicted compounds from the chemical databases to identify the interaction pattern with the respective targets. The crystal structure of the targets such as AChE enzyme (PDB ID: 4M0E), BACE1 enzyme (PDB ID: 4ivt), β-amyloid aggregation (PDB ID: 1IYT), BuChE enzyme (PDB ID: 6EZ2), Cyclin-dependent kinase 5 (PDB ID: 3O0G), Gamma-secretase enzyme (PDB ID: 6IYC), Glutaminyl Cyclase (QC) enzyme (PDB ID: 3PBB), GSK-3β enzyme (PDB ID: 5F94), MAO-B enzyme (PDB ID: 2V5Z), NMDA receptor (PDB ID: 1PBQ), and PDE 10A (PDB ID: 6MSA) were extracted from the protein databank (available from https://www.rcsb.org/). In the case of 5-HT6 protein, there are no experimental structures available in the protein data bank, so we have retrieved the predicted protein structure from the AlphaFold Protein Structure Database (Available from https://alphafold.ebi.ac.uk/entry/P50406) with the UniProt: P50406, Source organism: *Homo sapiens* (Human), and AlphaFold id: AF-P50406-F1-model_v2. To confirm the reliability of the predicted structure, we have validated the structure by Ramachandran plot server (Available from https://swift.cmbi.umcn.nl/servers/html/ramchk.html), which represents the good quality of the model (see Fig. S1 in the see Supporting Information SI-2). The molecular docking study was executed using the Biovia Discovery Studio client 4.1 [41] platform following the protocol discussed by Robertson et al. [42] and Kumar et al. [43]. Before molecular docking, the protein was prepared by checking for any missing residues, having explicit hydrogen added, and generating the active site. The active site was generated using the Biovia Discovery Studio client 4.1 [41] platform from the ligand binding domain of the bound ligand and generating the site 'from the current selection' program in the 'receptor-ligand interaction module' of the software. The bound ligand was taken out after active site generation for new molecule docking. In the case of 5-HT6 protein, we have predicted the multiple active sites at the surface of the protein using the Biovia discovery studio 4.1 client [41] platform

Table 1
Individual 2D QSAR models and their statistical validation matrics were obtained from the models.

Target	Equation	LV	Training Set				Test Set			
			Train	R^2	$Q^2_{(LOO)}$	PQ	Test	Q^2F_1	Q^2F_2	PQ
5-HT6 antagonist	pIC_{50} (nM) = $-1.671 + 0.449 \times minssNH + 0.155 \times F04[C-O] + 0.465 \times nCp + 1.710 \times nArNH2 - 0.226 \times MaxTD + 0.393 \times nBnz$	5	60	0.800	0.742	BAD	20	0.786	0.784	Moderate
AChE inhibitors	pIC_{50} (nM) = $-5.907 + 0.192 \times X2v + 0.30397 \times DBI + 0.56431 \times B07[N-N] + 0.00419 \times D/Dtr05 - 0.01895 \times NdsCH - 0.15889 \times C-016 - 0.28519 \times F04[O-O] - 0.25533 \times NsssN + 0.89374 \times F04[N-Cl] - 0.01257 \times T(N..Cl) + 1.44516 \times B05[C-N] - 1.08240 \times B01[N-N] - 3.08774 \times nROCON - 0.56172 \times B09[N-O] + 0.04808 \times minssO - 1.06153 \times totalcharge - 0.38974 \times CATS2D\_02\_AP - 0.57236 \times B02[N-N] + 1.05258 \times B04[O-S] - 0.07341 \times CATS2D\_08\_AL + 1.42829 \times nR\#CH/X - 0.41181 \times C-008 + 0.26843 \times CATS2D\_08\_AA - 1.01794 \times B06[O-S] + 1.95405 \times CATS2D\_01\_DD + 0.47598 \times CATS2D\_06\_AP - 1.58252 \times B09[O-F] + 0.36591 \times nArCO$	7	1325	0.635	0.621	–	408	0.678	0.678	Moderate
BACE1 inhibitors	pIC_{50} (nM) = $-11.3326 + 0.2442 \times SAscore + 0.2420 \times F08[N-N] + 2.0777 \times Ui + 0.4070 \times X5v - 0.1594 \times F04[O-O] + 0.0720 \times mindO + 0.3921 \times nR\#CH/X - 0.4098 \times B02[O-O] - 0.0706 \times SaaCH - 0.3112 \times F04[F-F] - 0.9326 \times minaaaC - 0.5385 \times O-058 - 0.0108 \times D/Dtr04 + 0.5356 \times nR04 + 0.1804 \times B06[O-O] + 0.1890 \times F09[C-Cl] + 0.9600 \times nN = C-N < -0.4325 \times B05[N-Cl]$	9	680	0.669	0.650	–	225	0.675	0.674	Moderate
β-amyloid inhibitors	pIC_{50} (nM) = $-9.3579 + 0.1369 \times SssCH2 - 0.2513 \times F02[N-N] - 0.1060 \times F05[C-O] + 20.6914 \times PW3 - 0.3449 \times MAXDN + 0.3534 \times F04[N-O]$	4	197	0.729	0.705	Moderate	65	0.844	0.844	Good
BuChE inhibitors	pIC_{50} (nM) = $-0.64834 + 0.10456 \times C-002 + 0.87920 \times N-070 - 0.02751 \times nCs - 0.18997 \times nArNHR + 0.55922 \times MaxaaaC + 1.07712 \times nArOCON - 0.60012 \times Psi\_e\_A - 0.08769 \times H-051 + 0.00608 \times MDEC-22 + 0.07105 \times F04[C-N] - 0.10931 \times MAXDP + 0.44817 \times B07[N-N] + 1.01229 \times N-077 + 1.95654 \times nArC = N - 2.11509 \times C-036 - 0.55208 \times B03[N-N] + 1.64662 \times C-035 + 3.47045 \times nN - N + 1.34851 \times BLI - 0.32650 \times F02[N-N] + 0.08134 \times SdsN + 0.02974 \times minsF + 0.76341 \times C-009 + 2.59201 \times nROCON - 1.20613 \times C-037 + 0.24057 \times B04[O-O] - 0.23948 \times nR09 - 0.19739 \times nCsp$	8	1882	0.689	0.668	Moderate	625	0.702	0.702	Moderate
CDK-5 inhibitors	pIC_{50} (nM) = $-2.82109 + 0.52376 \times NaaaC - 0.33618 \times PBF - 0.01011 \times T(S..S) + 0.21105 \times SaasC - 0.02307 \times T(N..F) - 1.65453 \times S-106 - 0.37066 \times F07[N-S]$	4	169	0.675	0.652	–	56	0.790	0.790	Good
Gamma-secretase inhibitors	pIC_{50} (nM) = $-15.4364 + 0.0155 \times P\_VSA\_MR\_7 - 1.1252 \times nOHp - 0.6246 \times B05[N-N] - 0.9356 \times nRCOOH - 0.5702 \times CATS2D\_07\_NL - 0.5050 \times F05[N-O] + 2.3424 \times ATS6s - 0.5725 \times B09[C-S] + 0.4953 \times F08[O-S] + 0.1089 \times VE3sign\_D + 0.4441 \times B10[O-F] + 0.1313 \times VE3sign\_D/Dt - 1.4220 \times B07[O-Cl] - 0.1858 \times F06[F-F] - 1.9204 \times nS(=O) 2 + 0.7022 \times C-029$	7	172	0.773	0.720	–	45	0.734	0.734	Moderate
Glutaminyl Cyclase inhibitors	pIC_{50} (nM) = $-1.24512 + 1.87780 \times C-034 - 6.43315 \times ETA\_Shape\_Y - 0.21722 \times F05[C-S] - 0.43309 \times PBF + 0.46170 \times B09[C-O] - 0.24517 \times F02[N-N] + 0.02157 \times T(N..S)$	5	99	0.944	0.934	Good	33	0.956	0.956	Good
GSK-3β inhibitors	pIC_{50} (nM) = $-5.65151 + 0.98713 \times nThiazoles + 0.09410 \times SaasC + 2.20185 \times PDI + 0.08050 \times SaaaC - 0.14490 \times mindsCH + 0.57896 \times B03[O-Br] + 0.21992 \times F06[N-O] - 0.31342 \times B03[C-O] - 0.00636 \times nPyrroles + 0.09118 \times B05[N-O]$	5	118	0.703	0.648	–	41	0.763	0.763	Moderate
MAO-B inhibitors	pIC_{50} (nM) = $-7.74797 - 0.48129 \times mindsCH + 1.46202 \times F02[O-O] + 0.09291 \times C\% + 0.63918 \times nRCN - 0.54786 \times B06[O-Cl] - 0.10809 \times ALOGP - 0.34911 \times F05[C-N] - 1.36772 \times B01[C-O] + 0.71091 \times B03[N-N] - 0.39179 \times minssCH2 + 1.04117 \times SAscore - 1.99700 \times ETA\_Eta\_B\_RC - 0.94480 \times nArNHR$	6	125	0.722	0.649	–	45	0.639	0.639	–
NMDA receptor antagonist	pIC_{50} (nM) = $0.81523 - 0.04386 \times SaaN - 0.87822 \times MDEN - 23 - 0.00678 \times TPSA(NO) - 1.66542 \times C-043 - 0.04382 \times ATSC2m + 0.00139 \times ATSC1p - 0.81419 \times MATS4i + 0.10333 \times C-028 + 0.00021 \times TPSA(Tot) - 0.12093 \times MDEC-24 - 0.92317 \times GATS2m - 0.39657 \times GATS8v - 0.11764 \times F05[C-O] + 0.19022 \times B10[C-O]$	7	267	0.740	0.708	Good	89	0.640	0.639	Good
Phosphodiester enzyme	pIC_{50} (nM) = $-7.87353 + 1.09932 \times N-070 + 0.00916 \times AMR + 0.99679 \times minsCH3 - 1.22178 \times MDEN - 23 - 0.06459 \times MaxDD - 0.34705 \times nFuranes + 7.19158 \times Eta\_epsi\_2 - 1.74389 \times B03[N-F] - 1.36496 \times F09[N-N] - 0.57580 \times nArNR2 - 0.82639 \times nRCOOR + 1.07168 \times B09[N-N] - 1.49240 \times NNRS - 0.52332 \times mindssC - 1.39784 \times B06[O-F]$	9	222	0.722	0.677	Moderate	67	0.739	0.730	Moderate

Table 2
Selectivity-based models and their statistical validation matrics were obtained from the models.

Target (Selectivity)	Equation	LV	Training Set				Test Set			
			Train	R^2	$Q^2_{(LOO)}$	PQ	Test	Q^2F_1	Q^2F_2	PQ
AChEI - BACE1	Selectivity (AChEI-BACE1) $= -0.42092 + 1.61441 \times F08[N-O] - 1.82470 \times B05[N-O] + 0.45021 \times X2v - 0.56114 \times F10[O-O] + 0.90094 \times SsssCH - 0.10198 \times F05[C-C]$	5	34	0.827	0.744	Good	9	0.627	0.624	Good
AChEI -β-amyloid	Selectivity (AChEI-β-Amyloid) $= -2.06170 + 0.70243 \times B07[C-N] - 1.34266 \times nArNH2 + 1.01867 \times F06[C-Cl] + 0.30819 \times SAscore - 1.12566 \times F10[C-Cl] + 1.04712 \times ETA\_Shape\_Y - 0.10865 \times PJI2$	6	62	0.787	0.679	Good	21	0.901	0.898	Good
AChEI -BuChEI	Selectivity (AChEI-BuChEI) $= 3.15944 + 0.81697 \times F06[N-O] - 3.26082 \times B03[N-N] - 0.71013 \times B03[N-O] - 1.80289 \times nROR + 1.63240 \times nR12 - 0.22007 \times N\% - 8.65969 \times GD - 1.24833 \times B02[N-N] + 0.20227 \times nDB - 0.21524 \times ICR - 0.04797 \times PHI$	6	90	0.785	0.717	Good	23	0.732	0.709	Good
AChEI - MAO-B	Selectivity (AChEI-MAO-B) $= 5.01821 - 1.81468 \times SsssCH - 0.53685 \times O\% - 0.23837 \times F10[C-C] - 0.47260 \times F10[C-N] + 0.17398 \times F09[C-C] + 1.97941 \times MaxsCH3 - 2.36195 \times minsCH3$	6	42	0.778	0.651	–	10	0.725	0.723	Moderate
BACE1 - GSK-3β	Selectivity (BACE1-GSK3B) $= 0.401 + 1.85 \times LipinskiFailures - 0.299 \times nRotB - 2.83 \times AATSC1s$	–	15	0.889	0.766	Good	5	0.876	0.870	Good
BuChEI - BACE1	Selectivity (BuChEI-BACE1) $= 0.260 + 0.877 \times nArCO - 3.90 \times nR = Ct + 0.176 \times NaasC + 2.51 \times minaaaC - 0.0159 \times ALOGP2 + 0.000332 \times ZM1Kup$	–	39	0.903	0.863	Good	12	0.866	0.828	Good
BuChEI - β-amyloid	Selectivity (BuChEI-β-amyloid) $= 3.69262 + 0.83031 \times B04[N-N] + 0.00558 \times D/Dtr10 - 0.46510 \times B10[N-O] - 3.56920 \times B10[C-N] \times 3.24186 \times B08[C-N] + 0.48410 \times B10[C-O] - 0.62086 \times B08[O-O] + 2.72960 \times ETA\_dEpsilon\_B - 0.30788 \times LOGP99 - 9.01720 \times ETA\_Shape\_Y - 2.20437 \times PJI2$	7	93	0.821	0.763	Moderate	30	0.696	0.656	Moderate
BuChEI - MAO-B	Selectivity (BuChEI-MAO-B) $= 2.92274 - 2.35759 \times SsssCH - 0.97195 \times B08[N-O] - 0.26699 \times NdssC + 0.72489 \times B06[N-O] - 1.99605 \times minsCH3 - 0.01919 \times SAacc + 0.94967 \times SsssN$	3	38	0.756	0.637	–	10	0.826	0.801	Moderate
[a]AChEI – GSK-3β	Selectivity (AChEI-GSK3B) $= -17.8 - 113 \times ETA\_BetaP\_ns\_d + 10.7 \times Psi\_i\_A - 0.143 \times F02[C-O]$	–	18	0.766	0.664	Moderate	3[a]	0.784	0.765	G
[a]BuChEI - GSK-3β	Selectivity (BuChEI-GSK3B) $= -6.54 - 2.39 \times ETA\_Beta\_ns\_d + 4.43 \times Psi\_i\_A + 0.867 \times nCL$	–	18	0.735	0.647	Moderate	3[a]	0.770	0.752	G

[a] Models were developed by SmallDataModeler_betaVersion

from the "define and edit binding site" using the module "generate active site from receptor cavities", and docked the ligand in each site to identify the favorable binding site (identified most favorable active site coordinate X: 18.945, Y: 0.896, Z: 11.313, the radius of sphere 19.299). To prepare ligands, the selected compounds were run through the Discovery Studio platform's 'small-molecule module', where several ligand conformers were formed. Each of these generated conformers was subsequently employed in the CDOCKER module for molecular docking using a CHARMm-based molecular dynamic scheme. The CDOCKER interaction energy parameter (kcal/mol) was examined for all receptor-ligand complexes, and the highest-scoring (more negative; hence favorable to binding) poses with only non-covalent interactions (ionic bonds, hydrophobic interactions, hydrogen bonds, etc.) were kept for future investigation. A graphic representation of the methodologies is shown in Fig. 1.

3. Results and discussion

The current research aimed at developing statistically significant 2D-QSAR models against 12 major targets with easily interpretable descriptors and using them to check the applicability domain of four chemical drug-like databases (ZINC[12], Asinex, NCI, and InterBioscreen databases) and providing prioritized set of compounds for experimental detection of their performance as anti-Alzheimer's drugs. The interpretations from the 2D-QSAR models were further confirmed by molecular docking strategies. The current work comprises five phases: (1) development of a well-validated 2D-QSAR model (individual, selectivity based, and QSAAR) against 12 major targets; (2) chemical Read-Across analysis; (3) prediction of inhibitory activity of four chemical compound databases against 12 major targets, using the developed 2D-QSAR models; (5) molecular docking of the most and least active compounds of the modeled datasets.

3.1. 2D QSAR analysis

In this analysis, we have developed PLS-based 2D-QSAR models against 12 major targets (AChE enzyme, BuChE enzyme, BACE1 enzyme, β-amyloid aggregation, 5-HT6, CDK-5, Gamma-secretase enzyme, Glutaminyl Cyclase enzyme, GSK-3β enzyme, MAO-B enzyme, NMDA receptor and phosphodiesterase enzyme (PDE 10A)) to search for novel anti-Alzheimer's agents and identify important structural features responsible for inhibiting the enzymes involved in AD. We have also developed the 10 Selectivity-based models (6-PLS based and 4-MLR based models) and 17 QSAAR-based models (15-PLS based and 2-MLR based models) for the identification of features with the dual inhibitory activity (see supplementary information SI-1, sheet 1–31). The details about the models and their various validation metrics are given in Tables 1, 2, and 3. Additionally, we have also developed models using the whole sets and performed the leave-many-out cross-validated correlation coefficient ($Q^2_{(LMO)}$) at different folds in the case of the small datasets (see sheets 13–31 in supplementary information SI-1). The statistical results indicated that all the models were acceptable in terms of stability, predictive ability, and fitness. The obtained features define the structural and functional requirements for compounds to improve their inhibitory activity against the respective enzymes. The scatter plot (Figs. 2–8) describes the closeness of the observed and predicted values for the modeled enzyme inhibitors. In the loading plot (see Figs. S2–34 in the Supporting Information SI-2), quantitative contributions of similar/dissimilar descriptors, as well as interrelationships among the X variables and the Y-response are illustrated. Moreover, we have also performed the Y-randomization test using the SIMCA-P 10.0 software [36] and MLRPlusValidation1.3 tool (available at http://dtclab.webs.com/software-tools) by randomly reshuffling (100 permutations) the dependent variable to ensure that the developed models were not obtained by any chance. The R^2 and Q^2 values for the random models

Table 3
QSAAR models and their statistical validation matrics were obtained from the models.

Variable		Equation	LV	Training Set				Test Set			
Y	X			Train	R^2	$Q^2_{(LOO)}$	PQ	Test	Q^2F_1	Q^2F_2	PQ
AChEI	BACE1	pIC_{50} (nM)_AChEI $= -9.09333 + 0.41826 \times nCrs \times 1.26840 \times MaxaasC + 0.00358 \times log\_BACE1 + 9.75454 \times PW3$	3	33	0.896	0.843	Good	10	0.746	0.744	Good
BACE1	AChEI	pIC_{50} (nM)_ BACE1 $= -3.37330 + 0.27402 \times X5v - 0.23660 \times C - 001 - 0.20158 \times nCconj - 0.55996 \times nArOH + 0.97466 \times B08[O - O] + 0.02858 \times log\_AChEI$	4	33	0.766	0.637	–	10	0.772	0.749	Moderate
AChEI	β-amyloid	pIC_{50} (nM)_AChEI $= -3.32884 + 0.85551 \times log\_Beta\ Amyloid + 0.34000 \times nArCO + 2.12019 \times B10[N - Cl] + 6.27610 \times PW3 + 0.70658 \times B07[C - N]$	3	63	0.962	0.949	Good	20	0.968	0.968	Good
β-amyloid	AChEI	pIC_{50} (nM)_ β-amyloid $= 3.16964 + 1.00643 \times log\_AChEI - 1.78556 \times ETA\_EtaP\_F\_L - 1.99421 \times B10[N - Cl] - 0.65971 \times B07[C - N] - 0.55678 \times SAscore$	4	63	0.947	0.924	Good	20	0.972	0.972	Good
AChEI	BuChEI	pIC_{50} (nM)_AChEI $= -5.82692 + 0.71516 \times NssNH + 0.40986 \times nR10 + 1.23241 \times ICR + 0.94089 \times B04[N - O] - 0.24740 \times F08[C - N] + 0.09780 \times log\_BuChEI + 0.61060 \times F06[N - O] - 1.26313 \times nROR + 1.04963 \times B04[N - Cl] + 0.51418 \times nCrt$	5	89	0.867	0.808	Good	29	0.749	0.736	Good
BuChEI	AChEI	pIC_{50} (nM)_ BuChEI $= -0.767384 + 0.126111 \times NaaCH - 0.753111 \times F06[N - O] - 0.883081 \times MaxsCH3 + 0.419744 \times DECC + 0.302945 \times log\_AChEI + 0.756970 \times B10[N - O] - 0.293615 \times ETA\_EtaP\_F - 0.124508 \times nCrs - 0.126071 \times Ram - 0.064782 \times nCb -$	6	89	0.789	0.724	–	29	0.767	0.756	Moderate
AChEI	GSK-3β	pIC_{50} (nM)_ AChEI $= -8.9587 - 0.6585 \times C - 026 + 0.6051 \times log\_GSK3B + 15.8263 \times ETA\_Epsilon\_1$	2	14	0.847	0.761	–	4	0.837	0.824	Good
GSK-3β	AChEI	pIC_{50} (nM)_ GSK-3β $= -4.25506 + 0.31626 \times log\_AChEI + 2.20735 \times MaxaasC$	1	14	0.891	0.858	–	4	0.943	0.939	Good
AChEI	MAO-B	pIC_{50} (nM)_ AChEI $= 37.6420 + 0.4495 \times F01[C - N] - 0.7457 \times N\% + 1.0504\ MaxdssC - 27.2159 \times ETA\_Epsilon\_5 - 4.0017 \times MaxaaCH + 0.3758 \times B05[N - Cl] - 2.2041 \times MAXDP + 0.1345 \times log\_MAO - B$	6	42	0.793	0.677	–	10	0.755	0.638	Moderate
BACE1	GSK-3β	pIC_{50} (nM)_ BACE1 $= -3.10715 + 0.11713 \times max\_conj\_path - 0.12209 \times minsOH + 0.62254 \times C - 025 + 0.81474 \times log\_\ GSK - 3\beta$	3	15	0.862	0.662	–	5	0.808	0.779	Good
GSK-3β	BACE1	pIC_{50} (nM)_ GSK-3β $= -2.50148 - 2.03802 \times MATS2e + 0.14716 \times log\_BACE1 - 0.02185 \times ZMIC4 + 0.10199 \times MAXDN - 0.04842 \times AATS3s$	3	15	0.808	0.569	–	5	0.668	0.666	–
BACE1	BuChEI	pIC_{50} (nM)_ BACE1 $= -3.5140 + 0.58485 \times MaxsCH3 - 0.13063 \times F01[C - O] + 0.94806 \times nFuranes + 0.06244 \times log\_BuChEI + 1.37864 \times B05[O - O] - 0.07915 \times nCIC$	5	39	0.943	0.882	Good	12	0.942	0.940	Good
BuChEI	BACE1	pIC_{50} (nM)_ BuChEI $= -3.12300 - 1.57604 \times nFuranes + 9.30594 \times RFD - 0.00055 \times F08[C - C] + 1.97203 \times nR09 + 0.09518 \times log\_BACE1 - 2.62602 \times B04[O - O]$	–	39	0.956	0.857	Good	12	0.942	0.941	Good
β-amyloid	BuChEI	pIC_{50} (nM)_ β-amyloid $= -2.12671 - 4.83670 \times ETA\_EtaP\_F + 0.13649 \times log\_BuChEI + 1.32042 \times minssCH2 + 0.16273 \times SaasC + 0.09384 \times F08[C - C] - 0.11038 \times F06[C - C] + 1.01759 \times Uc$	4	93	0.805	0.771	Moderate	30	0.776	0.776	Moderate
BuChEI	β-amyloid	pIC_{50} (nM)_BuChEI $= -2.17 + 0.00370 \times D/Dtr10 - 0.0654 \times nCp + 1.45 \times nCrt + 1.73 \times nRNHR + 0.403 \times log\_Beta\ Amyloid - 0.0696 \times nAB$	–	93	0.879	0.837	Good	30	0.910	0.910	Good
GSK-3β	BuChEI	pIC_{50} (nM)_ GSK-3β $= -3.38300 + 0.22068 \times N\% + 0.14374 \times log\_BuChEI + 0.30867 \times minssCH2$	2	14	0.946	0.858	Moderate	4	0.922	0.922	Good
MAO-B	BuChEI	pIC_{50} (nM)_ MAO-B $= -7.47908 + 1.04311 \times nCt + 2.00343 \times minsCH3 + 0.61755 \times B08[N - O] + 0.24486 \times F06[C - O] + 0.06530 \times log\_BuChEI + 1.33039 \times nROH + 2.42118 \times nRNHR - 0.67456 \times ETA\_Beta\_ns\_d$	6	39	0.758	0.638	–	12	0.771	0.771	–

(Y-axis) are plotted against the correlation coefficient between the original Y values and the permuted Y values (X-axis). The statistical results obtained from randomized models suggested that the developed models were not found by any chance, as shown in Figs. S35–67 (see supplementary information SI-2) and sheet 32–35 (see supplementary information SI-1). The variable importance plot (VIP) and regression coefficient plot, as depicted in Figs. S68–100 and Figs. S101–133 respectively (see Supplementary Information SI-2), are used to determine the significance and contribution of the obtained features in the models toward the inhibitory activity.

3.1.1. Mechanistic interpretation of the descriptors involved in the developed individual QSAR models

The details about all variables that appeared in this model, their contribution, meaning of descriptors, examples of compounds, and their corresponding descriptors values are depicted in Table S1 (see Supporting Information SI-2).

3.1.1.1. 5-Hydroxytryptamine receptor 6 (5-HT6). The first most important descriptor in this model is minssNH, as we can see from Table S1 (see Supporting Information SI-2), the value of minssNH is 2.844 for the most active compound **1** and 0 for the least active compound **74**. The positive regression coefficient of this descriptor indicates that the higher value of the descriptor leads to an improved inhibitory activity against the enzyme. The second most significant descriptor in this series is F04[C–O], as is shown in Table S1 (see Supporting Information SI-2), the value of F04[C–O] descriptor is 6 for the most active compound **3** and 0 for the least active compound **74**. The positive sign of

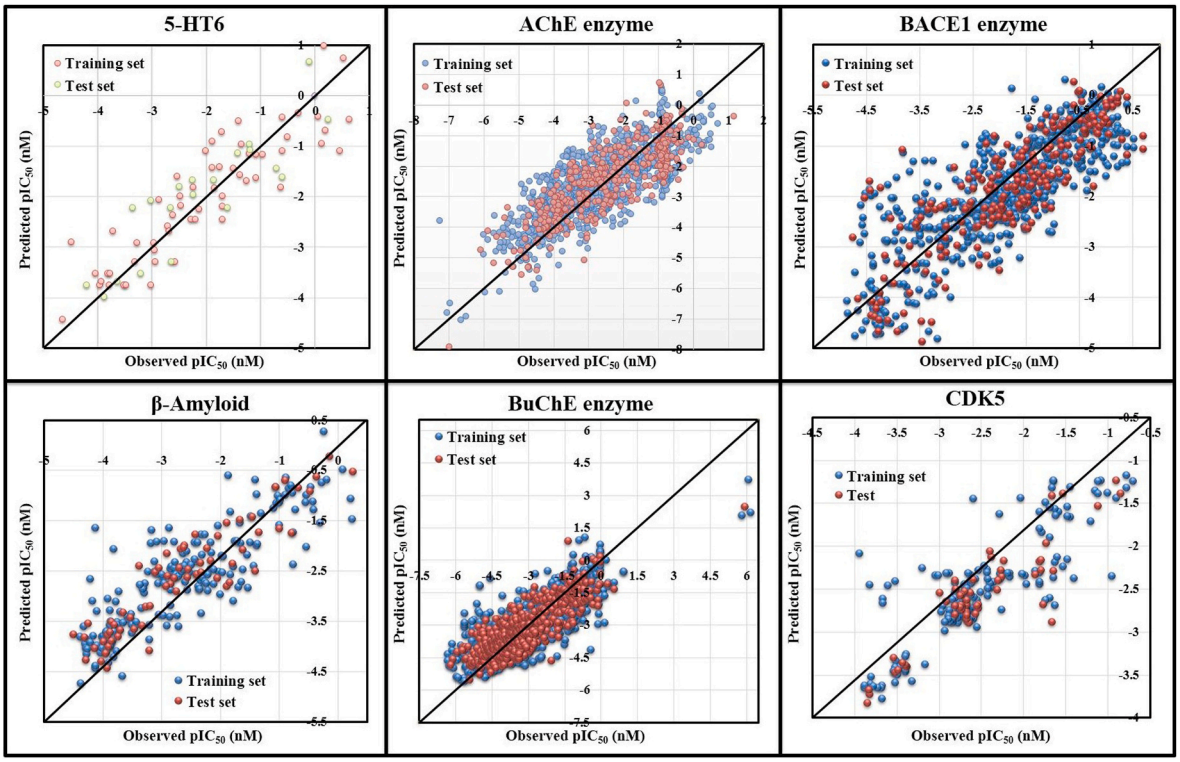


Fig. 2. Scatter plots of observed vs predicted values for 5-HT6, AChE, BACE1, β-amyloid, BuChE, and CDK-5 models.

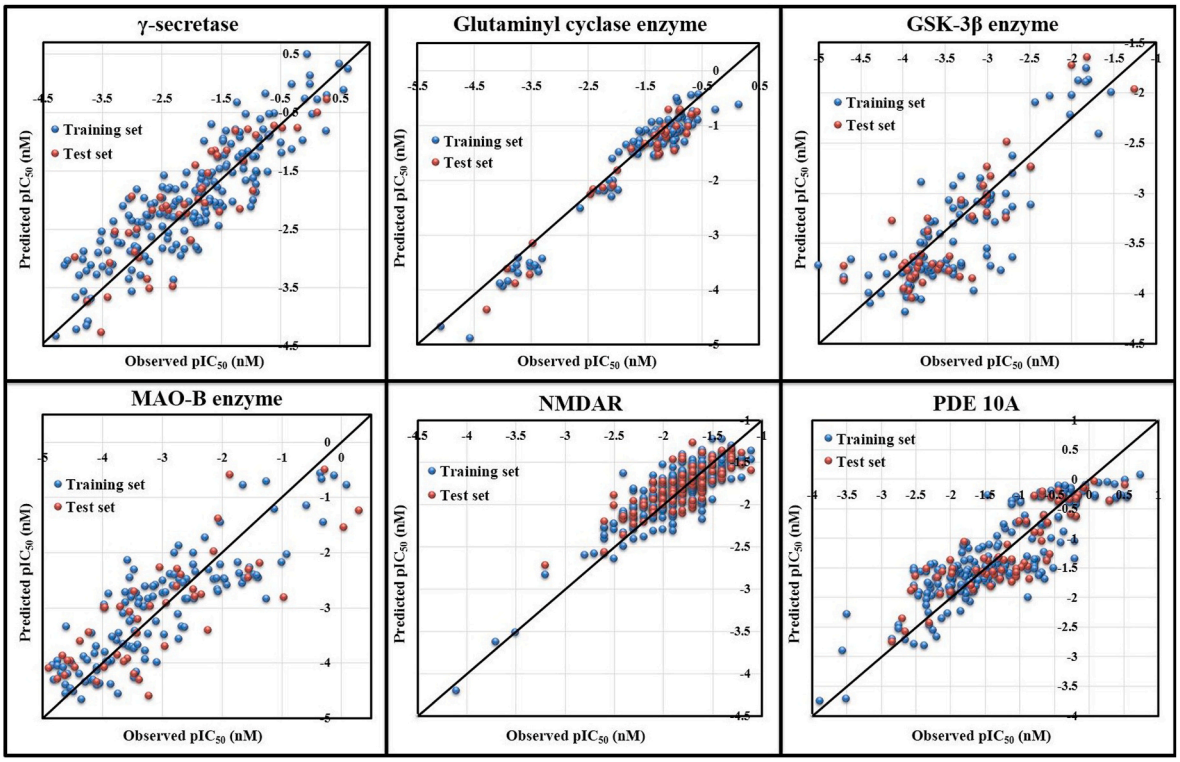


Fig. 3. Scatter plots of observed vs predicted values for γ-secretase, Glutaminyl Cyclase, GSK-3β, MAO-B, NMDAR, and PDE 10A models.

the coefficient of this descriptor confirms that larger the value of the descriptor higher the inhibitory activity towards the enzyme. According to the discussion made above, the minssNH, and F04[C–O] of the most active compounds must be higher than those of the less active compounds (see Fig. 10).

3.1.1.2. Acetylcholinesterase (AChE) enzyme. In this model, the first most significant descriptor is X2v which belongs to the class of connectivity index and is related to molecular branching and shape information. The positive regression coefficient sign confirms that the AChE enzyme inhibitory activity may be increased by an increase in molecular

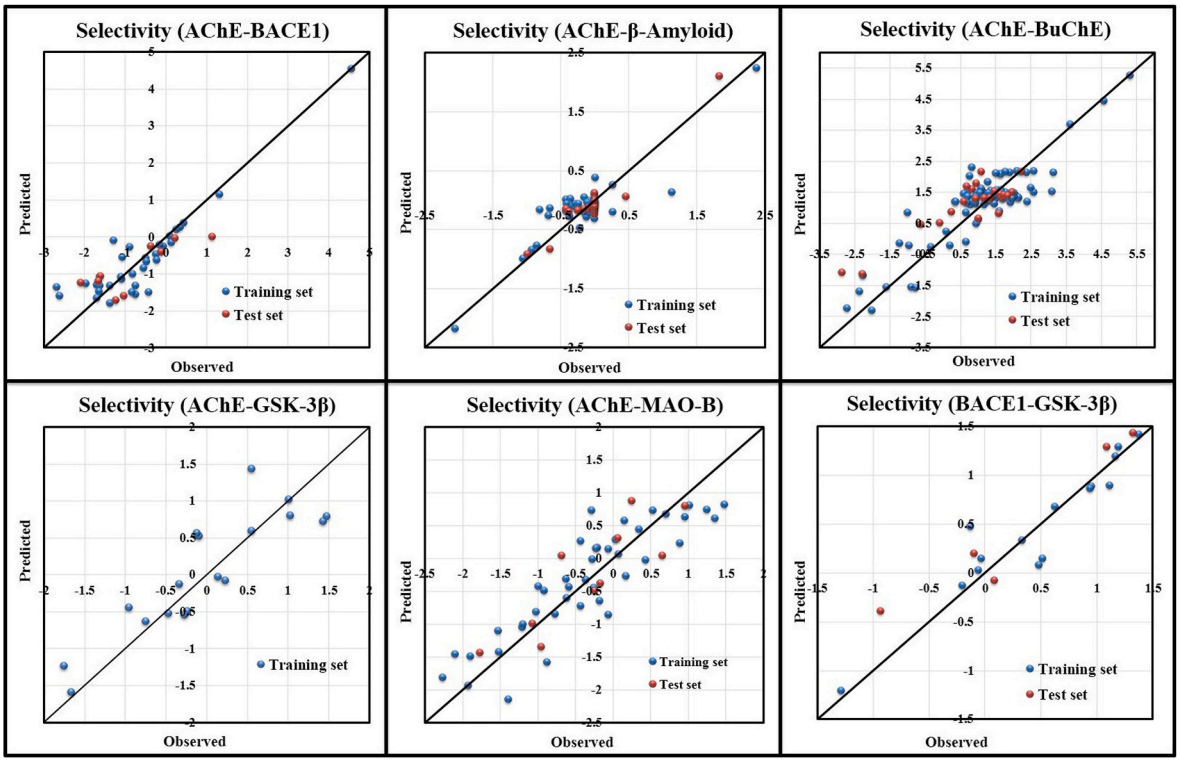


Fig. 4. Scatter plots of observed vs predicted values for selectivity based (AChE-BACE1, AChE-β-Amyloid, AChE-BuChE, AChE-GSK-3β, AChE-MAO-B, and BACE1-GSK-3β) models.

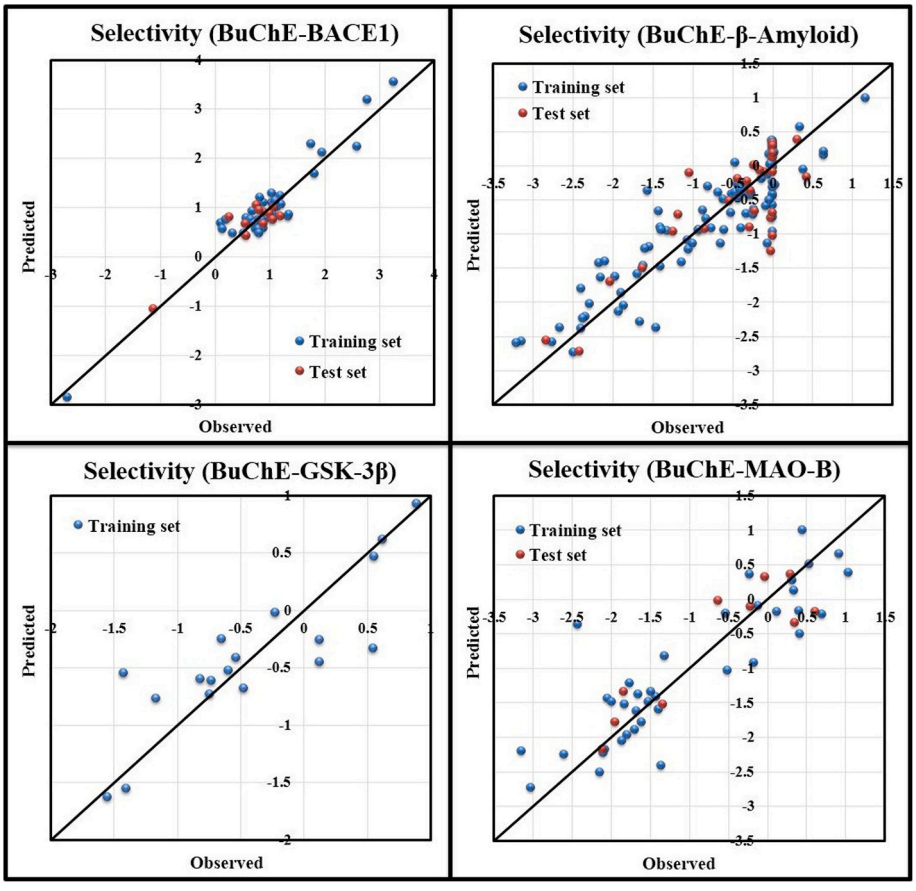


Fig. 5. Scatter plots of observed vs predicted values for selectivity-based (BuChE-BACE1, BuChE-β-Amyloid, BuChE-GSK-3β, and BuChE-MAO-B) models.

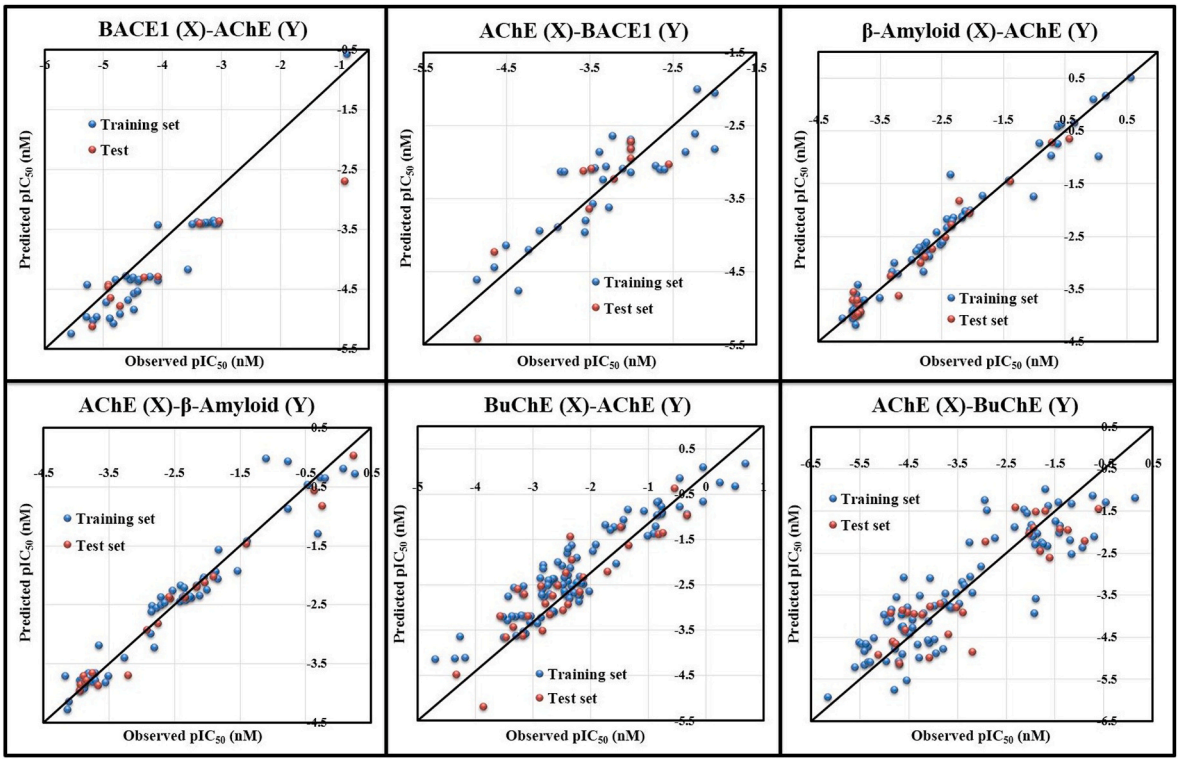


Fig. 6. Scatter plots of observed vs predicted values for QSAAR (BACE1 (X)-AChE (Y), AChE (X)-BACE1 (Y), β-Amyloid (X)-AChE (Y), AChE (X)-β-Amyloid (Y), BuChE (X)-AChE (Y) and AChE (X)-BuChE (Y)) models.

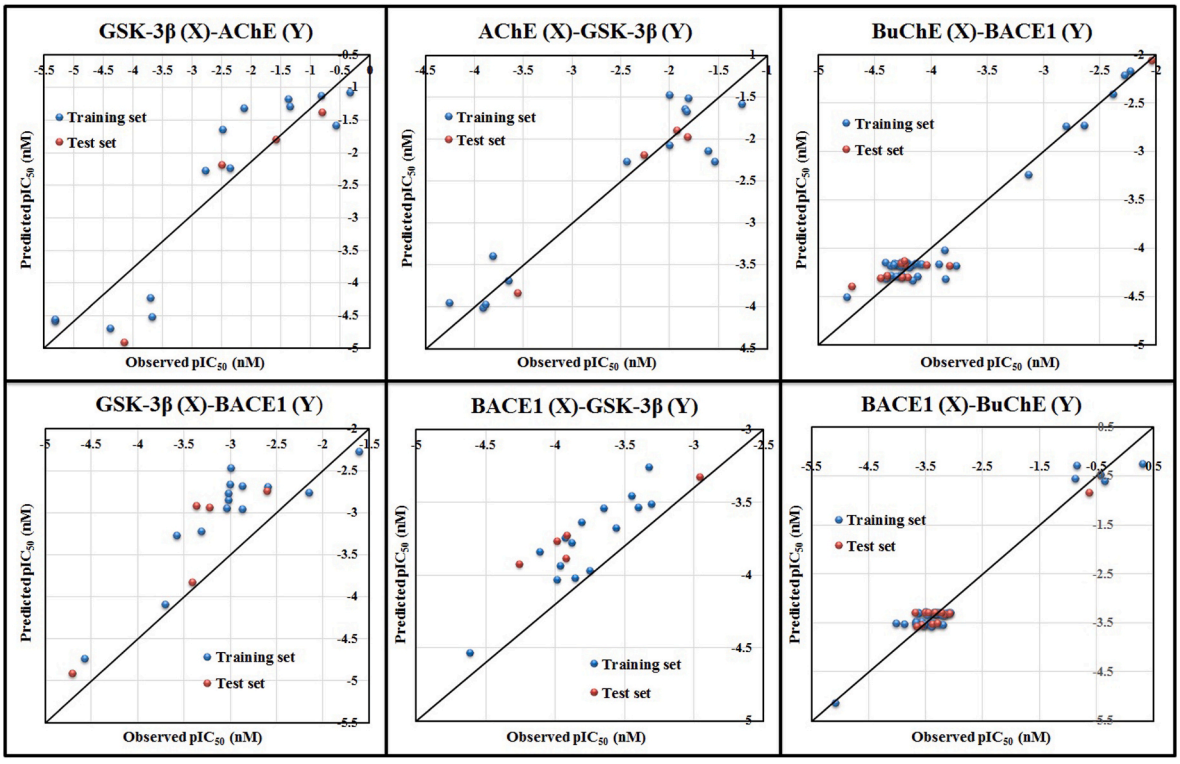


Fig. 7. Scatter plots of observed vs predicted values for QSAAR (GSK-3β (X)-AChE (Y), AChE (X)-GSK-3β (Y), BuChE (X)-BACE1 (Y), GSK-3β (X)-BACE1 (Y), BACE1 (X)-GSK-3β (Y) and BACE1 (X)-BuChE (Y)) models.

branching and shape (bond angles) in the compounds. For instance, we can see in Table S1 (see Supporting Information SI-2), the most active compound **1754** has a higher value leading to better inhibitory activity, while the least active compound **698** has the lowest value for this descriptor. Another positively correlated descriptor, DBI, represents the branching nature of the compound. With an increase in the branching

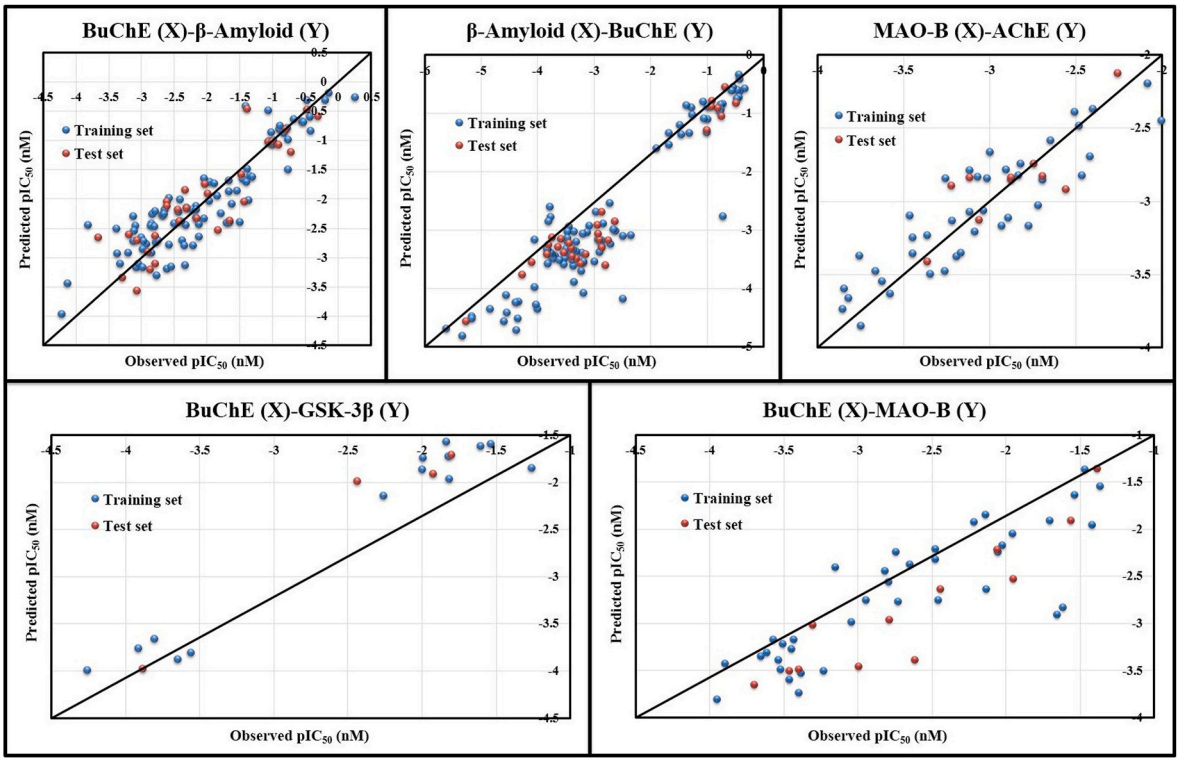


Fig. 8. Scatter plots of observed vs predicted values for QSAAR (BuChE (X)-β-Amyloid (Y), β-Amyloid (X)-BuChE (Y), MAO-B (X)-AChE (Y), BuChE (X)-GSK-3β (Y) and BuChE (X)-MAO-B (Y)) models.

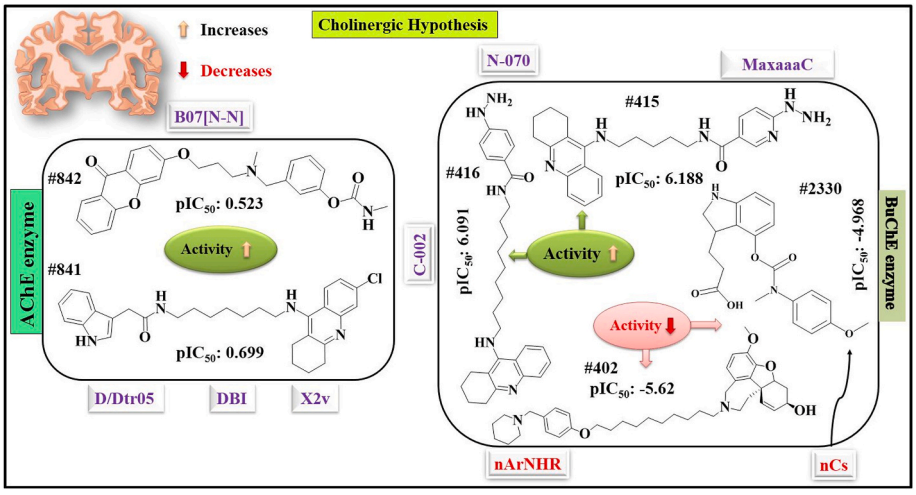


Fig. 9. Possible mechanistic interpretation of the most significant descriptors obtained from the models against AChE and BuChE enzymes.

index, the inhibitory activity will increase as observed in compound **3** (most active). The next positively correlated descriptor, B07[N–N], leading to better inhibitory activity against the enzyme as the value of descriptor is 1 for the most active compound 842 and zero for the compound **725** Table S1 (see Supporting Information SI-2). The ring descriptor, D/Dtr05, having a positive regression coefficient suggests that a higher numerical value of this descriptor leads to an improved inhibitory activity as verified by the compound **841** (most active) and the compounds with no such fragment show lower AChE enzyme inhibitory activity as found in the compound **710** (least active) (see Fig. 9).

3.1.1.3. Butyrylcholinesterase (BuChE) enzyme. Among the essential features enhancing BuChE enzyme inhibitory activity are nArOCON, N-070, MaxaaaC, C-002, F04[C–N], and MDEC-22. As lipophilicity is an important parameter for the AD drugs, it can be traced from variables nArOCON, C-002, MDEC-22 and MaxaaaC which contributed positively towards the inhibitory activity as evidenced by the compounds **593**, **416,** and **415** (most active) respectively. The enhanced concentration of electronegative atom count in a molecule has a direct impact on the improved inhibitory activity. This hypothesis can be confirmed by the presence of variables N-070 and F04[C–N] which contributed positively as evidenced by the compounds **415** and **416** (most active) respectively. Again, variables nCs, nArNHR, H-051, and Psi_e_A contributed

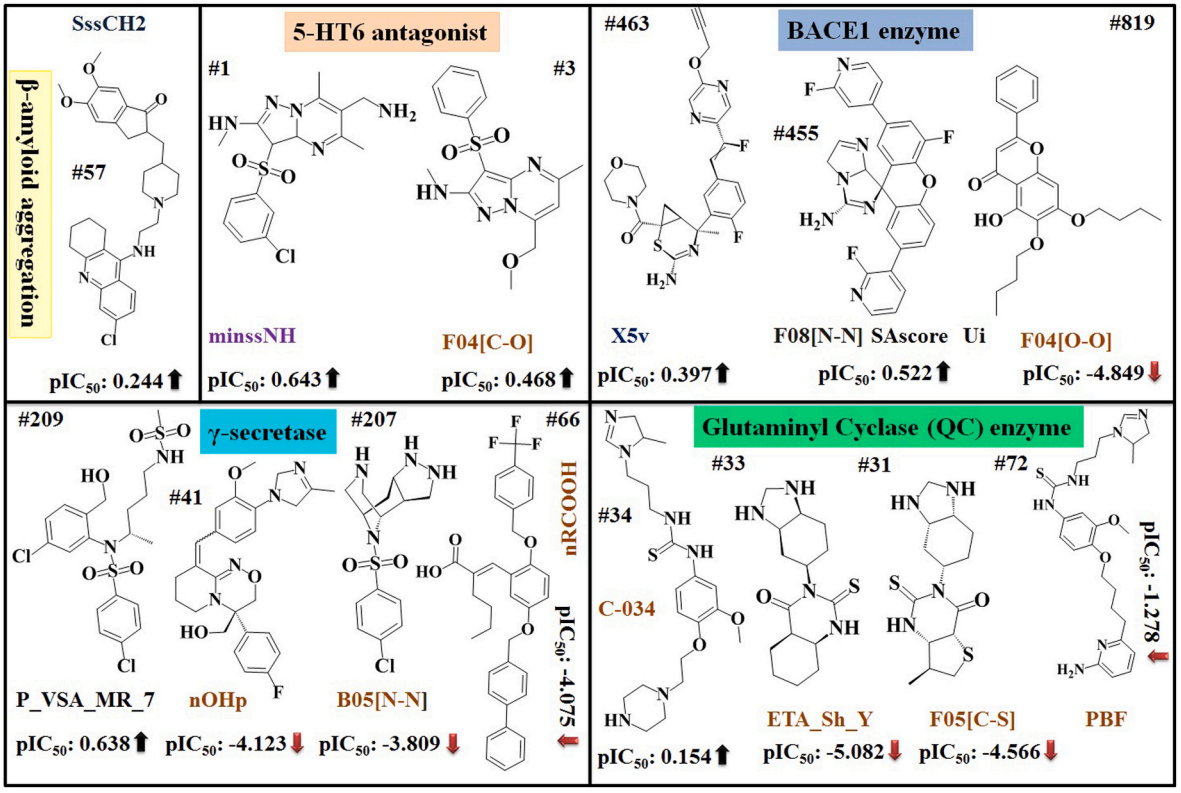


Fig. 10. Mechanistic interpretation of the most significant descriptors obtained from the models against β-amyloid aggregation, 5-HT6, BACE1, γ-Secretase, and Glutaminyl Cyclase.

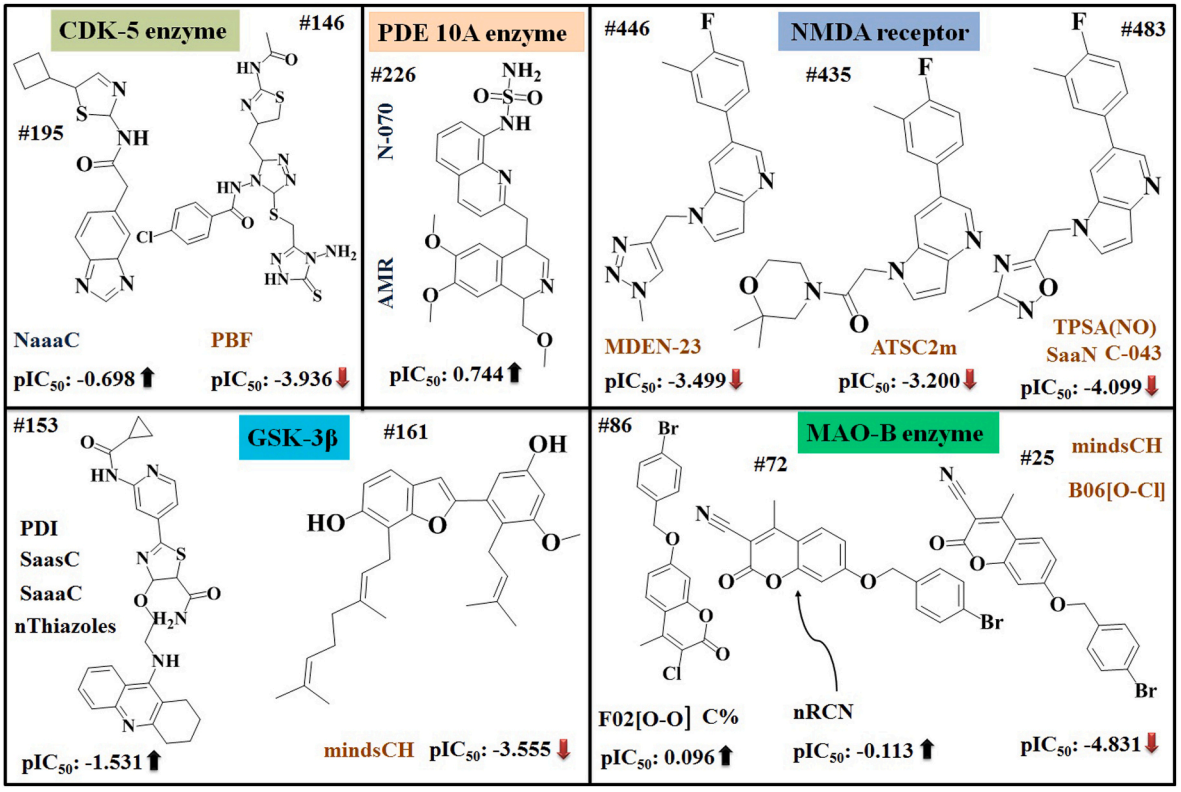


Fig. 11. Contributions of the most significant descriptors obtained from the models against CDK-5, PDE 10A, NMDAR, GSK-3β, and MAO-B enzymes.

negatively which means that the presence of these features in the molecules leads to lower inhibitory activity as observed in the compounds **2330**, **402**, **399**, and **411** (least active) respectively(see Fig. 9).

3.1.1.4. Beta-secretase 1 (BACE1) enzyme. The final equation comprises 28 descriptors, among which five descriptors (SA score, F08[N–N], Ui, X5v, and F04[O–O]) contributed most against the enzyme as shown by the VIP statistic (VIP score >1). Among them, four variable showed positive contribution towards an inhibitory activity which include SA score, F08[N–N], Ui, and X5v, which means that the presence of these features in the molecules leads to better inhibitory activity as found in the compounds **455** and **463** (in case of X5v). The negative contribution of F04[O–O] denotes the frequency of two oxygen atoms at topological distance 4, suggesting that a higher number of electronegative atom count in a molecule has a direct impact on the inhibitory activity as observed in the compound **819** (least active) (see Fig. 10).

3.1.1.5. β-Amyloid aggregation. The major group of positive contributing descriptors involved in the developed equation by PLS is subgroups like SssCH2, PW3, and F04[N–O]. The direct relationship of the descriptor SssCH2 suggested that the presence of the number of such –CH2- groups in the molecules leads to better β-amyloid aggregation inhibitory activity as longer chain compounds with a higher number –CH2- groups would be more lipophilic resulting in improved brain permeability. This assumption can be confirmed by compound **57** having a higher number of –CH2- groups in their structure, showing higher descriptor values leading to their higher range of inhibitory activity. Descriptor, PW3 can be considered as a shape descriptor whose value increases with increased branching in the vertices. The descriptor F04[N–O] denotes the frequency of nitrogen and oxygen atoms at the topological distance 4. These two contribute positively towards inhibitory activity as evident by the compound **59**.Again, descriptors F05[C–O], F02[N–N],and MAXDN contributed negatively which means that the presence of these features in the molecules shows the lower inhibitory activity as observed in the compounds **149**, **262** and **175** respectively (see Fig. 10).

3.1.1.6. Cyclin Dependent Kinase 5 (CDK-5) protein. In this model, NaaaC is the most contributing descriptor toward inhibitory activity. It denotes the number of atoms of type aaaC (π), aromatic fused carbons. The fragment has a positive contribution toward inhibitory activity against protease. A compound like **195** shows higher inhibitory activity due to the presence of a high number of aaaC fragments. The next most important descriptor PBF denotes the Plane of best fit in the compounds having a negative coefficient towards the inhibitory activity. Therefore, for an improved activity, their values must be augmented as low as possible. This statement can be confirmed by compound **146** having a higher descriptors values, showing their lower range of inhibitory activity (see Fig. 11).

3.1.1.7. Gamma-secretase enzyme. As per the VIP plot, P_VSA_MR_7 is the most significant with a positive coefficient feature towards the inhibitory activity in this model. P_VSA_MR_7 belongs to P_VSA-like descriptors that are calculated from the amount of van der Waals surface area (VSA), thus being related to the lipophilic feature. Compound **209** with a high value of P_VSA_MR_7 showed strong inhibitory activity (see Fig. 10). The variables nOHp, B05[N–N], and nRCOOH, contributed negatively suggesting that the presence of these features in the compounds will result in a lower inhibitory activity as observed in the compounds **41**, **207**, and **66** (least active) respectively(see Fig. 10).

3.1.1.8. Glutaminyl Cyclase (QCs) enzyme. The developed QSAR model showed that the features C-034, ETA_Sh_Y, F05[C–S], and PBF are important for defining the inhibitory activity against the QCs enzyme. Lipophilicity is a significant parameter for AD drugs as can be noticed from the variable C-034 which contributed positively towards the inhibitory activity as evidenced by the compound **34** (most active). C-034 enhancing lipophilicity represents the feature R-CR..X, where R is any group linked through carbon; X is an electronegative atom (O, N, S, P, Se, halogens); '-' is an aromatic bond as in benzene or delocalized bonds such as the N, O bond in a nitro group; .. denotes aromatic single bond). In this series, variables ETA_Sh_Y, F05[C–S], and PBF contributed negatively suggesting that the presence of these features in the compounds will result in a lower inhibitory activity as evidenced by the molecules **33**, **31,** and **72** (least active) respectively (see Fig. 10).

3.1.1.9. Glycogen synthase kinase-3β (GSK-3β) enzyme. The first most crucial feature important for the GSK-3β enzyme inhibitory activity was nThiazoles, which designates the number of thiazole rings present in the molecules. The lipophilicity accompanying molecular bulk was the next most significant feature responsible for enzyme inhibitory activity which appeared as SaaaC, SaasC, and PDI. SaaaC is an atom-type E-state index for carbons with three aromatic connections and SaasC is for aromatic carbons with an attached substituent atom. The positive coefficients of SaaaC and SaasC in the model were considered to be a consequence of the importance of aromatic rings in controlling the inhibitory activity of the compounds, both in determining the compound's hydrophobicity and its π-π interactions with the target as witnessed in the molecular docking study. The packing density index (PDI) is a molecular property descriptor. PDI is designated as the ratio between the McGowan volume and the total surface area [44]. The positive sign of the coefficient of these descriptors confirms that the larger the value of descriptors higher is the inhibitory activity towards the enzyme as witnessed by the compound **153** (most active). Again, the next significant descriptor mindsCH designates the minimum atom-type E-State: =CH- contributed negatively means the presence of this feature in the compounds shows the lower inhibitory activity as observed in the compound **161** (least active) (see Fig. 11).

3.1.1.10. Monoamine oxidase B (MAO-B) enzyme. As per the VIP score, mindsCH is the most significant descriptor that appeared in this equation which contributed negatively towards the inhibitory activity as observed in the compound **25** (least active). This feature also appeared in the previous model (GSK-3β) with a negative coefficient toward inhibitory activity. In this series next most significant descriptors are F02[O–O], C%, and nRCN. The positive regression coefficient sign confirms that the inhibitory activity increases with the presence of the above fragments which can be verified by the most active compounds **86** (in case of F02[O–O] and C%) and **72** (in case of nRCN) in their structure (see Fig. 11). The other significant features were polar oxygen groups (B06[O–Cl], B01[C–O]), molecular branching (Eta_B), secondary aromatic amine (nArNHR), minssCH2, logP, and F05[C–N] contributed negatively suggesting that the presence of these features in the compounds will result in lower inhibitory activity (see Table S1 in Supporting Information SI-2).

3.1.1.11. N-methyl-D-aspartate (NMDA) receptor. The higher concentration of electronegative atom count in a compound has a direct impact on the lowering antagonistic activity against the target protein. This hypothesis can be confirmed from the presence of variables SaaN, TPSA (NO), C-043, ATSC2m, and MDEN-23 contributed negatively as evidenced by the compound **483** and **446** (least active). According to the VIP scores, SaaN is the most contributing descriptor in the model which encodes information about both the topological environment of the particular atom and the electronic interactions due to all other atoms in the molecule. SaaN is the sum of E-State values of all nitrogen atoms with two aromatic bonds found in the molecule (see Fig. 11).

3.1.1.12. Phosphodiester 10A (PDE 10A) enzyme. The leading group of positive contributing descriptors involved in the developed model is

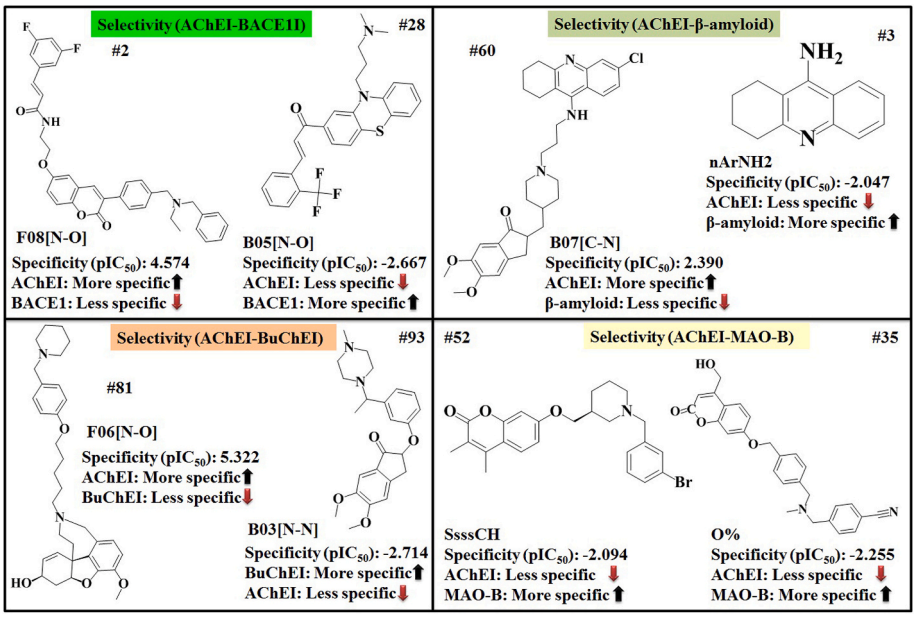


Fig. 12. Mechanistic interpretation of the most significant descriptors obtained from the selectivity-based (AChE-BACE1, AChE-β-Amyloid, AChE-BuChE, and AChE-MAO-B) models.

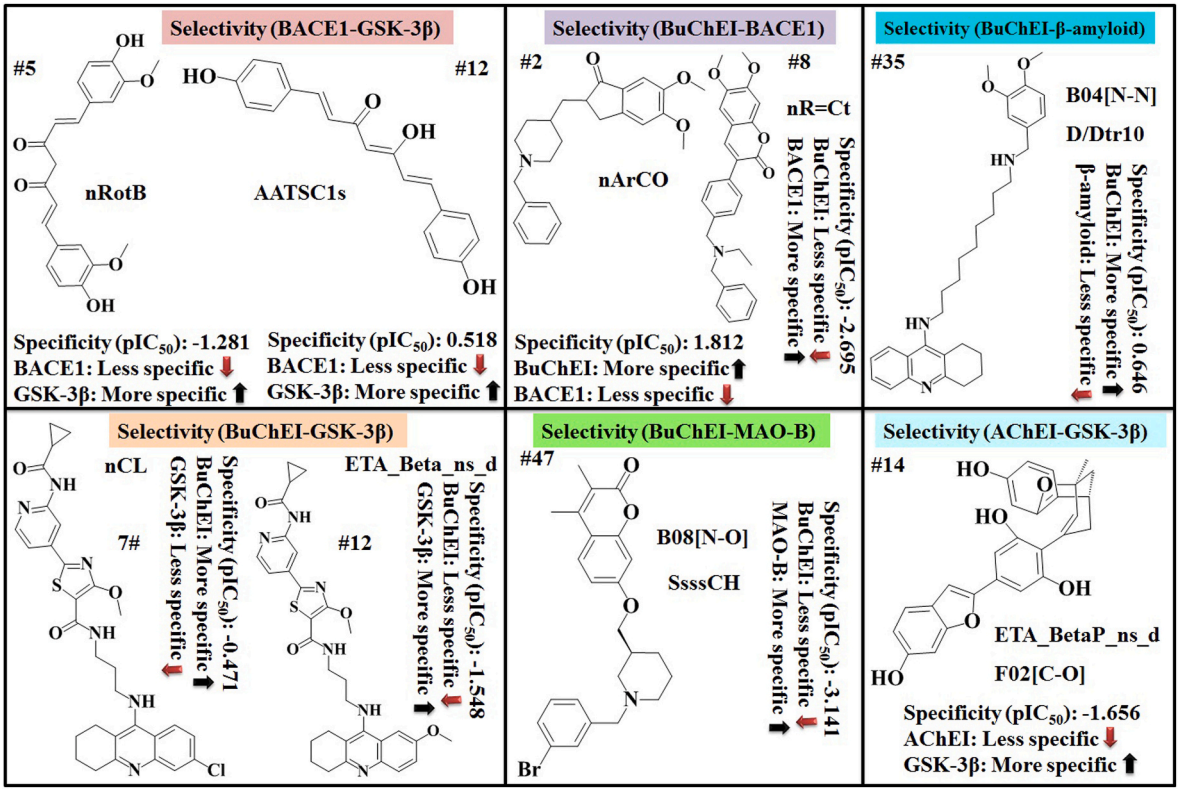


Fig. 13. Probable mechanistic interpretation of the most significant descriptors obtained from the selectivity-based (BACE1-GSK-3β, BuChE-BACE1, BuChE-β-Amyloid, BuChE-GSK-3β, BuChE-MAO-B, and AChE-GSK-3β) models.

subgroups like N-070, and AMR. The direct association of the descriptor N-070 suggested that the presence of the number of N atoms in the Ar-NH-Al group with Al representing aliphatic groups in the compounds leads to better enzyme inhibitory activity. The hydrophobicity associated with molecular bulk was the next most significant feature responsible for the inhibitory activity which appeared as AMR variables, as higher the features in the compounds higher would be the lipophilicity resulting in improved brain permeability. This concept can be confirmed from compound **226** (most active) (see Fig. 11).

3.1.2. Mechanistic interpretation of the descriptors involved in the development of selectivity-based models

The details about all variables that appeared in these models, their contribution, meaning of descriptors, examples of compounds, and their

corresponding descriptors values are given in Table S2 (see Supporting Information SI-2).

3.1.2.1. Selectivity of AChE and BACE1 enzyme inhibitors. According to the VIP score, F08 [N–O] is the most significant positively contributing descriptor in the model for more specific inhibition to AChE enzyme than BACE1 enzyme inhibitory activity. Thus, the higher number of this fragment leads to better AChE enzyme inhibitory activity as noticed in the case of compound **2**. Again, the variable B05[N–O] contributed negatively is more specific to BACE1 enzyme inhibitory activity than AChE enzyme inhibitory activity. So, the higher number of this fragment correlates with lower AChE enzyme inhibitory activity as observed in the case of compound **28** (see Fig. 12).

3.1.2.2. Selectivity of AChE enzyme and β-amyloid aggregation inhibitors. The important group of the positive contributing variables more specific to the AChE enzyme inhibitory activity involved in the developed model is subgroups like B07[C–N], F06[C–Cl], SAscore, and ETA_Shape_Y. In these features, ETA_Shape_Y signifies the measure of molecular shape. The straight connotation of these descriptors suggested that the presence of above groups in the compounds leads to better AChE enzyme inhibitory activity. This concept can be confirmed from compound **60** (most active). The next significant features of negative contributing more specific to β-amyloid aggregation inhibitory activity than AChE enzyme inhibitory activity appeared in the developed equation are subgroups like nArNH2, F10[C–Cl], and PJI2. In this series, variable nArNH2 signifies the presence of the number of primary aromatic amines in the compounds. The above specification can be confirmed from compounds **3** (in the case of nArNH2) and **10** respectively (see Fig. 12).

3.1.2.3. Selectivity of AChE and BuChE enzyme inhibitors. Among the crucial features improving AChE enzyme inhibitory activity are F06 [N–O] and nR12, as lipophilicity is an important parameter for the AD drugs and can be outlined from variable nR12. Thus, the higher number of these fragments leads to better AChE enzyme inhibitory activity as noticed in the case of compound **81** (most active). Again, variables constituting electronegative atoms (B03[N–N], B03[N–O], nROR, N%, and GD) contributed negatively suggesting that the presence of these features in the compounds more specific to BuChE inhibitory activity as observed in the compounds **93**, **92** and **94** respectively (see Fig. 12).

3.1.2.4. Selectivity of AChE and MOA-B enzyme inhibitors. The most significant features enhancing lipophilicity (SsssCH, and F10[C–C]), and constituting electronegative and polar atoms (O%, F10[C–N], and minsCH3) contributed negatively suggesting that the presence of these features in the compounds are more favorable for MOA-B enzyme inhibitory activity as evidenced by the compounds **52, 44, 25** and **35**. In this series, the next most significant descriptors that contributed positively are F09[C–C], and MaxsCH3, which are more specific to the AChE enzyme inhibitory activity than MOA-B enzyme inhibitory activity. Thus, the higher number of this fragment leads to better AChE enzyme inhibitory activity as observed in compounds **27** and **18** respectively (see Fig. 12).

3.1.2.5. Selectivity of BACE1 and GSK-3β enzyme inhibitors. The significant descriptors in this series are nRotB and AATSC1s which contributed negatively toward the inhibitory activity against BACE1 enzyme; accordingly, these features are more specific to GSK-3β enzyme inhibitory activity. Thus, the higher number of this fragment shows lower BACE1 enzyme inhibitory activity as seen in compounds **5** and **12** respectively (see Fig. 13).

3.1.2.6. Selectivity of BuChE and BACE1 enzyme inhibitors. In the MLR equation, variables NaasC, minaaaC, nArCO, and ZM1Kup contributing positively were the most significant descriptors toward the inhibitory activity against the BuChE enzyme, which means the presence of these features in the compounds are most specific to the BuChE enzyme inhibitory activity than BACE1 as observed in the molecules **7, 2,** and **3** respectively. While variables constituting lipophilicity (ALOGP2, and nR = Ct) contributed negatively suggesting that the presence of these features in the molecules is more specific to BACE1 inhibitory activity as observed in compound **8** (see Fig. 13).

3.1.2.7. Selectivity of BuChE enzyme and β-amyloid aggregation inhibitors. The higher concentration of electronegative atom count and aromaticity in a compound has a direct impact on the better inhibitory activity against the BuChE enzyme. This premise can be confirmed from the presence of variables D/Dtr10 (aromatic ring count) and B04[N–N], B10[C–O], B08[C–N], ETA_dEpsilon_B (electronegative atom counts) contributed positively as evidenced by the compounds **35, 1** and **4** (most active) respectively. The other features B10[C–N], LOGP99, B10[N–O], PJI2, B08[O–O], and ETA_Shape_Y contributed negatively to the model are more specific to the β-amyloid aggregation inhibitory activity than BuChE enzyme inhibitory activity as observed in the compounds **124, 118, 120, and 14** respectively(see Fig. 13).

3.1.2.8. Selectivity of BuChE and MOA-B enzyme inhibitors. The molecules constituting variables containing tertiary nitrogen atoms (SsssN) and nitrogen and oxygen atoms at the topological distance 6 (B06[N–O]) are more specific to the BuChE enzyme inhibitory activity than the MOA-B enzyme inhibitory activity as seen in the compound **23**. The variables imparting lipophilicity (minsCH3, SsssCH, NdssC, and SAacc) and electronegativity (B08[N–O]) of the compound are more specific to the MOA-B enzyme inhibitory activity than the BuChE enzyme inhibitory activity as observed in the compounds **4, 47**, and **51**(see Fig. 13).

3.1.2.9. Selectivity of AChE and GSK-3β enzyme inhibitors. The variable Psi_i_A contributed positively suggesting that the presence of such fragment in the molecule is more specific to the AChE enzyme inhibitory activity than GSK-3β enzyme inhibitory activity as seen in the compound **10**. The negatively contributing descriptors ETA_BetaP_ns_d, and F02 [C–O] are more specific to the GSK-3β enzyme inhibitory activity than the AChE enzyme inhibitory activity as observed in the compound **14** (see Fig. 13).

3.1.2.10. Selectivity of BuChE and GSK-3β enzyme inhibitors. As we can see in both selectivity-based models, we found similar features contributed accordingly to the inhibitory activity. The variables Psi_i_A and nCL suggested that the presence of such fragment in the compounds is more specific to the BuChE enzyme inhibitory activity as seen in compounds **13** and **7** respectively. As the previous model descriptor, ETA_Beta_ns_d contributed negatively, the same effect exerted in this equation also suggests that the presence of this feature in the molecule is more specific to the GSK-3β enzyme inhibitory activity than BuChE enzyme inhibitory activity as observed in the compound **12** (see Fig. 13).

3.1.3. Mechanistic interpretation of the descriptors involved in the development of QSAAR-based models

The QSAAR models are the *in silico* analysis of activity-activity correlations that can be used to evaluate the dual inhibitory activity of a molecule. The QSAAR analysis usually comprises the use of activity values of one of the endpoint as the response variable (i.e., Y-variable), whereas the activity to the other endpoint is used as one of the predictor variables (i.e., X-variable). If the activity values of the identified molecules for one endpoint correspond strongly to the values for another endpoint, the chemicals are likely to have a similar mode of action for both and vice versa. The importance of different descriptors (as specified by the VIP plots) is illustrated in Table 3 and provide a collection of equations for QSAAR models. The detailed description of modeled descriptors that appeared in the QSAAR models is given in Table S3 (See

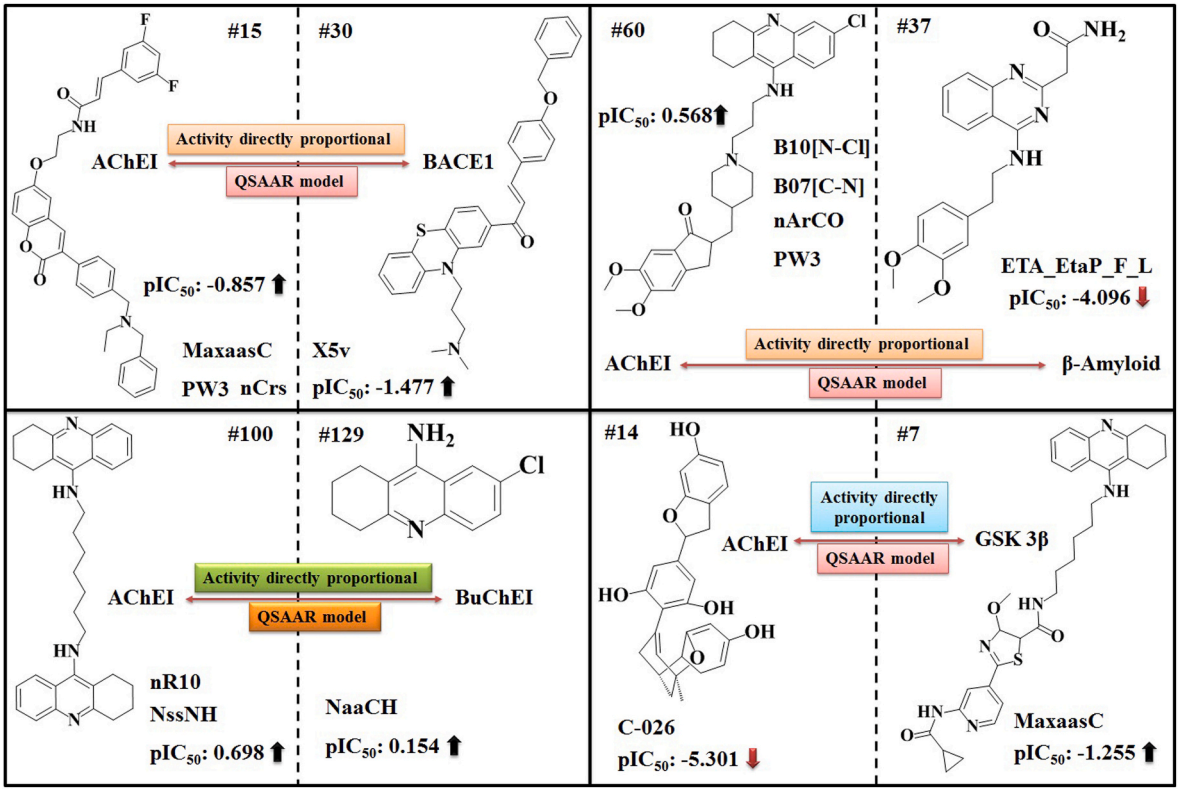


Fig. 14. Probable mechanistic interpretation of the most significant descriptors obtained from the QSAAR (AChE (X)-BACE1 (Y), BACE1 (X)-AChE (Y), β-Amyloid (X)-AChE (Y), AChE (X)-β-Amyloid (Y), BuChE (X)-AChE (Y) and AChE (X)-BuChE (Y), GSK-3β (X)-AChE (Y), and AChE (X)-GSK-3β (Y)) models.

Supporting Information SI-2). This section provides a concise and more relevant description of the mechanistic interpretation of various QSAAR models.

3.1.3.1. BACE1 enzyme inhibitory activity as a predictor (X) and AChE enzyme inhibitory activity as the response (Y) and vice versa. The activity to both targets is shown to be the most significant predictor variable, positively contributing to each other. Other important descriptors in the BACE1 (X)-AChE (Y) and AChE (X)-BACE1 (Y) models include nCrs, MaxaasC, PW3, X5v, B08[O–O] (having positive regression coefficients), C-001, nCconj and nArOH (having negative regression coefficients) respectively, as established by the VIP plots. The next most important inhibitory activity enhancing features identified were hydrophobicity, molecular bulk, and electronegativity. The descriptors proving this hypothesis were hydrophobicity variables (nCrs, and MaxaasC), molecular bulk (PW3), electronegativity (B08[O–O]), and size index (X5v). Thus, the presence of these features in the compounds leads to better inhibitory activity (see Fig. 14).

3.1.3.2. AChE enzyme inhibitory activity as a predictor (X) and β-amyloid inhibitory activity as a response (Y) and vice versa. The activity terms (both contributing positively) are the highest contributing descriptors by the corresponding VIP and loading plots. In the β-amyloid (X)-AChE (Y) model, all the identified variables (nArCO, B10[N–Cl], PW3, and B07[C–N]) contributed positively toward inhibitory activity. The other recognized significant features were B10[N–Cl] and B07[C–N] having positive regression coefficients in the β-amyloid (X)-AChE (Y) model while negative regression coefficients in AChE (X)-β-amyloid (Y) model, as per the VIP plot. The remaining two descriptors of the AChE (X)-β-amyloid (Y) models are SAscore and ETA_EtaP_F_L inversely correlated to the response as shown by their negative regression coefficients (see Fig. 14).

3.1.3.3. BuChE enzyme inhibitory activity as a predictor (X) and AChE enzyme inhibitory activity as the response (Y) and vice versa. The activities against both the targets are directly related (both having positive regression coefficients) and the most relevant predictor variables for the respective QSAAR models. In the BuChE (X)-AChE (Y) model, the essential features enhancing inhibitory activity were lipophilicity (nR10, and nCr), electronegative atoms (NssNH, B04[N–O], F06[N–O] and B04[N–Cl]) and other feature like ICR (radial centric information index). As per the regression coefficient plot, the presence of these features in the compounds leads to better inhibitory activity. The features contributing negatively were F08[C–N], and nROR which indicate that the decrease in the values of these variables affects an increase in the inhibitory activity. In the AChE (X)-BuChE (Y) model, the variable constituting the lipophilicity (NaaCH), shape (DECC) and electronegative atoms (B10[N–O]) are significant for the enhancing inhibitory activity. Apart from the above descriptors, the other identified variables were F06[N–O], MaxsCH3, ETA_EtaP_F, nCrs, Ram, and nCb-inversely correlated to the response as suggested by their negative regression coefficients (see Fig. 14).

3.1.3.4. GSK-3β enzyme inhibitory as a predictor (X) and AChE enzyme inhibitory activity as the response (Y) and vice versa. The responses against both the targets are shown to be the most significant predictor variables, contributing positively to each other. Other significant variables in the GSK-3β (X)-AChE(Y) and AChE (X)-GSK-3β (Y) models comprise C-026 (having negative regression coefficients), ETA_Epsilon_1 and MaxaasC (having positive regression coefficients) respectively, as established by the VIP plots. C-026 inversely correlated to the response. Furthermore, variables constituting fragments imparting lipophilicity (MaxaasC) and electronegative atoms (ETA_Epsilon_1) in the molecules lead to an improved inhibitory activity (see Fig. 14).

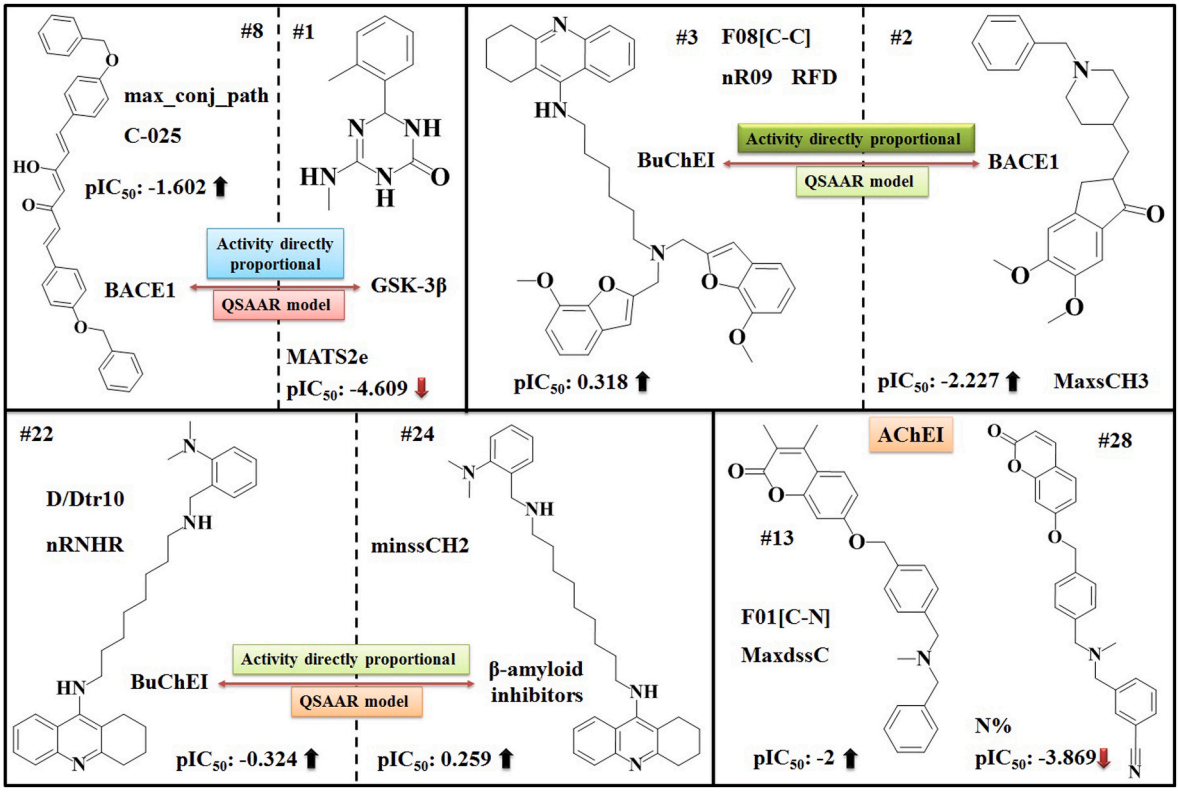


Fig. 15. Probable mechanistic interpretation of the most significant descriptors obtained from the QSAAR (GSK-3β (X)-BACE1 (Y), BACE1 (X)-GSK-3β (Y), BuChE (X)-BACE1 (Y), BACE1 (X)-BuChE (Y), BuChE (X)-β-Amyloid (Y), β-Amyloid (X)-BuChE (Y) and MAO-B (X)-AChE (Y)) models.

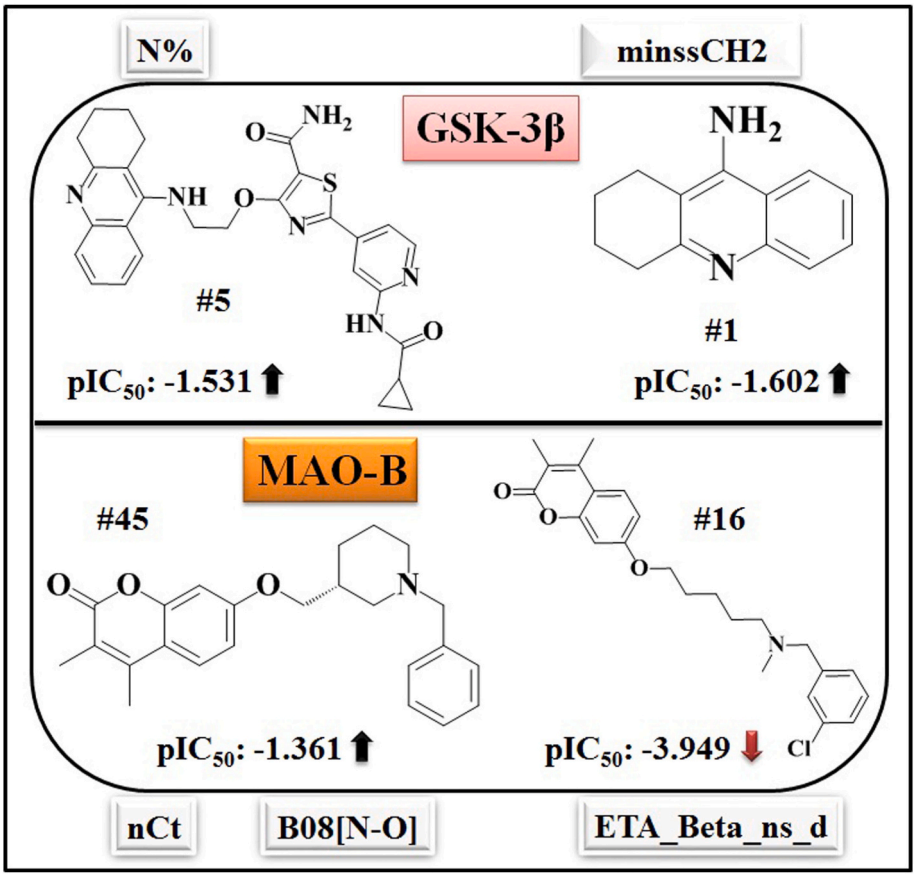


Fig. 16. Possible mechanistic interpretation of the most significant descriptors obtained from the QSAAR (BuChE (X)-GSK-3β (Y) and BuChE (X)-MAO-B (Y)) models.

3.1.3.5. MAO-B enzyme inhibitory as a predictor (X) and AChE enzyme inhibitory activity as the response (Y). The positive regression coefficient of the activity variable specifies that the MAO-B enzyme inhibitory value is proportionally related with the AChE enzyme inhibitory activity (response). In this series, the most significant variables (having positive regression coefficients) were F01[C–N], MaxdssC, and B05[N–Cl], enhancing the inhibitory activity. The other recognized significant features were N%, MAXDP, ETA_Epsilon_5, and MaxaaCH inversely associated with the response as suggested by their negative regression coefficients (see Fig. 15).

3.1.3.6. GSK-3β enzyme inhibitory as a predictor (X) and BACE1 enzyme inhibitory activity as the response (Y) and vice versa. The activity end points of both targets (with positive regression coefficients) contributed the most to the development of the respective models. The QSAAR model for GSK-3β (X)-BACE1 (Y) includes max_conj_path and C-025, which contributed positively towards inhibitory activity. The last variable in this model was minsOH having a negative regression coefficient towards the inhibitory activity. In the BACE1 (X)-GSK-3β (Y) model, the identified important features were MAXDN (having positive regression coefficient) signifying the maximal electrotopological negative variation in the compounds. The other recognized important variables were MATS2e, ZMIC4, and AATS3*s* inversely associated toward inhibitory activity (see Fig. 15).

3.1.3.7. BuChE enzyme inhibitory as a predictor (X) and BACE1 enzyme inhibitory activity as the response (Y) and vice versa. Similar to the aforementioned QSAAR models delineated above, the activity end points of both targets (with positive regression coefficients) contributed the most to the development of the respective models. The significant variables in the BuChE(X)-BACE1(Y) model includes MaxsCH3, nFuranes, and B05[O–O], having positive regression coefficients, thus, the presence of these features in the compounds leads to better inhibitory activity. Again, the variables F01[C–O], and nCIC have negative regression coefficients, and inversely affect the inhibitory activity. In the BACE1(X)-BuChE (Y) model, the second most significant inhibitory activity improving feature was identified as hydrophobicity. The variable evidencing this premise was hydrophobicity variables (RFD, F08[C–C], and nR09), thus, the presence of these fragments in the molecules leads to enhanced inhibitory activity (see Fig. 15).

3.1.3.8. BuChE enzyme inhibitory as a predictor (X) and β-amyloid inhibitory activity as the response (Y) and vice versa. The inhibitory activities against both the targets are shown to be the most important predictor variables, contributing positively to each other. Amongst the essential features enhancing inhibitory were hydrophobic moieties (SaasC, F08[C–C], Uc, and nCrt), size (D/Dtr10), constituting secondary aliphatic amines (nRNHR), and hybrid group (minssCH2), means the presence of these fragments in the molecules leads to improved inhibitory activity. In contrast, fragments with inhibitory activity lowering potential against the enzyme were ETA_EtaP_F, F06[C–C], nCp, and nAB (see Fig. 15).

3.1.3.9. BuChE enzyme inhibitory as predictor (X) and GSK-3β enzyme inhibitory activity as the response (Y). The positive regression coefficient of the activity variable indicates that the BuChE enzyme inhibitory value is proportionally related to the GSK-3 enzyme inhibitory activity (the endpoint). The other important variables in this model were N% and minssCH2, having positive regression coefficients towards the inhibitory activity (see Fig. 16).

3.1.3.10. BuChE enzyme inhibitory as predictor (X) and MAO-B enzyme inhibitory activity as the response (Y). The positive regression coefficient of the activity variable suggests that the BuChE enzyme inhibitory value is consistently connected to the MAO-B enzyme inhibitory activity (endpoint). The essential features improving inhibitory activity were hydrophobic moiety (nCt), polar moieties (nROH, F06[C–O], and B08 [N–O]), constituting secondary aliphatic amines (nRNHR) and hybrid group (minsCH3), which suggests that the presence of these features in the compounds leads to better inhibitory activity. On the other hand, fragments with inhibitory activity lowering potential against the enzyme is ETA_Beta_ns_d (see Fig. 16).

3.1.4. Applicability domain (AD) study

In the current investigation, we have applied the DModX approach [45,46] at the 99% confidence level using SIMCA-P version 10.0 software to perform the AD evaluation of established PLS models, whereas the AD of the developed MLR models was estimated using the standardization approach [47] with the help of tool developed in our laboratory (http://teqip.jdvu.ac.in/QSAR Tools/).

3.1.4.1. Individual QSAR models. The DModX outcomes of the PLS models revealed that most of the models (against 5-HT6, β-amyloid aggregation, GSK-3β enzyme, and MAO-B enzyme) had no compound in the test set outside the AD (see Figs. S134, 137, 142 and 143 in the Supporting Information SI-2). However, in case of AChE enzyme model (see Fig. S135 in the Supporting Information SI-2), we found 48 compounds (i.e. compounds **241, 300, 364, 412, 458, 520, 655, 720, 764, 765, 776, 814, 839, 1077, 1115, 1135, 1138, 1141, 1212, 1216, 1271, 1275, 1299, 1300, 1338, 1352, 1357, 1364, 1366, 1368, 1371, 1379, 1398, 1399, 1411, 1463, 1651, 1652, 1656, 1673, 1679, 1680, 1680, 1688, 1689, 1690, 1693,** and **1695**) in the test set are located outside the AD. In case of BuChE enzyme model (see Fig. S138 in the Supporting Information SI-2), we found 57 compounds (i.e. compounds **186, 210, 255, 257, 281, 294, 419, 438, 451, 478, 580, 635, 637, 641, 645, 651, 807, 879, 928, 933, 955, 966, 970, 972, 981, 1003, 1006, 1012, 1015, 1016, 1405, 1580, 1582, 1617, 1733, 1793, 1797, 1799, 1802, 1817, 1819, 1914, 1928, 2052, 2104, 2156, 2216, 2224, 2237, 2242, 2249, 2256, 2257, 2354, 2374, 2386, and 2387**) in the test set are placed outside the AD. In the case of the BACE1 enzyme model (see Fig. S136 in the Supporting Information SI-2), we found 16 compounds (i.e. compounds **7, 23, 33, 147, 336, 409, 464, 574, 684, 695, 788, 791, 820, 824, 830, and 857**) in the test set are positioned outside the AD. In the case of the CDK-5 protein model (see Fig. S139 in the Supporting Information SI-2), we found only 3 compounds (i.e. compounds **32, 184,** and **186**) in the test set are situated outside the AD. In the case of the Gamma-secretase enzyme model (see Fig. S140 in the Supporting Information SI-2), we found only 2 compounds (i.e. compounds **76** and **80**) in the test set are positioned outside the AD. In the case of the Glutaminyl Cyclase (QCs) enzyme model (see Fig. S141in the Supporting Information SI-2), we found only 1 compound (i.e. compound **3**) in the test set is located outside the AD. In the case of the NMDA receptor model (see Fig. S144 in the Supporting Information SI-2), we found only 3 compounds (i.e. compounds **107**, **338,** and **434**) in the test set are placed outside the AD. In the case of the PDE 10A enzyme model (see Fig. S145 in the Supporting Information SI-2), we found that 5 compounds (i.e. compounds **58, 79, 182, 204,** and **206**) in the test set are located outside the AD.

3.1.4.2. Selectivity-based QSAR models. The AD results of the models revealed that the majority of the models (BACE1-GSK-3β enzyme inhibitors, AChE-GSK-3β enzyme inhibitors, BuChE-GSK-3β enzyme inhibitors, AChE-BACE1 enzyme inhibitors, AChE enzyme-β-amyloid aggregation inhibitors, and AChE-MOA-B enzyme inhibitors) had no molecules in the test set outside the AD (see sheet 33 and Figs. S146, 147, 149 respectively in the Supporting Information SI-1and 2). However, in the case of the AChE-BuChE enzyme inhibitors model (see Fig. S148 in the Supporting Information SI-2), we found only 2 compounds (i.e. compounds **29** and **137**) in the test set are positioned outside the AD. In the case of the BuChE-BACE1 enzyme inhibitors

Table 4
List of identified top lead compounds from the databases with multitarget inhibitory activity using developed models.

Compound ID	Smiles	Targets and their predicted inhibitory activity (IC_{50} nM)			
STOCK7S-62250	CC(C1CCC(CC1)(C)c1cc2Cc3cc(c(cc3O)Cc3cc(O)c(cc3C3(C)CCC(CC3)C(C)C)Cc3cc(c(cc3O)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2O)c(c6)C1(C)CCC(CC1)C(C)C)O)c(O)c5)C1(C)CCC(CC1)C(C)C)c(c4)C1(C)CCC(CC1)C(C)C)O)cc3C1(C)CCC(CC1)C(C)C)O)C1(C)CCC(CC1)C(C)C)C1(C)CCC(CC1)C(C)C)C	**AChE** 0.000474	**BACE1** 2.663E-08	**CDK-5** 0.00243	**GSI** 0.0282
118199	O=CN(C)C(=S)S[Sb](SC(=S)N(C=O)C)SC(=S)Nc1ccc(cc1)S(=O)(=O)Nc1ccc(/N=N/c2ccccc2)cc1	**AChE** 0.00157	**BACE1** 2.794E-05	**GSI** 0.0097	**BuChE** 7.870
118200	S(C(=S)NC(=O)OCC)[Sb](SC(=S)NC(=O)OCC)SC(=S)Nc1ccc(cc1)S(=O)(=O)Nc1ccc(/N=N/c2ccccc2)cc1	0.00108	0.00061	7.97E-05	1.629
BAS 01060168	c12c3c(C(c4c5c6c(cc4)c(ccc6c(n[nH]5)C)OC)c4ccccc4)ccc1c(ccc2c(n[nH]3)C)OC	**5-HT6** 0.0119	**BuChE** 1.875	**CDK-5** 0.059	
BAS 01060169	C(c1c2c3c(cc1)c(ccc3c(n[nH]2)C)OC)(c1c2c3c(cc1)c(ccc3c(n[nH]2)C)OC)c1ccc(cc1)OC	0.0087	1.954	0.046	
STOCK7S-67258	CCCC[Sn](OC(=O)CCCC(N1C(=O)C(=C(C1=O)Cl)Cl)C(=O)O[Sn](CCCC)(CCCC)CCCC)(CCCC)CCCC	**AChE** 6.2906E-06	**BACE1** 6.96627E-07	**GSI** 0.0948	
STOCK7S-70271	CCCCCCC1c2cc3c4cc2OP(=S)(Oc2c1cc1c(c2)OP(=S)(Oc2c(C1CCCCCC)cc1C(c5cc(C3CCCCCC)c(OP(=S)(O4)N(C)C)cc5OP(=S)(Oc1c2)N(C)C)CCCCCC)N(C)C)N(C)C	0.00986	0.000221	2.486E-05	
127500	S(C(=S)NC(=O)OCC)[Sb](SC(=S)NC(=O)OCC)Sc1nc2c(s1)cc(/N=N/c1ccc(cc1)NS(=O)(=O)c1ccc(cc1)NC(=O)C(N1CC1)(N1CC1)N1CC1)cc2	8.182E-05	5.273E-06	0.001002	
381279	[Si](N([Si](C)(C)C)[Sn](C(Cl)(Cl)Cl)(Cl)N([Si](C)(C)C)[Si](C)(C)C)(C)(C)C	1.5761E-06	1.0843E-25	0.062	
633331	P1([Fe+2](P(c2ccccc2)(c2ccccc2)C=C1)([I-])([C-]#[O+])([C-](C(F)(F)F)(C(F)(F)F)F)[C-]#[O+])(c1ccccc1)c1ccccc1	2.878E-07	1.298E-23	0.1194	
633993	P1([Fe+2](P(CC1)(c1ccccc1)c1ccccc1)([C-]#[O+])([C-]#[O+])([I-])[C-](C(F)(F)F)(F)C(F)(F)F)(c1ccccc1)c1ccccc1	1.637E-07	2.149E-28	0.119	
635009	[Mn+]1([O-]C(=O)C(F)(F)F)([C-]#[O+])([C-]#[O+])([C-]#[O+])P(C=CP1(c1ccccc1)c1ccccc1)(c1ccccc1)c1ccccc1	0.00635	1.429E-19	0.1122	
643862	[Sn](CCC)(O[Sn](CCC)(OC(=O)c1c(nccc1)SC)CCC)(OC(=O)c1cccnc1SC)CCC	0.00505	6.882E-13	0.0073	
677704	c1c2c(ccc1[N+](=O)[O-])[O-]1[Cu+2]3([N](=C2)[N-]C(=[S]3)NCC)[O-]2c3c(cc(cc3)[N+](=O)[O-])C = [N]3[N-]C(=[S][Cu+2]123)NCC	0.00131	3.599E-09	0.0320	
677707	c1c2c(ccc1[N+](=O)[O-])[O-]1[Cu+2]3([N](=C2)[N-]C(=[S]3)N(CC)CC)[O-]2c3c(cc(cc3)[N+](=O)[O-])C = [N]3[N-]C(=[S][Cu+2]123)N(CC)CC	0.00021	1.3122E-09	0.0182	
28796–39–6	O=CN(C)C(=S)S[Sb](SC(=S)N(C=O)C)Sc1nc2c(s1)cc(/N=N/c1ccc(cc1)NS(=O)(=O)c1ccc(cc1)NC(=O)C(N1CC1)(N1CC1)N1CC1)cc2	0.00036	2.548E-07	0.112	
29878–72–6	[Sn](c1ccccc1)(c1ccccc1)([Sn](c1ccccc1)(c1ccccc1)OC(=O)CCl)OC(=O)CCl	4.6586E-07	3.2537E-27	0.089	
69272–27–1	c1c(c(c(nc1C)S[Sn](Cl)(Cl)Sc1c(C(=O)OCC)c(cc(n1)C)C)C(=O)OCC)C	8.9729E-08	0.00046	0.1039	
82475–53–4	ClCCN(P(=O)(O[Si](c1ccccc1)(C(C)(C)C)c1ccccc1)O[Si](c1ccccc1)(C(C)(C)C)c1ccccc1)CCCl	0.0034	1.1551E-06	0.026	
STOCK7S-64595	CCCCC1(CCCC1)c1cc2Cc3cc(c(cc3O)Cc3cc(O)c(cc3C3(CCCC)CCCC3)Cc3cc(c(cc3O)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2O)c(c6)C1(CCCC)CCCC1)O)c(O)c5)C1(CCCC)CCCC1)c(c4)C1(CCCC)CCCC1)O)cc3C1(CCCC)CCCC1)O)C1(CCCC)CCCC1)C1(CCCC)CCCC1	**BACE1** 1.7698E-07	**GSI** 0.024	**CDK-5** 0.00105	
STOCK7S-65432	Oc1cc2Cc3cc(O)c(cc3C3(C)CCCC3)Cc3cc(c(cc3O)Cc3cc(O)c(cc3C3(C)CCCC3)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2C1(C)CCCC1)c(O)c6)C1(C)CCCC1)c(c5)C1(C)CCCC1)O)c(O)c4)C1(C)CCCC1)cc3O)C1(C)CCCC1)C1(C)CCCC1	6.2808E-05	0.069	0.0028	
STOCK7S-67132	NC(=O)COc1cc2Cc3cc(OCC(=O)N)c(cc3C3(C)CCCCC3)Cc3cc(c(cc3OCC(=O)N)Cc3cc(OCC(=O)N)c(cc3C3(C)CCCCC3)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2C1(C)CCCCC1)c(OCC(=O)N)c6)C1(C)CCCCC1)c(c5)C1(C)CCCCC1)OCC(=O)N)c(OCC(=O)N)c4)C1(C)CCCCC1)cc3OCC(=O)N)C1(C)CCCCC1)C1(C)CCCCC1	0.00012	0.0248	0.00244	
STOCK7S-67182	Oc1cc2Cc3cc(O)c(cc3C3(C)CCCCC3)Cc3cc(cc(c3O)Cc3cc(c(Cc4c(cc(Cc5c(c(Cc1cc2C1(C)CCCCC1)cc(c5)C1(C)CCCCC1)O)c(O)c4)C1(C)CCCCC1)cc3O)C1(C)CCCCC1)C1(C)CCCCC1	0.000313	0.1742	0.0732	
STOCK7S-67536	Oc1cc2Cc3cc(O)c(cc3C3(C)CCCCC3)Cc3cc(c(cc3O)Cc3cc(O)c(cc3C3(C)CCCCC3)Cc3cc(c(Cc4c(cc(Cc5c(cc(Cc6c(cc(Cc1cc2C1(C)CCCCC1)c(O)c6)C1(C)CCCCC1)c(c5)C1(C)CCCCC1)O)c(O)c4)C1(C)CCCCC1)cc3O)C1(C)CCCCC1)C1(C)CCCCC1	9.8853E-07	0.0513	0.00241	
STOCK7S-70545	CCCCCCCC1c2cc3c4cc2OP(Oc2c1cc1c(c2)OP(OCCCCCCCC)Oc2c(C1CCCCCC)cc1C(c5cc(C3CCCCCC)c(OP(O4)OCCCCCCCC)cc5OP(Oc1c2)OCCCCCCCC)CCCCCCC)OCCCCCCCC	0.00174	0.1682	0.00034	
619178	n12ccc(c3c1c1n(ccc(c1cc3)c1ccccc1)[Co+3]132(n2ccc(c4c2c2n1ccc(c2cc4)c1ccccc1)c1ccccc1)n1ccc(c2c1c1n3ccc(c1cc2)c1ccccc1)c1ccccc1)c1ccccc1.O=C([O-])[C@@H]([O-])[C@@H]([O-])C(=O)[O-]	6.6985E-05	0.0053	0.043	
STOCK7S-65614	CCCCOc1ccc(cc1)/C/1 = C/2\C=CC(=N2)/C(=c\2/cc/c(=C(/C3=N/C(=C(\c4[nH]c1cc4)/c1ccc(cc1)OCCCC)/C=C3)\c1ccc(cc1)OCCCC)/[nH]2)/c1ccc(cc1)OCCCC	**GSI** 0.161	**CDK-5** 0.0044	**BuChE** 1.819	
332889	O(C(=O)Nc1nc(c(c(c1)Cl)N=C(c1ccccc1)c1ccccc1)N)CC.O(C(=O)Nc1nc(c(c(c1)Cl)N=C(c1ccccc1)c1ccccc1)N)CCC	5-HT6 0.00045	AChE 0.0062	GSI 2.848E-05	
618825	P(c1ccccc1)(c1ccccc1)(c1ccccc1)[Cu+](P(c1ccccc1)(c1ccccc1)c1ccccc1)(/[S] = c\1/cccc[nH]1)[Cl-]	**5-HT6** 0.0079	**AChE** 0.000171	**BACE1** 7.793E-17	
1262–78–8	[Sn](CC(c1ccccc1)(C)C)(CC(c1ccccc1)(C)C)(CC(c1ccccc1)(C)C)CC(c1ccccc1)(C)C	0.0082	0.0049	0.00045	
60042–87–7	c1c(cc2c(c1C(C)(C)C)O[Sb](N2)(c1ccccc1)(c1ccccc1)c1ccccc1)C(C)(C)C	9.0369E-06	0.0081	0.0164	
60042–88–8	C(c1cc(c2c(c1)N[Sb](O2)(c1ccccc1)(c1ccccc1)c1ccccc1)C(C)(C)C)(c1ccccc1)(c1ccccc1)c1ccccc1	6.0399E-06	0.00498	0.028	
81928–48–5	c1cc(c(cc1)[Sn](c1ccccc1C)(Cl)O[Sn](c1ccccc1C)(Cl)c1ccccc1C)C	0.0039	0.00059	3.674E-14	
5424–36–2	[Sn](c1cccc2ccccc12)(c1c2c(cccc2)ccc1)(c1cccc2ccccc12)c1c2ccccc2ccc1	**5-HT6** 0.0112	**BACE1** 0.0056	**CDK-5** 0.00155	
26246	C12 = C3C4 = C5[C-]1[Fe+2]16782345C2 = C7[C-]8C6 = C12.II	**AChE** 0.00134	**BACE1** 0.000212	**CDK-5** 0.0681	
176220	[C-]12C3 = C4C5 = C1[Fe+2]16782345[C-]2C1 = C6C7 = C82.II	0.00134	0.000212	0.0681	
118016		**AChE**	**CDK-5**	**GSI**	

(*continued on next page*)

Table 4 (*continued*)

	n1c(c2[nH]cnc2nc1)S[Sn](c1ccccc1)(c1ccccc1)c1ccccc1.n1c(c2ncn(c2nc1)[Sn](c1ccccc1)(c1ccccc1)c1ccccc1)S[Sn](c1ccccc1)(c1ccccc1)c1ccccc1	3.0972E-06	0.0058	0.00041
118017	n1c(c2[nH]cnc2nc1)S[Pb](c1ccccc1)(c1ccccc1)c1ccccc1.n1c(c2ncn(c2nc1)[Pb](c1ccccc1)(c1ccccc1)c1ccccc1)S[Pb](c1ccccc1)(c1ccccc1)c1ccccc1	6.6093E-09	0.0055	0.000409
403635	c1ccc(cc1CCCCCCCCCCCCCCC)OC(=O)Nc1cccc2c1cccc2.c1ccc(cc1CCCCCCC/C═C/CCCCCC)OC(=O)Nc1cccc2c1cccc2.c1ccc(cc1CCCCCCC/C═C/C/C═C/CCC)OC(=O)Nc1cccc2c1cccc2.c1ccc(cc1CCCCCCC/C═C/C/C═C/CC═C)OC(=O)Nc1cccc2c1cccc2	0.00381	0.000211	0.00644
700533	c12c(c(c3c(n1)c1c(CC3)ccc(c1)c1c3c(nc4c1cccc4)ccccc3)c1ccc(cc1)OCc1cc(cc(c1)OCc1cc(cc(c1)C(C)(C)C)C(C)(C)C)OCc1cc(cc(c1)C(C)(C)C)C(C)(C)C)CCc1c2cc(cc1)c1c2c(nc3c1cccc3)ccccc2	0.00815	5.435E-08	0.085
682371	[Pt+2]([Cl-])([Cl-])(n1ccc(cc1)/C/1 = c/2\cc/c(=C(\c3ccc(cc3)C)/C3=N/C(=C(\c4ccc(/C(=C\5/C═CC1=N5)/c1ccc(cc1)C)[nH]4)/c1ccc(cc1)C)/C═C3)/[nH]2)[O] = S(C)C	**BACE1** 0.00696	**CDK-5** 0.00593	**QC** 2.973E-05
1803–10–7	N(c1ccccc1)(c1ccccc1)C(=S)S[Sn](SC(=S)N(c1ccccc1)c1ccccc1)(c1ccccc1)c1ccccc1	**AChE** 0.00646	**BACE1** 1.9209E-07	**BuChE** 4.130
118195	S(C(=S)NC(=O)OCC)[Sb](SC(=S)NC(=O)OCC)Sc1nc2c(s1)cc(cc2)/N═N/c1ccc(cc1)NS(=O)(=O)c1ccc(cc1)NC(=O)C(Cl)Cl	**BACE1** 1.4771E-05	**GSI** 0.00182	**BuChE** 4.130
636592	c12c(c(c3c4c1c(=O)oc1c4c(c(=O)o3)c(c(c1O)O)c1c(c(O)c(O)cc1C(=O)OC[C@H]1[C@@H](OC(=O)c3c2c(c(O)c(O)c3)O)[C@H]2OC(=O)c3cc(O)c(O)c(O)c3c3c(C(=O)O[C@H]2[C@@H](O1)O)cc(O)c(O)c3O)O)O)O	**AChE** 0.00087	**GSI** 0.0297	**MAO-B** 3.672E-07
676818	c12c3c(c(c(c1O)O)O)c1c(cc(c(c1O)O)O)C(=O)O[C@H]1[C@H]([C@@H]([C@H]4[C@@H](C5=C([C@H]2[C@](C5=O)(C(=O)OCC)O)C(=O)O4)O)OC3=O)OC(=O)c2c(c(c(c(c2)O)O)O)c2c(c(c(cc2C(=O)OC1)O)O)O	0.00124	0.0230	5.799E-06
676822	c12c3c(c(c(c1O)O)O)c1c(cc(c(c1O)O)O)C(=O)O[C@H]1[C@H]([C@@H]([C@@H]4[C@H]5c6c7c(c(cc6O[C@]65[C@H](C2=C(C6=O)O)C(=O)O4)O)C[C@@H]([C@H](O7)c2cc(c(cc2)O)O)O)OC3=O)OC(=O)c2c(c(c(c(c2)O)O)O)c2c(c(c(cc2C(=O)OC1)O)O)O	0.000485	0.0199	0.00028
676825	[C@H]1([C@H](OC(=O)c2c(c3c(c(c(cc3C(=O)O1)O)O)O)c(c(c(c2)O)O)O)O)[C@H]1OC(=O)c2c(c(c(c(c2)O)O)O)O)c2c3c4c5c(c(=O)oc4c(c2O)O)c(c(c(c5oc3 = O)O)O)c2c(c(c(cc2C(=O)OC[C@@H]1O)O)O)O)O)C═O	0.00378	0.01570	4.694E-06
24312–00–3	c1(c(c2c3c(c1O)[C@@H]([C@H]([C@H]1OC(=O)c4c2c(c(c(c4c2c(c(c(cc2C(=O)O[C@H]2[C@H]1OC(=O)c1cc(c(c(c1c1c(c(c(cc1C(=O)OC2)O)O)O)O)O)O)O)O)O)O)O)O)OC3=O)O)O)O	0.000905	0.0338	0.000144
36001–47–5	c1c2c(c3c(cc(c(c3O)O)O)C(=O)O[C@H]3[C@H](COC2=O)OC(=O)c2cc(c(c(c2c2c4c(c5c6c(c(c(c5O)O)O)O)[C@@H]([C@H]([C@@H]3OC4 = O)OC6=O)O)c(c(c2O)O)O)O)O)O)c(c(c1O)O)O	0.000905	0.0338	0.000144
173121	[CH-]12[CH]3 = [CH]4[CH]5 = [CH]1[Co+]162345[CH]2 = [CH]1[S]6C2	**AChE** 0.00106	**BACE1** 3.5773E-12	**β-amyloid** 0.143
633340	[C]123 = C[Rh+]456783([C]1(=C8)CC(=C)CC[C]5(=C7)[C]4(=C6)CC2)[Cl-]	7.4994E-06	7.015E-16	0.528
81741–72–2	[Ge]([Ge](N(CC)CC)(CC)CC)(CC)(CC)CC	7.948E-06	1.1578E-07	0.353
81741–73–3	N(CC)(CC)[Ge](CC)(CC)[Ge](CC)(CC)N(CC)CC	1.0373E-05	3.3865E-11	0.457
993–62–4	[Ge](CC)(CC)(CC)[Ge](CC)(CC)CC	2.839E-06	0.00352	0.218
993–63–5	[Sn]([Sn](CC)(CC)CC)(CC)(CC)CC	5.587E-11	5.527E-12	0.236
847446–05–3	N(c1c(C)cc(C═CC#N)cc1C)c1nc(Nc2ccc(C#N)cc2)ncc1.C([C@H](OCP(=O)(O)O)C)n1c2c(nc1)c(N)ncn2.N(c1c2c(n(cn2)[C@@H]2C[C@H](CO)C═C2)nc(N)n1)C1CC1	**CDK-5** 0.1225	**5-HT6** 0.00424	**BACE1** 0.177

model (see sheet 32 in the Supporting Information SI-1), we found only 1 compound (i.e. compound **1**) in the test set is located outside the AD. In the case of the BuChE enzyme- β-amyloid aggregation inhibitors model (see Fig. S150 in the Supporting Information SI-2), we found only 2 compounds (i.e. compounds **62** and **76**) in the test set placed outside the AD. In the case of the BuChE -MOA-B enzyme inhibitors model (see Fig. S151 in the Supporting Information SI-2), we found only 1 compound (i.e. compound **9**) in the test set is situated outside the AD.

3.1.4.3. QSAAR based models. AD plots of the QSAAR models revealed that except for the β-amyloid (X)-BuChE (Y) model which contained one test set molecule (molecule **14**) and the AChE (X)-β-amyloid (Y) model which contain two test set compounds (i.e. compound **26** and **80**) outside the AD (see sheet 34in the SI-1 and Fig. S155 in the Supporting Information SI-2), all of the developed QSAAR models exhibited 100% domain of applicability for all potential combinations (see Figs. S152–154, 156–166 and sheet 35 in the Supporting Information SI-2 and 1).

3.1.5. Activity prediction using the developed 2D-QSAR models

In the current work, well-validated QSAR models were individually used to predict the inhibitory activity of four chemical drug-like databases, which had no reported quantitative experimental response values in their source file against the respective targets. Primarily, the established models were utilized to compute the predicted values of database compounds; the validated models were capable of precisely predicting the inhibitory activity of the majority of the molecules, as suggested by the 'alvaRunner version 2.0.4' tool (https://www.alvascience.com/alvarunner/). After prediction, we have arranged the compounds based on predicted values (highest to least active). In a detailed analysis of the predicted chemical databases, we have identified the top 56 lead compounds with multitarget inhibitory activity, which are stated in Table 4. Additionally, we have also identified the compounds with dual target inhibitory activity as given in the supplementary information SI-4. The detailed analysis of the predicted databases using the developed models are given in Table S4. The detailed external set evaluations are stated in the supplementary information SI-3, sheet 1–12. In the supplementary information, we have only listed the compounds from the estimated databases showing higher potent inhibitory activity than the respective modeled dataset compounds due to large data and file size.

3.2. Chemical read-across analysis

In the present investigation, we have performed the similarity-based quantitative read-across predictions using the same training and test set combinations as used in 2D-QSAR modeling. The current approach employs three distinct similarity-based measures: Laplacian kernel similarity-based (LK) predictions, Gaussian kernel similarity-based (GK), and Euclidean distance-based (ED) estimations, and after hyperparameter optimization, we found that the external validation results obtained from a quantitative read-across algorithm using Gaussian Kernel Similarity-based functions (in case of 5-HT-6, CDK-5, PDE 10A models) and Laplacian kernel similarity-based function (in case of AChE and BACE1 enzyme models) were better compared to the results obtained from the individual regression-based 2D-QSAR models. The details about the external validation results are given in sheets 36–40 (see supplementary information SI-1).

Table 5
Molecular docking results (5-HT6 antagonist, AChE, BACE1 enzyme, β-amyloid, BuChE enzyme, CDK-5 inhibitors) and correlation with 2D-QSAR models in this study.

Compound	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
5-hydroxytryptamine 6 (5-HT6) antagonist				
1 (most active)	35.318	ALA A:266, LYS A:262, HIS A:259, SER A:61, TYR A:134, ASP A:123, LUE A:127	Hydrogen Bond (conventional), Attractive charge, π-sulfur, Alkyl, π-Alkyl	F04[C–O], minssNH, nCp, nArNH2, nBnz
5 (most active)	32.544	SER A:61, ASP A:123, TYR A:134, HIS A:259, LYS A:262, LUE A:127	Hydrogen Bond (conventional and Carbon), Attractive charge, π-sulfur, π-Alkyl, π-π T-shaped	F04[C–O], minssNH, nCp, nArNH2, nBnz
70 (least active)	19.878	ASN A:62, ASN A:59, LYS A:262, HIS A:259	Hydrogen Bond (conventional and Carbon)	F04[C–O]
71 (least active)	17.953	LYS A:262, LEU A:127	Hydrogen Bond (Carbon), π-Alkyl	nBnz
Acetylcholinesterase (AChE) enzyme inhibitor				
841 (most active)	67.407	TYR A:72, TRP A:286, SER A:293, TYR A:341, TRP A:86, TYR A:337, GLU A:202, HIS A:447	Hydrogen Bond (conventional, Carbon and π-donor), π-π T-shaped, π-π stacked, π-Lone pair, Halogen (Cl, Br, I), π-Alkyl, π-Cation	D/Dtr05, DBI, X2v, F04 [N–Cl], B05[C–N]
842 (most active)	51.491	THR A:75, TRP A:286, SER A:293, TRP A:86, TYR A:337, TYR A:124, TYR A:341,	Hydrogen Bond (conventional and Carbon), π-π T-shaped, π-π stacked	X2v, DBI, B07[N–N], B05 [C–N], minssO, nArCO
714 (least active)	38.559	PHE A:297, TRP A:286, TYR A:72	Hydrogen Bond (conventional), π-Alkyl	minssO
721 (least active)	39.518	TYR A:341, TYR A:72	Hydrogen Bond (conventional), π-Alkyl	minssO
β-secretase 1 (BACE1) enzyme inhibitors				
458 (most active)	56.095	ALA A:335, TYR A:14, THR A:232, GLY A:230, ASP A:228, GLN A:73, TYR A:198, GLY A:34, SER A:35, TYR A:71	Hydrogen Bond (conventional, Carbon and π-donor), π-π T-shaped, π-Anion, π-Lone pair, Halogen (Fluorine), Alkyl, π-Alkyl	F08[N–N], Ui,
463 (most active)	51.796	LEU A:30, SER A:229, THR A:231, TYR A:14, LYS A:9, VAL A:170, GLU A:339, GLY A:13, GLY A:11, ALA A:335, THR A:232, GLY A:230, TYR A:71, THR A:72, ASP A:228, ASP A:32	Hydrogen Bond (conventional, Carbon), Salt bridge, Attractive charge, Sulfur-X, Halogen (Fluorine), Alkyl, π-Alkyl	F08[N–N], Ui, X5v, mindO, nR#CH/X
816 (least active)	31.749	ARG A:128, TYR A:198, ASP A:228	Hydrogen Bond (conventional), Attractive charge, π-Anion	B06[O–O]
831 (least active)	27.353	LYS A:107, GLY A:230, LEU A:30	Hydrogen Bond (conventional and carbon), π-Alkyl	mindO
β-amyloid inhibitors				
208 (most active)	30.398	LEU A:17, LYS A:16, HIS A:13, ALA A:21, VAL A:18	Hydrogen Bond (conventional and carbon), π-Alkyl, Alkyl, π-π T-shaped	SssCH2, PW3, F04[N–O]
124 (least active)	19.144	ALA A:21, LYS A:16	π-Alkyl, π-Cation	PW3
Butyrylcholinesterase (BuChE) enzyme inhibitors				
415 (most active)	62.027	TRP A:82, ALA A:328, PRO A:285, HIS A:438, TRP A:231, PHE A:329, LEU A:286, SER A:287	Hydrogen Bond (conventional, Carbon and π-donor), π-π T-shaped, π-π stacked, π-Cation, π-Alkyl	C-002, N-070, MaxaaaC, MDEC-22, F04[C–N], B07 [N–N], nN-N, BLI,
420 (most active)	64.064	ASP A:70, TYR A:332, GLU A:197, HIS A:438, TRP A:231, LEU A:286, ALA A:328, TRP A:430, TRP A:82	Hydrogen Bond (conventional, Carbon), π-π T-shaped, Attractive charge, Alkyl, π-Alkyl	C-002, N-070, MaxaaaC, MDEC-22, F04[C–N], B07 [N–N], nN-N, BLI
417 (least active)	31.984	HIS A:438, ASP A:70	Hydrogen Bond (conventional, Carbon)	C-002
2353 (least active)	38.996	GLU A:197, GLY A:115, ILE A:69	Hydrogen Bond (Carbon)	MDEC-22
Cyclin-dependent kinase 5 (CDK-5) inhibitors				
1 (most active)	54.534	LYS A:89, ASP A:86, LEU A:133, ALA A:31, VAL A:18, GLY A:11, GLU A:12, ILE A:10	Hydrogen Bond (conventional, Carbon), Salt bridge, π-Sigma, Alkyl, π-Cation, π-Alkyl	NaaaC, SaasC
2 (most active)	58.694	LYS A:89, ASP A:86, ALA A:31, LEU A:133, PHE A:80, VAL A:18, VAL A:64, CYS A:83, ILE A:10, LYS A:9	Hydrogen Bond (conventional, Carbon), π-π T-shaped, Attractive charge, Alkyl, π-Alkyl	NaaaC, SaasC
146 (least active)	47.848	LYS A:33, LEU A:133, VAL A:18	Hydrogen Bond (Carbon), π-Alkyl	SaasC
194 (least active)	30.142	LEU A:133, PHE A:82, ILE A:10, LYS A:89	Hydrogen Bond (Conventional), π-Alkyl, π-Sulfur	SaasC

3.3. Molecular docking analysis

In this exploration, we have performed molecular docking using the most and least active compounds from the initial datasets to explore the molecular interactions at the active pocket of the respective targets. The detailed description of molecular docking analysis and their corresponding Figs. S167–187 are given in the Supporting Information SI-2, page number 85–103. The evidence of docking interactions, CDocker interaction energy, and their correlation with the features obtained from the developed 2D-QSAR models are demonstrated in Tables 5 and 6. Moreover, we have also performed the molecular docking analysis using the top predicted compounds from the databases at the active pocket of the respective targets. The details about the docking analysis are demonstrated in Table S5 and Figs. S188–199, page number 104–109 (See supporting Information SI-2).

3.4. Comparisons of the proposed QSAR models with our previous published studies

In the present study, we have implemented a comparison of the best models of this study with our previously published models [4,16,48]. The details of different internal and external validation parameters

Table 6
Molecular docking results (γ-secretase, GSK-3β, MAO-B inhibitors, NMDA receptor antagonist, PDE 10A enzyme, Glutaminyl Cyclase (QC) inhibitors) and correlation with 2D-QSAR models in this study.

Compound	-CDocker interaction energy (kcal/mol)	Interacting residues	Interactions	Correlation with QSAR model
γ-secretase inhibitors				
180 (most active)	126.087	PRO A:424, PRO A:141, TYR A:453, ARG A:281, GLU A:333, ASN A:142, TYR A:337	Hydrogen Bond (Conventional and Carbon), Alkyl, π-Cation	P_VSA_MR_7, ATS6s, VE3sign_D, VE3sign_D/Dt
208 (least active)	52.566	HIS A:444, PRO A:424, ARG A:281, PRO A:141, TYR A:453, SER A:425	Hydrogen Bond (Carbon), π-Alkyl, π-Cation, Alkyl	P_VSA_MR_7, ATS6s
GSK-3β inhibitors				
41 (most active)	38.688	VAL A:70, ILE A:62, VAL A:135, ALA A:83, TYR A:134, LEU A:188, ASP A:133	Hydrogen Bond (Conventional, Carbon), π-π stacked, π-Alkyl	SaasC, PDI, SaaaC, F06 [N–O], nPyrroles, B05 [N–O]
112 (least active)	18.996	LEU A:188, ALA A:83	π-Alkyl	SaaaC
MAO-B inhibitors				
72 (most active)	47.999	LYS A:296, TRP A:388, VAL A:294, GLY A:57, CYS A:397, TYR A:435, ILE A:14, SER A:15, GLY A:13, THR A:426, ARG A:42, ALA A:439	Hydrogen Bond (Conventional, Carbon), Amide-π stacked, π-Alkyl, Alkyl	F02[O–O], C%
87 (most active)	44.213	ILE A:14, SER A:15, GLY A:13, THR A:426, ALA A:439, TYR A:435, ARG A:42, CYS A:397, GLY A:57, GLY A:434	Hydrogen Bond (Conventional, Carbon), Amide-π stacked, π-Alkyl, Alkyl, Halogen (Fluorine)	F02[O–O], C%
14 (least active)	37.933	TYR A:398, CYS A:397, ARG A:42, ILE A:14, ALA A:439, THR A:426	Hydrogen Bond (Carbon), π-Alkyl, Alkyl	C%
25 (least active)	32.254	GLY A:58, GLY A:40, ARG A:42, MET A:436, TYR A:435, TYR A:398, GLN A:206	Hydrogen Bond (Carbon), π-π T-shaped, π-Sigma, π-Alkyl, Alkyl, Halogen (Cl, Br, I)	C%
NMDA receptor antagonist				
485 (most active)	32.035	TYR A:184, GLN A:144, LEU A:146, SER A:180, THR A:126, HOH A:1036, HOH A:1111, GLN A:13, TRP A:223, PRO A:15, VAL A:227, ALA A:226, PHE A:16, ASP A:224	Hydrogen Bond (Conventional, Carbon and Water), π-π T-shaped, π-Lone pair, π-Sigma, π-Alkyl, Alkyl, Halogen (Fluorine)	ATSC1p, C-028, TPSA (Tot), B10[C–O]
486 (most active)	32.578	PHE A:92, LEU A:125, THR A:126, PRO A:124, HOH A:1078, HOH A:1036, ASP A:224	Hydrogen Bond (Conventional, Carbon and Water), π-Anion, π-Alkyl, Alkyl,	ATSC1p, C-028, TPSA (Tot)
398 (least active)	27.011	GLN A:144, ASP A:224, PRO A:124, PHE A:92, THR A:126	Hydrogen Bond (Carbon), π-Lone pair, π-Alkyl, Halogen (Fluorine)	C-028
455 (least active)	21.838	PHE A:92, PHE A:246, LEU A:146, TYR A:184	Hydrogen Bond (Carbon), π-π T-shaped, π-π stacked, π-Alkyl	TPSA(Tot)
Phosphodiester 10A (PDE 10A) enzyme inhibitors				
226 (most active)	51.832	TYR A:524, ASP A:674, SER A:677, ILE A:692, LEU A:675, GLN A:726, PHE A:729, PHE A:696, LEU A:635, ALA A:636, SER A:571, ASN A:572, GLU A:592	Hydrogen Bond (Conventional and Carbon), π-π T-shaped, π-π stacked, π-Alkyl	N-070, AMR, minsCH3, Eta_epsi_2
228 (most active)	57.044	LEU A:675, SER A:677, PHE A:729, GLN A:726, ILE A:692, MET A:713, PHE A:696, PHE A:570, HIS A:525, GLU A:592, THR A:633, LEU A:635, ASP A:674	Hydrogen Bond (Conventional and Carbon), Salt bridge, π-π T-shaped, π-π stacked, π-Sigma, Alkyl, π-Alkyl, π-Cation, Attractive charge	N-070, AMR, minsCH3, Eta_epsi_2
217 (least active)	36.980	LEU A:635, HIS A:567, PHE A:696, PHE A:729, ILE A:692, GLN A:726	Hydrogen Bond (Conventional and π-Donor), π-π T-shaped, π-π stacked, π-Alkyl	Eta_epsi_2, AMR
221 (least active)	35.608	LEU A:675, ILE A:692, PHE A:729, TYR A:524, HIS A:525, ASP A:634, LEU A:635	Hydrogen Bond (Carbon), π-π T-shaped, π-π stacked, π-Alkyl, π-Cation	Eta_epsi_2, AMR
Glutaminyl Cyclase (QC) inhibitors				
87 (most active)	131.348	TRP A:329, LYS A:144, HIS A:330, ASP A:159, GLU A:202, GLU A:201, SER A:160, MET A:167, PRO A:163, CYS A:139, LEU A:137, LEU A:246, ALA A:138, LEU A:247, ASP A:248, ASP A:305	Hydrogen Bond (Conventional, Carbon and π-Donor), Attractive charge, Alkyl, π-π T-shaped, π-Anion, π-Cation, π-Alkyl	C-034, B09[C–O], T(N..S)
88 (most active)	128.118	TRP A:329, TRP A:207, HIS A:330, ASP A:159, GLU A:201, SER A:160, MET A:167, PRO A:163, CYS A:139, LEU A:137, VAL A:245, LEU A:247, ASP A:248, ASP A:305	Hydrogen Bond (Conventional and Carbon), Attractive charge, Alkyl, π-Sigma, π-Anion, π-Alkyl	C-034, B09[C–O], T(N..S)
31 (least active)	15.879	TRP A:329, TRP A:207, ILE A:303	Alkyl, π-Alkyl	T(N..S)
33 (least active)	17.698	GLN A:304, TRP A:207	Hydrogen Bond (Conventional), π-Cation	T(N..S)

obtained from this study and those obtained from previous models are given in Table 7.

4. Conclusion and prospects

The present investigation includes the development of *in silico*-based predictive 2D-QSAR models against the twelve major targets of AD using a PLS regression approach for the exploration of the structural features responsible for the inhibitory activity toward respective targets using simple and easily interpretable 2D descriptors. We have also developed the 17 QSAAR and 10 selectivity-based models by screening the common inhibitors from the primary datasets. All models are extensively validated and found to be robust enough to satisfy the acceptance criteria. From the insights obtained from the models, it can be inferred that features contributing to hydrophobicity as the significant parameter for the AD drugs, molecular bulk, and electronegativity of the compounds may enhance the inhibitory activity of the compounds. The identified features/properties are responsible for enhancing brain permeability and an entropically more favorable binding to the receptor and intermolecular interactions by strong H bonds for the improvement of the inhibitory activity against the targets. The developed individual models were used to check the applicability domain of the four chemical druglike databases (18,963,690 compounds) for the search for novel inhibitors with multitarget inhibitory activity. In a detailed analysis of

Table 7
Comparisons of proposed QSAR models with our previous published studies.

Sources		E. L.	LV	Model	Training set			Test set		
					n	R^2	Q^2	n	Q^2F_1	Q^2F_2
Models in this study	AChE model	28	7	PLS	1325	0.635	0.621	408	0.678	0.678
	BuChE model	28	8	PLS	1882	0.689	0.668	625	0.702	0.702
	BACE1 model	18	9	PLS	680	0.669	0.650	225	0.675	0.674
	β-Amyloid model	6	4	PLS	197	0.729	0.705	65	0.844	0.844
Published models	AChE model [4]	15	6	PLS	798	0.662	0.645	199	0.661	0.660
	BuChE model [4]	13	5	PLS	603	0.674	0.656	158	0.663	0.660
	BACE1 model [48]	5	3	PLS	76	0.826	0.795	22	0.846	0.846
	β-Amyloid model [16]	12	6	PLS	252	0.664	0.621	62	0.765	0.763

the predicted chemical databases, we have identified the top 56 lead compounds with multitarget inhibitory activity that could act as inhibitors for the treatment of AD. Furthermore, we have also implemented chemical Read-Across predictions using the Read-Across-v3.1 tool (https://dtclab.webs.com/software-tools); the results for the external validation parameters were found to be better than the 2D-QSAR-derived predictions. Finally, molecular docking analysis has been implemented to identify the interactions between target proteins and inhibitors, and the results showed the active compounds formed hydrogen bonding and hydrophobic π interactions with active amino acid residues that lead to explaining the influences of different features which appeared in the 2D-QSAR models. Moreover, the evidences obtained from molecular docking analysis well corroborated with the features obtained from the 2D-QSAR analysis. Lastly, we may conclude that the developed QSAR, QSAAR, and selectively based models might be enormously valuable as guides for investigators to predict the inhibitory activity of novel compounds against respective targets and the evidence obtained from the 2D-QSAR analysis and molecular docking studies can be useful for the design of next-generation Alzheimer's treatments.

CRediT author statement

VK: Data curation, Formal analysis, Writing – Original draft.
KR: Conceptualization, Supervision, Funding, Writing – Review and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

VK thanks the Indian Council of Medical Research (ICMR), New Delhi (File No: BMI/11(03)/2022, IRIS Cell No.: 2021–8243, dated: 13/05/2022) for financial support in the form of a Research Associateship (RA).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2022.104734.

References

[1] J. Therriault, E.R. Zimmer, A.L. Benedet, T.A. Pascoal, S. Gauthier, P. Rosa-Neto, Staging of Alzheimer's disease: past, present, and future perspectives, Trends Mol. Med. 28 (9) (2022) 726–741, https://doi.org/10.1016/j.molmed.2022.05.008.
[2] S. Gauthier, P. Rosa-Neto, J.A. Morais, C. Webster, World Alzheimer Report 2021: Journey through the Diagnosis of Dementia, Alzheimer's disease International, London, England, 2021.
[3] N. Koutsodendris, M.R. Nelson, A. Rao, Y. Huang, Apolipoprotein e and alzheimer's disease: findings, hypotheses, and potential mechanisms, Annu. Rev. Pathol. 17 (2022) 73–99, https://doi.org/10.1146/annurev-pathmechdis-030421-112756.
[4] V. Kumar, A. Saha, K. Roy, In silico modeling for dual inhibition of acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) enzymes in Alzheimer's disease, Comput. Biol. Chem. 88 (2020), 107355, https://doi.org/10.1016/j.compbiolchem.2020.107355.
[5] S.A. Tatulian, Challenges and Hopes for Alzheimer's Disease, Drug Discov, 2022, https://doi.org/10.1016/j.drudis.2022.01.016.
[6] M. Vaz, S. Silvestre, Alzheimer's disease: recent treatment strategies, European J. Pharm. Pharmacol. 887 (2020), 173554, https://doi.org/10.1016/j.ejphar.2020.173554.
[7] K.G. Yiannopoulou, A.I. Anastasiou, V. Zachariou, S.H. Pelidou, Reasons for failed trials of disease-modifying treatments for Alzheimer disease and their contribution in recent research, Biomedicines 7 (4) (2019) 97, https://doi.org/10.3390/biomedicines7040097.
[8] B. Zhang, J. Zhao, Z. Wang, P. Guo, A. Liu, G. Du, Identification of multi-target anti-AD chemical constituents from traditional Chinese medicine formulae by integrating virtual screening and in vitro validation, Front. Pharmacol. 12 (2021), 709607, https://doi.org/10.3389/fphar.2021.709607.
[9] M.S. Nadeem, J.A. Khan, U. Rashid, Fluoxetine and sertraline based multitarget inhibitors of cholinesterases and monoamine oxidase-A/B for the treatment of Alzheimer's disease: synthesis, pharmacology and molecular modeling studies, Int. J. Biol. Macromol. 193 (2021) 19–26, https://doi.org/10.1016/j.ijbiomac.2021.10.102.
[10] L. Brunetti, R. Leuci, A. Carrieri, M. Catto, S. Occhineri, G. Vinci, L. Piemontese, Structure-based design of novel donepezil-like hybrids for a multi-target approach to the therapy of Alzheimer's disease, Eur. J. Med. Chem. 237 (2022), 114358, https://doi.org/10.1016/j.ejmech.2022.114358.
[11] A. Ajala, A. Uzairu, G.A. Shallangwa, S.E. Abechi, 2D QSAR, design, docking study and ADMET of some N-aryl derivatives concerning inhibitory activity against Alzheimer disease, Future J. Pharm. Sci. 8 (1) (2022) 1–14, https://doi.org/10.1186/s43094-022-00420-w.
[12] P. Ambure, J. Bhat, T. Puzyn, K. Roy, Identifying natural compounds as multi-target-directed ligands against Alzheimer's disease: an in silico approach, J. Biomol. Struct. Dyn. 37 (5) (2019) 1282–1306, https://doi.org/10.1080/07391102.2018.1456975.
[13] P. Zhang, S. Xu, Z. Zhu, J. Xu, Multi-target design strategies for the improved treatment of Alzheimer's disease, Eur. J. Med. Chem. 176 (2019) 228–247, https://doi.org/10.1016/j.ejmech.2019.05.020.
[14] N. Kumar, V. Kumar, P. Anand, V. Kumar, A.R. Dwivedi, V. Kumar, Advancements in the development of multi-target directed ligands for the treatment of Alzheimer's disease, Bioorg. Med. Chem. (2022), 116742, https://doi.org/10.1016/j.bmc.2022.116742.
[15] M.K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Res. 44 (D1) (2016) D1045–D1053, https://doi.org/10.1093/nar/gkv1072. Available from: www.bindingdb.org.
[16] V. Kumar, P.K. Ojha, A. Saha, K. Roy, Cheminformatic modelling of β-amyloid aggregation inhibitory activity against Alzheimer's disease, Comput. Biol. Med. 118 (2020), 103658, https://doi.org/10.1016/j.compbiomed.2020.103658.
[17] M. González-Medina, J.J. Naveja, N. Sánchez-Cruz, J.L. Medina-Franco, Open chemoinformatic resources to explore the structure, properties and chemical space of molecules, RSC Adv. 7 (85) (2017) 54153–54163, https://doi.org/10.1039/C7RA11831G.
[18] D. Gadaleta, A. Lombardo, C. Toma, E. Benfenati, A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications, J. Cheminf. 10 (1) (2018) 1–13, https://doi.org/10.1186/s13321-018-0315-6.
[19] F.P. Steinmetz, C.L. Mellor, T. Meinl, M.T. Cronin, Screening chemicals for receptor-mediated toxicological and pharmacological endpoints: using public data

to build screening tools within a KNIME workflow, Mol. Inform. 34 (2-3) (2015) 171–178, https://doi.org/10.1002/minf.201400188.
[20] A. Rácz, D. Bajusz, K. Héberger, Multi-level comparison of machine learning classifiers and their performance metrics, Molecules 24 (15) (2019) 2811, https://doi.org/10.3390/molecules24152811.
[21] A. Rácz, D. Bajusz, K. Héberger, Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification, Molecules 26 (4) (2021) 1111, https://doi.org/10.3390/molecules26041111.
[22] A. Rácz, D. Bajusz, K. Héberger, Consistency of QSAR models: correct split of training and test sets, ranking of models and performance parameters, SAR QSAR Environ. Res. 26 (7–9) (2015) 683–700, https://doi.org/10.1080/1062936X.2015.1084647.
[23] P. Ambure, A. Gajewicz-Skretna, M.N.D. Cordeiro, K. Roy, New workflow for QSAR model development from small data sets: small dataset curator and small dataset modeler. integration of data curation, exhaustive double cross-validation, and a set of optimal model selection techniques, J. Chem. Inf. Model. 59 (10) (2019) 4070–4076, https://doi.org/10.1021/acs.jcim.9b00476.
[24] P. De, K. Roy, QSAR and QSAAR modeling of nitroimidazole sulfonamide radiosensitizers: application of small dataset modeling, Struct. Chem. 32 (2) (2021) 631–642, https://doi.org/10.1007/s11224-021-01734-w.
[25] R.K. Mukherjee, V. Kumar, K. Roy, Ecotoxicological QSTR and QSTTR modeling for the prediction of acute oral toxicity of pesticides against multiple avian species, Environ. Sci. Technol. 56 (1) (2021) 335–348, https://doi.org/10.1021/acs.est.1c05732.
[26] I.N.C. Minitab, Minitab 16 Statistical Software, Minitab Inc., State College, PA, 2010. Available from: https://www.minitab.com/en-us/products/minitab/.
[27] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, H. Kabir, Comparative studies on some metrics for external validation of QSPR models, J. Chem. Inf. Model. 52 (2) (2012) 396–408, https://doi.org/10.1021/ci200520g.
[28] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inform. 29 (6-7) (2010) 476–488, https://doi.org/10.1002/minf.201000061.
[29] P. Király, P. Kiss, D. Kovács, A. Ballaj, G. Tóth, The relevance of goodness-of-fit, robustness and prediction validation categories of OECD-QSAR principles with respect to sample size and model type, Mol. Inform. 41 (11) (2022), 2200072, https://doi.org/10.1002/minf.202200072.
[30] D. Kovács, P. Király, G. Tóth, Sample-size dependence of validation parameters in linear regression models and in QSAR, SAR QSAR Environ. Res. 32 (4) (2021) 247–268, https://doi.org/10.1080/1062936X.2021.1890208.
[31] P. Gramatica, A. Sangion, A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology, J. Chem. Inf. Model. 56 (6) (2016) 1127–1131, https://doi.org/10.1021/acs.jcim.6b00088.
[32] A. Golbraikh, A. Tropsha, Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, Mol. Divers. 5 (4) (2000) 231–243, https://doi.org/10.1023/A:1021372108686.
[33] V. Consonni, D. Ballabio, R. Todeschini, Evaluation of model predictive ability by external validation techniques, J. Chemom. 24 (3–4) (2010) 194–201, https://doi.org/10.1002/cem.1290.
[34] S. Shayanfar, A. Shayanfar, Comparison of various methods for validity evaluation of QSAR models, BMC chem 16 (1) (2022) 1–9, https://doi.org/10.1186/s13065-022-00856-4.
[35] V. Kumar, P. De, P.K. Ojha, A. Saha, K. Roy, A multi-layered variable selection strategy for QSAR modeling of butyrylcholinesterase inhibitors, Curr. Top. Med. Chem. 20 (18) (2020) 1601–1627, https://doi.org/10.2174/1568026620666200616142753.
[36] SIMCA-P 10.0, Umetrics. Available from: https://umetrics.com/products/simca (accessed January 10, 2022).
[37] P. De, V. Kumar, S. Kar, K. Roy, J. Leszczynski, Repurposing FDA approved drugs as possible anti-SARS-CoV-2 medications using ligand-based computational approaches: sum of ranking difference-based model selection, Struct. Chem. (2022) 1–13, https://doi.org/10.1007/s11224-022-01975-3.
[38] M. Chatterjee, K. Roy, Application of cross-validation strategies to avoid overestimation of performance of 2D-QSAR models for the prediction of aquatic toxicity of chemical mixtures, SAR QSAR Environ. Res. 33 (6) (2022) 463–484, https://doi.org/10.1080/1062936X.2022.2081255.
[39] A. Banerjee, M. Chatterjee, P. De, K. Roy, Quantitative predictions from chemical read-across and their confidence measures, Chemometr. Intell. Lab. Syst. 227 (2022), 104613, https://doi.org/10.1016/j.chemolab.2022.104613.
[40] A. Banerjee, K. Roy, First report of q-RASAR modeling toward an approach of easy interpretability and efficient transferability, Mol. Divers. 26 (5) (2022) 2847–2862, https://doi.org/10.1007/s11030-022-10478-6.
[41] Discovery studio predictive science application | dassault systèmes BIOVIA. https://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/(accessed November 10, 2021).
[42] G. Wu, D.H. Robertson, C.L. Brooks III, M. Vieth, Detailed analysis of grid-based molecular docking: a case study of CDOCKER—a CHARMm-based MD docking algorithm, J. Comput. Chem. 24 (13) (2003) 1549–1562, https://doi.org/10.1002/jcc.10306.
[43] V. Kumar, S. Kar, P. De, K. Roy, J. Leszczynski, Identification of potential antivirals against 3CLpro enzyme for the treatment of SARS-CoV-2: a multi-step virtual screening study, SAR QSAR Environ. Res. 33 (5) (2022) 357–386, https://doi.org/10.1080/1062936X.2022.2055140.
[44] A. Pirovano, S. Brandmaier, M.A. Huijbregts, A.M. Ragas, K. Veltman, A. J. Hendriks, The utilisation of structural descriptors to predict metabolic constants of xenobiotics in mammals, Environ. Toxicol. Pharmacol. 39 (1) (2015) 247–258, https://doi.org/10.1016/j.etap.2014.11.025.
[45] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, QSAR applicability domain estimation by projection of the training set in descriptor space: a review, ATLA 33 (5) (2015) 445–459, https://doi.org/10.1177/026119290503300508.
[46] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2) (2001) 109–130, https://doi.org/10.1016/S0169-7439(01)00155-1.
[47] K. Roy, S. Kar, P. Ambure, On a simple approach for determining applicability domain of QSAR models, Chemometr. Intell. Lab. Syst. 145 (2015) 22–29, https://doi.org/10.1016/j.chemolab.2015.04.013.
[48] V. Kumar, P.K. Ojha, A. Saha, K. Roy, Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease, SAR QSAR Environ. Res. 31 (2) (2020) 87–133, https://doi.org/10.1080/1062936X.2019.1695226.