DECIPHERING THE MECHANISMS OF ALZHEIMER'S AND PARKINSON'S DISEASES USING NETWORK BIOLOGY AND A FUNCTIONAL GENOMICS APPROACH

Thesis Submitted
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in
Biotechnology

by

Rahul Tripathi
(2K18/PHDBT/503)

Under the Supervision of

Prof. Pravir Kumar, PhD
Professor and Dean (International Affairs)
Department of Biotechnology
Delhi Technological University, Delhi

To the
Department of Biotechnology

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042. India

2025

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I, **Rahul Tripathi**, hereby certify that the work which is being presented in the thesis entitled **"Deciphering mechanism of Alzheimer's and Parkinson's diseases using network biology and functional genomics approach"** in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy, submitted in the Department of Biotechnology, Delhi Technological University is an authentic record of my own work carried out during the period from **December, 2018 to March, 2025** under the supervision of **Prof. Pravir Kumar**, Department of Biotechnology, Delhi Technological University, Delhi.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Rahul Tripathi

Date: 26/06/2025

Candidate's Signature
Rahul Tripathi
2K18/PHDBT/503

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis, and the statement made by the candidate is correct to the best of our knowledge.

Pravinkumar

Signature of Supervisor(s)

Suman

Signature of External Examiners

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

This is to certify that **Rahul Tripathi** (2K18/PHDBT/503) has carried out the research work presented in this thesis entitled **"Deciphering the mechanisms of Alzheimer's and Parkinson's diseases using network biology and a functional genomics approach**" for the award of **Doctor of Philosophy** from Department of Biotechnology, Delhi Technological University, Delhi, under our supervision. The thesis embodies results of original work, and studies are carried out by the student himself, and the contents of the thesis do not form the basis for the award of any other degree to the candidate or anybody else from this or any other University/Institution.

Date: 02.07.25

Pravir Kumar

Prof. Pravir Kumar
Professor and Dean (IA)
Department of Biotechnology
Delhi Technological University
Shahbad Daulatpur, Bawana Road,
Delhi-110042

02.07.25

Prof. Yasha Hasija
HoD (Biotechnology) & DRC Chairperson
Department of Biotechnology
Delhi Technological University
Shahbad Daulatpur, Bawana Road,3
Delhi-110042

Deciphering the mechanisms of Alzheimer's and Parkinson's diseases using network biology and a functional genomics approach

Rahul Tripathi

ABSTRACT

Neurodegenerative disorders are known to exhibit genetic overlap and shared pathophysiology. This study aims to find the shared genetic architecture of Alzheimer's disease (AD) and Parkinson's disease (PD), two major age-related progressive neurodegenerative disorders. The gene expression profiles of GSE67333 (containing samples from AD patients) and GSE114517 (containing samples from PD patients) were retrieved from the Gene Expression Omnibus (GEO) functional genomics database managed by the National Center for Biotechnology Information (NCBI). The web application GREIN (GEO RNA-seq Experiments Interactive Navigator) was used to identify differentially expressed genes (DEGs). 617 DEGs (239 upregulated and 379 downregulated) were identified from the GSE67333 dataset. Likewise, 723 DEGs (378 upregulated and 344 downregulated) were identified from the GSE114517 dataset. The protein-protein interaction (PPI) networks of the differentially expressed genes (DEGs) were constructed, and the top 50 hub genes were identified from the network of the respective dataset. Of the 4 common hub genes between the two datasets, CXCR4 was selected due to its gene expression signature profile and the same direction of differential expression between the two datasets. Mavorixafor was chosen as the reference drug due to its known inhibitory activity against CXCR4 and its ability to cross the blood-brain barrier. Molecular docking and molecular dynamics simulation of 51 molecules having structural similarity with Mavorixafor were performed to find two novel molecules, ZINC49067615 and ZINC103242147. Natural compounds are gaining prominence in the therapy of neurodegenerative disorders due to their biocompatibility and potential neuroprotective properties, including their ability to modulate CXCR4 expression. Recent advancements in artificial intelligence (AI) and machine learning (ML) algorithms have opened new avenues for drug discovery research across various therapeutic areas, including neurodegenerative disorders. We produced an ML model using cheminformatics-guided machine learning algorithms using data of compounds with known CXCR4 activity, retrieved from the Binding Database, to analyse diverse physicochemical attributes of natural compounds obtained from the COCONUT Database and predict their inhibitory activity against CXCR4.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to all those who have supported and guided me throughout the process of completing this thesis. Your encouragement and assistance have been invaluable, and I am truly thankful for your presence in this journey. Without your unwavering support and belief in my abilities, this achievement would not have been possible.

First and foremost, I am profoundly grateful to my supervisor, **Prof. Pravir Kumar**, for his invaluable guidance, insightful feedback, and unwavering support. His expertise and encouragement have been instrumental in shaping this research and helping me navigate various challenges. His deep knowledge and passion for the subject have inspired me to push the boundaries of my understanding and strive for excellence. His dedication to my progress has been a constant source of motivation.

I extend my heartfelt thanks to the members of my Student Research Committee: **Prof. Mukesh Kumar** (ICMR, Delhi), **Dr. Md. Imtaiyaz** (Jamia Millia Islamia, Delhi), **Prof. S.G. Warkar** (Department of Applied Chemistry, DTU), and **Dr. Asmita Das** (Department of Biotechnology, DTU). Their constructive criticism and suggestions have been invaluable in broadening my understanding and refining my research. Each member's diverse perspectives and expertise have significantly contributed to the depth and quality of my work.

I would like to acknowledge the support of the former Vice-Chancellors, **Prof. Yogesh Singh** and **Prof. Jai Prakash Saini**, as well as the current Vice-Chancellor, **Prof. Prateek Sharma**. Their leadership has been instrumental in fostering a conducive research environment and promoting academic excellence at the university. Their vision and commitment to research have greatly benefited my academic journey. Their encouragement and guidance have been invaluable, providing me with the confidence and resources needed to pursue my research goals. I am deeply grateful for their contributions to my academic success.

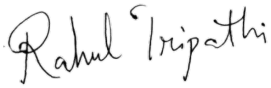
Special thanks to **Prof. Yasha Hasija**, Head and DRC Chairman, Department of Biotechnology, DTU, for her guidance and support throughout this journey. Her insights and encouragement have been crucial in overcoming various hurdles. I also extend my gratitude to **Prof. Jai Gopal Sharma**, former Head of the Department, for his valuable contributions and support during the initial stages of my research. His early guidance set a strong foundation for my work. I would also like to extend my gratitude to **Dr. Rashmi Ambasta** (Vanderbilt

University, USA) for her valuable insights and support. Her expertise and encouragement have been instrumental in refining my research and broadening my understanding of the subject.

I am deeply appreciative of all the faculty members and office staff for their unwavering support and assistance. Their dedication and hard work behind the scenes have ensured a smooth and efficient research process. Their commitment has not only facilitated my progress but also inspired me to strive for excellence. I owe a significant part of my achievements to their invaluable support and encouragement. I am also thankful to Delhi Technological University and the Department of Biotechnology for providing the necessary resources and support to conduct this research.

On a personal note, I am eternally grateful to my family and friends for their unwavering support and encouragement. To my mother, **Jaya Tripathi**, and my father, **Gopalji Tripathi**, thank you for your unwavering love and belief in me. Your teachings have been a guiding light in difficult times, providing me with the strength and resilience to persevere. Lastly, I would like to thank my friend, **Fallon Anastasia**, whose unwavering support and belief in my abilities have been a constant source of encouragement and strength. Her friendship has been a pillar of support throughout this journey.

Thank you all for your support and encouragement.



Rahul Tripathi

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD: Alzheimer's Disease

ADME: Absorption, Distribution, Metabolism and Excretion

AI: Artificial Intelligence

Aβ: Amyloid-β

BBB: Blood-Brain Barrier

CNS: Central Nervous System

CXCL12: CXC motif chemokine 12

CXCR4: CXC Motif Chemokine Receptor 4

DEG: Differentially Expressed Gene

DNN: Deep Neural Network

GPCR: G protein-coupled receptor

IC50: Half-maximal inhibitory concentration

LGBM: Light Gradient Boosting Machine

MD: Mahalanobis distance

ML: Machine Learning

NDD: Neurodegenerative Disorders

PAINS: Pan Assay Interference Compounds

PD: Parkinson's Disease

PDB: Protein Data Bank

RF: Random Forest

SMILES: Simplified Molecular Input Line Entry System

XGB: Extreme Gradient Boosting

α-Syn: α-Synuclein

CHAPTER 1

INTRODUCTION

CHAPTER 1
INTRODUCTION

Neurodegenerative disorders are reported to share a common pathophysiology. Genome-wide association data (GWAS) have identified genetic overlap and revealed common biological pathways between neurodegenerative diseases, such as Alzheimer's disease (AD) and Parkinson's disease (PD) (Bonham et al., 2018). Discovering common genetic architecture among different neurodegenerative disorders may help determine underlying shared disease mechanisms and facilitate early diagnosis and treatment strategies. AD and PD are age-related progressive neurodegenerative disorders with an enormous emotional and economic impact on patients and caregivers. AD is characterised by intracellular neurofibrillary tangles (NFT) composed of aggregated hyperphosphorylated tau protein in the neurons and glial cells (Dickson, Rademakers, & Hutton, 2007; Kovacs, 2015) and extracellular amyloid plaques consisting of aggregated amyloid-β (Aβ) (Dubois et al., 2007; Hyman et al., 2012). Even though PD is conventionally characterised by the aggregation of α-synuclein (α-Syn), tau protein, and NFTs have also been reported to modify PD symptomatology and disease risk (D. J. Irwin, Lee, & Trojanowski, 2013; Nalls et al., 2011; Simón-Sánchez et al., 2009). Besides their activity in the immune system, chemokines and their receptors are remarkably expressed in the central nervous system and modulate cell migration and neurotransmission.

The CXC motif chemokine receptor 4 (CXCR4) belongs to the G protein-coupled receptor (GPCR) protein superfamily. CXCR4 and its ligand CXCL12 are usually linked with hematopoiesis. But, CXCL12 has also been reported to be expressed in other tissues such as the brain, heart, kidney, lung, thymus, liver, and spleen. Likewise, CXCR4-mediated biological pathways are associated with crucial cellular processes like cell proliferation, transport, and stress resistance (Britton, Poznansky, & Reeves, 2021). CXCR4 performs an array of regulatory functions in the immune system and neurodevelopment (Klein & Rubin, 2004; Kokovay et al., 2010; Zou, Kottman, Kuroda, Taniuchi, & Littman, 1998). CXCR4 has been reported to modulate axon guidance and apoptosis via microglial activation and astroglial signalling (Bezzi et al., 2001). CXCR4 has also been reported to be involved in cell cycle regulation via p53 and Rb proteins (M. Z. Khan et al., 2008; Muhammad Zafrullah Khan et al., 2003). CXCR4 and functionally related genes have been associated with increased risk for various age-associated neurodegenerative diseases, such as AD and PD. CXCR4 expression has been observed to be elevated in AD and PD patients as compared to controls (H. Li &

Wang, 2017; Yuanyuan Li et al., 2019). The multifaceted activity of CXCR4 is explained by the complex mechanisms regulating its biological functions, such as receptor crosstalk, isoforms of receptor and ligands, non-canonical ligands, and signalling break-off. Impairment of the CXCR4/CXCL12 signalling and consequent biological processes are associated with an array of disease pathologies, including neurodegenerative disorders, autoimmune diseases, immunodeficiency disorders, developmental abnormalities, and malignancy (Britton et al., 2021).

Due to technological advancements, a large amount of genomic, transcriptomic and proteomic information is now available for discovering new drug targets and screening new lead compounds (Anderson, 2012). Computer-aided drug discovery (CADD) is emerging as a powerful tool that employs this information to accelerate drug design, through advanced computational methods and mathematical modelling to simulate receptor-drug interactions and predict binding affinities to screen potential compounds without prior purchase or synthesis, thereby saving a significant amount of time and money (Baig, Ahmad, Rabbani, Danishuddin, & Choi, 2018). CADD is usually classified into two categories: structure-based drug discovery (SBDD) and ligand-based drug discovery (LBDD). SBDD aims to design new lead compounds by analysing 3D structure information of receptors (proteins) to determine binding sites and interactions vital for their biological activity. While LBDD utilises physicochemical properties and biological activity of known ligands to optimise existing drugs or design new drugs with enhanced activity (Yu & Mackerell, 2017). Some of the commonly utilised CADD methodologies are receptor (protein) identification, molecular docking and simulation studies, QSAR (quantitative structure activity relationship), ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties, molecule design and lead optimisation (Sehgal, Hammad, Tahir, Akram, & Ahmad, 2018). Molecular docking is the most frequently used SBDD methodology facilitated by exponential growth in the number of easily accessible 3D structures of receptors (proteins) and small molecule compounds (ligands), and availability of computational power to analyse them (Stanzione, Giangreco, & Cole, 2021). It is used to predict the orientation and affinity of interaction between a ligand and a protein in its binding site, and provide a docking score that signifies the binding strength between the ligand and the target protein (Tiwari & Singh, 2022).

Neurodegenerative disorder refers to an array of diverse conditions that result in progressive degeneration and loss of neurons, leading to a decline in cognitive and motor functions (Lamptey et al., 2022; Wood, Winslow, & Strasser, 2015). There is a severe lack of effective diagnostic and treatment options for neurogenerative disorders due to the complexity of the molecular mechanisms of their pathophysiology and the heterogeneity of the patient population (Myszczynska et al., 2020). Machine learning (ML), a branch of artificial intelligence (AI) and computer science, involves using data and algorithms to learn and make predictions (Chetty, Hallinan, Ruz, & Wipat, 2022). This can be used for the development of approaches for the diagnosis and treatment of complex conditions, including neurodegenerative disorders, such as Alzheimer's disease (AD) and Parkinson's disease (PD) (Khaliq, Oberhauser, Wakhloo, & Mahajani, 2023). Due to the availability of large amounts of experimental data, AI and ML have shown potential to be an indispensable tool to derive valuable insights and help in decision-making during drug development (Abouchekeir et al., 2022). Thus, the advancements in AI/ML algorithms have presented an unprecedented potential for accelerating drug development through AI/ML-driven approaches (Vatansever et al., 2021).

In our previous study, we observed the upregulation of the CXC motif chemokine receptor 4 (CXCR4), a member of the G protein-coupled receptor (GPCR) protein superfamily, in AD and PD, by analyzing gene expression patterns of AD and PD patients compared to healthy controls of similar age, using publicly available RNA sequencing datasets (Tripathi & Kumar, 2023). Other studies have also reported upregulation of the CXCR4 gene in neurodegenerative disorder patients, including AD and PD, as compared to healthy age-matched controls (H. Li & Wang, 2017; Yuanyuan Li et al., 2019). CXCR4 and its ligand C-X-C motif chemokine ligand 12 (CXCL12) are usually associated with the bone marrow niche and haematopoiesis, the production of blood cellular components. However, CXCL12 is also expressed in various other tissues, like the brain, kidney, lung, heart, etc. Similarly, CXCR4 signalling pathways are crucial for various cellular processes, such as cellular proliferation and transport (Britton et al., 2021). It has been suggested that dysregulation in the expression of CXCR4 and related microglial genes may play a role in age-related neurodegenerative disorders, including AD and PD (Bonham et al., 2018).

In this study, we aim to find natural compounds that can potentially be used for the treatment of age-related neurodegenerative disorders, like AD and PD, through inhibition of

the CXCR4 signalling pathway. Natural compounds have become increasingly more popular as lead compounds for drug development, providing an alternative to synthetic compounds due to their biocompatibility and neuroprotective properties(Andrade et al., 2023). Natural compounds have been suggested to offer protection against excitotoxicity, neuroinflammation, oxidative stress and proteinopathies (Howes & Simmonds, 2014; P. Kumar, Khanum, Khanum, & Khanum, 2012). Various compounds, such as curcumin, quercetin, resveratrol, ginsenosides and rosmarinic acid, are commonly found in plant species used in traditional medicine or food products like fruits, herbs and spices. However, their poor bioavailability and delivery in the central nervous system (CNS) limit their utilisation for therapeutic purposes. For example, even though curcumin has proven benefits in the treatment of AD, it is limited due to low absorption and bioavailability (Sharifi-Rad et al., 2020). New pharmaceutical strategies like nanoparticles and nanocarriers are required for their targeted delivery and controlled release, to increase blood-brain barrier (BBB) permeability and improve stability against metabolic degradation, resulting in elevated bioavailability and reduced toxic side effects (Enrico, 2019).

CHAPTER 2

MATERIAL AND METHODS

CHAPTER 2
MATERIAL AND METHODS

2.1. Data Curation

The datasets with accession numbers GSE67333 and GSE114517 were retrieved from the Gene Expression Omnibus (GEO) (Barrett et al., 2005) database (https://www.ncbi.nlm.nih.gov/geo/). GSE67333 RNA-seq dataset contains 4 samples from late-onset Alzheimer's disease (LOAD) patients and 4 samples from age-matched healthy controls. The dataset is a comprehensive list of transcriptomics alterations and warrants a holistic approach, including both coding and non-coding RNAs in functional studies to understand the pathophysiology of LOAD. Additionally, the GSE114517 RNA-seq dataset contains 17 samples from Parkinson's disease patients and 13 samples from age-matched healthy controls. In GSE114517, for performing RNA-seq, the samples, acquired from the Netherlands Brain Bank, were taken post-mortem from PD and non-PD control donors. The CXCR4 protein structure (PDB Code: 3ODU) was retrieved from Protein Data Bank (www.rcsb.org) (Berman et al., 2000). The acquired CXCR4 protein has 2 chains with a sequence length of 502 and a resolution of 2.50 Å. The structure reveals a consistent homodimer with an interface including helices V and VI. All undesired molecules (water molecules, ligands, and cofactors) were removed, and hydrogen atoms were added. The structure of the reference drug, Mavorixafor, was retrieved from PubChem (www.pubchem.ncbi.nlm.nih.gov/compound/amd-070). ZINC database (www.zinc.docking.org) (J. J. Irwin et al., 2020) was used for structure similarity search, and SDF files of 51 molecules with more than 40% similarity were selected for further analysis. Avogadro software (Hanwell et al., 2012) was used to convert the SDF format to the MOL2 format.

This study utilises the collection and analysis of the data from the Collection of Open Natural Products (COCONUT) Database (Sorokina, Merseburger, Rajan, Yirik, & Steinbeck, 2021) and the Binding Database (BindingDB) (Gilson et al., 2016). The COCONUT Database (http://www.coconut.naturalproducts.net), a large public resource of natural products, was the test data source. To enhance the relevance of the test data, the molecular structure of Mavorixafor, a blood-brain barrier (BBB) permeable inhibitor of CXCR4, was obtained from PubChem (https://www.pubchem.ncbi.nlm.nih.gov/compound/amd-070) and used as a reference to select 10,000 structurally similar compounds (Tanimoto Coefficient > 65%) from

the COCONUT Database. Similarity Search of Coconut Database ranks compounds according to S_{AB} similarity score (defined as similarity between compound A and B) and produces a list of compounds with S_{AB} score greater than a threshold (0.65 in our case). This search produced 10,000 compounds (maximum possible) with the last compound having a similarity score of 65.04%. In our study, we have used a relatively low threshold of 0.65 to increase chemical search space as done in previous studies (Szilágyi et al., 2021). A threshold of 0.65 was also evaluated to be most optimal for the prediction of compound-protein interaction in another study (Mulia, Kusuma, & Afendi, 2018). Simultaneously, the training data containing 1529 compounds with known activity against CXCR4 was sourced from the BindingDB (https://www.bindingdb.org). RDKit (https://www.rdkit.org), used for curation of this data, is an open-source cheminformatics and machine learning toolkit initially developed by Rational Discovery and presently maintained at Novartis Institutes for BioMedical Research (Tosco, Stiefl, & Landrum, 2014). Utilising the RDKit toolkit, Simplified Molecular Input Line Entry System (SMILES) strings of test and training data were standardised into canonical SMILES. This step was crucial in ensuring a consistent and comparable format for the chemical structures within the training and test data. Redundant and duplicate entries were removed from the training data. This resulted in a final, high-quality training data containing 1266 compounds. These compounds were then classified according to their inhibitory activity as inhibitors (IC50 $\leq$ 100 nM) and non-inhibitors (IC50 > 100 nM) for subsequent predictive modelling.

2.2. Molecular docking analysis

The docking analyses were performed using AutoDock Vina (Eberhardt, Santos-Martins, Tillack, & Forli, 2021). To predict the predominant binding mode of identified natural compounds and estimate their binding affinity as inhibitors, we conducted their molecular docking analyses with CXCR4, utilising Vina-GPU 2.0 (Ding et al., 2023). Using graphics processing units (GPUs), Vina-GPU 2.0 accelerates molecular docking techniques (such as AutoDock Vina, QuickVina 2, and QuickVina-W) to facilitate a speedy analysis. Our specific choice, Vina-GPU+ (www.github.com/DeltaGroupNJUPT/Vina-GPU-2.0), optimises the acceleration of Vina-GPU and facilitates batch docking for multiple ligands against a single receptor. The X-ray crystal structure (resolution: 2.50 Å) of our target protein, CXCR4 (PDB Code: 3ODU), was obtained from the Protein Data Bank (https://www.rcsb.org) (Berman et al., 2000). To ensure precision in our analyses, the CXCR4 protein structure was optimised to

remove water molecules, ligands, and cofactors. Hydrogen atoms were then added to the cleaned protein structure using Avogadro software (Hanwell et al., 2012). In the preparation of ligands for docking, Open Babel was employed to convert the Structure Data Format (SDF) files to the MOL2 format (O'Boyle et al., 2011). Subsequently, the refined CXCR4 protein was subjected to molecular docking with Mavorixafor, a reference compound, and the test compounds to evaluate their binding affinities. The determination of potential inhibitors rested on the identification of compounds exhibiting high binding affinities. The interactions between natural compounds and specific residues of the CXCR4 protein were recorded and compared with existing literature. Identification of natural compounds with interactions in line with previous studies increases the reliability of the study.

2.3. Prediction of pharmacokinetic/toxicity (ADME/T) properties

2.5.1. Bioavailability radar

The bioavailability radar is a tool that considers six different physicochemical properties (size, solubility, lipophilicity, flexibility, polarity, and saturation) to predict a molecule's drug-likeness. Each axis of the radar represents one of these properties, whose range is determined by different descriptors (Lovering, Bikker, & Humblet, 2009; Ritchie, Ertl, & Lewis, 2011) and depicted as a pink zone where the radar plot of the molecule must fall for it to be deemed as drug-like.

2.5.2. Physicochemical properties

The physicochemical properties consist of various physical and chemical descriptors like the number of particular atom types, the number of specific bond types, molecular weight (MW), fraction of sp3 hybridized carbons (Fraction Csp3), molecular r efractivity (MR) and topological polar surface area (TPSA) and are utilized to predict absorption and blood-brain permeability. The PSA is calculated using the fragmental technique called TPSA, considering sulfur and phosphorus as polar atoms. This has proven a useful descriptor in many models and rules to quickly estimate some ADME properties, especially concerning biological barrier crossing, such as absorption and brain access (Daina & Zoete, 2016; Lovering et al., 2009).

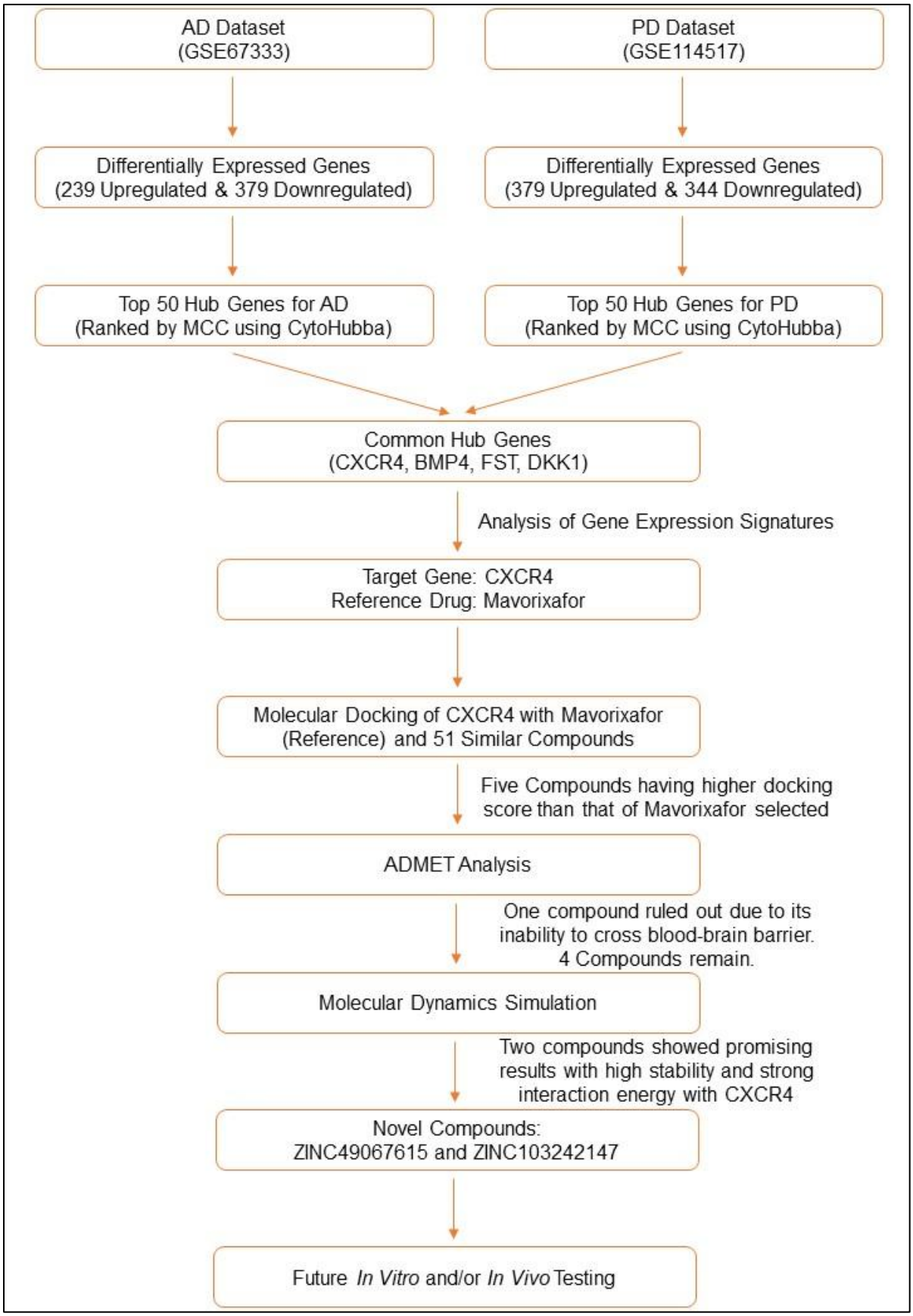


Figure 2.1: Methodology for Virtual Screening of CXCR4 Inhibitors as Potential Therapeutic Agents for Alzheimer's and Parkinson's Diseases. The datasets with accession numbers GSE67333 and GSE114517 were

retrieved from the Gene Expression Omnibus (GEO) database. To determine the differentially regulated gene (DEG) profiles for AD and PD patients, the datasets GSE67333 and GSE114517 were analyzed with the GEO RNA-seq Experiments Interactive Navigator (GREIN) web platform. STRING, a database of known and predicted interacting genes/proteins of a wide range of organisms, was searched for potential protein-protein interactions between the proteins encoded by the DEGs obtained from the datasets GSE67333 (AD) and GSE114517 (PD). Top 50 hub genes were screened separately from the datasets GSE67333 (AD) and GSE114517 (PD) using CytoHubba plugin for Cytoscape. Virtual screening was performed via molecular docking to identify novel potential therapeutic agents, by computing their binding affinity score and interacting residues of our target protein, CXCR4. 3D structure of selected protein CXCR4 (PDB ID: 3ODU) was retrieved from Protein Data Bank. Mavorixafor and all five test compounds were analysed to predict their ADME properties. Five test compound complexes with CXCR4 displaying the highest docking scores were chosen for molecular dynamics simulations to validate the molecular docking results.

2.5.3. Lipophilicity

The lipophilicity of drugs is measured by evaluating the octanol/water partition coefficient (log $P_{o/w}$). Many computational methods for log Po/w estimation were developed with diverse performance on various chemical sets. Common practice is to use multiple predictors either to select the most accurate methods for a given chemical series or to generate consensus estimation. SwissADME (www.swissadme.ch) calculates five predictors (iLOGP, MLOGP, WLOGP, XLOGP3, and SILICOS-IT) to compute the consensus log $P_{o/w}$ (average of all five predictors) (Arnott & Planey, 2012; Mannhold, Poda, Ostermann, & Tetko, 2009). Here, XLOGP3, an atomistic method including corrective factors and knowledge-based library29; WLOGP, our own implementation of a purely atomistic method based on the fragmental system of Wildman and Crippen. MLOGP, an archetype of topological method relying on a linear relationship with 13 descriptors, whereas, SILICOS-IT, a hybrid method relying on 27 fragments and 7 descriptors.

2.5.4. Water solubility

The water solubility of molecules is calculated by SwissADME using three methods – ESOL (Estimated Solubility) (Delaney, 2004), Ali (Ali, Camilleri, Brown, Hutt, & Kirton, 2012), and SILICOS-IT – as Log S values that are the decimal logarithm (i.e., logarithm with base 10) of their molar aqueous solubility and is calculated in mol/l and mg/ml. Their qualitative solubility class is also provided.

2.5.5. Pharmacokinetics

SwissADME utilizes various specialized models to assess the ADME properties of molecules. The first model predicts the blood-brain barrier (BBB) permeability and passive

gastrointestinal absorption (Daina & Zoete, 2016). The second model predicts whether a molecule is a substrate of permeability glycoprotein (Pgp). It assesses a molecule's ability to permeate through membranes such as the gastrointestinal wall and the brain (Montanari & Ecker, 2015). The third model predicts the interaction of a molecule with major cytochrome P450 (CYP) isoforms (CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP3A4) that play a crucial role in drug elimination by utilizing biotransformation. Furthermore, their inhibition leads to drug interactions (Huang et al., 2008), resulting in toxicity or other undesired ramifications. The last model predicts the skin permeability coefficient (Kp) that shows a linear correlation between molecular mass and lipophilicity (Pecoraro et al., 2019). An increase in the negative value of log Kp (cm/s) corresponds to a decrease in skin permeability. SwissADME applied the support vector machine algorithm (SVM) on meticulously cleansed large datasets of known substrates/non-substrates or inhibitors/non-inhibitors.

2.5.6. Drug-likeness

It qualitatively analyses the potential of a molecule to act as a therapeutic agent regarding its oral bioavailability by comparing the physicochemical properties of a molecule with its biopharmaceutical nature inside the human body (Bickerton, Paolini, Besnard, Muresan, & Hopkins, 2012). SwissADME utilizes multiple filters consisting of different ranges of descriptors that must be fulfilled for a molecule to be considered drug-like. There filters are Lipinski (Lipinski, Lombardo, Dominy, & Feeney, 2001), Ghose (Ghose, Viswanadhan, & Wendoloski, 1999), Veber (Veber et al., 2002), Egan (Egan, Merz, & Baldwin, 2000), and Muegge (Muegge, Heald, & Brittelli, 2001). Additionally, a bioavailability score (Martin, 2005) is given to the molecules that describe them in four classes with oral bioavailability of 0.11, 0.17, 0.55, or 0.85. It seeks to predict the probability of a compound to have at least 10% oral bioavailability in rat or measurable Caco-2 permeability. Multiple estimations allow consensus views or selection of methods best fitting the end-user's specific needs in terms of chemical space or project-related demands. Any violation of any rule described here appears explicitly in the output panel.

2.5.7. Medicinal chemistry friendliness

SwissADME removes PAINS (pan assay interference compounds) using two pattern recognition filters. These compounds can potentially cause problems in assays irrespective of the target protein and can also lead to false positive results (Baell & Holloway, 2010). The

problematic functional groups or moieties, which can cause metabolic imbalance, toxicity, or inferior pharmacokinetics, are also displayed as "Structural Alerts". Lead-likeness can be similar to drug-likeness, but it focuses more on the physicochemical properties that must be fulfilled for a molecule to be regarded as a good lead (Brenk et al., 2008). The "Synthetic Accessibility" score that lies between 1 (easy synthesis) and 10 (complex synthesis) provides an estimation of the ease of synthesis of drug-like molecules. It allows the selection of promising virtual screened molecules for synthesis and further biological testing (Ertl & Schuffenhauer, 2009).

2.5.8. Blood-brain barrier permeability prediction and ADME analysis

Assessing a compound's potential to traverse the blood-brain barrier (BBB) is crucial for developing drug candidates targeting the central nervous system (CNS) (D. Roy, Hinge, & Kovalenko, 2019). Our analysis of BBB permeability employed three tools—DeePred-BBB, LightBBB, and SwissADME. DeePred-BBB (www.github.com/12rajnish/DeePred-BBB), utilizing SMILES notations, predicted BBB permeability using a deep neural network (DNN)-based model, boasting superior accuracy (98.07%) compared to one-dimensional convolutional neural network (97.44%) and convolutional neural network (97.61%) (R. Kumar et al., 2022). LightBBB (http://ssbio.cau.ac.kr/software/bbb), a publicly available web server, also predicted BBB permeability through the SMILES query format (Shaker et al., 2021). SwissADME assessed physicochemical properties, pharmacokinetics, drug-likeness, and medicinal chemistry friendliness (Daina, Michielin, & Zoete, 2017). Compounds identified as BBB+ by all three tools underwent further evaluation of their absorption, distribution, metabolism, and excretion (ADME) properties. Drug-likeness, gauging the resemblance of a compound to existing drugs, is crucial for the drug development process (Lee, Jang, Seo, Lim, & Kim, 2022). Medicinal chemistry friendliness is essential to identify fragments that could impede drug development. Pan assay interference compounds (PAINS) denote substances with substructures leading to false positive biological outcomes (Baell & Holloway, 2010). While Brenk alerts point to fragments with potential toxicity, reactivity, or properties detrimental to pharmacokinetics. Compounds with zero violations for drug-likeness filters (Lipinski, Ghose, Veber, Egan, and Muegge) and zero medicinal chemistry friendliness alerts (PAINS and Brenk) constituted the test dataset for machine learning analysis, focusing on their inhibitory activity against CXCR4.

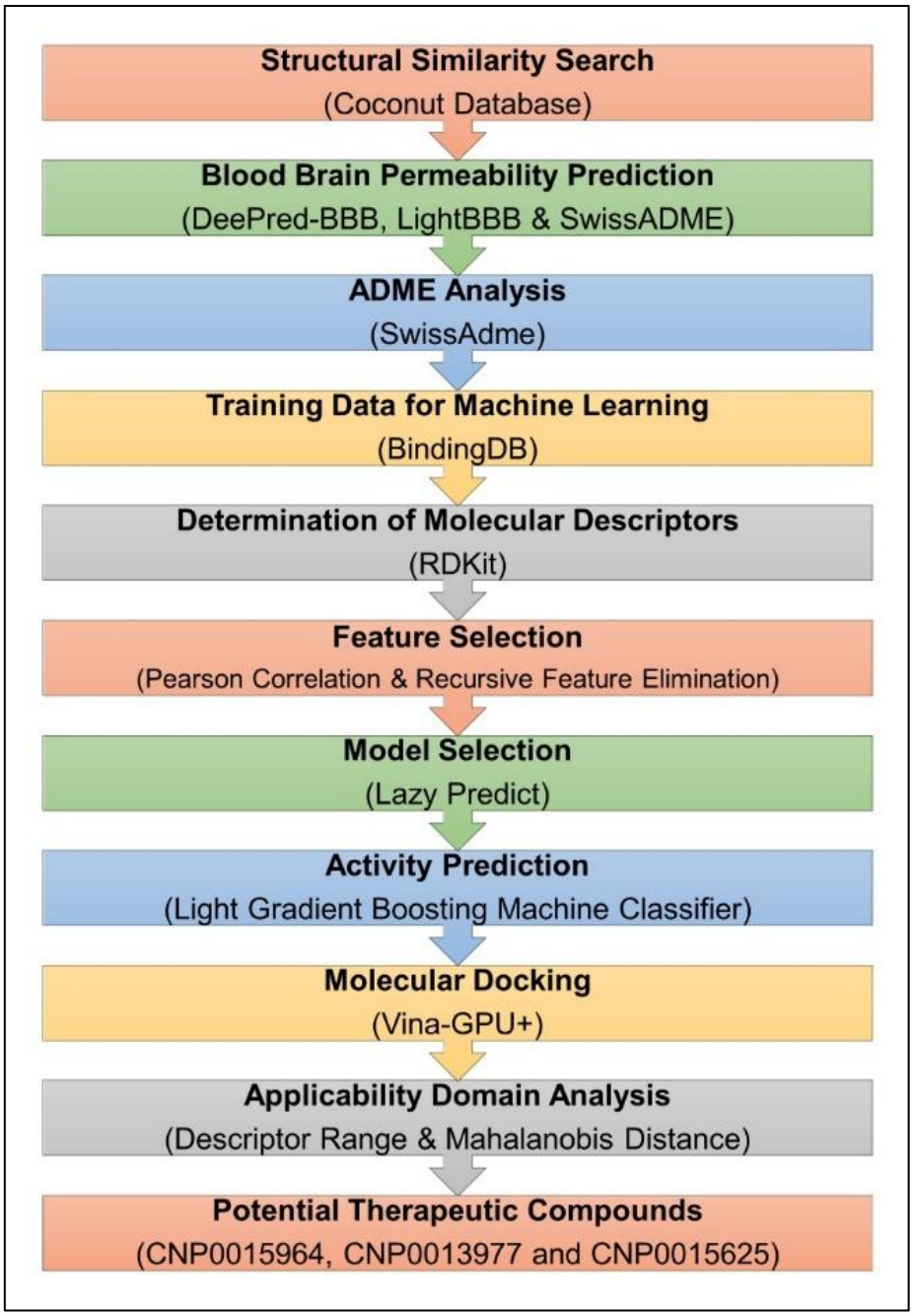


Figure 2.2: Methodology for prediction of CXCR4 inhibitors using machine learning. We initiated the process by extracting 10,000 natural compounds structurally similar to Mavorixafor from the COCONUT Database, focusing on a Tanimoto coefficient greater than 0.65. Subsequent assessments included evaluating the blood-brain barrier (BBB) permeability using DeePred-BBB, LightBBB, and SwissADME, followed by ADME analysis to evaluate drug-likeness and medicinal chemistry friendliness. Compounds with zero violations for different drug-likeness filters and zero medicinal chemistry alerts were selected to prepare the test set. On the other hand, compounds with known activity against CXCR4 were retrieved from BindingDB and used to prepare a training set. Molecular descriptors for both the test dataset and training dataset were calculated using RDKit. Less important descriptors were eliminated using feature selection methods. Lazy Predict aided in selecting the best machine learning model, Light Gradient Boosting Machine (LGBM) Classifier. Applying this model to the test dataset enabled the estimation of CXCR4 inhibitor activity. Compounds predicted to be inhibitors were evaluated for binding affinity through molecular docking and analysed for applicability domain using descriptor ranges and Mahalanobis distance. Compounds with the highest docking scores and presence inside the applicability domain were selected as potential therapeutic compounds.

2.4. Molecular dynamics (MD) simulations

The 50 ns molecular dynamics simulation was done for the docked reference and test compound structures. The best-docked structures were selected for 50 ns MD simulation using GROMACS package (Van Der Spoel et al., 2005). The protein and ligands were split into separate structure files, and their topology of protein was produced using CHARMM36 force field and default (TIP3P) water conditions. Ligand topology was created using the CGenFF webserver by feeding the Mol2 file for ligand after arranging the bond orders. The solvation was performed by defining the unit cell and filling it with a solvent like water. The system was neutralized by introducing the appropriate number of counter ions. The steepest descent method was utilized for energy minimization and subsequently positioned restraining the ligand. Later, the system was equilibrated in a constant volume (NVT) and constant pressure (NPT) environment for 100 ps each. V-rescale temperature coupling was used to maintain 300 K temperature, while C-rescale pressure coupling was used to hold 1 bar pressure. Finally, the production MD was run after releasing the restraints for data collection. The produced trajectories were evaluated for Root Mean Square Deviations (RMSD), Radius of Gyration (Rg), Root Mean Square Fluctuation (RMSF) per residue, and Interaction Energies of Coulomb and Lennard-Jones.

2.5. Machine Learning

Our approach was aimed at using machine learning methodologies to classify our test compounds as either inhibitors or non-inhibitors of CXCR4. For this classification process, we constructed a training and test datasets.

2.5.1. Training Dataset

The training dataset was constructed by retrieving compounds with known CXCR4 activity from BindingDB. To analyze the molecular characteristics of these compounds, we employed RDKit, a sophisticated and versatile cheminformatics toolkit. Molecular descriptors, quantitative representations of chemical features, were calculated for each compound within the training dataset, providing an understanding of their structural and physicochemical properties. Feature selection methods were employed utilizing SciPy (Virtanen et al., 2020) and Scikit-learn (Pedregosa et al., 2011) Python libraries to identify the most relevant features for machine learning. Initial screening involved discarding descriptors with more than 50% zeroes, enhancing the efficiency and relevance of subsequent analyses. Additionally, outliers

were eliminated according to the z-score values of their features using a threshold of 3 standard deviations above and below the mean. Subsequently, we examined the relationships between different descriptors by calculating Pearson correlation coefficients using the pearsonr function from the scipy.stats module of SciPy. Pairs exhibiting correlations beyond a certain threshold underwent refinement, with one descriptor from each pair being eliminated. Visualization of these correlations was achieved through the utilization of Seaborn (Waskom, 2021), a powerful Python library adept at rendering statistical data into insightful heatmaps. Finally, Recursive Feature Elimination (RFE) function from sklearn.feature_selection module of Scikit-learn was utilised to rank the remaining descriptors based on their importance as features for predicting the activity of natural compounds. After the selection of the most relevant features, the Lazy Predict Python library (https://lazypredict.readthedocs.io/en/latest) was used to identify and select the most suitable models appropriate for our training data, optimizing the subsequent phases of analysis. With our chosen models in place, we proceeded to train them using the training dataset.

2.5.2. Test Dataset

The test dataset containing BBB+ natural compounds with desirable ADME (Absorption, Distribution, Metabolism, and Excretion) properties was created from compounds retrieved from COCONUT Database. DeePred-BBB, LightBBB, and SwissADME were used to select BBB+ compounds. Subsequently, SwissADME was used to evaluate violations for drug-likeness filters (Lipinski, Ghose, Veber, Egan, and Muegge) and zero medicinal chemistry friendliness alerts (PAINS and Brenk) and compounds with zero violations were selected for further processing. Just like training dataset, RDKit was employed to analyze the molecular descriptors of these test compounds. SciPy and Scikit-learn Python libraries were again employed to verify that features obtained using training dataset are also relevant to our training data. Relationships between different features in test dataset were evaluated by calculating Pearson correlation coefficients using the pearsonr function from the scipy.stats module of SciPy. Pairs exhibiting correlations beyond a certain threshold underwent refinement, with one descriptor from each pair being eliminated. Visualization of these correlations was achieved through the utilization of Seaborn, a powerful Python library adept at rendering statistical data into insightful heatmaps. The ranges of each feature in training dataset were then calculated and test compounds outside this range were eliminated. This process in essential to ensure the reliability of machine learning predictions. Thus, the remaining compounds constituted the test

dataset for machine learning analysis. Finally, natural compounds within the test dataset were classified with selected training models as either inhibitors or non-inhibitors of CXCR4.

2.6. Performance Metrics of Selected Models

For classification tasks, the results can be summarized into four categories, depending on their true and predicted labels (Varoquaux & Colliot, 2023):

1. True Positives (TP) are compounds with the true and predicted labels as 1.
2. True Negatives (TN) are compounds with the true and predicted labels as 0.
3. False Positives (FP) are compounds with the true label as 0 and the predicted label as 1.
4. False Negatives (FN) are compounds with the true label as 1 and the predicted label as 0.

In our case, '1' was the label given to the inhibitor and '0' was the label given to the non-inhibitor. The following performance metrics were calculated using the above four categories:

1. Accuracy: A fraction of the compounds correctly predicted.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2. Precision: A fraction of the positively predicted compounds that are indeed positive.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall: A fraction of positive compounds correctly predicted as positive.

$$Recall = \frac{TP}{TP + FN}$$

4. Specificity: A fraction of negative compounds correctly predicted as negative.

$$Specificity = \frac{TN}{TN + FP}$$

5. F_1 score: Harmonic mean of Precision and Recall.

$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2TP}{2TP + FP + FN}$$

6. Area under the receiver operating characteristic curve (ROC AUC).

2.7. Applicability Domain

The Applicability Domain, in machine learning, denotes the part of the chemical space where the model can give reliable predictions for compounds present (Ain et al., 2014). Various methods have been developed to evaluate the reliability of the machine learning predictions, including range-based, distance-based, descriptor-based, and fragment-based approaches (Sushko et al., 2010). Mahalanobis distance (MD) is one such distance-based metric that accounts for correlations between different variables, making it particularly effective in multivariate scenarios and is, therefore, widely used for applicability domain assessment (Sahigara et al., 2012). It is defined as the measure of distance between a compound and the distribution of compounds in the training dataset. The calculation of MD is done by computing the mean vector of training data (consisting of the mean of all molecular descriptors) and the inverse of the covariance matrix (which represents the covariance between the descriptors). This allows us to know how far a compound is from the centre or mean of training compounds in a multidimensional chemical space created from molecular descriptors while considering the correlation between different descriptors, explaining how similar a compound is compared to training data (Cabana, Lillo, & Laniado, 2021). In cheminformatics, it is significantly valuable for determining the similarity of compounds to the training data, which allows it to be used for assessment of the applicability domain. The aim is to find a similarity threshold that can be used to predict the inhibitory activity of new compounds reliably. For this purpose, the applicability domain was defined by the calculation of the average (d) and standard deviation (std) of MD of all compounds in the training dataset. These values were calculated utilizing the *mahalanobis* function of *scipy.spatial.distance* module from the SciPy library. The threshold (t) to determine the border of the applicability domain was set as:

$$Threshold\ (t) = Distance\ (d) + Standard\ Deviation\ (std) \times z$$

Where z is an arbitrary parameter (default value is 0.5) (Sahigara et al., 2012). Since MD accounts for the scale, correlation and shape of variables, it is considered better than Euclidean and Manhattan distance-based methods for assessment of the applicability domain. Using compounds structurally similar to Mavorixafor, a compound present in the training set, also improved their applicability to the training model. Furthermore, test data outside the descriptor range of training data was removed before activity prediction to remove compounds that may fall outside the applicability domain. The overall methodology of our study, encompassing data retrieval, preparation, machine learning, molecular docking and applicability domain analysis, has been elucidated in detail in Fig. 2.2.

CHAPTER 3

RESULTS AND DISCUSSION

CHAPTER 3

RESULT AND DISCUSSION

3.1. Differentially regulated genes concurrently expressed in AD and PD patients

The differential expression was measured by comparing the expression level of genes in AD or PD patients with that of age-matched healthy controls (in respective datasets). Differentially expressed genes (DEGs) were identified using two parameters: the fold-change of gene expression (FC) and the statistical significance (p-value). To acquire the list of significant DEGs from the individual datasets, $|\log_2 FC| > 0.8$ and p-value < 0.05 were taken as cutoff values, where DEGs were upregulated if $\log_2 FC > 0.8$ and downregulated if $\log_2 FC < -0.8$. From the GSE67333 dataset of AD patients and aged-matched healthy controls, 617 DEGs were identified, out of which 239 genes were upregulated (p-value < 0.05, $\log_2 FC > 0.8$) and 379 genes were downregulated (p-value < 0.05, $\log_2 FC < -0.8$) (Fig. 3.1). Likewise, from GSE114517 dataset of PD patients and aged-matched healthy controls, 723 DEGs were identified, out of which 378 genes were upregulated (p-value < 0.05, $\log_2 FC > 0.8$) and 344 genes were downregulated (p-value < 0.05, $\log_2 FC < -0.8$) (Fig. 3.1). Finally, to identify common DEGs between AD and PD, two datasets were run on Venn Diagram, and 23 common DEGs were confirmed, out of which 9 genes were upregulated (p-value < 0.05, $\log_2 FC > 0.8$) and 14 genes were downregulated (p-value < 0.05, $\log_2 FC < -0.8$) (Table 3.1 and Fig. 3.1).

Table 3.1: Common Differentially Regulated Genes (DEGs) between datasets GSE67333 and GSE114517:

Gene		GSE67333		GSE114517	
		Log_2FC	**p-value**	**Log_2FC**	**p-value**
Upregulated Genes	Bone Morphogenetic Protein 4 (BMP4)	2.151	0.00023	1.206	0.00766
	Cyclin Dependent Kinase Inhibitor 2B (CDKN2B)	0.882	0.01028	0.999	0.01369
	C-X-C Motif Chemokine Receptor 4 (CXCR4)	0.958	0.03434	0.933	0.00854
	Gonadotropin-Releasing Hormone 1 (GNRH1)	0.986	0.01160	1.420	0.00231
	Glutamate Metabotropic Receptor 2 (GRM2)	1.222	0.00398	1.046	0.04115
	Potassium Calcium-Activated Channel Subfamily M Regulatory Beta Subunit 3 (KCNMB3)	0.896	0.00805	0.903	0.04042
	Solute Carrier Family 13 Member 4 (SLC13A4)	1.594	0.01058	1.304	0.00105
	Solute Carrier Family 26 Member 4 (SLC26A4)	1.063	0.00123	0.928	0.04888
	Solute Carrier Family 5 Member 5 (SLC5A5)	2.996	0.00010	1.050	0.01544
Downregulated Genes	Complement C1q Like 3 (C1QL3)	-0.970	0.01215	-1.626	0.02612
	DCC Netrin 1 Receptor (DCC)	-0.889	0.01781	-1.076	0.01732
	Follistatin (FST)	-1.311	0.00604	-1.939	0.03004
	Granulysin (GNLY)	-1.256	0.00473	-0.984	0.03550
	G Protein-Coupled Receptor 26 (GPR26)	-1.087	0.00291	-1.348	0.02741
	NLR Family Pyrin Domain Containing 2 (NLRP2)	-0.933	0.00145	-2.197	0.00452
	Nuclear Receptor Subfamily 4 Group A Member (NR4A2)	-1.151	0.00624	-0.990	0.03180

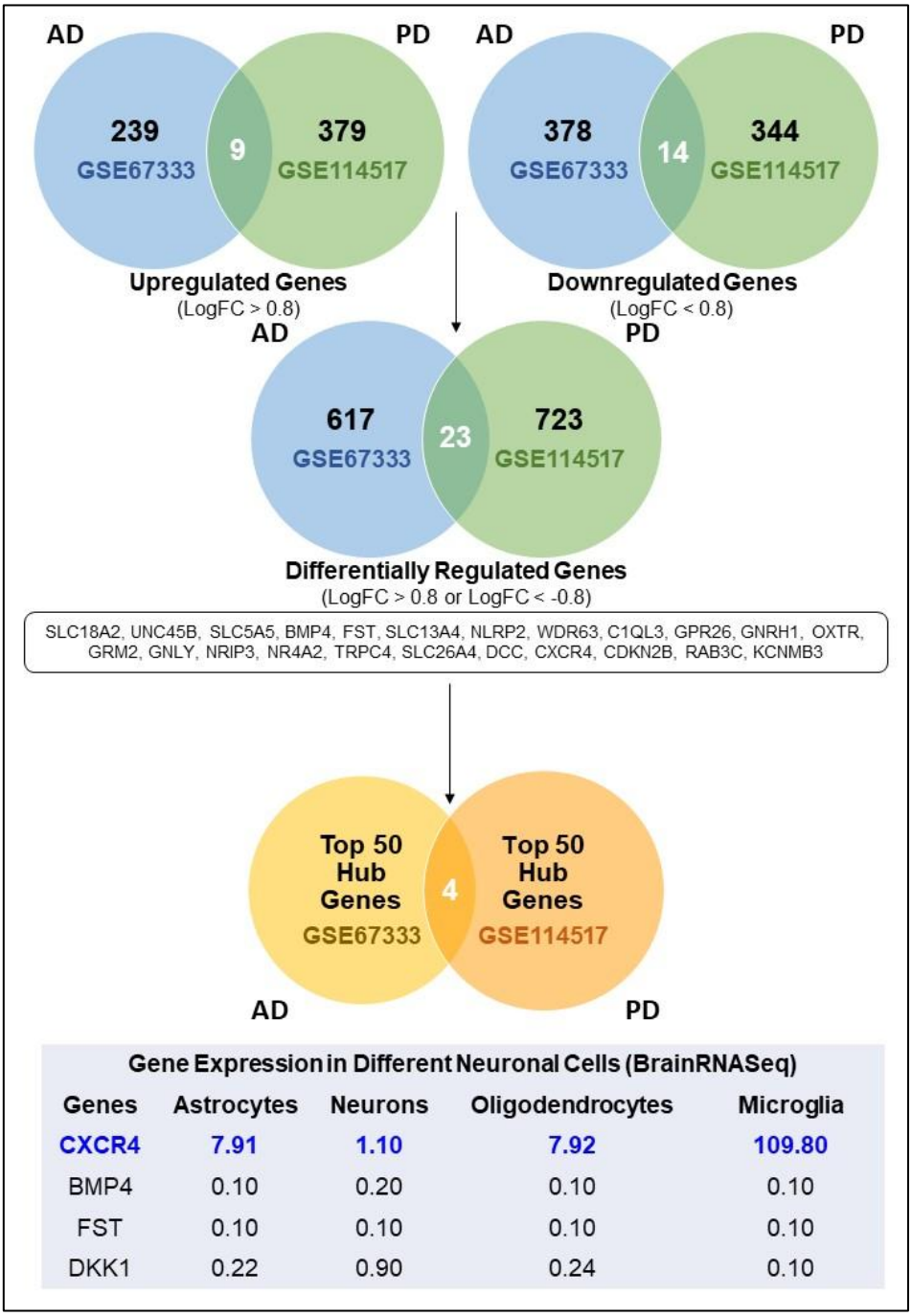
	Nuclear Receptor Interacting Protein 3 (NRIP3)	-1.208	0.00787	-0.990	0.01343
	Oxytocin Receptor (OXTR)	-1.289	0.00787	-1.036	0.00387
	RAB3C, Member RAS Oncogene Family (RAB3C)	-0.803	0.03211	-1.056	0.03173
	Solute Carrier Family 18 Member A2 (SLC18A2)	-2.972	0.00002	-2.379	0.00806
	Transient Receptor Potential Cation Channel Subfamily C Member 4 (TRPC4)	-0.984	0.01605	-1.076	0.03033
	Unc-45 Myosin Chaperone B (UNC45B)	-2.166	0.00101	-1.463	0.00963
	WD Repeat Domain 63 (WDR63)	-1.698	0.00037	-1.099	0.01050

3.2. Protein-protein interaction network and hub genes analysis

Protein-protein interactions play a crucial role in predicting target proteins' biological function in healthy and diseased individuals (Rao, Srinivas, Sujini, & Kumar, 2014). STRING (version 11.5) (Szklarczyk et al., 2023), a database of known and predicted interacting genes/proteins of a wide range of organisms, was searched for potential protein-protein interactions between the proteins encoded by the DEGs obtained from the datasets GSE67333 (AD) and GSE114517 (PD). These interactions can be either physical or functional in nature, and derived from co-expression analysis, text-mining of the scientific literature or computational prediction. Significant interactions, with the minimum confidence score of 0.4, were selected to build PPI networks in order to identify the most important genes and proteins that may play a vital role in development of AD and PD, respectively. The resulting PPI networks were obtained from the STRING database and exported as short tabular text output (.tsv) files. Cytoscape (www.cytoscape.org) (version 3.9.1) was used to analyze these PPI networks and identify the hub genes (Shannon et al., 2003). CytoHubba (www.apps.cytoscape.org/apps/cytohubba) plugin for Cytoscape was used to determine top hub genes graded by Maximal Clique Centrality (MCC) Score (Chin et al., 2014). Top 50 hub genes were screened separately from the datasets GSE67333 (AD) and GSE114517 (PD). 4 common hub genes, CXCR4, DKK1, BMP4 and FST, were found (Fig. 3.1) between the top 50 hub genes obtained from datasets GSE67333 (AD) and GSE114517 (PD). Studies have demonstrated that in AD and PD, CXCR4 modulate TLR4 signaling pathway that leads to MAPK activation and the production of inflammatory cytokines (Yan, Su, & Zhang, 2022). Further, inhibition of Wnt signaling, modulation of Tau phosphorylation and induction of neuronal cell death by DKK1 (Scali et al., 2006). Similarly, Dun et al., 2012 demonstrated that inhibition of canonical Wnt pathway by DKK contributes to the etiology of PD rat model (Dun et al., 2012). Additionally, studies have found the upregulation of TGFβ superfamily ligand, BMP4 in CNS after KA-induced neurodegeneration (Abdipranoto-Cowley et al., 2009). FST a secreted extracellular glycoprotein expressing widely in nervous system. Downregulation of FST disrupts synaptic transmission in hippocampus and leads to cognitive impairments (S. Xiang et al., 2020).

3.3. Analyzing gene expression signatures of screened genes in different cell types

To confirm expression of selected hub genes, BrainRNASeq (www.brainrnaseq.org) was used, as it contains experimental RNA sequencing data of different brain cells, viz. astrocytes, neurons, oligodendrocytes and microglia. In BrainRNASeq, single-cell RNA-Seq profiles for microglia and brain myeloid cells (1816 cells total passed quality control) from different developmental stages (E14.5, P7 and P60) were generated, which were grouped into 15 clusters using Seurat package. A gene was considered to be significantly expressed if its FPKM (Fragments Per Kilobase of transcript per Million) value was more than 0.5. BMP4 was not significantly expressed in any cell type and had FPKM values of 0.10 (astrocytes), 0.20 (neurons), 0.10 (oligodendrocytes) and 0.10 (microglia). FST was also not significantly expressed in any cell type and had FPKM values of 0.10 (astrocytes), 0.10 (neurons), 0.10 (oligodendrocytes) and 0.10 (microglia). DKK1 was not significantly expressed in any cell type besides neurons and had FPKM values of 0.22 (astrocytes), 0.90 (neurons), 0.24 (oligodendrocytes) and 0.10 (microglia). But DKK1 was upregulated in AD while downregulated in PD and, therefore, was ruled out. The only remaining gene, CXCR4, was significantly expressed (FPKM > 0.5) in all cell types, and had FPKM values of 7.91 (astrocytes), 1.10 (neurons), 7.92 (oligodendrocytes) and 109.80 (microglia). It was also upregulated in both AD and PD. Thus, CXCR4 was selected for the next part of the study. Previous studies have reported the upregulation of CXCR4 in AD (Sanfilippo, Castrogiovanni, Imbesi, Nunnari, & Di Rosa, 2020) and PD (Bagheri, Khorramdelazad, Hassanshahi, Moghadam-Ahmadi, & Vakilian, 2019), which is in line with our results. For example, it was found out that CXCR4/CXCL12 based mechanism suggests that Aβ plaques attract microglia to activate the inflammatory cascade by which CXCL12 stimulates CXCR4-dependent signaling both in microglia and in astrocytes to release pro-inflammatory cytokines such as tumor necrosis factor α. The Involvement of Ca^{2+} cascade is also suggested to be involved based on this signaling mechanism which ends in activation of kinases, phosphorylation and further excitotoxicity cascade triggered by excessive stimulation by glutamate (Gavriel, Rabinovich-Nikitin, & Solomon, 2022). Likewise, Tian et al., 2022 demonstrated that HMGB1 A Box protects TH+ neurons by binding CXCR4 to inhibit the migration/infiltration of T cells and macrophages to substantia nigra mediated by HMGB1/CXCL12 complex formed by neuron derived HMGB1, and thus, prevents neuronal damage in PD mice model (Tian et al., 2022).



Gene Expression in Different Neuronal Cells (BrainRNASeq)

Genes	Astrocytes	Neurons	Oligodendrocytes	Microglia
CXCR4	**7.91**	**1.10**	**7.92**	**109.80**
BMP4	0.10	0.20	0.10	0.10
FST	0.10	0.10	0.10	0.10
DKK1	0.22	0.90	0.24	0.10

Figure 3.1: Analysis of Differential Gene Expression. GEO RNA-seq Experiments Interactive Navigator (GREIN) web platform was used to find differentially regulated genes from datasets GSE67333 (AD) and GSE114517 (PD). To acquire the list of significant DEGs from the individual datasets, $|log2FC| > 0.8$ and $p < 0.05$ were taken as cutoff values, where DEGs were upregulated if $log2FC > 0.8$ and downregulated if $log2FC < -0.8$. From GSE67333 Dataset, 617 DEGs were identified, 239 upregulated genes ($p < 0.05$, $log2FC > 0.8$) and 379 downregulated genes ($p < 0.05$, $log2FC < -0.8$). From GSE114517 Dataset, 723 DEGs were identified, 378 upregulated genes ($p < 0.05$, $log2FC > 0.8$) and 344 downregulated genes ($p < 0.05$, $log2FC < -0.8$). The top 50 Hub Genes were identified from both datasets, and common 4 hub genes were selected. To confirm expression of selected hub genes, BrainRNASeq was used, as it contains experimental RNA sequencing data of different brain cells, viz. astrocytes, neurons, oligodendrocytes and microglia. A gene was considered to be significantly expressed if its FPKM value was more than 0.5. Only CXCR4 gene was significantly expressed (FPKM > 0.5) in all cell types and had FPKM values of 7.91 (astrocytes), 1.10 (neurons), 7.92 (oligodendrocytes) and 109.80 (microglia). It was also upregulated in both AD and PD. Thus, CXCR4 was selected for the next part of the study.

3.4. Molecular Docking of Small Molecules (ZINC Database)

Virtual screening was performed via molecular docking to screen novel potential therapeutic agents, by computing their binding affinity score and interacting residues of our target protein, CXCR4. 3D structure of selected protein CXCR4 (PDB ID: 3ODU) was retrieved from Protein Data Bank. Mavorixafor was selected as a reference drug as it is a known inhibitor of CXCR4 and can pass through the blood-brain barrier. Its 3D structure was retrieved from PubChem. The ZINC database was searched to find molecules showing at least 40% similarity with Mavorixafor to find molecules that could have similar biological activity. 52 such molecules were selected as test compounds to perform molecular docking. Mavorixafor and these test compounds were docked with the CXCR4 (PDB ID: 3ODU) using AutoDock Vina. Mavorixafor, our reference drug, displayed docking score or binding affinity (ΔG) of −8.3 kcal/mol. Among our 52 test compounds, five molecules were selected that displayed higher docking score or binding affinity (ΔG) than that of Mavorixafor. These molecules were ZINC103242147 (-8.9 kcal/mol), ZINC1353043237 (-8.6 kcal/mol), ZINC49069258 (-8.5 kcal/mol), ZINC49067615 (-8.5 kcal/mol), ZINC49069264 (-8.5 kcal/mol). This indicates that these molecules could show higher inhibitory activity towards CXCR4 as compared to Mavorixafor. The binding affinity scores and interacting residues for reference and all selected molecules have been listed in Table 3.2.

Table 3.2: Similarity score and docking score of Mavorixafor (Reference) and five selected compounds:

Compound	Formula	Similarity	Docking Score	Residues
Mavorixafor (Reference)	$C_{21}H_{27}N_5$	-	8.3	TYR45, AL112, ILE185, ASP187, GLU288
ZINC49069258	$C_{22}H_{25}N_5$	0.77	8.5	TRP94, ASP97, TRP102, VAL112, TYR116, TYR255, GLU288
ZINC49067615	$C_{21}H_{25}N_5$	0.72	8.5	LEU41, TRP94, ASP97, VAL112, ARG183
ZINC49069264	$C_{21}H_{25}N_5O$	0.70	8.5	TRP94, ASP97, ALA98, HIS113, TYR116, CYS186
ZINC103242147	$C_{23}H_{32}N_4$	0.61	8.9	GLU32, LEU41, TRP94, ASP97, ALA98, VAL112, HIS113, TYR116, CYS186, GLU288
ZINC1353043237	$C_{23}H_{32}N_4$	0.56	8.6	TRP94, ASP97, ALA98, VAL112, HIS113, ASP187

3.5. ADME Analysis of Small Molecules (ZINC Database)

Mavorixafor and all five test compounds were predicted to be orally bioavailable (Fig. 3.2). The radar plot of all six compounds falls in the pink zone and thus can be deemed to be drug-like, satisfying all six criteria: XLOGP3 between −0.7 and +5.0, molecular weight between 150 and 500 g/mol, TPSA between 20 and 130 Å, log S not higher than 6, fraction of carbons in the sp3 hybridization not less than 0.25, and no more than 9 rotatable bonds. The molecular and physicochemical properties of Mavorixafor and five test compounds were evaluated (Table 3.3). Mavorixafor and five test compounds were tested for the five predictors of lipophilicity (iLOGP, XLOGP3, WLOGP, MLOGP, Silicos-IT Log P) and the consensus log $P_{o/w}$ (Table 3.3). The water solubility of Mavorixafor and five test compounds was analyzed by three methods (Ali, ESOL, and SILICOS-IT). The resulting log S values (where S is molar solubility in water) and respective qualitative solubility classes have been mentioned in Table 3.3. Blood-brain barrier (BBB) permeability, gastrointestinal absorption, skin permeation, and effect on permeability glycoprotein (Pgp) and major isoforms of cytochromes P450 (CYP) were predicted for all molecules (Table 3.3). To evaluate the drug-likeness of all molecules were tested by Lipinski, Ghose, Veber, Egan, and Muegge methods (Table 3.3). No molecule exhibited any violation in any of the five methods. The bioavailability score of all molecules was calculated to be 0.55 for all molecules. No undesirable moieties were observed under the PAINS filter for any molecule. Besides ZINC49069258, no other molecule exhibited undesirable moieties under the Brenk filter. ZINC49067615, ZINC49069264, and ZINC103242147 showed violations related to lead-likeness (Table 3.3).

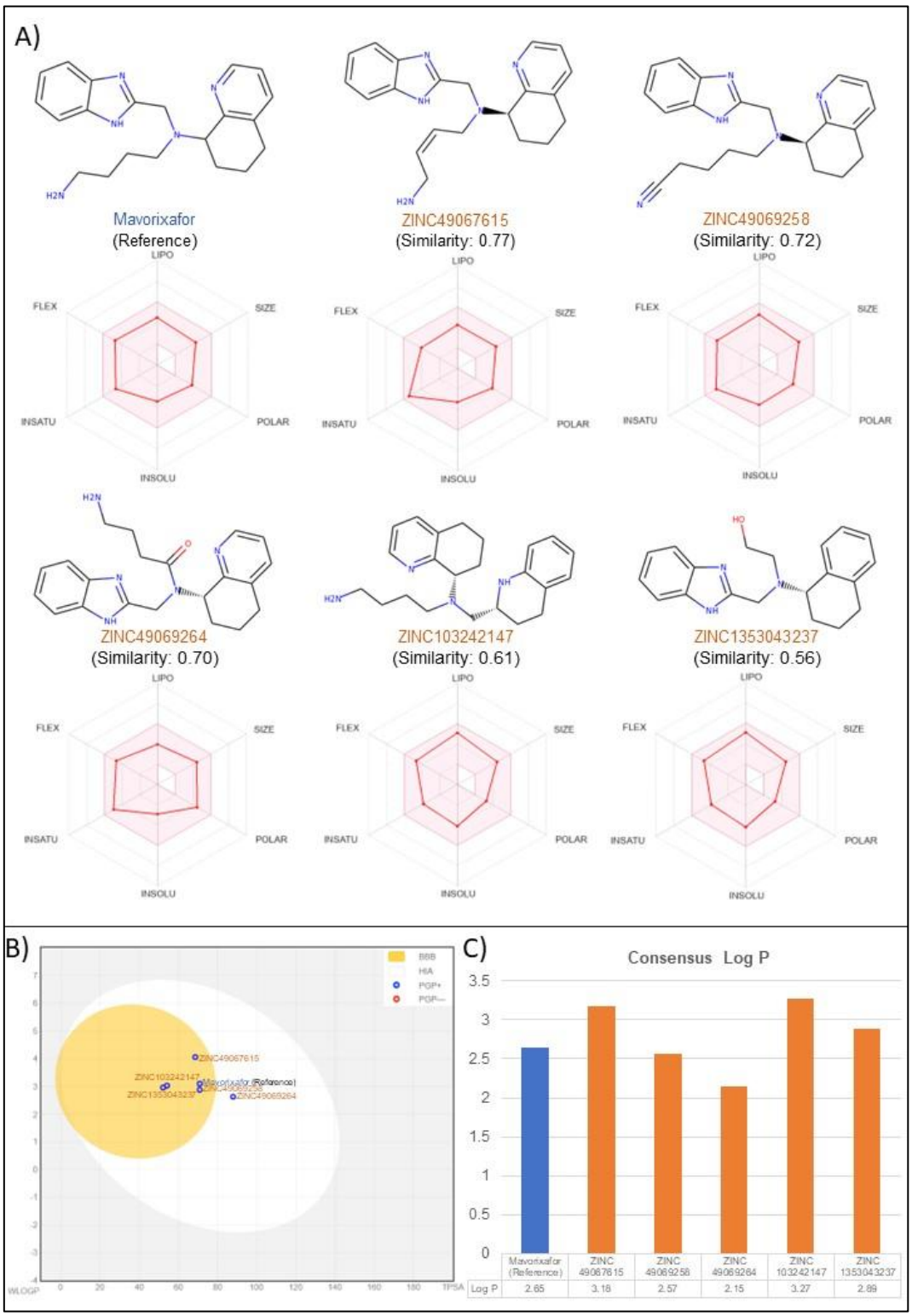


Figure 3.2: ADME Analysis of Mavorixafor (Reference Drug) and Five Test Compounds. A) Molecular structure, bioavailability radar, and similarity score with the reference drug. The pink zone denotes the optimal range for each property (lipophilicity: XLOGP3 b/w −0.7 and +5.0, size: molecular weight b/w 150 and 500 g/mol, polarity: TPSA b/w 20 and 130 Å, solubility: log S not higher than 6, saturation: fraction of carbons in the sp3 hybridization not less than 0.25, and flexibility: no more than 9 rotatable bonds. B) BOILED-Egg prediction of blood-brain barrier penetration and gastrointestinal absorption. C) SwissADME calculates five predictors (iLOGP, MLOGP, WLOGP, XLOGP3, and SILICOS-IT) to compute the consensus log Po/w (mean of all five predictors).

Table 3.3: Different descriptors for ADME properties of Mavorixafor and 5 test compounds:

Descriptors		Mavorixafor (Reference)	ZINC 49067615	ZINC 49069258	ZINC 49069264	ZINC 103242147	ZINC 1353043237
Physicochemical Properties	Formula	$C_{21}H_{27}N_5$	$C_{22}H_{25}N_5$	$C_{21}H_{25}N_5$	$C_{21}H_{25}N_5O$	$C_{23}H_{32}N_4$	$C_{20}H_{23}N_3O$
	MW	349.47	359.47	347.46	363.46	364.53	321.42
	#Heavy atoms	26	27	26	27	27	24
	#Aromatic heavy atoms	15	15	15	15	12	15
	Fraction Csp3	0.43	0.41	0.33	0.38	0.52	0.35
	#Rotatable bonds	7	7	6	7	7	5
	#H-bond acceptors	4	4	4	4	3	3
	#H-bond donors	2	1	2	2	2	2
	MR	105.45	107.29	104.97	105.65	115.54	96.49
	TPSA	70.83	68.6	70.83	87.9	54.18	52.15
Lipophilicity	iLOGP	2.2	2.33	2.67	2.04	3.54	2.3
	XLOGP3	2.38	3.01	1.97	1.41	3.42	2.95
	WLOGP	3.1	4.06	2.88	2.63	3.03	2.96
	MLOGP	1.91	2.06	1.84	1.43	2.66	2.52
	Silicos-IT Log P	3.68	4.44	3.5	3.22	3.71	3.72
	Consensus Log P	2.65	3.18	2.57	2.15	3.27	2.89
Water Solubility	ESOL Log S	-3.47	-3.91	-3.27	-2.93	-4.12	-3.82
	ESOL Solubility (mg/ml)	1.18E-01	4.38E-02	1.88E-01	4.26E-01	2.76E-02	4.82E-02
	ESOL Solubility (mol/l)	3.38E-04	1.22E-04	5.42E-04	1.17E-03	7.56E-05	1.50E-04
	ESOL Class	Soluble	Soluble	Soluble	Solubl e	Moderately soluble	Soluble
	Ali Log S	-3.51	-4.12	-3.08	-2.86	-4.24	-3.71
	Ali Solubility (mg/ml)	1.08E-01	2.76E-02	2.87E-01	5.01E-01	2.11E-02	6.30E-02
	Ali Solubility (mol/l)	3.10E-04	7.67E-05	8.26E-04	1.38E-03	5.78E-05	1.96E-04

	Ali Class	Soluble	Moderately soluble	Soluble	Solubl e	Moderately soluble	Soluble
	Silicos-IT LogSw	-6.9	-7.33	-6.18	-6.43	-7.02	-6.26
	Silicos-IT Solubility (mg/ml)	4.44E-05	1.67E-05	2.31E-04	1.35E-04	3.46E-05	1.77E-04
	Silicos-IT Solubility (mol/l)	1.27E-07	4.63E-08	6.64E-07	3.72E-07	9.48E-08	5.50E-07
	Silicos-IT class	Poorly soluble	Poorly soluble	Poorly soluble	Poorly soluble	Poorly soluble	Poorly soluble
Pharmacokinetics	GI absorption	High	High	High	High	High	High
	BBB permeant	Yes	Yes	Yes	No	Yes	Yes
	Pgp substrate	Yes	Yes	Yes	Yes	Yes	Yes
	CYP1A2 inhibitor	Yes	Yes	No	No	No	Yes
	CYP2C19 inhibitor	Yes	Yes	Yes	Yes	No	No
	CYP2C9 inhibitor	No	Yes	No	No	No	No
	CYP2D6 inhibitor	Yes	Yes	Yes	Yes	Yes	Yes
	CYP3A4 inhibitor	Yes	Yes	Yes	Yes	Yes	Yes
	log Kp (cm/s)	-6.74	-6.36	-7.02	-7.52	-6.1	-6.17
Drug-likeness	Lipinski #violations	0	0	0	0	0	0
	Ghose #violations	0	0	0	0	0	0
	Veber #violations	0	0	0	0	0	0
	Egan #violations	0	0	0	0	0	0
	Muegge #violations	0	0	0	0	0	0
	Bioavailability Score	0.55	0.55	0.55	0.55	0.55	0.55
Medicinal Chemistry Friendliness	PAINS #alerts	0	0	0	0	0	0
	Brenk #alerts	0	0	1	0	0	0
	Lead-likeness #violations	0	1	0	1	1	0
	Synthetic Accessibility	3.38	3.36	3.45	3.31	3.99	2.91

3.6. MD Simulation

Five test compound complexes with CXCR4 displaying the highest docking scores were chosen for molecular dynamics simulations. The RMSD values offered an understanding of these complexes' stability, with the docked structure values lying between 0.15 nm and 0.45 nm (Fig. 3.3A). A low fluctuation in the RMSD values implies that inhibitors achieved stability in the docked structure. RMSF values per residue for all molecules are mostly less than 0.4 nm (Fig. 3.3B), indicating the stability of complexes. Additionally, the elastic stability of the studied complexes was examined as Radius of Gyration (Rg) values varied between 2.1 nm and 2.4 nm, indicating the system was able to achieve stability after 15 ns (Fig. 3.3C). ZINC103242147 showed a stronger Coulomb potential of interaction (Fig. 3.3D), while ZINC49067615 a stronger Lennard-Jones potential of interaction energy (Fig 3.3E). This suggests that these two molecules could show more significant inhibitory activity than Mavorixafor. The molecular dynamics trajectory was divided into 10 ns intervals to analyze the minimum distance and interactions between protein residues and ligand molecules. The first 10 ns was excluded as the duration needed for the system to reach stability. The minimum distance was calculated using the "gmx mindist" command of gromacs, and residues with an average distance of 0.35 nm or less have been shown in Fig 3.4. The heatmap was generated for residues showing the amount of time distance was less than 0.35 nm. The green color signifies that the distance was less than 0.35 nm for 100% of the duration, while the yellow indicates that the distance was less than 0.35 nm for 0% of the duration. Specific residues like TYR45, TRP94, and ASP97 can be seen to be present for most of the duration of the simulation. Interaction between ligand and protein residues at 10 ns, 20 ns, 30 ns, 40 ns, and 50 ns have been shown in Fig 3.5. BIOVIA Discovery Studio was used to visualise these interactions. Here again, specific residues like TYR45, TRP94, and ASP97 can be seen to be interacting with ligands at different timeframes of the simulation, signifying the importance of these residues for receptor-ligand interaction.

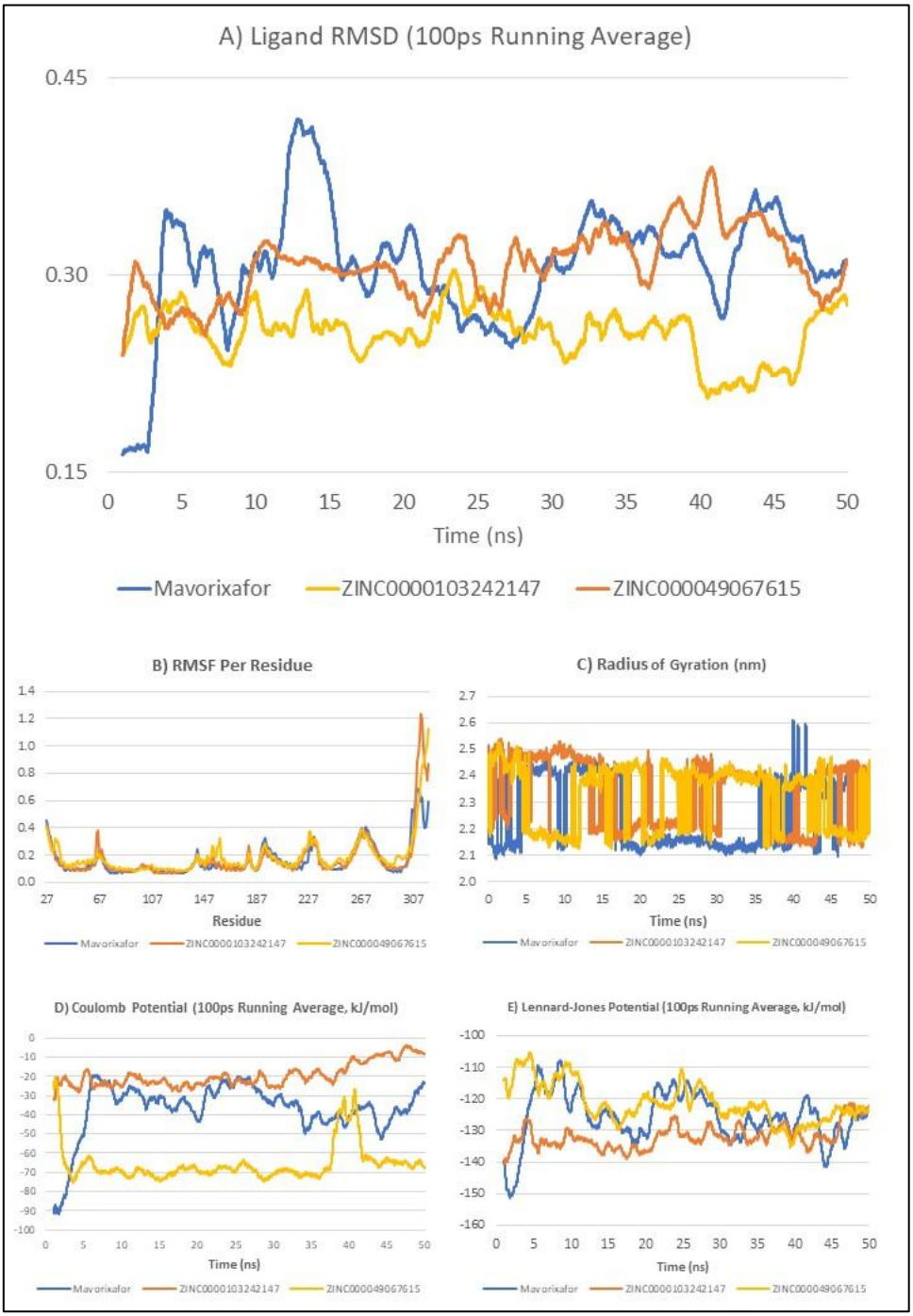


Figure 3.3: Results of Molecular Dynamics Simulation. A) RMSD of ligand in terms of 100ps running averages. The RMSD values of these complexes lied between 0.15 nm and 0.45 nm. A low fluctuation in the RMSD values implies that inhibitors achieved stability in the docked structure. B) Root Mean Square Fluctuation (RMSF) per residue. RMSF values per residue for all molecules are mostly less than 0.4 nm, indicating the stability of complexes. C) Radius of gyration of the protein-ligand complexes. The elastic stability of the studied complexes was examined as Radius of Gyration (Rg) values varied between 2.1 nm and 2.4 nm, indicating the system was able to achieve stability after 15 ns. D) Interaction energy between protein and ligand – Coulomb Potential. ZINC103242147 showed a stronger Coulomb potential of interaction. E) Interaction energy between protein and ligand – Lennard-Jones Potential. ZINC49067615 a stronger Lennard-Jones potential of interaction energy.

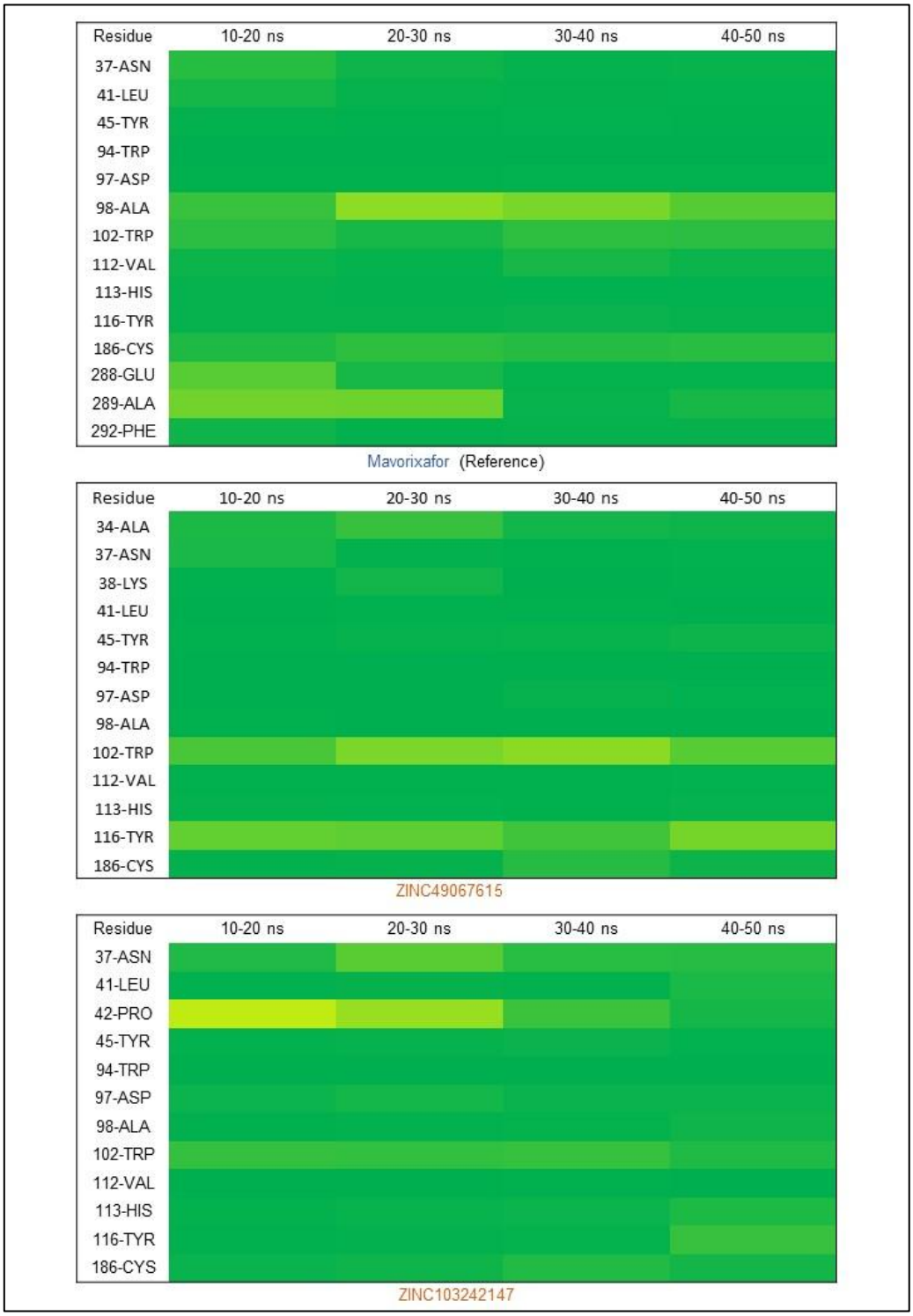


Figure 3.4: Minimum distance between ligand and protein residues. The minimum distance was calculated using the "gmx mindist" command of gromacs, and residues with an average distance of 0.35 nm or less have been shown. The first 10 ns of the simulation were excluded as the duration needed for the system to reach stability. The simulation was divided into 10 ns intervals, and a heatmap was generated by showing the amount of time distance was less than 0.35 nm. Green signifies that the distance was less than 0.35 nm for 100% of the duration, while yellow indicates that the distance was less than 0.35 nm for 0% of the duration. Specific residues like TYR45, TRP94, and ASP97 can be seen to be present for most of the duration of the simulation.

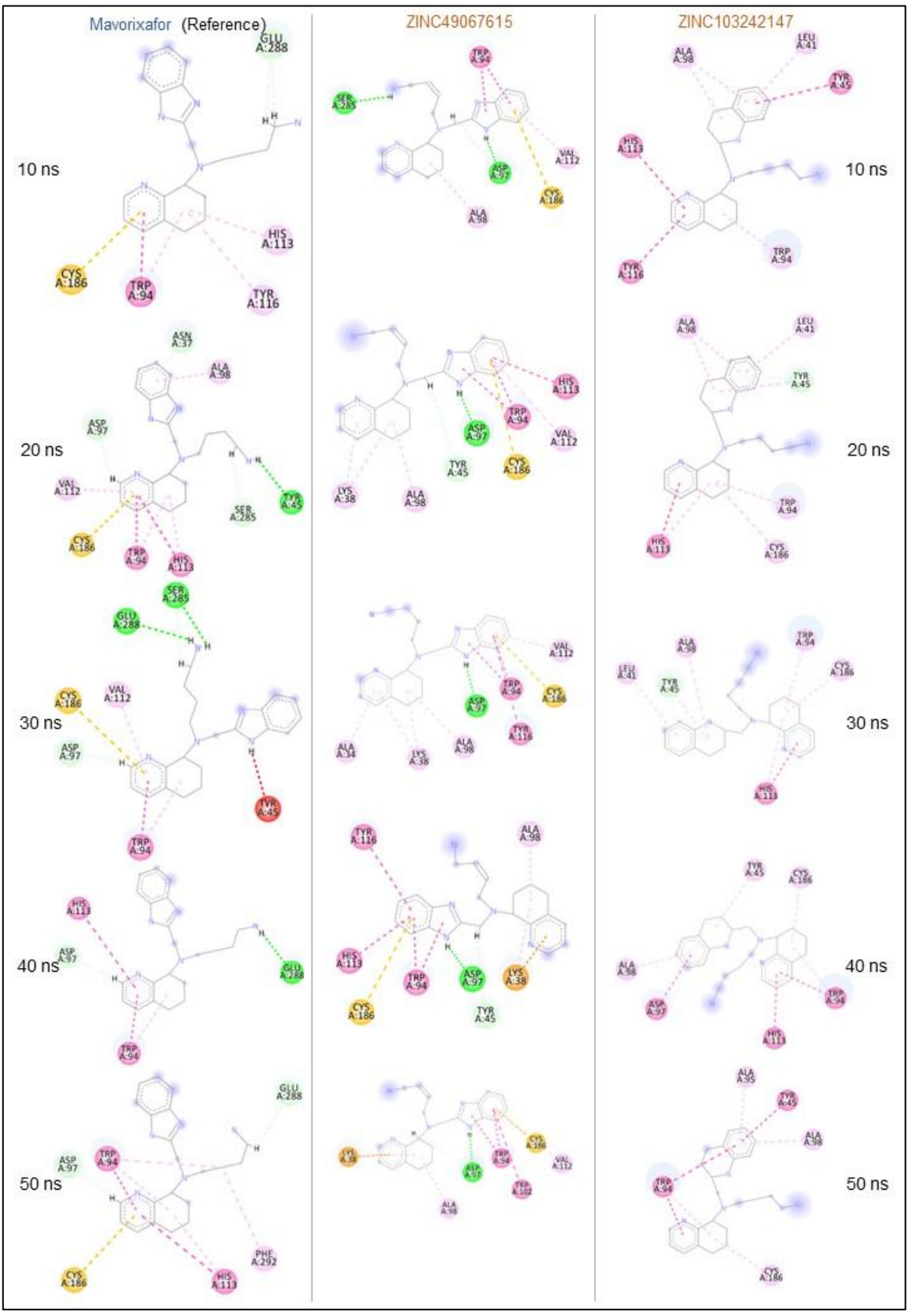


Figure 3.5: Interaction between ligand and protein residues at 10 ns, 20 ns, 30 ns, 40 ns, and 50 ns. The first 10 ns of the simulation were excluded as the duration needed for the system to reach stability. Certain residues like TYR45, TRP94, and ASP97 can be seen to be present throughout the different time points (10 ns, 20 ns, 30 ns, 40 ns, and 50 ns) of the simulation. BIOVIA Discovery Studio was used to visualise these interactions.

3.7. Blood-Brain Barrier Permeability and ADME Analysis of Natural Compounds

The significance of structural similarity is fundamental in the domain of drug design and discovery. This significance is based on the core assumption that compounds with similar structures are likely to manifest similar biological activities (A. Kumar & Zhang, 2018). In the specific context of our study, our primary objective was to identify natural compounds sharing structural similarity with Mavorixafor, an inhibitor of CXCR4 that is permeable through the blood-brain barrier (BBB). Thus, we retrieved 10,000 natural compounds (from the COCONUT Database) exhibiting a structural similarity exceeding 65% (using the Tanimoto coefficient) to Mavorixafor. Since the goal of our study is to identify CXCR4 inhibitors for the treatment of neurodegenerative disorders, an important criterion of our selection process was the BBB permeability of compounds as this barrier selectively allows only certain molecules to penetrate the brain while preventing the transport of neurotoxins (O'Connor et al., 2020; Pardridge, 2012). To address this, we analyzed all 10,000 natural compounds using three distinct tools (DeePred-BBB, LightBBB, and SwissADME) and selected 2,712 compounds that were predicted to be BBB+ by all three tools. Subsequently, these compounds were analyzed for their pharmacokinetic properties. The parameters of absorption, distribution, metabolism, and excretion (ADME) serve as pivotal metrics in comprehending a drug's pharmacokinetics (Doogue & Polasek, 2013). ADME describes how our body processes a drug while it executes its therapeutic function and, thereby, provides an understanding of its molecular mechanisms, toxicity and drug-drug interactions (Yuhua Li et al., 2019; X. Wang et al., 2013). In an era before the formal integration of ADME testing into the drug discovery pipeline, a staggering 40% of compounds failed due to suboptimal pharmacokinetic properties. Thanks to advancements in this field, the failure rate has been now significantly reduced to less than 10% (Kola & Landis, 2004). SwissADME was used to evaluate the ADME properties of BBB+ compounds, and those found to violate any of the five drug-likeness filters (Lipinski, Ghose, Veber, Egan, and Muegge) were systematically excluded. These are widely accepted guidelines in drug discovery and development to evaluate 'drugability', the ability of a compound to be used commercially as a pharmaceutical drug (Avti, Singh, Dahiya, & Khanduja, 2023). Similarly, compounds exhibiting one or more alerts for medicinal chemistry unfriendliness filters (PAINS and Brenk) were also eliminated. These filters are used to identify fragments or substructures of a compound that can lead to unfavourable chemical properties such as toxicity or non-specificity (Alam & Khan, 2019). The removal of these compounds with poor pharmacokinetic properties is vital for increasing the reliability of the study. Thus, in the final refinement, 1,566 compounds devoid of drug-likeness violations and medicinal chemistry unfriendliness alerts, were selected to form the test set of natural compounds (see Fig. 3.6).

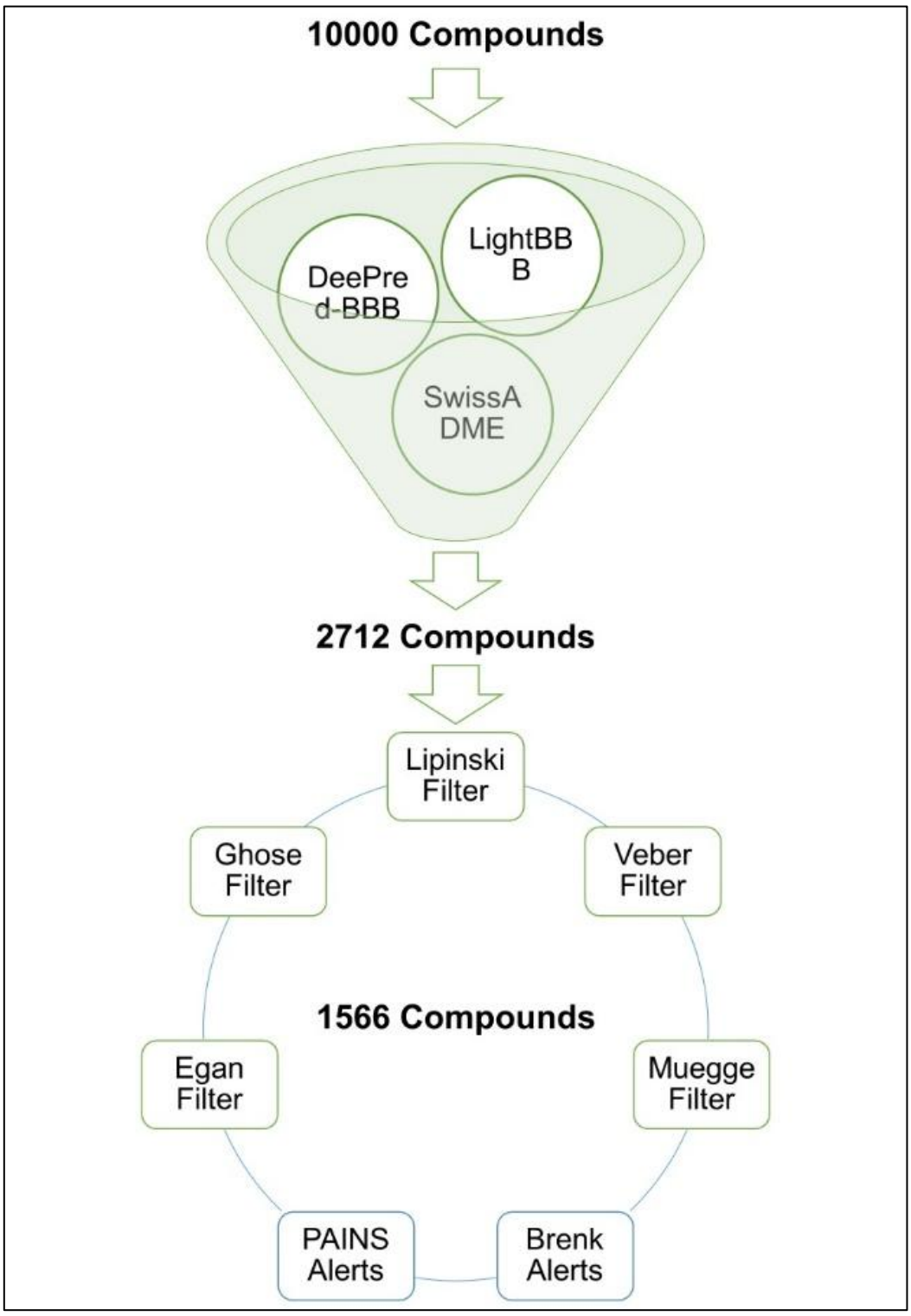
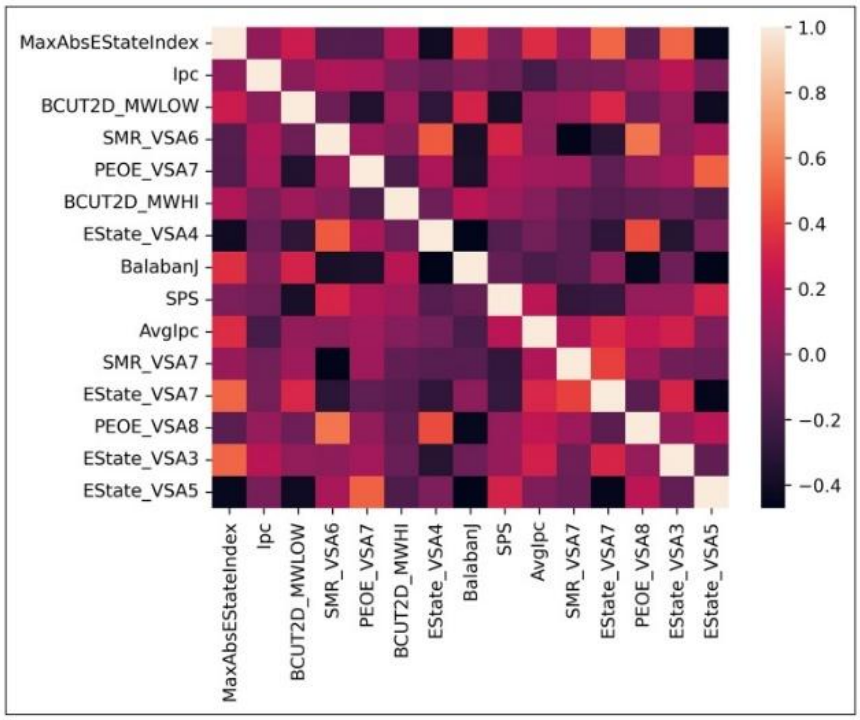


Figure 3.6: Prediction of blood-brain permeability and ADME analysis of natural compounds. 10,000 compounds obtained from the COCONUT Database were evaluated for their blood-brain barrier (BBB) permeability using three distinct tools, namely DeePred-BBB, LightBBB, and SwissADME, to ensure a comprehensive assessment. Subsequently, 2,712 compounds emerged as promising candidates, being predicted as BBB+ by all three algorithms. The selected compounds underwent further scrutiny to determine their drug-likeness and medicinal chemistry friendliness. The compounds were evaluated using five drug-likeness filters (Lipinski, Ghose, Veber, Egan, and Muegge) and two medicinal chemistry filters (PAINS and Brenk). Following this analysis, 1,566 compounds displayed zero violations for the drug-likeness filters and zero alerts for medicinal chemistry filters. Thus, compounds with positive blood-brain permeability, optimal drug-likeness and favourable medicinal chemistry attributes were selected for further analysis.

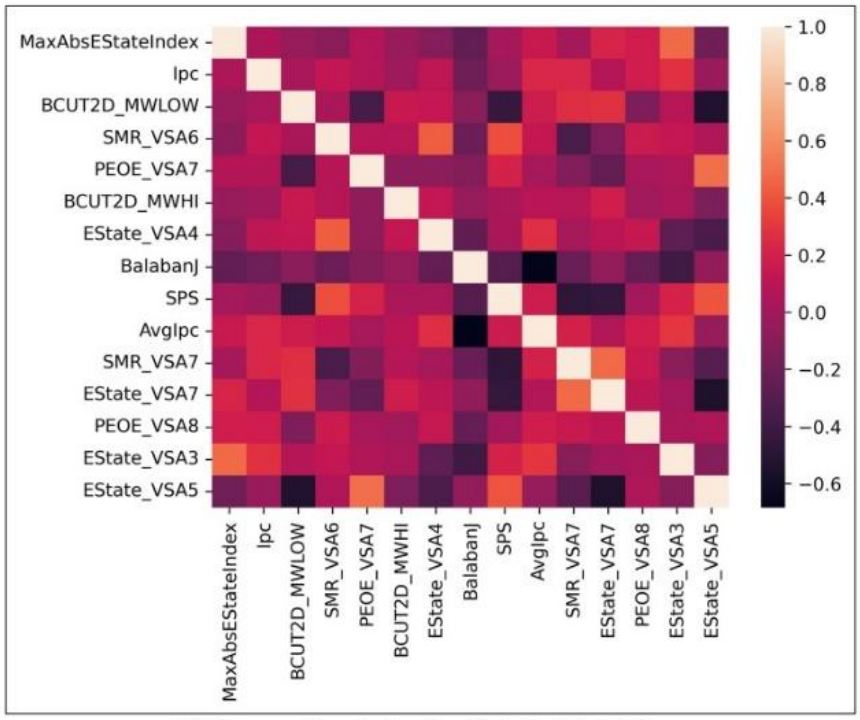
3.8. Calculation of Molecular Descriptors and Feature Selection

To predict the activity of compounds against CXCR4, we created a training set using compounds retrieved with known activity against CXCR4 from the BindingDB, recording their IC50 values and SMILES notations. After data curation that involved standardization of data and removal of duplicates 1266 compounds were left in training data. Compounds with IC50 values ≤ 100 nM were categorized as inhibitors, while those with IC50 values > 100 nM were labelled as non-inhibitors. Recognizing the intrinsic link between a compound's molecular structure and its biological activity (Todeschini & Consonni, 2010), we evaluated molecular descriptors — depicting various physical and chemical characteristics — to enhance the efficiency of our drug design and discovery process. Utilizing the Descriptors module of RDKit, an open-source tool for cheminformatics and machine learning, we calculated descriptors for both the training and test datasets. To increase the speed and accuracy of machine learning model creation, redundant data was removed, and most crucial features were selected for machine learning. Firstly, descriptors with more than 50% zero values were eliminated from either dataset, followed by the removal of outliers.

In the training set, any outlier with a z-score of 3 standard deviations above and below the mean for any feature was removed, finally resulting in 674 compounds in the training data, which included 322 inhibitors and 352 non-inhibitors. Subsequently, we scrutinized the remaining descriptors for significant correlations, utilizing the Pearson correlation coefficient computed with SciPy. If a pair of descriptors had the Pearson correlation coefficient exceeding 0.60 in the training dataset, one of the descriptors was removed. This process led to the selection of 17 descriptors out of the original 209 for further analysis. The Recursive Feature Elimination (RFE) algorithm of Scikit-learn ranked these 17 descriptors based on their feature importance by eliminating the least important features one by one, resulting in a list of descriptors ranked by their significance. A threshold of 0.60 was selected by using Grid Search on different thresholds and using RFE to select the threshold with the best model accuracy. The heatmaps of the Pearson correlation coefficients of the top 15 descriptors ranked by RFE for both the training and test dataset was generated using the Seaborn Python library (Fig. 3.7).



A) Pearson Correlation Coefficient of Training Set



B) Pearson Correlation Coefficient of Test Set

Figure 3.7: Selection of molecular descriptors as best features for generation of machine learning model. Molecular descriptors were first evaluated for both the training dataset and test set. To eliminate less important descriptors, feature selection was performed. Descriptors with more than 50% zero values were eliminated. Pearson Correlation Test was performed for the rest of the descriptors and if the Pearson Correlation Coefficient of two descriptors was more than 0.60, one descriptor of the pair was eliminated. Finally, 17 descriptors were selected for further analysis. A heatmap of Pearson Correlation Coefficients was prepared for both A) the Training dataset and B) the Test dataset, using the top 15 descriptors, predicted by Recursive Feature Elimination, that displayed maximum model accuracy.

3.9. Model Selection and Activity Prediction

After the creation of training and test datasets, the Lazy Predict Python library was used to identify the most optimum machine learning algorithm, as it can fit and evaluate all the machine learning models contained in the scikit-learn Python library with a minimal amount of coding in a single step. It is composed of a Lazy Classifier and Lazy Regressor to perform the classification and regression tasks, respectively. We used Lazy Classifier to find a model for the classification of natural compounds as CXCR4 inhibitors or non-inhibitors by building and evaluating 27 different machine learning models using our training dataset. We used Lazy Classifier to select the best models for different numbers of descriptors removing the lowest RFE ranked descriptors one by one. After building models for different numbers of descriptors, we compared the highest accuracy obtained with each set of descriptors. The highest accuracy obtained for each set of descriptors has been shown in Fig. 3.8(A). Additionally, the names and accuracy of the top three models for different numbers of descriptors have been mentioned in Table 3.4. The highest model accuracy was achieved at the set of top 15 ranked descriptors for Light Gradient Boosting Machine (LGBM) Classifier with 81.94% model accuracy, followed by eXtreme Gradient Boosting (XGB) Classifier with 79.86% model accuracy and Random Forest (RF) Classifier with 77.48% model accuracy. All the models and their accuracy with the set of top 15 ranked descriptors have been shown in Fig. 3.8(B).

Table 3.4: Comparison of accuracy of top 3 models with different number of descriptors:

Features	Best Model	Accuracy	2nd Best Model	Accuracy	3rd Best Model	Accuracy
1	ExtraTreeClassifier	67.26%	RandomForestClassifier	67.06%	ExtraTreesClassifier	66.37%
2	ExtraTreeClassifier	67.26%	RandomForestClassifier	66.27%	KNeighborsClassifier	66.17%
3	ExtraTreesClassifier	75.40%	XGBClassifier	75.10%	DecisionTreeClassifier	73.81%
4	ExtraTreesClassifier	78.27%	LGBMClassifier	76.79%	RandomForestClassifier	76.69%
5	LGBMClassifier	75.99%	RandomForestClassifier	75.89%	ExtraTreesClassifier	75.00%
6	ExtraTreesClassifier	79.66%	RandomForestClassifier	76.79%	XGBClassifier	76.79%
7	XGBClassifier	77.28%	ExtraTreesClassifier	75.89%	LGBMClassifier	75.10%
8	LGBMClassifier	78.27%	XGBClassifier	78.17%	ExtraTreesClassifier	76.59%
9	LGBMClassifier	77.38%	RandomForestClassifier	75.30%	ExtraTreesClassifier	73.61%
10	RandomForestClassifier	79.07%	LGBMClassifier	76.79%	XGBClassifier	74.40%
11	RandomForestClassifier	78.17%	LGBMClassifier	75.99%	XGBClassifier	75.30%
12	RandomForestClassifier	78.97%	LGBMClassifier	77.48%	ExtraTreesClassifier	77.18%
13	LGBMClassifier	77.48%	RandomForestClassifier	77.38%	XGBClassifier	76.79%

14	LGBMClassifier	79.76%	XGBClassifier	79.17%	RandomForestClassifier	76.49%
15	**LGBMClassifier**	**81.94%**	XGBClassifier	79.86%	RandomForestClassifier	77.48%
16	RandomForestClassifier	79.86%	LGBMClassifier	79.76%	ExtraTreesClassifier	76.59%
17	LGBMClassifier	79.56%	XGBClassifier	78.47%	RandomForestClassifier	78.08%

Finally, hyperparameter tuning was performed for all three models to find the optimum set of parameters.

LGBM Classifier again showed the best model accuracy of 0.8222 (or 82.22%) with the parameters:

$$colsample\_bytree: 0.6, learning\_rate: 0.1, max\ \_depth: 7, min\_child\_samples: 20$$

$$n\_estimators: 100, num\_leaves: 20, subsample: 0.6$$

XGB Classifier showed an accuracy of 0.8074 (or 80.74%) with the following parameters:

$$colsample\_bytree: 0.8, learning\_rate: 0.2, max\_depth: 3, n\_estimators: 100, subsample: 0.7$$

Random Forest Classifier showed an accuracy of 0.8 (or 80.00%) with the parameters:

$$max\_depth: None, min\_samples\_leaf: 1, min\_samples\_split: 10, n\_estimators: 300$$

Along with Model Accuracy, these three models were also tested for various performance metrics such as Precision, Recall, Sensitivity, F1-score and ROC AUC Score. To calculate all these performance metrics, including accuracy, 20% of training dataset was used as the validation dataset to perform holdout validation of data. In the holdout validation approach, training set is split into two parts, one for training (70-80%) and one for validation (20-30%). Subsequently, 70-80% training set is used to tune the model and remaining 20-30% is used to evaluate the model (Allgaier & Pryss, 2024). Precision signifies the fraction of positively predicted compounds are actually positive. LGBM Classifier show precision of 0.9 (or 90%). Specificity signifies the fraction of negatively compounds that are actually predicted as negative. Specificity of LGBM Classifier was calculated to be 0.9048 (or 90.48%). Both of these metrics shows that our model (LGBM Classifier) is highly accurate in classifying negatives as true negatives instead of false positives. Therefore, the LGBM Classifier was used for the prediction of the activity of test data using the top 15 ranked descriptors. Performance metrics (Accuracy, Precision, Recall, Sensitivity, F1-score and ROC AUC Score) of top 3 models have been mentioned in the Table 3.5.

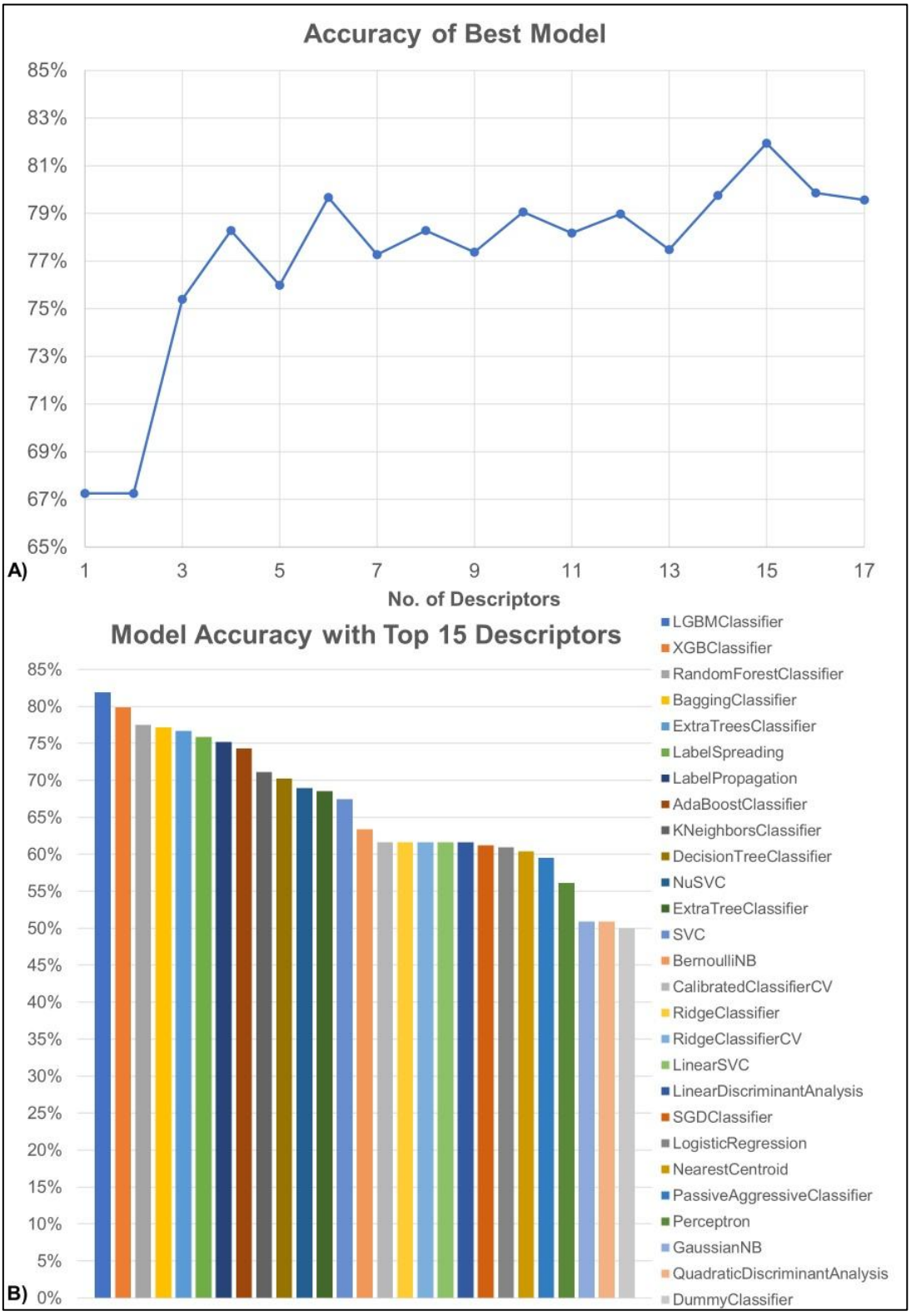


Figure 3.8: Selection of best machine learning model for classification of activity of natural compounds as inhibitors and non-inhibitors of CXCR4. 17 descriptors obtained after the Pearson Correlation Test were ranked by their feature importance using the Recursive Feature Elimination (RFE) method to improve model performance by removing less important features. Lazy Predict was used to select the best model for different numbers of descriptors removing lowest lowest-ranked descriptors one by one. Best model accuracy was achieved at 15 descriptors with 81.94% accuracy for the Light Gradient Boosting Machine (LGBM) Classifier. A) Comparison of balanced accuracy of best models at the different number of descriptors. B) Comparison of accuracy of different models with a dataset of 15 highest ranked descriptors.

Table 3.5: Performance metrics of top 3 models selected through Lazy Predict and hyperparameter tuning:

Classifier	Parameter	Accuracy	Precision	Recall	Specificity	F1 Score	ROC AUC Score
Light Gradient-Boosting Machine (LGBM)	colsample_bytree: 0.6 learning_rate: 0.1 max_depth: 7 min_child_samples: 20 n_estimators: 100 num_leaves: 20 subsample: 0.6	0.8222	0.9000	0.7500	0.9048	0.8182	0.8274
Extreme Gradient Boosting (XGB)	colsample_bytree: 0.8 learning_rate: 0.2 max_depth: 3 n_estimators: 100 subsample: 0.7	0.8074	0.8594	0.7639	0.8571	0.8088	0.8105
Random Forest (RF)	max_depth: None min_samples_leaf: 1 min_samples_split: 10 n_estimators: 300	0.8000	0.8571	0.7500	0.8571	0.8000	0.8036

Before using our trained model for activity prediction, it was important to remove outliers from our test data to only consider molecules within the descriptor space of training data to ensure consistency between both datasets. Thus, the range of features of training data was evaluated and outliers in the test data that were outside the range for corresponding features of training data were identified and removed. This was done to mitigate the data discrepancies between training and test datasets, thereby improving the reliability and applicability of the developed machine learning model. By aligning the range of features in both datasets and removing outliers that could adversely affect model accuracy, the model performance on unseen data is expected to improve. This resulted in 975 compounds in our test data that we can use for activity prediction. Violin plots were generated using the Seaborn python library to visualize feature distribution of test and training data (Fig. 3.9). The LGBM Classifier machine learning model was then used to classify our test compounds as inhibitors and non-inhibitors of CXCR4. Finally, 20 compounds that were classified as CXCR4 inhibitors were analyzed for their binding affinity by molecular docking and applicability to the trained model through analysing the Applicability Domain.

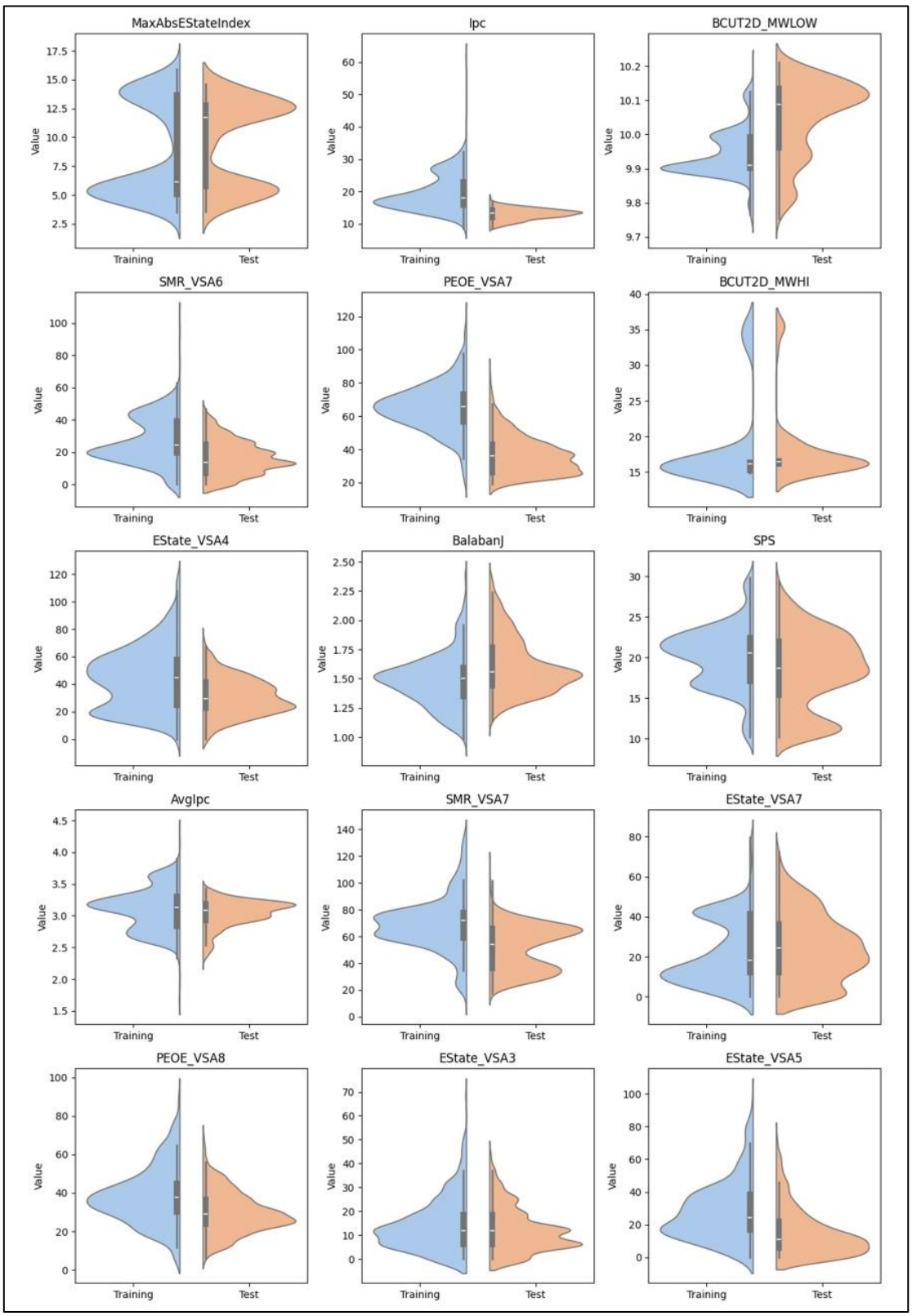


Figure 3.9: Feature distribution of training and test data. Outliers in test data that were outside the feature range of training data were removed. Subsequently, violin plots were generated for each feature to compare the distribution of the top 15 features (ranked according to RFE) in training and test data. Each subplot contains the individual feature, with the value of the feature mentioned in the y-axis and the data type (training or test) mentioned in the x-axis. The Ipc feature was transformed using a logarithmic scale to improve visibility. The feature distribution shows that the range of features in test data is within that of training data, ensuring the integrity of the predicted models.

3.10. Validation: Molecular Docking and Applicability Domain

After the LGBM Classifier was determined as the best model for our training data through Lazy Predict, test data was evaluated for their activity against CXCR4. 20 compounds, mentioned in Table 3.6, were predicted to possess inhibitory activity against CXCR4. Mavorixafor and 20 compounds classified as CXCR4 inhibitors, by LGBM Classifier, after machine learning were docked to evaluate their binding affinity with CXCR4 using Vina-GPU+. It is one of the three docking methods of Vina-GPU 2.0 and allows accelerated docking using a GPU and batch processing of multiple ligands with a single receptor molecule. Our reference drug (Mavorixafor) had a binding score of -7.4 kcal/mol. 19 out of 20 compounds exhibited better docking scores than Mavorixafor, the lowest being -7.5 kcal/mol and the highest being -11.1 kcal/mol. One compound had a slightly lower binding score (-7.3 kcal/mol) than Mavorixafor. The next step was to analyze these compounds for their applicability to the trained model. Mahalanobis distance (MD) was used as a metric for this purpose, as it allows us to check whether a compound falls in the descriptor space of training data. The applicability domain was defined by the calculation of the average (d) and standard deviation (std) of MD of all compounds in the training data. These values were calculated utilizing SciPy and Scikit-learn Python libraries.

The threshold (t) to determine the border of the applicability domain was set as:

$$threshold\ (t) = d + st \times z$$

For our training data, values of d, std and t were determined to be 3.63, 1.34 and 4.30, respectively. 7 compounds with MD less than 4.30 were, therefore, observed to fall in the applicability domain of our training data. However, one compound (CNP0399717) had a slightly lower binding score as mentioned above. All 20 compounds were visualized for their MD and their presence inside or outside of the applicability domain using the pyplot interface of matplotlib (Hunter, 2007), as shown in Fig 3.10 (A). Compounds with green colour refer to those that are inside the applicability domain. While compounds with red colour refer to those that are outside the applicability domain.

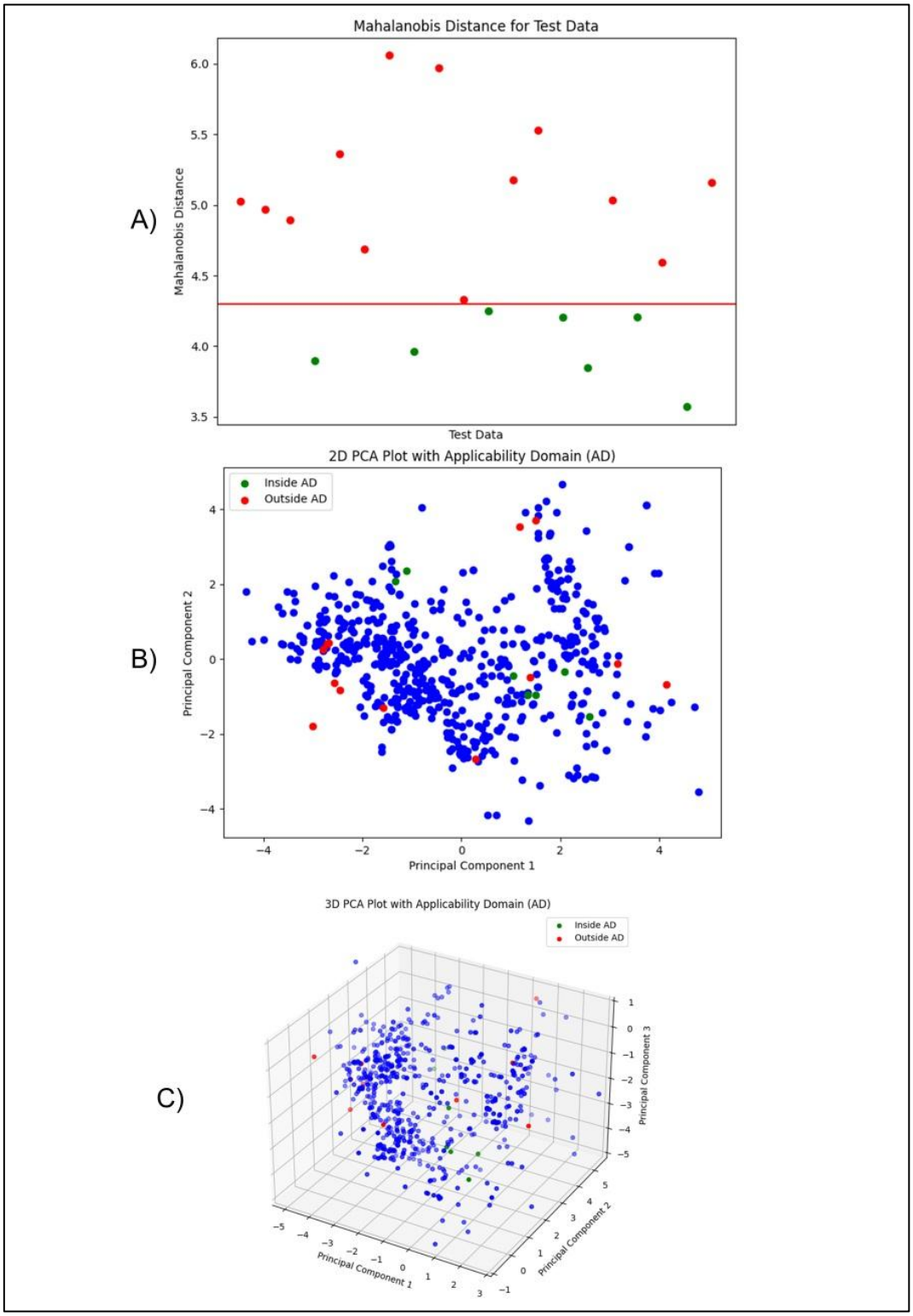


Figure 3.10: Applicability Domain and PCA Analysis. After Lazy Predict determined the top machine learning algorithms (LGBM Classifier), compounds in the test dataset were evaluated for their activity against CXCR4. After the prediction of 20 compounds as inhibitors, these compounds were tested if they fell in the applicability domain. Mahalanobis distance was used as a metric to analyze the presence of our selected test compounds falling in the descriptor space of training data. To define the applicability domain, we calculated the average (d) and the standard deviation (std) of the Mahalanobis distance. The threshold was set as d + std x z, where z is the arbitrary parameter (default value is 0.5). Using this approach 7 compounds were observed to fall in the applicability domain. Compounds with green colour refer to those that are inside the applicability domain. While compounds with red colour refer to those that are outside the applicability domain.

Since we are dealing with 15 descriptors for our study, we used Principal component analysis (PCA) to reduce the dimensionality of our data while minimizing information loss. This is done by the creation of new uncorrelated variables, known as the principal components, defined according to the given data, which makes it an adaptive data analysis technique (Jollife & Cadima, 2016). Depending on the number of principal components being selected while visualization, we can generate 2D PCA with two principal components or 3D PCA with three principal components. 2D PCA and 3 PCA plots were visualized while using the pyplot interface of matplotlib (Hunter, 2007) and predicted test compounds were observed in the PCA space of training data along with their status for applicability domain in Fig. 3.10 (A) and 3.10 (B), respectively. After accessing the compounds for AD, their interaction with different residues of CXCR4 was analysed. CXCR4 protein contains seven transmembrane (TM) helices, TM1 to TM7 that are arranged in a barrel-like structure with the inside of the barrel being more hydrophilic and the outside being more hydrophobic (Tegler et al., 2020). Besides TM helices, CXCR4 contains an extracellular N-terminal domain, three extra-cellular loops (ECL), three intracellular loops (ICL) and an intracellular C-terminal domain (Bianchi & Mezzapelle, 2020).

Important interacting residues of CXCR4 are Glu32 in N-terminal domain; Phe36/Asn37/Leu41/Tyr45 in TM1; Trp94/Asp97/Ala98 in TM2; Val112/His113/Tyr116 in TM3; His281/Ile284/Ser285/Glu288 in TM7; Trp102 in ECL1; and Cys186/Arg188 in ECL2 (Das et al., 2015). Most of these residues can be seen as interacting residues of our selected compounds (Fig. 3.11). TRP94, VAL112 and HIS113 were observed to be the most prominent interaction residues across all the molecules. Thus, our results are in line with existing literature about the interaction of small molecules with CXCR4. CNP0015964 (benzimidazole derivative), CNP0013977 (pyridinylpiperazine derivative), and CNP0015625 (aminopyridine derivative) were the top 3 compounds with the highest docking scores that were within the applicability domain of our trained model. All of the most prominent interacting residues, TRP94, VAL112 and HIS113, are observed in their interactions with CXCR4 (Fig. 3.11). BIOVIA Discovery Studio was used to visualise these interactions. Furthermore, they had 44.59%, 24.32% and 21.62% higher docking scores, respectively, as compared to Mavorixafor. Thus, these compounds can potentially be more effective inhibitors than Mavorixafor and may be used for the treatment of neurodegenerative disorders.

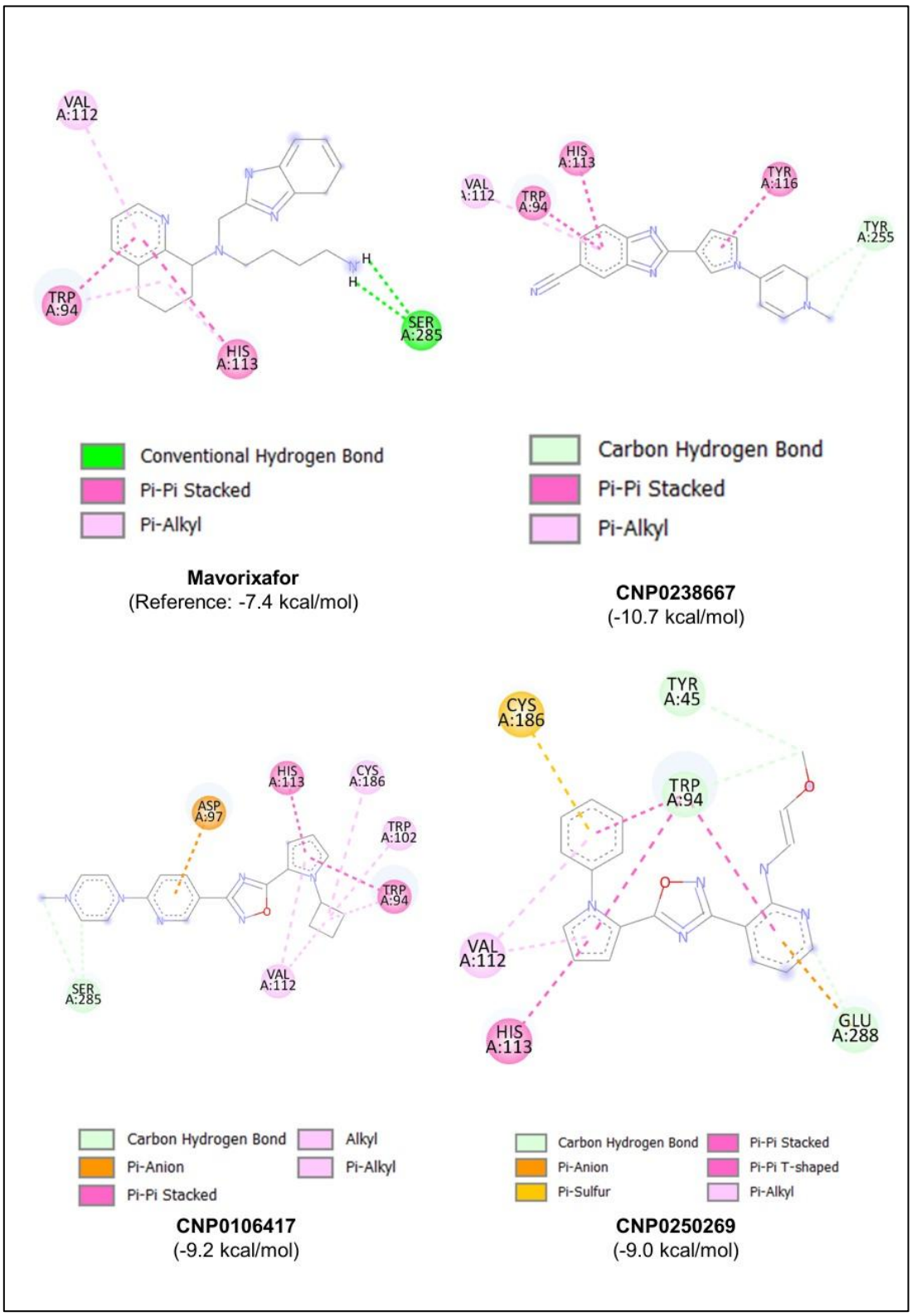


Figure 3.11: Receptor-Ligand Interactions of Mavorixafor and top 3 compounds predicted by machine learning. Different types of binding interactions of the top 3 compounds and Mavorixafor with CXCR4 have been shown. Important interacting residues of CXCR4 are Glu32 in N-terminal domain; Phe36/Asn37/Leu41/Tyr45 in TM1; Trp94/Asp97/Ala98 in TM2; Val112/His113/Tyr116 in TM3; His281/Ile284/Ser285/Glu288 in TM7; Trp102 in ECL1; and Cys186/Arg188 in ECL2. Certain residues, like TRP94, VAL112 and HIS113, have been observed to be binding residues in all three selected molecules and Mavorixafor. These results about interacting residues complement existing literature. BIOVIA Discovery Studio was used to visualise these interactions.

Table 3.6: List of 20 natural compounds predicted as CXCR4 inhibitors by LGBM Classifier with their docking scores and their presence in Applicability Domain (AD):

S.No.	ID	Mahalanobis Distance	Within AD	Docking Score (kcal/mol)
1	CNP0004424	4.68	No	-11.1
2	**CNP0015964**	**3.89**	**Yes**	**-10.7**
3	CNP0319279	5.97	No	-10
4	CNP0094414	5.02	No	-9.8
5	CNP0165909	4.33	No	-9.7
6	**CNP0013977**	**3.57**	**Yes**	**-9.2**
7	CNP0450948	5.16	No	-9.1
8	**CNP0015625**	**4.20**	**Yes**	**-9**
9	CNP0014385	4.89	No	-8.8
10	**CNP0127043**	**4.25**	**Yes**	**-8.7**
11	CNP0015957	5.17	No	-8.5
12	**CNP0314177**	**4.20**	**Yes**	**-8.4**
13	CNP0015859	5.36	No	-8.3
14	CNP0408379	5.03	No	-8
15	CNP0360705	6.06	No	-7.8
16	CNP0397928	5.53	No	-7.8
17	CNP0122007	4.59	No	-7.7
18	CNP0216455	4.97	No	-7.6
19	**CNP0015823**	**3.85**	**Yes**	**-7.5**
20	**CNP0399717**	**3.96**	**Yes**	**-7.3**

3.11. Discussion

AD pathophysiology is characterized by abnormal Aβ aggregation and tau protein hyperphosphorylation (Rajmohan & Reddy, 2017), while PD pathophysiology is characterized by abnormal α-Syn (Gómez-Benito et al., 2020). Inhibition of accumulation or promoting clearance Aβ and α-Syn are potential therapeutic targets for treatment of AD and PD, respectively (Fields, Bengoa-Vergniory, & Wade-Martins, 2019; Nalivaeva & Turner, 2019). CXCR4 is a receptor present on the surface of B lymphocytes, which secretes antibodies that target toxic proteins in AD and PD (Sim, Im, & Park, 2020). Downregulation of CXCR4 is essential for the migration of immature B-cells into the peripheral blood from the bone marrow (Beck, Gomes, Cyster, & Pereira, 2014). CXCR4 upregulation in AD patients results in

decreased migration of B-cells and decline in anti-Aβ antibody secretion, leading to Aβ aggregation (Q.-L. Wang, Fang, Huang, & Xue, 2022). Microarray analysis has revealed upregulation of CXCR4 in the inferior parietal lobe (IPL) brain sections of AD patients, which was also confirmed by real-time PCR and immunohistochemistry (Weeraratna et al., 2007). AMD3100, a CXCR4 inhibitor, has been shown to improve cognitive performance, reduce neuroinflammation, and alleviate AD pathophysiology (Gavriel, Rabinovich-Nikitin, Ezra, Barbiro, & Solomon, 2020). However, normal levels of CXCR4 expression is required for cognitive performance and normal development of hippocampal dentate gyrus (M. Lu, Grove, & Miller, 2002). Chronic administration of AMD3100 has been shown to result in learning and memory dysfunction in young mice (Parachikova & Cotman, 2007). Thus, crux of the treatment lies in maintaining CXCR4 levels. In another study, MPTP-mediated CXCR4 upregulation preceded the loss of DA neurons, suggesting role of CXCR4 in the PD etiology and its potential as a new target molecule for PD treatment (Shimoji, Pagan, Healton, & Mocchetti, 2009). AMD3100 also increased life span and improved motor function in the $SOD1^{G93A}$ mice model of ALS (Rabinovich-Nikitin, Ezra, Barbiro, Rabinovich-Toidman, & Solomon, 2016). Likewise, Mason et al., 2021 demonstrated that AMD3100 pre-treated animals displayed reduced AMPH-induced locomotor activity in comparison to SAL pre-treated animals. Postmortem analyses of brain tissue revealed elevated CXCR4 protein levels in the striatum of all experimental groups. The results implicated that CXCR4 inhibition with AMD3100 attenuates amphetamine induced locomotor activity (Mason et al., 2021). Furthermore, CXCR4 knockout prevents degeneration of dopamine neurons of MPTP-lesioned mice through microglial and astroglial activation. CXCR4 knockout also protected blood-brain barrier (BBB) from MPTP-induced damage (Ma et al., 2023). Therefore, downregulation of CXCR4 can be a promising approach for the treatment for AD and PD.

However, drug discovery is an expensive and time-consuming process with the average approved drug requiring 10 to 15 years to develop with an estimated cost of 0.8–2 billion USD. Various FDA-approved drugs, such as aliskiren, captopril, dorzolamide, oseltamivir, and nolatrexed, were all optimized using CADD (Talele, Khedkar, & Rigby, 2010), and a large number of publications describe the successful design and discovery of leads/drugs using CADD (W. Lu, Zhang, Jiang, Zhang, & Luo, 2018). The drugs that were optimized or designed (Espinoza-Moraga, Caballero, Gaube, Winckler, & Santos, 2012; Fjelldal et al., 2019; Ha, Fatima, & Gaurav, 2015; S. Kumar, Chowdhury, & Kumar, 2017; Mishra et al., 2017; Popugaeva et al., 2019; Remya, Dileep, Tintu, Variyar, & Sadasivan, 2013; Samadi et al., 2012;

Tadayon & Garkani-Nejad, 2019; Thomas & Grossberg, 2009) (Vancraenenbroeck et al., 2014; Y. Wang, Lv, Jin, & Liang, 2020) for AD or PD and later confirmed through *in vitro* or *in vivo* studies (De Andrade Teles et al., 2018; Du et al., 2016; "Memantine for Treatment of Cognitive Impairment in Patients With Parkinson's Disease and Dementia - Full Text View - ClinicalTrials.gov," n.d.; K. K. Roy et al., 2012; Samadi et al., 2013; Varadaraju et al., 2013; Venkata et al., 2017; Wei et al., 2013; West, 2017) have been listed in Table 3.7. The main goal of CADD is to reduce these timescales and costs without affecting quality (Kapetanovic, 2008). Importantly, CADD can be used in most stages of drug development: from target identification to target validation, from lead discovery to optimization, and in preclinical studies. Thus, It is estimated that CADD could lower the drug development cost by up to 50% (Macalino, Gosu, Hong, & Choi, 2015; M. Xiang, Cao, Fan, Chen, & Mo, 2012). Therefore, our study aims to find potential leads for the treatment of AD and PD among the large number of small molecule compounds available in online databases.

We search for molecules similar to Mavorixafor, a known CXCR4 inhibitor, in Zinc Database and found 52 such compounds. Due to similarity in structure, these compounds may have similar bioactivity, which we later validated using molecular docking and molecular dynamics simulation. Through molecular docking we found 5 molecules that had better binding affinity than Mavorixafor. The prediction of the fate of a drug and the effects caused by a drug inside the body, such as how much drug is absorbed if administered orally and how much is absorbed in the gastrointestinal tract, is an indispensable part of drug discovery. In a similar way, if the absorption is poor, its distribution and metabolism would be affected, which can lead to causing neurotoxicity and nephrotoxicity. Ultimately, the study is to understand the disposition of a drug molecule within an organism. Thus, ADME study is one of the most essential parts of computational drug design. Therefore, we validated the molecules through their ADME profile and found 4 out of 5 of these molecules were suitable as lead compounds, only exception being ZINC49069264, which as not blood-brain barrier permeable. Then we employed molecular dynamics simulation to validate stability of protein-ligand complexes. We found two molecules, ZINC49067615 and ZINC103242147, that were stable in their interaction with CXCR4 and had stronger binding energies than Mavorixafor. Certain residues such as TYR45, TRP94, ASP97, etc. were found to be involved in binding between these molecules and CXCR4 through molecules docking as well as molecular dynamics simulation. This is in line with the residues that have been reported to be important for stability of

interactions with known CXCR4 inhibitors (Neves, Simões, & Sá e Melo, 2010; Vinader, Ahmet, Ahmed, Patterson, & Afarinkia, 2013), further confirming our results.

Table 3.7. Neurodegenerative diseases with their diagnostic targets and respective drugs that have been identified using Computer-aided drug discovery. The experimental validation, either in vitro or in vivo, performed for each drug is also indicated.

Disease	Targets	Molecules	Software	In Vitro or In Vivo Study
Alzheimer's Disease	N-methyl-D-aspartate (NMDA) Receptor	1-Benzyl-1,2,3,4-Tetrahydro-β-Carboline	ICM	Cell-based assay
	Acetylcholinesterase (AChE) and Butyrylcholinesterase (BuChE)	6-Chloro-pyridonepezils	Autodock Vina, QikProp	*In vitro* blood–brain barrier (BBB) model
	Nucleoside hydrolase	Flavonoids	AutoDock	Mice and rat models
	N-methyl-D-aspartate (NMDA) Receptor	Ifenprodil	Schrödinger Suite	Chicken embryo forebrain cultures (E10)
	AChE, BuChE, BACE 1, MAO and NMDA	Memantine	Glide	Clinical Trial
	AChE	Morin	Glide	APPswe/PS1dE9 mice
	ABCG2 enzyme	2,4-disubstituted pyridopyrimidine derivatives	Autogrid, Autodock, and GROMACS	In vitro enzyme inhibitory model
	AChE	Pyridonepezil	Autodock Vina	In vitro blood–brain barrier model
	Transient receptor potential canonical 6 (TRPC6)	Piperazine derivatives	PASS software	Ellman's method
	Human islet amyloid polypeptide (hIAPP)	Rutin	AutoDock and AutoDock Vina	Neuroblastoma cells (IMR32) and Wistar rats
Parkinson's Disease	Leucine-rich repeat kinase 2 (LRRK2)	9-methyl-N-phenylpurine-2,8-diamine, N-phenylquinazolin-4-amine, and 1,3-dihydroindol-2-one)	MOE	In vitro and in vivo studies

Several studies have reported that CXCR4, a chemokine receptor predominantly expressed on B-cell surface, has diverse regulatory functions within the immune system, cell proliferation, and neurodevelopment (Bianchi & Mezzapelle, 2020). CXCR4 downregulation is a critical event that facilitates the migration of immature B cells from the bone marrow into the bloodstream (Beck et al., 2014). However, in neurodegenerative disorders, an aberrant upregulation of CXCR4 has been implicated in a cascade of events leading to reduced B cell migration, decreased antibody secretion, and the consequential aggregation of toxic proteins, such as Aβ in AD and α-Syn in PD (Sierks et al., 2011; Q.-L. Wang et al., 2022). In the absence of effective treatment options, natural compounds have recently gained significant attention due to their inherent biocompatibility, low toxicity and neuroprotective properties, making

them attractive candidates for therapeutic alternatives to synthetic chemical compounds (Jamshidi, Rostami, Shojaei, Taherkhani, & Taherkhani, 2024; Mohd Sairazi & Sirajudeen, 2020). Notably, among their diverse protective effects, natural compounds can counter excitotoxicity, reduce neuroinflammation, alleviate oxidative stress, and mitigate proteinopathies (Bagli, Goussia, Moschos, Agnantis, & Kitsos, 2016).

The main objective of this study was to classify natural compounds as inhibitors (IC50 ≤ 100 nM) or non-inhibitors (IC50 > 100 nM) based on their activity against CXCR4 so that they can be ultimately used as a potential therapeutic option for neurodegenerative disorders, particularly AD and PD. To create a test dataset, we selected 10,000 natural compounds from the COCONUT Database, ensuring that they were structurally similar (Tanimoto Coefficient > 65%) to Mavorixafor. We then screened these compounds for their ability to cross the blood-brain barrier (BBB), an important factor for drug efficacy in central nervous system (CNS) disorders (Wu et al., 2023). Only 2712 compounds that passed the initial screening by DeePred-BBB, LightBBB, and SwissADME were selected for further analysis. We also assessed the drug-likeness and medicinal chemistry friendliness of these compounds to eliminate those with suboptimal absorption, distribution, metabolism, and excretion (ADME) properties, which can hinder drug development (A. P. Li, 2001). This resulted in the identification of 1675 compounds that exhibited zero violations for Lipinski, Ghose, Veber, Egan, and Muegge filters, as well as zero Pan Assay Interference Compounds (PAINS) and Brenk alerts. These compounds were considered suitable for further investigation. Finally, we calculated the molecular descriptors of these compounds using RDKit software to create a test dataset for machine learning.

Simultaneously, a robust training dataset was formulated using molecules with documented activity against CXCR4 from the BindingDB. Lazy Predict facilitated the exploration of various classification algorithms, ultimately leading to the selection of the LGBM Classifier, XGB Classifier and Random Forest Classifier, all of which demonstrated superior accuracy against the training set. LGBM Classifier offering the best model accuracy after hyperparameter tuning was ultimately used for predicting the activity of test compounds and identified 20 compounds as potential inhibitors (Table 3.6). Then, we delved into a detailed assessment of the binding affinity of these 20 compounds employing Vina-GPU 2.0 and docked these compounds with CXCR4. Remarkably, 19 out of 20 compounds exhibited superior

binding scores compared to Mavorixafor. The last compound had only a slightly inferior binding score compared to Mavorixafor. Finally, these compounds were tested for their applicability to our training model. For this purpose, Mahalanobis distance, a distance-based metric employed to define the applicability domain (Sahigara et al., 2012), was used to analyze the presence of predicted test compounds in the feature space of our training data. Furthermore, test compounds outside the feature range of training data were already eliminated before activity prediction, reducing the likelihood of predicted test compounds falling outside the applicability domain of training data. Therefore, 7 compounds were observed to fulfil the applicability domain criteria and prediction for these compounds by our machine learning model can be suggested to be reliable.

Three compounds, CNP0015964 (benzimidazole derivative), CNP0013977 (pyridinylpiperazine derivative), and CNP0015625 (aminopyridine derivative), which emerged as the top three CXCR4 inhibitors were within the applicability domain of the model used for training and displayed higher docking scores than Mavorixafor by 44.59%, 24.32% and 21.62%, respectively. Many benzimidazole, pyridinylpiperazine and aminopyridine derivatives have been reported for their neuroprotective properties (Imran et al., 2021; Kikuoka et al., 2020; Strupp et al., 2017), suggesting the potential of these compounds as therapeutic options for neurodegenerative disorders. Analysis of the binding interactions between ligands and receptors revealed specific residues TRP94, VAL112, and HIS113 as potential determinants for CXCR4 inhibition, which have been reported in previous studies (Hung, Lee, Chen, & Chen, 2014; Senthil Kumar, Kishore, Elumalai, & Gupta, 2023; Tripathi & Kumar, 2023). Additionally, it is worth mentioning that these residues were consistently observed as binding residues in all three selected molecules as well as Mavorixafor. This observation strengthens the idea that CNP0015964, CNP0013977, and CNP0015625 have therapeutic value as CXCR4 inhibitors and could be potentially effective in the management of neurodegenerative disorders. The significance of our findings extends beyond the identification of potential therapeutic molecules as it not only unravels the structural motifs that influence CXCR4 inhibition but also establishes a robust computational framework that seamlessly integrates cheminformatics and machine learning to expedite the discovery of CXCR4 inhibitors from natural compound libraries. The cost-effectiveness and efficiency associated with this strategy position it as a compelling avenue for drug development, particularly in the context of inflammatory and immune-related disorders.

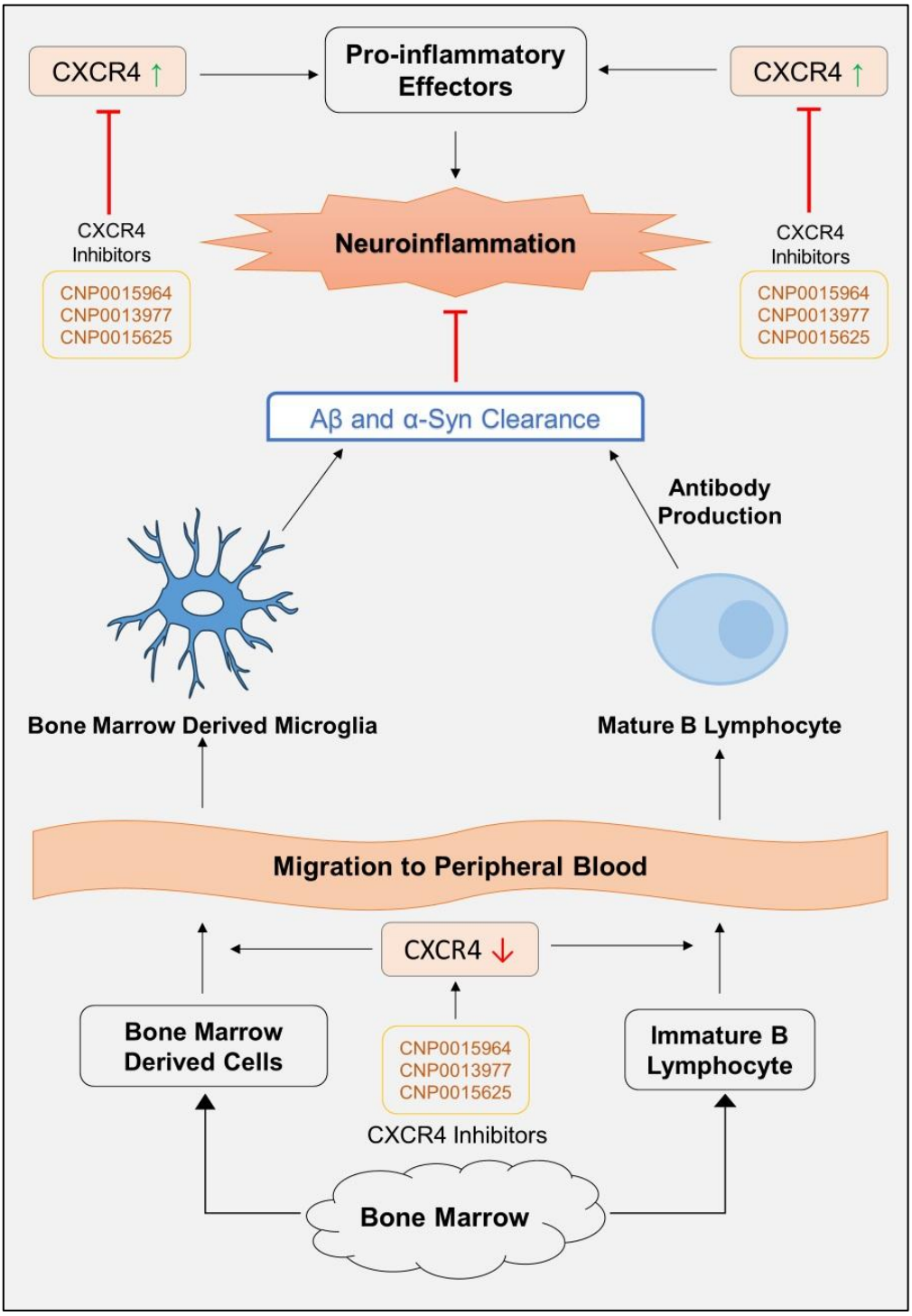


Figure 3.12: Fall in CXCR4 expression is vital for the migration of immature B lymphocytes and bone marrow-derived cells from the bone marrow into the peripheral blood. Mature B lymphocytes secrete antibodies that cross the blood-brain barrier and target toxic proteins, like Aβ and α-Syn. Inhibition of CXCR4 expression through natural compounds may be helpful for Aβ and α-Syn clearance, thereby helping in treating neurodegenerative disorders caused by aggregation of such toxic proteins, namely AD and PD. Similarly, an increase in CXCR4 expression activates effectors (such as Akt, ERK, c-Jun, NF-kB, p38, mTOR, etc.), as it is a pro-inflammatory chemokine, and results in neuroinflammation, which is an important characteristic of the pathophysiology of neurodegenerative disorders, such as AD and PD. Thus, CXCR4 inhibitors may help in regulating CXCR4 expression to normal range, thereby providing protection against neuroinflammation and cell death.

CHAPTER 4

SUMMARY, CONCLUSION

AND

FUTURE SCOPE

CHAPTER 4

SUMMARY, CONCLUSION AND FUTURE SCOPE

4.1. Summary

In this study, we explore the common molecular signatures of Alzheimer's and Parkinson's disease, outlined shared transcriptional changes in both, and catalogued specific ways in which gene expression results implicate mitochondrial pathways, synaptic degeneration, or inflammatory pathways in both diseases. These signals also highlight functioning mechanisms across these neurodegenerative diseases. We took an innovative, and computational drug-repurposing approach to screen the compounds from the COCONUT library to discover potential drug-treatment targets for the chemokine receptor CXCR4, a compound that is shown to play an important role in both neuroinflammation, and neurodegeneration. This was comprised of several steps: virtual screening, molecular docking, and pharmacokinetic profiling to select compounds that will work effectively and develop into good drugs. We identified several different novel compounds with some degree of binding affinity to CXCR4, and relevant levels of stability and pharmacokinetic properties. In silico test methods were utilized to screen, predict, and confer efficacy of discovered compounds. Using molecular dynamics approaches, they were evaluated based on the length of time the compounds will adhere with some degree of stability. In determining the interaction between the novel drug discovered, and previous knowledge of interaction sites, there are key discovered residues in the binding pocket of the CXCR4 chemokine receptor, which provide potential for future research and gives insight to rational drug design towards CXCR4. Structural dynamics between the compound and CXCR4 receptor at the binding pocket facilitate a highly selective bond with key residues for inhibition of receptor CXCR4 actions in cellular chemistry. Specifically mapping the chemokine receptor protein and residues that are capable for binding, we can begin to understand the design space optimally binding to CXCR4. Alongside the study of synthetic compounds, we expanded our investigation to natural compounds as potential therapeutic agents for Alzheimer's and Parkinson's diseases. Several natural compounds showed high levels of inhibition against CXCR4 and some of those compounds had the potential via their pharmacokinetic properties. These compounds are also derived from plants and are preferable because they are likely more natural and also safer than synthetic drugs. Due to their ability to cross the blood-brain barrier and their drug-likeness we focused our further investigation on testing these natural compounds. To confirm that selected compounds could be usable as therapeutic agents, we conducted experiments to establish the ability of these compounds to cross the blood-brain barrier—a crucial characteristic for drug development

when the target is neurodegenerative diseases. The ability to cross the blood-brain barrier was not the only important attribute verified, as we also established their drug-likeness and that they were friendly to medicinal chemistry—two characteristics indicative of properties often required for drug development. Only compounds displaying the most desirable characteristics of both drug-likeness and capability of crossing the blood-brain barrier were considered for additional studies. The potential of the selected compounds to function as inhibiting agents was independently tested via prediction algorithms utilizing machine learning. The LightGBM Classifier, which was trained based on known biological activity against CXCR4 from an archived dataset of compounds, was used as a predictive algorithm for success in novel compound comparisons. The LightGBM Classifier produced positive results with respect to accurate prediction of all the compounds and represented the strongest predictive capacity of the data models. This technique emphasizes the use of computational methods with experimental results when determining alternative therapies for disease. The interaction of the selected compounds with CXCR4 and supporting residues in the protein-ligand complex has been described in detail. These interactions form the foundation for compound activity as follows. With regard to an experimental perspective the authors further detail ways of rationalizing the compounds on a molecular basis in order to leverage future compounds that are more selective and better derived from data resourced biological remedies. To verify the predictions of the compounds were of relevance and reliable, we performed a rigorous assessment using the Mahalanobis distance for the applicability to the trained model, establishing it was in the chemical space of known CXCR4 inhibitors. The Mahalanobis distance metric provides more support for our predictions and the potential therapeutic value of the selected compounds.

Our research has highlighted common transcriptional signatures in the context of neurodegenerative disorders, specifically Alzheimer's and Parkinson's, suggesting overlapping molecular mechanisms that might be manipulated therapeutically. From our computational analysis, we have found novel and natural compounds that bested others with good, strong inhibition against CXCR4, good pharmacokinetic properties, and able to cross the blood-brain barrier. Machine diagnostic models are further suggestion the potential efficacy of the compounds. The interaction insights detailed in the present study and prompting the applicability to underline the potential these compounds can act as therapeutic approaches in neurodegenerative disorders. Our results open pathways for future research and development in effective therapies for Alzheimer and Parkinson's disease. This is promising news for people who suffer with neurodegenerative diseases.

4.2. Conclusion

The molecular docking and simulation were used to screen the molecules showing structural similarity with Mavorixafor (CXCR4 inhibitor) to find novel molecules potentially regulating the CXCL12/CXCR4 pathway. After the virtual screening of the molecules using molecular docking, five test compounds exhibiting high docking scores were selected for further analysis. After ADME analysis, ZINC49069264 was ruled out as it was observed to be unable to pass through the blood-brain barrier. All other molecules were selected for further research. After MD simulation, ZINC49067615 and ZINC103242147 were stable in docked configuration with CXCR4 and exhibited high interaction energy. The ADME profiles of these compounds were also suitable for being used a lead compounds, passing all five, Lipinski, Ghose, Veber, Egan, and Muegge, descriptors of drug-likeness with no violations. Thus, this preliminary study revealed that these molecules can be potential therapeutic agents and be further tested in biological studies, either in vitro or in vivo, to analyze their potential in the modulation of the CXCL12/CXCR4 pathway to treat AD and PD (Fig. 3.12). However, the study accompanies the limitations, such as *in vitro and in vivo* validation of the identified compounds in respect to AD and PD pathology.

Artificial intelligence (AI) and machine learning (ML) algorithms have made significant advances over the past decade. These computational tools, along with the widespread availability of biological data, have become indispensable for drug design and discovery. They not only expedite the drug development process but also offer a cost-effective way of identifying more potent treatment options for various diseases, including neurodegenerative disorders. In this study, we utilized Lazy Predict Python libraries to semi-automate machine learning to classify natural compounds as either inhibitors or non-inhibitors of CXCR4. Our methodology integrated machine learning with structural similarity, ADME (Absorption, Distribution, Metabolism, and Excretion), and molecular docking, leading to the identification of optimal candidates for lead discovery. This versatile approach can be applied to other drug targets beyond CXCR4. By utilizing existing data to generate machine learning models, our methodology accelerates the discovery of potential lead compounds for various targets. This not only improves the efficiency of the drug discovery pipeline but also offers a robust framework that can be used in conjunction with traditional experimental approaches. As the landscape of drug development continues to evolve, the amalgamation of computational and experimental techniques stands poised to redefine the efficiency and success of therapeutic interventions.

4.3. Future Scope

- Future studies may examine other molecular targets in addition to CXCR4 that are recognized to have physiological roles in both Alzheimer's and Parkinson's diseases. This consideration could advance more robust treatment strategies that specifically target a number of physiological roles in these complex diseases and disorders.

- Additionally, investigational potential benefit of CKX4 inhibitors combined with methods of traditional treatment using an increasingly good success rates will additively benefit to treatment options. Combined additive benefit will reduce the incidence and/or disease progression more effectively than other therapies.

- A crucial consideration in these future directions should involve extensive clinical trials to safely and efficaciously establish the identified compounds for use in humans, which will provide greater confidence in the possibility of use within the context of developing effective treatments for neurodegenerative disorders.

- An important consideration in these future areas of research would be to develop personalized treatment modalities based on genetic and physical characteristics, thinking differently about each individual patient will help in developing a treatment with greater response toward positive outcomes.

- Lastly, advanced computational modelling and machine learning algorithms will be increasingly helpful in predicting the efficacy and optimizing design for compound used in therapeutic treatments and optimizing accuracy within the compound development.

- Relatedly, expanded use of research in support of discovering natural compounds for therapeutic efficacy through the investigation of different medicinal natural sources, including plants and traditional medicines, will aid in the discovery of neuroprotective agents. Natural compounds provide advantageous properties that are not always present within synthetic candidates.

- It will be particularly important to investigate and study new routes of drug delivery through the blood-brain barrier. Drug delivery to the brain is essential in the treatment of brain disorders and overcoming the blood-brain barrier will be of utmost importance in increased efficacy for interventions combatting neurodegeneration, developing drugs and control methods with improved efficacy.

REFERENCES

Abdipranoto-Cowley, A., Jin, S. P., Croucher, D., Daniel, J., Henshall, S., Galbraith, S., … Vissel, B. (2009). Activin A Is Essential for Neurogenesis Following Neurodegeneration. *Stem Cells*, *27*(6), 1330–1346. https://doi.org/10.1002/STEM.80

Abouchekeir, S., Vu, A., Mukaidaisi, M., Grantham, K., Tchagang, A., & Li, Y. (2022). Adversarial deep evolutionary learning for drug design. *Biosystems*, *222*, 104790. https://doi.org/https://doi.org/10.1016/j.biosystems.2022.104790

Ain, Q. U., Méndez-Lucio, O., Ciriano, I. C., Malliavin, T., Van Westen, G. J. P., & Bender, A. (2014). Modelling ligand selectivity of serine proteases using integrative proteochemometric approaches improves model performance and allows the multi-target dependent interpretation of features. *Integrative Biology*, *6*(11), 1023–1033. https://doi.org/10.1039/C4IB00175C

Alam, S., & Khan, F. (2019). 3D-QSAR, Docking, ADME/Tox studies on Flavone analogs reveal anticancer activity through Tankyrase inhibition. *Scientific Reports*, *9*(1), 5414. https://doi.org/10.1038/s41598-019-41984-7

Ali, J., Camilleri, P., Brown, M. B., Hutt, A. J., & Kirton, S. B. (2012). Revisiting the general solubility equation: in silico prediction of aqueous solubility incorporating the effect of topographical polar surface area. *Journal of Chemical Information and Modeling*, *52*(2), 420–428. https://doi.org/10.1021/CI200387C

Allgaier, J., & Pryss, R. (2024). Cross-Validation Visualized: A Narrative Guide to Advanced Methods. *Machine Learning and Knowledge Extraction*. https://doi.org/10.3390/make6020065

Anderson, A. C. (2012). Structure-based functional design of drugs: From target to lead compound. *Methods in Molecular Biology*, *823*, 359–366. https://doi.org/10.1007/978-1-60327-216-2_23/COVER

Andrade, S., Nunes, D., Dabur, M., Ramalho, M. J., Pereira, M. C., & Loureiro, J. A. (2023). Therapeutic Potential of Natural Compounds in Neurodegenerative Diseases: Insights from Clinical Trials. *Pharmaceutics*, *15*(1). https://doi.org/10.3390/PHARMACEUTICS15010212/S1

Arnott, J. A., & Planey, S. L. (2012). The influence of lipophilicity in drug discovery and design. *Expert Opinion on Drug Discovery*, *7*(10), 863–875. https://doi.org/10.1517/17460441.2012.714363

Avti, P. K., Singh, J., Dahiya, D., & Khanduja, K. L. (2023). Dual functionality of pyrimidine

and flavone in targeting genomic variants of EGFR and ER receptors to influence the differential survival rates in breast cancer patients. *Integrative Biology*, *15*. https://doi.org/10.1093/INTBIO/ZYAD014

Baell, J. B., & Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry*, *53*(7), 2719–2740. https://doi.org/10.1021/JM901137J

Bagheri, V., Khorramdelazad, H., Hassanshahi, G., Moghadam-Ahmadi, A., & Vakilian, A. (2019). CXCL12 and CXCR4 in the Peripheral Blood of Patients with Parkinson's Disease. *Neuroimmunomodulation*, *25*(4), 201–205. https://doi.org/10.1159/000494435

Bagli, E., Goussia, A., Moschos, M. M., Agnantis, N., & Kitsos, G. (2016). Natural Compounds and Neuroprotection: Mechanisms of Action and Novel Delivery Systems. *In Vivo*, *30*(5), 535 LP – 547. Retrieved from http://iv.iiarjournals.org/content/30/5/535.abstract

Baig, M. H., Ahmad, K., Rabbani, G., Danishuddin, M., & Choi, I. (2018). Computer Aided Drug Design and its Application to the Development of Potential Drugs for Neurodegenerative Disorders. *Current Neuropharmacology*, *16*(6), 740. https://doi.org/10.2174/1570159X15666171016163510

Barrett, T., Suzek, T. O., Troup, D. B., Wilhite, S. E., Ngau, W. C., Ledoux, P., … Edgar, R. (2005). NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Research*, *33*(suppl_1), D562–D566. https://doi.org/10.1093/NAR/GKI022

Beck, T. C., Gomes, A. C., Cyster, J. G., & Pereira, J. P. (2014). CXCR4 and a cell-extrinsic mechanism control immature B lymphocyte egress from bone marrow. *Journal of Experimental Medicine*, *211*(13), 2567–2581. https://doi.org/10.1084/JEM.20140457/VIDEO-6

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/NAR/28.1.235

Bezzi, P., Domercq, M., Brambilla, L., Galli, R., Schols, D., De Clercq, E., … Volterra, A. (2001). CXCR4-activated astrocyte glutamate release via TNFα: amplification by microglia triggers neurotoxicity. *Nature Neuroscience 2001 4:7*, *4*(7), 702–710. https://doi.org/10.1038/89490

Bianchi, M. E., & Mezzapelle, R. (2020). The Chemokine Receptor CXCR4 in Cell

Proliferation and Tissue Regeneration. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/FIMMU.2020.02109/PDF

Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., & Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. *Nature Chemistry*, *4*(2), 90–98. https://doi.org/10.1038/NCHEM.1243

Bonham, L. W., Karch, C. M., Fan, C. C., Tan, C., Geier, E. G., Wang, Y., … Singleton, A. B. (2018). CXCR4 involvement in neurodegenerative diseases. *Translational Psychiatry 2017 8:1*, *8*(1), 1–10. https://doi.org/10.1038/s41398-017-0049-7

Brenk, R., Schipani, A., James, D., Krasowski, A., Gilbert, I. H., Frearson, J., & Wyatt, P. G. (2008). Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem*, *3*(3), 435–444. https://doi.org/10.1002/CMDC.200700139

Britton, C., Poznansky, M. C., & Reeves, P. (2021). Polyfunctionality of the CXCR4/CXCL12 axis in health and disease: Implications for therapeutic interventions in cancer and immune-mediated diseases. *FASEB Journal*, *35*(4). https://doi.org/10.1096/FJ.202001273R

Cabana, E., Lillo, R. E., & Laniado, H. (2021). Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Statistical Papers*, *62*(4), 1583–1609. https://doi.org/10.1007/S00362-019-01148-1/METRICS

Chetty, M., Hallinan, J., Ruz, G. A., & Wipat, A. (2022). Computational intelligence and machine learning in bioinformatics and computational biology. *Biosystems*, *222*, 104792. https://doi.org/https://doi.org/10.1016/j.biosystems.2022.104792

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., & Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, *8*(Suppl 4), S11. https://doi.org/10.1186/1752-0509-8-S4-S11

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports 2017 7:1*, *7*(1), 1–13. https://doi.org/10.1038/srep42717

Daina, A., & Zoete, V. (2016). A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem*, *11*(11), 1117–1121. https://doi.org/10.1002/CMDC.201600182

Das, D., Maeda, K., Hayashi, Y., Gavande, N., Desai, D. V., Chang, S. B., … Mitsuya, H. (2015). Insights into the mechanism of inhibition of CXCR4: Identification of

piperidinylethanamine analogs as anti-HIV-1 inhibitors. *Antimicrobial Agents and Chemotherapy*, *59*(4), 1895–1904. https://doi.org/10.1128/AAC.04654-14/SUPPL_FILE/ZAC003153800SO1.PDF

De Andrade Teles, R. B., Diniz, T. C., Costa Pinto, T. C., De Oliveira, R. G., e Silva, M. G., De Lavor, É. M., … Da Silva Almeida, J. R. G. (2018). Flavonoids as therapeutic agents in Alzheimer's and Parkinson's diseases: A systematic review of preclinical evidences. *Oxidative Medicine and Cellular Longevity*, *2018*. https://doi.org/10.1155/2018/7043213

Delaney, J. S. (2004). ESOL: estimating aqueous solubility directly from molecular structure. *Journal of Chemical Information and Computer Sciences*, *44*(3), 1000–1005. https://doi.org/10.1021/CI034243X

Dickson, D. W., Rademakers, R., & Hutton, M. L. (2007). Progressive supranuclear palsy: Pathology and genetics. *Brain Pathology*, *17*(1), 74–82. https://doi.org/10.1111/J.1750-3639.2007.00054.X

Ding, J., Tang, S., Mei, Z., Wang, L., Huang, Q., Hu, H., … Wu, J. (2023). Vina-GPU 2.0: Further Accelerating AutoDock Vina and Its Derivatives with Graphics Processing Units. *Journal of Chemical Information and Modeling*, *63*(7), 1982–1998. https://doi.org/10.1021/acs.jcim.2c01504

Doogue, M. P., & Polasek, T. M. (2013). The ABCD of clinical pharmacokinetics. *Therapeutic Advances in Drug Safety*, *4*(1), 5. https://doi.org/10.1177/2042098612469335

Du, Y., Qu, J., Zhang, W., Bai, M., Zhou, Q., Zhang, Z., … Miao, J. (2016). Morin reverses neuropathological and cognitive impairments in APPswe/PS1dE9 mice by targeting multiple pathogenic mechanisms. *Neuropharmacology*, *108*, 1–13. https://doi.org/10.1016/J.NEUROPHARM.2016.04.008

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., … Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *The Lancet Neurology*, *6*(8), 734–746. https://doi.org/10.1016/S1474-4422(07)70178-3

Dun, Y., Li, G., Yang, Y., Xiong, Z., Feng, M., Wang, M., … Ma, R. (2012). Inhibition of the canonical Wnt pathway by Dickkopf-1 contributes to the neurodegeneration in 6-OHDA-lesioned rats. *Neuroscience Letters*, *525*(2), 83–88. https://doi.org/10.1016/J.NEULET.2012.07.030

Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical*

Information and Modeling, *61*(8), 3891–3898. https://doi.org/10.1021/ACS.JCIM.1C00203/SUPPL_FILE/CI1C00203_SI_002.ZIP

Egan, W. J., Merz, K. M., & Baldwin, J. J. (2000). Prediction of drug absorption using multivariate statistics. *Journal of Medicinal Chemistry*, *43*(21), 3867–3877. https://doi.org/10.1021/JM000292E

Enrico, C. (2019). Nanotechnology-Based Drug Delivery of Natural Compounds and Phytochemicals for the Treatment of Cancer and Other Diseases. *Studies in Natural Products Chemistry*, *62*, 91–123. https://doi.org/10.1016/B978-0-444-64185-4.00003-4

Ertl, P., & Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, *1*(1). https://doi.org/10.1186/1758-2946-1-8

Espinoza-Moraga, M., Caballero, J., Gaube, F., Winckler, T., & Santos, L. S. (2012). 1-Benzyl-1,2,3,4-Tetrahydro-β-Carboline as Channel Blocker of N-Methyl-d-Aspartate Receptors. *Chemical Biology & Drug Design*, *79*(4), 594–599. https://doi.org/10.1111/J.1747-0285.2012.01317.X

Fields, C. R., Bengoa-Vergniory, N., & Wade-Martins, R. (2019). Targeting Alpha-Synuclein as a Therapy for Parkinson's Disease. *Frontiers in Molecular Neuroscience*, *12*. https://doi.org/10.3389/FNMOL.2019.00299

Fjelldal, M. F., Freyd, T., Evenseth, L. M., Sylte, I., Ring, A., & Paulsen, R. E. (2019). Exploring the overlapping binding sites of ifenprodil and EVT-101 in GluN2B-containing NMDA receptors using novel chicken embryo forebrain cultures and molecular modeling. *Pharmacology Research & Perspectives*, *7*(3), e00480. https://doi.org/10.1002/PRP2.480

Gavriel, Y., Rabinovich-Nikitin, I., Ezra, A., Barbiro, B., & Solomon, B. (2020). Subcutaneous Administration of AMD3100 into Mice Models of Alzheimer's Disease Ameliorated Cognitive Impairment, Reduced Neuroinflammation, and Improved Pathophysiological Markers. *Journal of Alzheimer's Disease : JAD*, *78*(2), 653–671. https://doi.org/10.3233/JAD-200506

Gavriel, Y., Rabinovich-Nikitin, I., & Solomon, B. (2022). Inhibition of CXCR4/CXCL12 signaling: a translational perspective for Alzheimer's disease treatment. *Neural Regeneration Research*, *17*(1), 108. https://doi.org/10.4103/1673-5374.314303

Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of*

Combinatorial Chemistry, *1*(1), 55–68. https://doi.org/10.1021/CC9800071

Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., & Chong, J. (2016). BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research*, *44*(D1), D1045–D1053. https://doi.org/10.1093/NAR/GKV1072

Gómez-Benito, M., Granado, N., García-Sanz, P., Michel, A., Dumoulin, M., & Moratalla, R. (2020). Modeling Parkinson's Disease With the Alpha-Synuclein Protein. *Frontiers in Pharmacology*, *11*, 1. https://doi.org/10.3389/FPHAR.2020.00356

Ha, C. H. H., Fatima, A., & Gaurav, A. (2015). In silico investigation of flavonoids as potential trypanosomal nucleoside hydrolase inhibitors. *Advances in Bioinformatics*, *2015*. https://doi.org/10.1155/2015/826047

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeerschd, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, *4*(8), 1–17. https://doi.org/10.1186/1758-2946-4-17/FIGURES/14

Howes, M. J. R., & Simmonds, M. S. J. (2014). The role of phytochemicals as micronutrients in health and disease. *Current Opinion in Clinical Nutrition and Metabolic Care*, *17*(6), 558–566. https://doi.org/10.1097/MCO.0000000000000115

Huang, S. M., Strong, J. M., Zhang, L., Reynolds, K. S., Nallani, S., Temple, R., … Lesko, L. J. (2008). New era in drug interaction evaluation: US Food and Drug Administration update on CYP enzymes, transporters, and the guidance process. *Journal of Clinical Pharmacology*, *48*(6), 662–670. https://doi.org/10.1177/0091270007312153

Hung, T. C., Lee, W. Y., Chen, K. B., & Chen, C. Y. C. (2014). Lead screening for CXCR4 of the human HIV infection receptor inhibited by traditional Chinese medicine. *BioMed Research International*, *2014*. https://doi.org/10.1155/2014/809816

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., … Montine, T. J. (2012). National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's and Dementia*, *8*(1), 1–13. https://doi.org/10.1016/J.JALZ.2011.10.007

Imran, M., Shah, F. A., Nadeem, H., Zeb, A., Faheem, M., Naz, S., … Li, S. (2021). Synthesis and Biological Evaluation of Benzimidazole Derivatives as Potential Neuroprotective

Agents in an Ethanol-Induced Rodent Model. *ACS Chemical Neuroscience*, *12*(3), 489–505. https://doi.org/10.1021/ACSCHEMNEURO.0C00659/ASSET/IMAGES/MEDIUM/CN0C00659_0010.GIF

Irwin, D. J., Lee, V. M. Y., & Trojanowski, J. Q. (2013). Parkinson's disease dementia: convergence of α-synuclein, tau and amyloid-β pathologies. *Nature Reviews Neuroscience 2013 14:9*, *14*(9), 626–636. https://doi.org/10.1038/nrn3549

Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., … Sayle, R. A. (2020). ZINC20 - A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *Journal of Chemical Information and Modeling*, *60*(12), 6065–6073. https://doi.org/10.1021/ACS.JCIM.0C00675/ASSET/IMAGES/LARGE/CI0C00675_0007.JPEG

Jamshidi, S., Rostami, A., Shojaei, S., Taherkhani, A., & Taherkhani, H. (2024). Exploring natural anthraquinones as potential MMP2 inhibitors: A computational study. *Biosystems*, *235*, 105103. https://doi.org/https://doi.org/10.1016/j.biosystems.2023.105103

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/RSTA.2015.0202

Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chemico-Biological Interactions*, *171*(2), 165–176. https://doi.org/10.1016/J.CBI.2006.12.006

Khaliq, F., Oberhauser, J., Wakhloo, D., & Mahajani, S. (2023). Decoding degeneration: The implementation of machine learning for clinical detection of neurodegenerative disorders. *Neural Regeneration Research*, *18*(6), 1235–1242. https://doi.org/10.4103/1673-5374.355982

Khan, M. Z., Brandimarti, R., Shimizu, S., Nicolai, J., Crowe, E., & Meucci, O. (2008). The chemokine CXCL12 promotes survival of postmitotic neurons by regulating Rb protein. *Cell Death and Differentiation*, *15*(10), 1663–1672. https://doi.org/10.1038/CDD.2008.95

Khan, Muhammad Zafrullah, Brandimarti, R., Musser, B. J., Resue, D. M., Fatatis, A., & Meucci, O. (2003). The chemokine receptor CXCR4 regulates cell-cycle proteins in neurons. *Journal of Neurovirology*, *9*(3), 300–314. https://doi.org/10.1080/13550280390201010

Kikuoka, R., Miyazaki, I., Kubota, N., Maeda, M., Kagawa, D., Moriyama, M., … Asanuma, M. (2020). Mirtazapine exerts astrocyte-mediated dopaminergic neuroprotection. *Scientific Reports 2020 10:1*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-77652-4

Klein, R. S., & Rubin, J. B. (2004). Immune and nervous system CXCL12 and CXCR4: parallel roles in patterning and plasticity. *Trends in Immunology*, *25*(6), 306–314. https://doi.org/10.1016/J.IT.2004.04.002

Kokovay, E., Goderie, S., Wang, Y., Lotz, S., Lin, G., Sun, Y., … Temple, S. (2010). Adult SVZ lineage cells home to and leave the vascular niche via differential responses to SDF1/CXCR4 signaling. *Cell Stem Cell*, *7*(2), 163–173. https://doi.org/10.1016/J.STEM.2010.05.019

Kola, I., & Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery 2004 3:8*, *3*(8), 711–716. https://doi.org/10.1038/nrd1470

Kovacs, G. G. (2015). Invited review: Neuropathology of tauopathies: principles and practice. *Neuropathology and Applied Neurobiology*, *41*(1), 3–23. https://doi.org/10.1111/NAN.12208

Kumar, A., & Zhang, K. Y. J. (2018). Advances in the development of shape similarity methods and their application in drug discovery. *Frontiers in Chemistry*, *6*(JUL), 383861. https://doi.org/10.3389/FCHEM.2018.00315/BIBTEX

Kumar, P., Khanum, F., Khanum, F., & Khanum, F. (2012). Neuroprotective Potential of Phytochemicals. *Pharmacognosy Reviews*, *6*(12), 81–90. https://doi.org/10.4103/0973-7847.99898

Kumar, R., Sharma, A., Alexiou, A., Bilgrami, A. L., Kamal, M. A., & Ashraf, G. M. (2022). DeePred-BBB: A Blood Brain Barrier Permeability Prediction Model With Improved Accuracy. *Frontiers in Neuroscience*, *16*, 858126. https://doi.org/10.3389/FNINS.2022.858126/BIBTEX

Kumar, S., Chowdhury, S., & Kumar, S. (2017). In silico repurposing of antipsychotic drugs for Alzheimer's disease. *BMC Neuroscience*. https://doi.org/10.1186/s12868-017-0394-8

Lamptey, R. N. L., Chaulagain, B., Trivedi, R., Gothwal, A., Layek, B., & Singh, J. (2022). A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics. *International Journal of Molecular Sciences 2022, Vol. 23, Page 1851*, *23*(3), 1851. https://doi.org/10.3390/IJMS23031851

Lee, K., Jang, J., Seo, S., Lim, J., & Kim, W. Y. (2022). Drug-likeness scoring based on unsupervised learning. *Chemical Science*, *13*(2), 554.

https://doi.org/10.1039/D1SC05248A

Li, A. P. (2001). Screening for human ADME/Tox drug properties in drug discovery. *Drug Discovery Today*, *6*(7), 357–366. https://doi.org/10.1016/S1359-6446(01)01712-3

Li, H., & Wang, R. (2017). A focus on CXCR4 in Alzheimer's disease. *Brain Circulation*, *3*(4), 199. https://doi.org/10.4103/BC.BC_13_17

Li, Yuanyuan, Niu, M., Zhao, A., Kang, W., Chen, Z., Luo, N., … Liu, J. (2019). CXCL12 is involved in α-synuclein-triggered neuroinflammation of Parkinson's disease. *Journal of Neuroinflammation*, *16*(1), 1–14. https://doi.org/10.1186/S12974-019-1646-6/FIGURES/6

Li, Yuhua, Meng, Q., Yang, M., Liu, D., Hou, X., Tang, L., … Bi, H. (2019). Current trends in drug metabolism and pharmacokinetics. *Acta Pharmaceutica Sinica. B*, *9*(6), 1113. https://doi.org/10.1016/J.APSB.2019.10.001

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *46*(1–3), 3–26. https://doi.org/10.1016/S0169-409X(00)00129-0

Lovering, F., Bikker, J., & Humblet, C. (2009). Escape from flatland: increasing saturation as an approach to improving clinical success. *Journal of Medicinal Chemistry*, *52*(21), 6752–6756. https://doi.org/10.1021/JM901241E

Lu, M., Grove, E. A., & Miller, R. J. (2002). Abnormal development of the hippocampal dentate gyrus in mice lacking the CXCR4 chemokine receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(10), 7090–7095. https://doi.org/10.1073/PNAS.092013799/ASSET/8B21CCD3-76BF-4D42-8024-C7E5C2D667F1/ASSETS/GRAPHIC/PQ0100137006.JPEG

Lu, W., Zhang, R., Jiang, H., Zhang, H., & Luo, C. (2018). Computer-aided drug design in epigenetics. *Frontiers in Chemistry*, *6*(MAR), 348111. https://doi.org/10.3389/FCHEM.2018.00057/BIBTEX

Ma, J., Dong, L., Chang, Q., Chen, S., Zheng, J., Li, D., … Li, X. (2023). CXCR4 knockout induces neuropathological changes in the MPTP-lesioned model of Parkinson's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1869*(2), 166597. https://doi.org/10.1016/J.BBADIS.2022.166597

Macalino, S. J. Y., Gosu, V., Hong, S., & Choi, S. (2015). Role of computer-aided drug design in modern drug discovery. *Archives of Pharmacal Research 2015 38:9*, *38*(9), 1686–1701.

https://doi.org/10.1007/S12272-015-0640-5

Mannhold, R., Poda, G. I., Ostermann, C., & Tetko, I. V. (2009). Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. *Journal of Pharmaceutical Sciences*, *98*(3), 861–893. https://doi.org/10.1002/JPS.21494

Martin, Y. C. (2005). A bioavailability score. *Journal of Medicinal Chemistry*, *48*(9), 3164–3170. https://doi.org/10.1021/JM0492002

Mason, B., Calhoun, C., Woytowicz, V., Pina, L., Kanda, R., Dunn, C., … Donaldson, S. T. (2021). CXCR4 inhibition with AMD3100 attenuates amphetamine induced locomotor activity in adolescent Long Evans male rats. *PLOS ONE*, *16*(3), e0247707. https://doi.org/10.1371/JOURNAL.PONE.0247707

Memantine for Treatment of Cognitive Impairment in Patients With Parkinson's Disease and Dementia - Full Text View - ClinicalTrials.gov. (n.d.). Retrieved July 10, 2023, from https://classic.clinicaltrials.gov/ct2/show/NCT00294554

Mishra, C. B., Kumari, S., Manral, A., Prakash, A., Saini, V., Lynn, A. M., & Tiwari, M. (2017). Design, synthesis, in-silico and biological evaluation of novel donepezil derivatives as multi-target-directed ligands for the treatment of Alzheimer's disease. *European Journal of Medicinal Chemistry*, *125*, 736–750. https://doi.org/10.1016/J.EJMECH.2016.09.057

Mohd Sairazi, N. S., & Sirajudeen, K. N. S. (2020). Natural Products and Their Bioactive Compounds: Neuroprotective Potentials against Neurodegenerative Diseases. *Evidence-Based Complementary and Alternative Medicine*. Hindawi Limited. https://doi.org/10.1155/2020/6565396

Montanari, F., & Ecker, G. F. (2015). Prediction of drug-ABC-transporter interaction--Recent advances and future challenges. *Advanced Drug Delivery Reviews*, *86*, 17–26. https://doi.org/10.1016/J.ADDR.2015.03.001

Muegge, I., Heald, S. L., & Brittelli, D. (2001). Simple selection criteria for drug-like chemical matter. *Journal of Medicinal Chemistry*, *44*(12), 1841–1846. https://doi.org/10.1021/JM015507E

Mulia, I., Kusuma, W. A., & Afendi, F. M. (2018). Algorithm for Predicting Compound Protein Interaction Using Tanimoto Similarity and Klekota-roth Fingerprint. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *16*(4), 1785–1792. https://doi.org/10.12928/TELKOMNIKA.V16I4.5916

Myszczynska, M. A., Ojamies, P. N., Lacoste, A. M. B., Neil, D., Saffari, A., Mead, R., … Ferraiuolo, L. (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature Reviews Neurology 2020 16:8*, *16*(8), 440–456. https://doi.org/10.1038/s41582-020-0377-8

Nalivaeva, N. N., & Turner, A. J. (2019). Targeting amyloid clearance in Alzheimer's disease as a therapeutic strategy. *British Journal of Pharmacology*, *176*(18), 3447. https://doi.org/10.1111/BPH.14593

Nalls, M. A., Plagnol, V., Hernandez, D. G., Sharma, M., Sheerin, U. M., Saad, M., … Wood, N. W. (2011). Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet*, *377*(9766), 641–649. https://doi.org/10.1016/S0140-6736(10)62345-8

Neves, M. A. C., Simões, S., & Sá e Melo, M. L. (2010). Ligand-guided optimization of CXCR4 homology models for virtual screening using a multiple chemotype approach. *Journal of Computer-Aided Molecular Design*, *24*(12), 1023–1033. https://doi.org/10.1007/s10822-010-9393-x

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An Open chemical toolbox. *Journal of Cheminformatics*, *3*(10), 1–14. https://doi.org/10.1186/1758-2946-3-33/TABLES/2

O'Connor, B. B., Grevesse, T., Zimmerman, J. F., Ardoña, H. A. M., Jimenez, J. A., Bitounis, D., … Parker, K. K. (2020). Human brain microvascular endothelial cell pairs model tissue-level blood–brain barrier function. *Integrative Biology*, *12*(3), 64–79. https://doi.org/10.1093/INTBIO/ZYAA005

Parachikova, A., & Cotman, C. W. (2007). Reduced CXCL12/CXCR4 results in impaired learning and is downregulated in a mouse model of Alzheimer disease. *Neurobiology of Disease*, *28*(2), 143. https://doi.org/10.1016/J.NBD.2007.07.001

Pardridge, W. M. (2012). Drug transport across the blood–brain barrier. *Journal of Cerebral Blood Flow & Metabolism*, *32*(11), 1959. https://doi.org/10.1038/JCBFM.2012.126

Pecoraro, B., Tutone, M., Hoffman, E., Hutter, V., Almerico, A. M., & Traynor, M. (2019). Predicting Skin Permeability by Means of Computational Approaches: Reliability and Caveats in Pharmaceutical Studies. *Journal of Chemical Information and Modeling*, *59*(5), 1759–1771. https://doi.org/10.1021/ACS.JCIM.8B00934

Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning*

Research, *12*(85), 2825–2830. Retrieved from http://jmlr.org/papers/v12/pedregosa11a.html

Popugaeva, E., Chernyuk, D., Zhang, H., Postnikova, T. Y., Pats, K., Fedorova, E., … Bezprozvanny, I. (2019). Derivatives of piperazines as potential therapeutic agents for Alzheimer's disease. *Molecular Pharmacology*, *95*(4), 337–348. https://doi.org/10.1124/MOL.118.114348/-/DC1

Rabinovich-Nikitin, I., Ezra, A., Barbiro, B., Rabinovich-Toidman, P., & Solomon, B. (2016). Chronic administration of AMD3100 increases survival and alleviates pathology in SOD1(G93A) mice model of ALS. *Journal of Neuroinflammation*, *13*(1), 123. https://doi.org/10.1186/s12974-016-0587-6

Rajmohan, R., & Reddy, P. H. (2017). Amyloid Beta and Phosphorylated Tau Accumulations Cause Abnormalities at Synapses of Alzheimer's disease Neurons. *Journal of Alzheimer's Disease : JAD*, *57*(4), 975. https://doi.org/10.3233/JAD-160612

Rao, V. S., Srinivas, K., Sujini, G. N., & Kumar, G. N. S. (2014). Protein-Protein Interaction Detection: Methods and Analysis. *International Journal of Proteomics*, *2014*, 1–12. https://doi.org/10.1155/2014/147648

Remya, C., Dileep, K. V., Tintu, I., Variyar, E. J., & Sadasivan, C. (2013). Design of potent inhibitors of acetylcholinesterase using morin as the starting compound. *Http://Dx.Doi.Org/10.1080/21553769.2013.815137*, *6*(3–4), 107–117. https://doi.org/10.1080/21553769.2013.815137

Ritchie, T. J., Ertl, P., & Lewis, R. (2011). The graphical representation of ADME-related molecule properties for medicinal chemists. *Drug Discovery Today*, *16*(1–2), 65–72. https://doi.org/10.1016/J.DRUDIS.2010.11.002

Roy, D., Hinge, V. K., & Kovalenko, A. (2019). To Pass or Not to Pass: Predicting the Blood-Brain Barrier Permeability with the 3D-RISM-KH Molecular Solvation Theory. *ACS Omega*, *4*(16), 16774–16780. https://doi.org/10.1021/ACSOMEGA.9B01512/SUPPL_FILE/AO9B01512_SI_002.XLSX

Roy, K. K., Tota, S., Tripathi, T., Chander, S., Nath, C., & Saxena, A. K. (2012). Lead optimization studies towards the discovery of novel carbamates as potent AChE inhibitors for the potential treatment of Alzheimer's disease. *Bioorganic & Medicinal Chemistry*, *20*(21), 6313–6320. https://doi.org/10.1016/J.BMC.2012.09.005

Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., & Todeschini, R. (2012).

Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules 2012, Vol. 17, Pages 4791-4810*, *17*(5), 4791–4810. https://doi.org/10.3390/MOLECULES17054791

Samadi, A., Estrada, M., Pérez, C., Rodríguez-Franco, M. I., Iriepa, I., Moraleda, I., … Marco-Contelles, J. (2012). Pyridonepezils, new dual AChE inhibitors as potential drugs for the treatment of Alzheimer's disease: Synthesis, biological assessment, and molecular modeling. *European Journal of Medicinal Chemistry*, *57*, 296–301. https://doi.org/10.1016/J.EJMECH.2012.09.030

Samadi, A., Revenga, M. D. L. F., Pérez, C., Iriepa, I., Moraleda, I., Rodríguez-Franco, M. I., & Marco-Contelles, J. (2013). Synthesis, pharmacological assessment, and molecular modeling of 6-chloro-pyridonepezils: New dual AChE inhibitors as potential drugs for the treatment of Alzheimer's disease. *European Journal of Medicinal Chemistry*, *67*, 64–74. https://doi.org/10.1016/J.EJMECH.2013.06.021

Sanfilippo, C., Castrogiovanni, P., Imbesi, R., Nunnari, G., & Di Rosa, M. (2020). Postsynaptic damage and microglial activation in AD patients could be linked CXCR4/CXCL12 expression levels. *Brain Research*, *1749*, 147127. https://doi.org/10.1016/J.BRAINRES.2020.147127

Scali, C., Caraci, F., Gianfriddo, M., Diodato, E., Roncarati, R., Pollio, G., … Caricasole, A. (2006). Inhibition of Wnt signaling, modulation of Tau phosphorylation and induction of neuronal cell death by DKK1. *Neurobiology of Disease*, *24*(2), 254–265. https://doi.org/10.1016/J.NBD.2006.06.016

Sehgal, S. A., Hammad, M. A., Tahir, R. A., Akram, H. N., & Ahmad, F. (2018). Current Therapeutic Molecules and Targets in Neurodegenerative Diseases Based on in silico Drug Design. *Current Neuropharmacology*, *16*(6), 649–663. https://doi.org/10.2174/1570159X16666180315142137

Senthil Kumar, G., Kishore, N., Elumalai, E., & Gupta, K. K. (2023). Identification of CXCR4 inhibitors as a key therapeutic small molecule in renal fibrosis. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2023.2246575

Shaker, B., Yu, M. S., Song, J. S., Ahn, S., Ryu, J. Y., Oh, K. S., & Na, D. (2021). LightBBB: computational prediction model of blood–brain-barrier penetration based on LightGBM. *Bioinformatics*, *37*(8), 1135–1139. https://doi.org/10.1093/BIOINFORMATICS/BTAA918

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T.

(2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Sharifi-Rad, M., Lankatillake, C., Dias, D. A., Docea, A. O., Mahomoodally, M. F., Lobine, D., … Sharifi-Rad, J. (2020). Impact of Natural Compounds on Neurodegenerative Disorders: From Preclinical to Pharmacotherapeutics. *Journal of Clinical Medicine*, *9*(4), 1061. https://doi.org/10.3390/jcm9041061

Shimoji, M., Pagan, F., Healton, E. B., & Mocchetti, I. (2009). CXCR4 and CXCL12 expression is increased in the nigro-striatal system of Parkinson's disease. *Neurotoxicity Research*, *16*(3), 318–328. https://doi.org/10.1007/S12640-009-9076-3/METRICS

Sierks, M. R., Chatterjee, G., McGraw, C., Kasturirangan, S., Schulz, P., & Prasad, S. (2011). CSF levels of oligomeric alpha-synuclein and beta-amyloid as biomarkers for neurodegenerative disease. *Integrative Biology*, *3*(12), 1188–1196. https://doi.org/10.1039/C1IB00018G

Sim, K. Y., Im, K. C., & Park, S. G. (2020). The Functional Roles and Applications of Immunoglobulins in Neurodegenerative Disease. *International Journal of Molecular Sciences*, *21*(15), 1–19. https://doi.org/10.3390/IJMS21155295

Simón-Sánchez, J., Schulte, C., Bras, J. M., Sharma, M., Gibbs, J. R., Berg, D., … Gasser, T. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature Genetics 2009 41:12*, *41*(12), 1308–1312. https://doi.org/10.1038/ng.487

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics*, *13*(1), 1–13. https://doi.org/10.1186/S13321-020-00478-9/FIGURES/4

Stanzione, F., Giangreco, I., & Cole, J. C. (2021). Use of molecular docking computational tools in drug discovery. *Progress in Medicinal Chemistry*, *60*, 273–343. https://doi.org/10.1016/BS.PMCH.2021.01.004

Strupp, M., Teufel, J., Zwergal, A., Schniepp, R., Khodakhah, K., & Feil, K. (2017). Aminopyridines for the treatment of neurologic disorders. *Neurology: Clinical Practice*, *7*(1), 65–76. https://doi.org/10.1212/CPJ.0000000000000321/ASSET/FFE6382A-84CF-4DDB-A9A7-FCE8A2214C1B/ASSETS/GRAPHIC/9FFU1.JPEG

Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Cherkasov, A., Li, J., … Tetko, I. V. (2010). Applicability domains for classification problems: Benchmarking of distance to models for ames mutagenicity set. *Journal of Chemical Information and Modeling*,

50(12), 2094–2111. https://doi.org/10.1021/CI100253R/SUPPL_FILE/CI100253R_SI_001.XLS

Szilágyi, K., Flachner, B., Hajdú, I., Szaszkó, M., Dobi, K., Lőrincz, Z., … Dormán, G. (2021). Rapid Identification of Potential Drug Candidates from Multi-Million Compounds' Repositories. Combination of 2D Similarity Search with 3D Ligand/Structure Based Methods and In Vitro Screening. *Molecules 2021, Vol. 26, Page 5593*, *26*(18), 5593. https://doi.org/10.3390/MOLECULES26185593

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., … von Mering, C. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(D1), D638–D646. https://doi.org/10.1093/NAR/GKAC1000

Tadayon, M., & Garkani-Nejad, Z. (2019). In silico study combining QSAR, docking and molecular dynamics simulation on 2,4-disubstituted pyridopyrimidine derivatives. *Https://Doi.Org/10.1080/10799893.2019.1641821*. https://doi.org/10.1080/10799893.2019.1641821

Talele, T., Khedkar, S., & Rigby, A. (2010). Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Current Topics in Medicinal Chemistry*, *10*(1), 127–141. https://doi.org/10.2174/156802610790232251

Tegler, L., Corin, K., Pick, H., Brookes, J., Skuhersky, M., Vogel, H., & Zhang, S. (2020). The G protein coupled receptor CXCR4 designed by the QTY code becomes more hydrophilic and retains cell signaling activity. *Scientific Reports 2020 10:1*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-77659-x

Thomas, S. J., & Grossberg, G. T. (2009). Memantine: a review of studies into its safety and efficacy in treating Alzheimer’s disease and other dementias. *Clinical Interventions in Aging*, *4*, 367–377. https://doi.org/10.2147/CIA.S6666

Tian, Y., Jiang, L., Zhang, S., Cao, Y., Liu, F., Xia, L., & Su, Z. (2022). HMGB1 A Box binds to CXCR4 to inhibit HMGB1/CXCL12 mediating macrophage and T cell infiltration and prevents neuronal damage in Parkinson's Disease. *Authorea Preprints*. https://doi.org/10.22541/AU.166126875.56478352/V1

Tiwari, A., & Singh, S. (2022). Computational approaches in drug designing. *Bioinformatics: Methods and Applications*, 207–217. https://doi.org/10.1016/B978-0-323-89775-4.00010-9

Todeschini, R., & Consonni, V. (2010). *Molecular Descriptors for Chemoinformatics*.

Molecular Descriptors for Chemoinformatics (Vol. 2). Wiley Blackwell. https://doi.org/10.1002/9783527628766

Tosco, P., Stiefl, N., & Landrum, G. (2014). The integration of Open3DTOOLS into the RDKit and KNIME. *Journal of Cheminformatics 2014 6:1*, *6*(1), 1–1. https://doi.org/10.1186/1758-2946-6-S1-P8

Tripathi, R., & Kumar, P. (2023). Preliminary study to identify CXCR4 inhibitors as potential therapeutic agents for Alzheimer's and Parkinson's diseases. *Integrative Biology*, *15*. https://doi.org/10.1093/INTBIO/ZYAD012

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16), 1701–1718. https://doi.org/10.1002/JCC.20291

Vancraenenbroeck, R., De Raeymaecker, J., Lobbestael, E., Gao, F., De Maeyer, M., Voet, A., … Taymans, J. M. (2014). In silico, in vitro and cellular analysis with a kinome-wide inhibitor panel correlates cellular LRRK2 dephosphorylation to inhibitor activity on LRRK2. *Frontiers in Molecular Neuroscience*, *7*(JUNE), 88442. https://doi.org/10.3389/FNMOL.2014.00051/BIBTEX

Varadaraju, K. R., Kumar, J. R., Mallesha, L., Muruli, A., Mohana, K. N. S., Mukunda, C. K., & Sharanaiah, U. (2013). Virtual screening and biological evaluation of piperazine derivatives as human acetylcholinesterase inhibitors. *International Journal of Alzheimer's Disease*, *2013*. https://doi.org/10.1155/2013/653962

Varoquaux, G., & Colliot, O. (2023). Evaluating Machine Learning Models and Their Diagnostic Value. *Neuromethods*, *197*, 601–630. https://doi.org/10.1007/978-1-0716-3195-9_20

Vatansever, S., Schlessinger, A., Wacker, D., Kaniskan, H. Ü., Jin, J., Zhou, M. M., & Zhang, B. (2021). Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. *Medicinal Research Reviews*, *41*(3), 1427–1473. https://doi.org/10.1002/MED.21764

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, *45*(12), 2615–2623. https://doi.org/10.1021/JM020017N

Venkata, G., Ramalingayya, Cheruku, S. P., Nayak, P. G., Kishore, A., Shenoy, R., … Krishnadas, N. (2017). Rutin protects against neuronal damage in vitro and ameliorates doxorubicin-induced memory deficits in vivo in Wistar rats. *Drug Design, Development*

and Therapy, *11*, 1011–1026. https://doi.org/10.2147/DDDT.S103511

Vinader, V., Ahmet, D. S., Ahmed, M. S., Patterson, L. H., & Afarinkia, K. (2013). Discovery and Computer Aided Potency Optimization of a Novel Class of Small Molecule CXCR4 Antagonists. *PLOS ONE*, *8*(10), e78744. https://doi.org/10.1371/JOURNAL.PONE.0078744

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods 2020 17:3*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, Q.-L., Fang, C.-L., Huang, X.-Y., & Xue, L.-L. (2022). Research progress of the CXCR4 mechanism in Alzheimer's disease. *Ibrain*, *8*(1), 3–14. https://doi.org/10.1002/IBRA.12026

Wang, X., Xu, X., Li, Y., Li, X., Tao, W., Li, B., … Yang, L. (2013). Systems pharmacology uncovers Janus functions of botanical drugs: activation of host defense system and inhibition of influenza virus replication. *Integrative Biology*, *5*(2), 351–371. https://doi.org/10.1039/C2IB20204B

Wang, Y., Lv, Y., Jin, L., & Liang, G. (2020). Revealing the Mechanism of EGCG, Genistein, Rutin, Quercetin, and Silibinin Against hIAPP Aggregation via Computational Simulations. *Interdisciplinary Sciences: Computational Life Sciences*, *12*(1), 59–68. https://doi.org/10.1007/s12539-019-00352-9

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/JOSS.03021

Weeraratna, A. T., Kalehua, A., DeLeon, I., Bertak, D., Maher, G., Wade, M. S., … Taub, D. D. (2007). Alterations in immunological and neurological gene expression patterns in Alzheimer's disease tissues. *Experimental Cell Research*, *313*(3), 450–461. https://doi.org/10.1016/J.YEXCR.2006.10.028

Wei, H., Wu, G., Chen, J., Zhang, X., Xiong, C., Lei, Y., … Ruan, J. (2013). (2S)-5, 2′, 5′-trihydroxy-7-methoxyflavanone, a natural product from abacopteris penangiana, presents neuroprotective effects in vitro and in vivo. *Neurochemical Research*, *38*(8), 1686–1694. https://doi.org/10.1007/S11064-013-1070-8/FIGURES/5

West, A. B. (2017). Achieving neuroprotection with LRRK2 kinase inhibitors in Parkinson disease. *Experimental Neurology*. https://doi.org/10.1016/j.expneurol.2017.07.019

Wood, L. B., Winslow, A. R., & Strasser, S. D. (2015). Systems biology of neurodegenerative

diseases. *Integrative Biology*, *7*(7), 758–775. https://doi.org/10.1039/C5IB00031A

Wu, D., Chen, Q., Chen, X., Han, F., Chen, Z., & Wang, Y. (2023). The blood–brain barrier: structure, regulation, and drug delivery. *Signal Transduction and Targeted Therapy 2023 8:1*, *8*(1), 1–27. https://doi.org/10.1038/s41392-023-01481-w

Xiang, M., Cao, Y., Fan, W., Chen, L., & Mo, Y. (2012). Computer-Aided Drug Design: Lead Discovery and Optimization. *Combinatorial Chemistry & High Throughput Screening*, *15*(4), 328–337. https://doi.org/10.2174/138620712799361825

Xiang, S., Zhang, Y., Jiang, T., Ke, Z., Shang, Y., Ning, W., … Zhang, T. (2020). Knockdown of Follistatin-like 1 disrupts synaptic transmission in hippocampus and leads to cognitive impairments. *Experimental Neurology*, *333*, 113412. https://doi.org/10.1016/J.EXPNEUROL.2020.113412

Yan, Y., Su, J., & Zhang, Z. (2022). The CXCL12/CXCR4/ACKR3 Response Axis in Chronic Neurodegenerative Disorders of the Central Nervous System: Therapeutic Target and Biomarker. *Cellular and Molecular Neurobiology*, *42*(7), 2147–2156. https://doi.org/10.1007/S10571-021-01115-1/TABLES/1

Yu, W., & Mackerell, A. D. (2017). Computer-Aided Drug Design Methods. *Methods in Molecular Biology (Clifton, N.J.)*, *1520*, 85. https://doi.org/10.1007/978-1-4939-6634-9_5

Zou, Y. R., Kottman, A. H., Kuroda, M., Taniuchi, I., & Littman, D. R. (1998). Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature 1998 393:6685*, *393*(6685), 595–599. https://doi.org/10.1038/31269

PUBLICATIONS

- **Tripathi R** and Kumar P (2025). Identification of CXCR4 Inhibitory Activity in Natural Compounds using Cheminformatics-guided Machine Learning Algorithms. Integrative Biology 17, zyaf004. DOI: 10.1093/intbio/zyaf004 (IF: 1.5)

- **Tripathi R** and Kumar P (2023). Preliminary study to identify CXCR4 inhibitors as potential therapeutic agents for Alzheimer's and Parkinson's diseases. Integrative Biology 15, zyad012. DOI: 10.1093/intbio/zyad012 (IF: 1.5)

- **Tripathi R**, Gupta R, Sahu M, Srivastava D, Das A, Ambasta RK and Kumar P (2022). Free radical biology in neurological manifestations: mechanisms to therapeutics interventions. Environmental Science and Pollution Research 29, 62160–62207. DOI: 10.1007/s11356-021-16693-2 (IF: 5.19)

- Gupta R, Kumari S, **Tripathi R**, Ambasta RK, Kumar P (2023). Unwinding the modalities of necrosome activation and necroptosis machinery in neurological diseases. Ageing Research Reviews 86, 101855. DOI: 10.1016/j.arr.2023.101855. (IF: 12.5)

- Gupta R, Sahu M, **Tripathi R**, Ambasta RK and Kumar P (2022). Protein S-sulfhydration: Unraveling the prospective of hydrogen sulfide in the brain, vasculature and neurological manifestations. Ageing Research Reviews 76, 101579. DOI: 10.1016/j.arr.2022.101579. (IF: 12.5)

- Sahu M, **Tripathi R**, Jha NK, Jha SK, Ambasta RK, Kumar P (2022). Cross talk mechanism of disturbed sleep patterns in neurological and psychological disorders. Neuroscience & Biobehavioral Reviews 140, 104767. DOI: 10.1016/j.neubiorev.2022.104767 (IF: 7.6)

- Advani D, Sharma S, **Tripathi R**, Gupta R, Jaiswal A, Ambasta RK, Kumar P (2021). Mitochondrial dysfunction in metabolic disorders. Mitochondrial Dysfunction and Nanotherapeutics, 91-137. DOI: 10.1016/B978-0-323-85666-9.00015-2.

- Advani D, Gupta R, **Tripathi R**, Sharma S, Ambasta RK, Kumar P (2020). Protective role of anticancer drugs in neurodegenerative disorders: A drug repurposing approach. Neurochemistry International 140, 104841. DOI: 10.1016/j.neuint.2020.104841 (IF: 4.4)

CONFERENCES & WORKSHOPS

- International Conference on “Antimicrobial Resistance, Novel Drug Discovery and Vaccine Development: Challenges and Opportunities” organised by SRM University Delhi-NCR, Sonepat. 18 to 20 March 2024
- DST STUTI Hands-On Training Program on “Biological Electron Microscopy” organised by Sophisticated Analytical Instrumentation Facility, All India Institute of Medical Sciences, New Delhi. 29 August to 4 September 2022.
- One-day national symposium on “Neurochemical Legacy of Neurological Disorders: Brainstorming of Novel Approaches” organised by Society for Neurochemistry, India (SNCI) and Department of Toxicology, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi. 9 March 2022.
- Workshop on “Neurological Disorders: Advances in Research Techniques and Translational Applications” organised by Society of Neurochemistry, India (SNCI) and Department of Toxicology, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi. 13 to 19 October 2019.
- International Conference on “Frontiers in Neuroscience and Neurochemistry: Dynamic Challenges and Approaches” organised by the Society of Neurochemistry, India (SNCI) and Department of Toxicology, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi. 10 to 12 October 2019.

RAHUL TRIPATHI

Qualification: Ph.D. Biotechnology
Supervisor: Prof. Pravir Kumar
Delhi Technological University

Address:
House No.4052, Street No. 108,
Sant Nagar, Burari, Delhi-110084
Mobile: 9990675355
Email: alaneuschevalier@gmail.com

Achievements:

- DBT-JRF Category I (2016)
- GATE-BT (2016)
- JGEEBILS (2016)
- GATE-XL (2014)
- JGEEBILS (2014)

Publication Details:

1. **Tripathi R** and Kumar P. Identification of CXCR4 Inhibitory Activity in Natural Compounds using Cheminformatics-guided Machine Learning Algorithms. Integrative Biology (IF: 1.5)
2. **Tripathi R** and Kumar P. Preliminary study to identify CXCR4 inhibitors as potential therapeutic agents for Alzheimer's and Parkinson's diseases. Integrative Biology 15, zyad012. DOI: 10.1093/intbio/zyad012 (IF: 1.5)
3. **Tripathi R**, Gupta R, Sahu M, Srivastava D, Das A, Ambasta RK and Kumar P. Free radical biology in neurological manifestations: mechanisms to therapeutics interventions. Environmental Science and Pollution Research 29, 62160–62207. DOI: 10.1007/s11356-021-16693-2 (IF: 4.23)
4. Gupta R, Kumari S, **Tripathi R**, Ambasta RK, Kumar P. Unwinding the modalities of necrosome activation and necroptosis machinery in neurological diseases. *Ageing Res Rev*. 2023; 86:101855. DOI: 10.1016/j.arr.2023.101855
5. Gupta R., Sahu M, **Tripathi R**, Ambasta RK and Kumar P. Protein S-sulfhydration: Unraveling the prospective of hydrogen sulfide in the brain, vasculature and neurological manifestations. Ageing Research Reviews 76, 101579. DOI: 10.1016/j.arr.2022.101579. (IF: 10.895)
6. Sahu M, Tripathi R, Jha NK, Jha SK, Ambasta RK, Kumar P. Cross talk mechanism of disturbed sleep patterns in neurological and psychological disorders. Neuroscience & Biobehavioral Reviews. 2022:104767. DOI: 10.1016/j.neubiorev.2022.104767 (IF: 9.052)
7. Advani D, Gupta R, **Tripathi R**, Sharma S, Ambasta RK, Kumar P. Protective role of anticancer drugs in neurodegenerative disorders: A drug repurposing approach. Neurochemistry International 2020, 140:104841. DOI: 10.1016/j.neuint.2020.104841 (IF: 3.921)
8. Advani D, Sharma S, **Tripathi R**, Gupta R, Jaiswal A, Ambasta RK, Kumar P. Mitochondrial dysfunction in metabolic disorders. Mitochondrial Dysfunction and Nanotherapeutics 2021, 91-137. DOI: https://doi.org/10.1016/B978-0-323-85666-

Work Experience:

- Senior Research Fellow at IIWBR, Karnal: December 2017 to December 2018

Conferences and Workshops:

- International Conference on "Frontiers in Neuroscience and Neurochemistry: Dynamic Challenges and Approaches" organized by Society of Neurochemistry, India (SNCI) and Department of Toxicology, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi. 10th to 12th October, 2019.
- Workshop on "Neurological Disorders: Advances in Research Techniques and Translational Applications" organized by Society of Neurochemistry, India (SNCI) and Department of Toxicology, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi. 13th to 19th October, 2019.
- One-day national symposium on "Neurochemical Legacy of Neurological Disorders: Brainstorming of Novel Approaches" organized by Society for Neurochemistry, India (SNCI) and Department of Toxicology, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi. 9th March, 2022.
- DST STUTI Hands-On Training Program on "Biological Electron Microscopy" organized by Sophisticated Analytical Instrumentation Facility, All India Institute of Medical Sciences, New Delhi. 29th August to 4th September, 2022.

M. Tech. Projects:

- **Title:** "Molecular Cloning and Overexpression of Lignin Biosynthesis Pathway Genes"
- **Duration and Organization:** 1 year, Motilal Nehru National Institute of Technology (Allahabad)

B. Tech. Projects:

- **Title:** "Effect of Pesticides on the Growth of Mosquito Larva – A Preliminary Study"
- **Duration and Organization:** 6 months, N.C. College of Engineering, Israna (Panipat)

- **Title:** "Utilization of Whey to Produce Value Added Beverages"
- **Duration and Organization:** 6 months, N.C. College of Engineering, Israna (Panipat)

Summer Training/Internship/Experience:

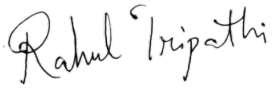
- **Organization:** Institute of Cytology and Preventive Oncology
  - **Title**: Basic Molecular Biology Techniques
  - **Duration**: 1 Month, Year-2009
  - **Brief Description:** Hands-on training in basic molecular biology techniques used in cancer research.

Educational Qualifications:

Qualification	Institution	Year	Percentage
M.Tech. (Biotechnology)	MNNIT Allahabad	2016	8.50 (CGPA)
B.Tech. (Biotechnology)	N.C. College of Engineering (Kurukshetra University)	2010	71.97
XII (C.B.S.E)	Apex Public School	2005	71.40
X (C.B.S.E)	D.A.V. Public School	2003	69.40

I hereby declare that all the information given is correct to the best of my knowledge.

Date: 26/06/2025



Place: Delhi (RAHUL TRIPATHI)

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis: **Deciphering the Mechanisms of Alzheimer's and Parkinson's Diseases Using Network Biology and a Functional Genomics Approach**

Total Pages: **89**

Name of the Scholar: **Rahul Tripathi**

Supervisor: **Prof. Pravir Kumar**

Department of Biotechnology, Delhi Technological University, Delhi-110042

This is to report that the above thesis was scanned for similarity detection. Process and outcome are given below:

Software used: **Turnitin**, Similarity Index: **10%**, Total Word Count: **23,173**

Date: 26-06-2025

Rahul Tripathi

Candidate's Signature

Pravir Kumar

Signature of Supervisor(s)

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Exclusions

- 6 Excluded Sources

Match Groups

- 123 Not Cited or Quoted 10%
  Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

7% Internet sources
7% Publications
6% Submitted works (Student Papers)

Integrity Flags

1 Integrity Flag for Review

Replaced Characters
41 suspect characters on 19 pages
Letters are swapped with similar characters from another alphabet.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Pravinkumar

Rahul

Tripathi R and Kumar P (2023). Preliminary study to identify CXCR4 inhibitors as potential therapeutic agents for Alzheimer's and Parkinson's diseases. Integrative Biology 15, zyad012. DOI: 10.1093/intbio/zyad012

OXFORD

Integrative Biology, 2023, **15**, 1–19
https://doi.org/10.1093/intbio/zyad012
Advance access publication date 19 August 2023
Original Article

Preliminary study to identify CXCR4 inhibitors as potential therapeutic agents for Alzheimer's and Parkinson's diseases

Rahul Tripathi and Pravir Kumar*

Department of Biotechnology, Molecular Neuroscience and Functional Genomics Laboratory, Delhi Technological University (Formerly Delhi College of Engineering), Delhi, India
*Corresponding author. E-mail: pravirkumar@dtu.ac.in
https://www.ncbi.nlm.nih.gov/myncbi/1DW465XG9bp5p/bibliography/public/; https://scholar.google.co.in/citations?user=WVLI4i4AAAAJ&hl=en;
Scopus ID: 14831447800; **ISI Research ID:** B-2164-2015

Abstract

Neurodegenerative disorders (NDDs) are known to exhibit genetic overlap and shared pathophysiology. This study aims to find the shared genetic architecture of Alzheimer's disease (AD) and Parkinson's disease (PD), two major age-related progressive neurodegenerative disorders. The gene expression profiles of GSE67333 (containing samples from AD patients) and GSE114517 (containing samples from PD patients) were retrieved from the Gene Expression Omnibus (GEO) functional genomics database managed by the National Center for Biotechnology Information. The web application GREIN (GEO RNA-seq Experiments Interactive Navigator) was used to identify differentially expressed genes (DEGs). A total of 617 DEGs (239 upregulated and 379 downregulated) were identified from the GSE67333 dataset. Likewise, 723 DEGs (378 upregulated and 344 downregulated) were identified from the GSE114517 dataset. The protein–protein interaction networks of the DEGs were constructed, and the top 50 hub genes were identified from the network of the respective dataset. Of the four common hub genes between two datasets, C-X-C chemokine receptor type 4 (CXCR4) was selected due to its gene expression signature profile and the same direction of differential expression between the two datasets. Mavorixafor was chosen as the reference drug due to its known inhibitory activity against CXCR4 and its ability to cross the blood–brain barrier. Molecular docking and molecular dynamics simulation of 51 molecules having structural similarity with Mavorixafor was performed to find two novel molecules, ZINC49067615 and ZINC103242147. This preliminary study might help predict molecular targets and diagnostic markers for treating Alzheimer's and Parkinson's diseases.

Insight Box
Our research substantiates the therapeutic relevance of CXCR4 inhibitors for the treatment of Alzheimer's and Parkinson's diseases. We would like to disclose the following insights about this study. We found common signatures between Alzheimer's and Parkinson's diseases at transcriptional levels by analyzing mRNA sequencing data. These signatures were used to identify putative therapeutic agents for these diseases through computational analysis. Thus, we proposed two novel compounds, ZINC49067615 and ZINC103242147, that were stable, showed a strong affinity with CXCR4, and exhibited good pharmacokinetic properties. The interaction of these compounds with major residues of CXCR4 has also been described.

Keywords: CXCR4; Alzheimer's disease; Parkinson's disease; neurodegenerative disorders; molecular docking; molecular dynamics simulation

INTRODUCTION

Neurodegenerative disorders are reported to share a common pathophysiology. Genome-wide association data have identified genetic overlap and revealed common biological pathways between neurodegenerative diseases, such as Alzheimer's disease (AD) and Parkinson's disease (PD) [1]. Discovering common genetic architecture among different neurodegenerative disorders may help determine underlying shared disease mechanisms and facilitate early diagnosis and treatment strategies. AD and PD are age-related progressive neurodegenerative disorders with an enormous emotional and economic impact on patients and caregivers. AD is characterized by intracellular neurofibrillary tangles (NFT) composed of aggregated hyperphosphorylated tau protein in the neurons and glial cells [2, 3] and extracellular amyloid plaques consisting of aggregated amyloid-β (Aβ) [4, 5]. Even though PD is conventionally characterized by the aggregation of α-synuclein (α-Syn), tau protein and NFT have also been reported to modify PD symptomatology and disease risk [6–8]. Besides their activity in the immune system, chemokines and their receptors are remarkably expressed in the central nervous system (CNS) and modulate cell migration and neurotransmission.

Received: December 15, 2022. **Revised:** July 10, 2023. **Accepted:** August 8, 2023

Tripathi R and Kumar P (2025). Identification of CXCR4 Inhibitory Activity in Natural Compounds using Cheminformatics-guided Machine Learning Algorithms. Integrative Biology 17, zyaf004. DOI: 10.1093/intbio/zyaf004

OXFORD

Integrative Biology, 2025, **17**, zyaf004
https://doi.org/10.1093/intbio/zyaf004
Original Article

Identification of CXCR4 inhibitory activity in natural compounds using cheminformatics-guided machine learning algorithms

Rahul Tripathi and Pravir Kumar *,†

Molecular Neuroscience and Functional Genomics Laboratory, Delhi Technological University, Shahabad Daulatpur, Bawana Road, Delhi 110042, India

*Corresponding authors. Department of Biotechnology, Molecular Neuroscience and Functional Genomics Laboratory, Delhi Technological University, Shahabad Daulatpur, Bawana Road, Delhi 110042, India. E-mail: pravirkumar@dtu.ac.in; kpravir@gmail.com

†Professor and Dean (International Affairs) Delhi Technological University (formerly Delhi College of Engineering)
Former Faculty, Tufts University School of Medicine, Boston, MA, USA
Room# FW4TF3, Mechanical Engineering Building
Shahbad Daulatpur, Bawana Road, Delhi 110042; Tel: +91-9818898622
Email: pravirkumar@dtu.ac.in; kpravir@gmail.com
https://www.ncbi.nlm.nih.gov/myncbi/1DW465XG9bp5p/bibliography/public
https://scholar.google.co.in/citations?user=WVLI4i4AAAAJ&hl=en
Scopus ID: 14831447800; ISI Research ID: B-2164-2015

Abstract

Neurodegenerative disorders are characterised by progressive damage to neurons that leads to cognitive impairment and motor dysfunction. Current treatment options focus only on symptom management and palliative care, without addressing their root cause. In our previous study, we reported the upregulation of the CXC motif chemokine receptor 4 (CXCR4), in Alzheimer's disease (AD) and Parkinson's disease (PD). We reached this conclusion by analysing gene expression patterns of AD and PD patients, compared to healthy individuals of similar age. We used RNA sequencing data from Gene Expression Omnibus to carry out this analysis. Herein, we aim to identify natural compounds that have potential inhibitory activity against CXCR4 through cheminformatics-guided machine learning, to aid drug discovery for neurodegenerative disorders, especially AD and PD. Natural compounds are gaining prominence in the treatment of neurodegenerative disorders due to their biocompatibility and potential neuroprotective properties, including their ability to modulate CXCR4 expression. Recent advances in artificial intelligence (AI) and machine learning (ML) algorithms have opened new avenues for drug discovery research across various therapeutic areas, including neurodegenerative disorders. We aim to produce an ML model using cheminformatics-guided machine learning algorithms using data of compounds with known CXCR4 activity, retrieved from the Binding Database, to analyse various physicochemical attributes of natural compounds obtained from the COCONUT Database and predict their inhibitory activity against CXCR4.

Received: April 23, 2024. **Revised:** October 22, 2024. **Accepted:** February 11, 2025

International Conference on "Frontiers in Neuroscience and Neurochemistry: Dynamic Challenges and Approaches" organised by the Society of Neurochemistry, India (SNCI) and Department of Toxicology, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi. 10 to 12 October 2019.

Celebrating 30 Years of Excellence in Professional Education and Research with Empathy for all

JAMIA HAMDARD

(Deemed to be University), Accredited 'A' Grade by NAAC,
Hamdard Nagar, New Delhi - 110062

Department of Toxicology, SCLS

In association with

Society for Neurochemistry, India (SNCI)

organizes

International Conference

on

Frontiers in Neuroscience and Neurochemistry:
Dynamic Challenges and Approaches

&

33rd Annual Meeting of Society for Neurochemistry, India (SNCI)

Certificate of Participation

We have the pleasure to confirm the participation of

Prof./Dr./Mr./Ms. Rahul Tripathi

in the International Conference on "Frontiers in Neuroscience and Neurochemistry: Dynamic Challenges and Approaches" during 10th to 12th, October 2019.

Dr. M. Mumtaz Alam
Jt. Organizing Secretary
SNCI-2019

Prof. (Dr.) P. Prakash Babu
Secretary General (H.Q.)
Society for Neurochemistry, India

Prof. (Dr.) M. K. Thakur
President
Society for Neurochemistry, India

Prof. (Dr.) Suhel Parvez
Organizing Secretary
SNCI-2019

Department of Science and Technology
Ministry of Science and Technology
Government of India
DST-FIST
PROMOTION OF UNIVERSITY RESEARCH AND SCIENTIFIC EXCELLENCE
Research and Development Infrastructure Division
(DST-PURSE)

MHRD
Govt. of India
Recommended by Ministry of Human Resource & Development
(MHRD) Panel for 'Institution of Eminence'
(5th Highest Ranking Deemed/Private University)

QS WORLD UNIVERSITY RANKINGS 2019

NAAC

UGC-SAP DRS II

nirf NATIONAL INSTITUTIONAL RANKING FRAMEWORK
Ministry of Human Resource Development
Government of India

1st Rank (Pharmacy Discipline)
18th Rank (University Category)

One-day national symposium on "Neurochemical Legacy of Neurological Disorders: Brainstorming of Novel Approaches" organised by Society for Neurochemistry, India (SNCI) and Department of Toxicology, School of Chemical and Life Sciences, Jamia Hamdard, New Delhi. 9 March 2022.

JAMIA HAMDARD

Deemed to be University
Accredited 'A' Grade by NAAC
Hamdard Nagar, New Delhi 110062

DEPARTMENT OF TOXICOLOGY, SCLS
in association with
SOCIETY FOR NEUROCHEMISTRY, INDIA (SNCI)
organizes one day symposium on

NEUROCHEMICAL LEGACY OF NEUROLOGICAL DISORDERS: BRAINSTORMING OF NOVEL APPROACHES

Certificate of Participation

We have the pleasure to confirm the participation of

Prof./Dr./Mr./Ms. Rahul Tripathi, Delhi Technological University

in the one day symposium held on 9th March, 2022. He/she also presented a Paper/Poster

Prof. (Dr.) M. Mumtaz Alam
Jt. Organizing Secretary
SNCI-2022

Prof. (Dr.) Suhel Parvez
Organizing Secretary
SNCI-2022

Prof. (Dr.) P. Prakash Babu
Secretary General (H.Q.)
Society of Neurochemistry

International Conference on "Antimicrobial Resistance, Novel Drug Discovery and Vaccine Development: Challenges and Opportunities" organised by SRM University Delhi- NCR, Sonepat. 18 to 20 March 2024

SRM UNIVERSITY
DELHI-NCR, SONEPAT

3rd International Conference
On
"Antimicrobial Resistance, Novel Drug Discovery and Vaccine Development: Challenges and Opportunities"
Organized by
SRM University Delhi - NCR, Sonepat
March 18 - 20, 2024

CERTIFICATE

This is to certify that ..... Rahul Tripathi ..... has delivered an ~~oral~~ / poster presentation in the 3rd International Conference held on March 18-20, 2024 at India Habitat Centre, New Delhi-110003, India

Prof. (Dr.) V. Samuel Raj
Organizing Secretary

Prof. (Dr.) Paramjit S. Jaswal
Vice-Chancellor