

Development and Application of a Long Short-Term Memory-Based Drug-likeness Molecular Designer Algorithm

A THESIS

*Submitted in partial fulfillment of the
requirements for the award of the degree*

of

DOCTOR OF PHILOSOPHY

by

SRINIVASARAO MANDE

17CHPH28



SCHOOL OF CHEMISTRY

UNIVERSITY OF HYDERABAD

**C. R. Rao Road, Central University P.O., Gachibowli
Hyderabad, Telangana-500046, INDIA**

JUNE 2023

Dedicated to My Beloved Family and Teachers

DECLARATION

I, Srinivasarao Mande, hereby declare that the matter embodied in the thesis entitled “**Development and Application of a Long Short-Term Memory-Based Drug-likeness Molecular Designer Algorithm**” is the result of my investigation carried out in the School of Chemistry, University of Hyderabad, Hyderabad, India, under the supervision of Dr. K. V. Jovan Jose.

In keeping with the general practice of reporting scientific observation, due acknowledgments have been made wherever the work described is based on the finding of other investigators. Any omission, which might have occurred by oversight error, is regretted. This research work is free from plagiarism. I hereby agree that my thesis can be deposited in Shodhganga/INFLIBNET. A report on plagiarism statistics from the University library is enclosed.

Srinivasa Rao. M

Srinivasarao Mande

17CHPH28

School of Chemistry
University of Hyderabad
Hyderabad – 500046 Telangana
India



CERTIFICATE

This is to certify that the thesis entitled **“Development and Application of a Long Short-Term Memory-Based Drug-likeness Molecular Designer Algorithm”** submitted by Srinivasarao Mande bearing Reg. No. 17CHPH28, in partial fulfillment of the requirements for the award of Doctor of Philosophy in Chemistry in the School of Chemistry, is a bonafide work carried out by him under my supervision and guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Research Publications:

1. Srinivasarao, Mande; K. V. Jovan Jose. A new model system preparation approach for ONIOM: Accurate estimation of energies and forces of ten molecules with Pi-bonds. *J. Indian Chem. Soc.* 2019, 96, 1003-1009.
2. Srinivasarao, Mande; Siva, Chiriki; K. V. Jovan Jose. LSTM-Enabled Onsite Recurrent Molecular Designer (LORD) for Predicting Drug-likeness Molecules on COVID-19 Protein Target (manuscript preparation).
3. Srinivasarao, Mande; K. V. Jovan Jose. A Novel Recurrent Neural Network (RNN)-Based Algorithm to Design Potential Drug-likeness Molecules to Alzheimer's Disease (Manuscript preparation).
4. Srinivasarao, Mande; K. V. Jovan Jose. Predictions of potential drug-likeness molecules to inhibit MAO- B in Parkinson's disease (Manuscript preparation).
5. Srinivasarao, Mande; K. V. Jovan Jose. Design of potential drug-likeness molecules for Tuberculosis. (Manuscript preparation).
6. Srinivasarao, Mande; K. V. Jovan Jose. Identification of New and Potential HIV-1 Reverse Transcriptase (RT) inhibitory by LORD molecular Designer Strategy. (Manuscript preparation).
7. Srinivasarao, Mande; K. V. Jovan Jose. LSTM Enabled Onsite Recurrent De Novo Drug (LORD) Designer Application to Breast Cancer. (Manuscript preparation).

Workshops Attended:

1. Participated in a workshop on “protein structure and drug design” (PSDD) at the School of life sciences, University of Hyderabad, August 27-September 05, 2017.
2. Participated in a workshop on “New vistas in drug discovery” (NVDD) at the School of Life Sciences, University of Hyderabad, March 16-17, 2019.
3. Participated in a workshop on Molecular Docking Studies in Drug Discovery at the Department of Zoology, Kakatiya University, Vidyarnyapuri, Warangal, May 29-31, 2019.

- Participated in a symposium and hand session on “ML for sciences: Symposium and discussion meeting” at the Centre for Computational Natural Sciences and Bioinformatics, IIIT Hyderabad, November 29-30, 2019.
- Participated in the winter school on Computational Chemistry and Biology at the School of Chemistry, University of Hyderabad, December 9-27, 2019.

Seminars: Oral and Poster Presentations

- Presented a poster presentation entitled “Developing an Artificial Neural Network-Based Algorithms for *De Novo* Computer Aided Drug Designing” and participated in NMSTC-2017 at the University of Hyderabad. October 13-14, 2017.
- Presented a poster presentation entitled “A novel method for computer-aided drug design” and participated in the 8th International Collaborative and Cooperative Chemistry (ICCCS-8) at the University of Hyderabad, December 18-19, 2017.
- Presented a poster presentation entitled “A machine learning approach to a fragment-based method in drug designing” and participated in the 15th Annual In-House symposium at the University of Hyderabad, March 9-10, 2018.
- Presented a poster presentation entitled “Developing computational methods for designing drug molecules in Alzheimer’s disease” and participated in the 16th Annual In-House symposium at the University of Hyderabad, February 22-23, 2019.
- Presented a poster presentation entitled “Development of a Novel Method for Potential Drug Likeness Molecules Combat COVID-19 by Recurrent Neural Network (RNN)” participated in 17th Theoretical Chemistry Symposium (TCS-2021) at IISER Kolkata, December 11-14, 2021.
- Presented an Oral or Poster Presentation entitled “Molecular Designer: A Novel Recurrent Neural Network (RNN)-Based Algorithm to Design Potential Drug Likeness Molecules” participated in CHEMFEST-2022 at the School of Chemistry, University of Hyderabad April 22-23, 2022.

The student has passed the following courses toward fulfilling the coursework requirement for Ph.D.

Course NO.	Title of Course	Credits	Result
CY 801	Research Proposal	4	Pass
CY 805	Instrumental Method-A	4	Pass
CY 806	Instrumental Method-B	4	Pass


30/06/2023
Dean

School of Chemistry
Dean
SCHOOL OF CHEMISTRY
University of Hyderabad
Hyderabad-500 046

V


Supervisor

School of Chemistry
JOVAN JOSE K. V.
Assistant Professor
School of Chemistry
University of Hyderabad
Hyderabad - 500 046.

ACKNOWLEDGEMENTS

Success is the progressive realization of a worthy goal. Many able people fail to achieve anything worthwhile because they have not been adequately guided. The success of any project depends solely on the support, guidance, and encouragement received from the guide and well-wishers.

I sincerely thank my research supervisor Dr. K. V. Jovan Jose, for his constant support and guidance during my Ph.D. tenure. I thank him for his continuous advice and encouragement throughout my Ph.D. tenure.

I thank my research doctoral committee (DRC) members, Professor Lalitha Guruprasad and Professor Debhasis Barik, for their insightful comments and encouragement on my thesis work.

I thank the Dean and former Deans of the School of Chemistry, University of Hyderabad, during my stay here, for providing the necessary facilities to carry out my research work. I thank all the faculty members of the School of Chemistry for their inspiring teaching, and all the non-teaching staff of the school are acknowledged for their support.

I thank my colleagues Dudam Praveen, Sridatri Nandy, and Gunjan Ramteke for creating a pleasant working atmosphere in the lab. I thank my project juniors, Peena Guguloth, V. Sridhar, Prabin Yash, and Ganesh.

It gives me immense pleasure to thank my seniors Dr. Surender Reddy and Dr. Arun Kumar. K, Dr. Narasimha, Dr. Siva, Dr. Laxman, Dr, Anupam dey, and my batch mates V. Arun Kumar, Dr. Ajaneyulu. B, Dinesh, Naresh, Vamshi, vinay, Nitai-Giri, Hafizur, Suraj, Shabbir, Asif, Nooral, Mujhid, Ishfaq, Intazar Ali, Isha Mishra, Archana, Keerthi, Daradi and Anju joseph.

Friends are the essential ingredient in the recipe of life. I consider myself lucky to have my life flavored with my all-dearest friends Thunga Ramesh, Venaknna Dabbakatla, Chandra Shekar Jalagam, Praneeth, J. Venkatesh, Jagadessh, Chand Pasha, Jahangir, Nares, Vamshi Krishna Reddy, Lakshaman, Ramesh D., Upender, Ramana, Rajshekar, Naveen, Raju G., Raju P. and, Rakesh, Siddu.

I thank M. Pharm. friends Rajasekhar, Srinivas, Saidul, Mounika, Mouna Sree, Dhana Laxmi, Sravanthi, Varalaxmi, P.S.P Laxmi, Pranitha, and N.R.S hostel friends Sanyasi Rao, Rama Naidu, Santhosh, Rana deep, Shashi, Karthik, Vinay, Shyam, Shameer, and Preety, who have showered me with much affection, and care and have been very encouraging during my struggling times.

I thank DSU- HCU friends Dr. Suman Damera (Mizoram Central University), Dr. Sampangi Shankar (OU), Kumar Raja Kuchipudi, Venkatadri Silapaka, Vamsi Surapogu, Sree Charan, Venu, Kamalakar, Balu, Chennuri Ramesh, Karthik Jirra, Birun, Vardhan Teja, Jayaraj, Nikhil and friends from Anthropology department A. Sharath Kumar, Sambaiah A., Venugopal, Dinesh, and Nagarju Bathini, and special thanks to Gokhan Akmaz (Ankara University, Turkey).

I express my gratitude to my grandfather M. Appaiah, D. Gopaiah; grandmother M. Pichamma; father, M. Nageswar Rao; mother, M. Ramadevi; brothers, M. Narasimharao and M. Ramarao; Sister-in-laws M. Sravani and M. Priyanka; Sister Rajeswari, for their love, affection care, encouragement to complete this research work.

My heartily and special thanks to Prof. J. Venkateshwara Rao, principal of Talla Padmavathi College of Pharmacy; Prof. Bethi Srinivas, HOD of Pharm. Chemistry, Tallapadmavathi College of Pharmacy; Mr. Nagarjuna (Sai Labs.), Srikanth (Saraca laboratory), Srinu and Jagadeesh (Aurobindo Pharma) and Vijay Bhaskar Reddy (R&D manager Saraca Laboratory), Ganga Reddy (GM Saraca Laboratory), Mr. Mahender (Sub-Inspector of Police Yousufguda).

I thank all those who directly or indirectly extended involved in completing my research work.

Srinivasarao Mande

TABLE OF CONTENTS

Chapter 1: Introduction to Computer-Aided Drug Design (CADD)	1
1.1 Overview of CADD drug designing	2
1.2 Brief history of CADD	3
1.3 The impact of technology on CADD	4
1.4 Drug-likeness molecules	6
1.5 Main categories of CADD approaches	7
1.6 Molecular docking and algorithms.....	8
1.7 Scoring functions (SF).....	11
1.8 Molecular dynamics (MD)simulation and binding free energy calculations	12
1.8.1 Force filed	13
1.8.2 Root mean square deviations (RMSD)	14
1.8.3 Binding free energy	14
1.9 Artificial intelligence (AI)	15
1.9.1 AI in drug discovery and applications.....	17
1.9.2 Feed-forward neural network	21
1.9.3 Recurrent neural networks (RNNs)	21
1.9.4 Long short-term memory (LSTM)	23
1.10 List of diseases.....	24
1.10.1 Overview of Alzheimer’s disease	24
1.10.2 Overview of COVID-19	25
1.10.3 Overview of Parkinson’s disease	26
1.10.4 Overview of Tuberculosis disease.....	26
References.....	28

Chapter 2: Methodology of Recurrent Neural Network Method to Design Novel Potential Drug-likeness Molecules	38
2.1 Computational details	39
2.2 Potential binding site Analysis by scanning MESP scanning	39
2.2.1 MESP calculation identification of binding sites in target protein.....	39
2.3 Data Preparation	40
2.4 Input preparation for LORD	41
2.5 LORD output preparation	42
2.6 Block diagram for LORD designer algorithm	43
2.7 LSTM network architecture and training	44
2.8 LORD molecular drug predictor	45
2.9 CHARMM force field in GROMACS	45
2.10 Free energy calculation (g_mmpbsa)	47
References.....	49
 Chapter 3: Potential Drug-Likeness Molecules for Alzheimer’s Disease	 50
3.1 Introduction to Alzheimer’s disease.	51
3.2 Results and Discussions.....	54
3.2.1 MESP calculation for identification of binding sites in Alzheimer's target protein.	54
3.2.2 Design of potential drug-likeness molecules	56
3.2.3 Interactions of stereo-isomers with target molecules	57
3.2.4 Protein-ligand complex and their interaction studies of the top four drug-likeness molecules at selected four binding sites.....	59
3.2.5 Binding site analysis of protein-ligand complexes.....	61
3.2.6 Cavity-drug-likeness molecules interaction matrix.....	62
3.2.7 Physiochemical properties and ADME properties	65

3.2.8 Correlation of potential drug-likeness molecules vs. site-wise binding energy	66
3.2.9 Molecular dynamics (MD) simulations	67
3.2.10 Free energy calculations	69
3.3 Conclusions.....	70
References.....	71
 Chapter 4: Potential Drug-likeness Molecules for COVID-19	 74
4.1 Introduction to COVID-19.....	78
4.2. Results and Discussions.....	78
4.2.1 MESP calculation identification of binding sites M^{pro} in COVID-19 target.....	78
4.2.2 Design of potential candidate molecules.....	79
4.2.3 Stereochemistry and protein-ligand complex interaction studies.....	83
4.2.4 Protein-ligand complex and their interaction studies of the top four drug-likeness molecules at selected four binding sites.....	85
4.2.5 Binding site analysis of protein-ligand complexes.....	86
4.2.6 Cavity-drug-likeness molecules interaction matrix	88
4.2.7 Physiochemical properties and ADME properties	90
4.2.8 Correlation of potential drug-likeness molecules vs. site-wise binding energy.....	92
4.2.9 Molecular dynamics (MD) simulations.....	93
4.2.10 Free energy calculations	94
4.3. Conclusions.....	95
References.....	97
 Chapter 5: Design of Potential Drug-likeness Molecules for Parkinson's Disease	 100
5.1 Introduction to Parkinson's disease.....	101
5.2 Results and Discussions.....	104

5.2.1 MESP calculation for identification binding sites in target protein	104
5.2. Design of potential candidate molecules.....	105
5.2.3 Stereochemistry and protein-ligand complex interaction studies	107
5.2.4 Protein-ligand complex and their interaction studies of the top four drug-likeness molecules at selected four binding sites.....	108
5.2.5 Binding site analysis from protein-ligand complexes.....	110
5.2.6 Cavity-drug-likeness molecules interaction matrix.....	111
5.2.7 Physiochemical properties and ADME properties	113
5.2.8 Correlation of potential drug-likeness molecules vs. site-wise binding energy....	115
5.2.9 Molecular dynamics (MD) simulations	116
5.2.10 Free energy calculations	117
5.3 Conclusions.....	118
References.....	119
 Chapter 6: Potential Drug-likeness Molecules for Mycobacterium Tuberculosis	122
6.1 Introduction to Tuberculosis.....	123
6.2 Results and Discussions.....	126
6.2.1 MESP calculation for identification of binding site in target protein	126
6.2.2 Design of potential drug-likeness molecules.....	127
6.2.3 Stereochemistry and protein-ligand complex interaction studies	129
6.2.4 Protein-ligand complex and their interaction studies of the top four drug-likeness molecules at selected four binding sites.....	131
6.2.5 Binding sites analysis from protein-ligand complex.....	132
6.2.6 Cavity-drug-likeness molecules interaction matrix.....	133
6.2.7 Physiochemical and ADME properties	135
6.2.8 Correlation of potential drug-likeness molecules vs. site-wise binding energy.....	136

6.2.9 Molecular dynamics (MD) simulations.....	137
6.2.10 Free energy calculations.	138
6.3 Conclusions.	139
References.....	141
 Chapter 7: Conclusions and Scope for Future Work	 143

LIST OF ABBREVIATIONS

CADD	:	Computer-Aided Drug Design
ADME	:	Absorption, Distribution, Metabolism, Excretion.
QSAR	:	Quantitative Structure-Activity Relationships
SBDD	:	Structure-Based Drug Design
LBDD	:	Ligand-Based Drug Design
SF	:	Scoring Function
AI	:	Artificial Intelligence
ANN	:	Artificial Neural Network
ML	:	Machine Learning
RNN	:	Recurrent Neural Network
LSTM	:	Long Short-Term Memory
LORD	:	LSTM-Enabled On-site Recurrent Molecular Designer
BE	:	Binding Energy
CBE	:	Corrected Binding Energy
FDA	:	Food and Drug Administration
EMBL	:	European Molecular Biology Laboratory
PDB	:	Protein Data Bank
GROMACS	:	Groningen Machine for Chemical Simulations
MD	:	Molecular Dynamics
RMSD	:	Root-Mean-Square Deviation
FE	:	Free Energy
MW	:	Molecular Weight
HBA	:	Hydrogen Bond Acceptor
HBD	:	Hydrogen Bond Donor
LogP	:	Lipophilicity
TPSA	:	Total Polar Surface Area

RBC	:	Rotatable Bond Count
AD	:	Alzheimer's Disease
ACh	:	Acetylcholine
AChE	:	Acetylcholinesterase
COVID-19	:	Coronavirus disease-2019
M ^{pro}	:	Mainprotease
PD	:	Parkinson's disease
MAO-B	:	Monoamine oxidase B
TB	:	Tuberculosis
RpsA	:	Ribosomal protein S1
C ₁	:	Site 1
C ₂	:	Site 2
C ₃	:	Site 3
C ₄	:	Site 4

CHAPTER 1

Introduction to Computer-Aided Drug Design

1.1 Overview of Computer-Aided Drug Design

Computer-aided drug design (CADD) involves discovering and developing new and effective pharmaceuticals (or) drugs utilizing computational methods and techniques.¹ According to reports, the standard drug discovery cycle, illustrated in Figure 1.1, from lead identifications to clinical trials, requires 10–15 years and over 500–800 million USD before leading to the market.^{2,3} For this reason, drug research and commercialization are drawn-out, challenging, and expensive procedures; CADD is frequently employed in the pharmaceutical industry to expedite the designing process in the best possible way.^{4,5} Utilizing computational tools that use AI and ML technologies in the hit-to-lead process has several benefits, including finding promising leads more rapidly and exploring a more expansive chemical space. With the CADD techniques, fewer compounds will need to be synthesized and examined in vitro, which can minimize the extensive trial-and-error testing required.⁶ As a result of the significant expansion of CADD over the past several years, which has enhanced awareness of complicated and demanding biological processes, some CADD-related technologies, such as data science, wearable technology, AI, and ML technologies, appear to be revolutionizing evidence-based medicine, providing a fascinating look into the future deep medicine.⁷ Technologies made it possible to find novel pharmacologically active drug molecules swiftly.

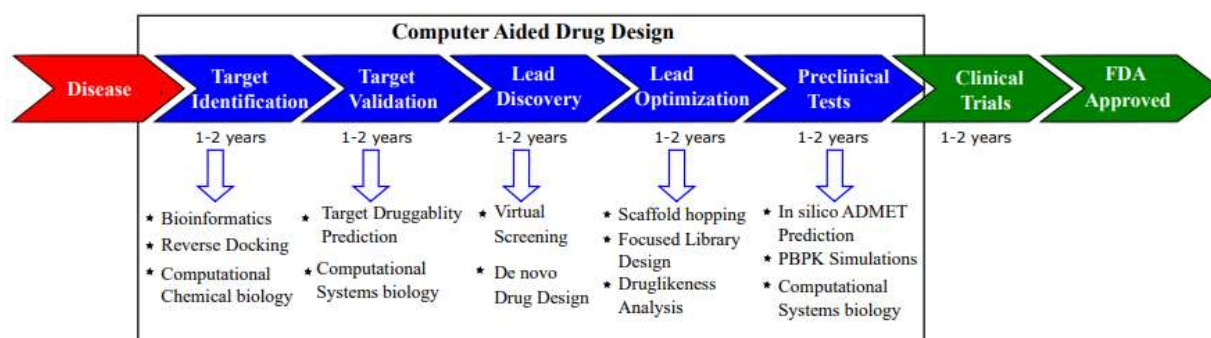


Figure 1.1 The steps and the timeline of the standard drug discovery cycle and the role of computer-aided drug design (CADD) in the modern drug discovery pipeline.

An excellent achievement of CADD has been the design and discovery of inhibitors for various critical illnesses, such as those caused by viruses,^{8,9,10,11} bacterial infections,¹² cancer,¹³ diabetes,¹⁴ and neurological disorders, etc.^{15,16,17} From time immemorial, a few examples of drugs that were developed utilizing the CADD are presented in Table 1.1 The current emphasis in CADD is on creating new computer software or computer-based programmers to create large compound libraries that include several pharmacologically active biomolecules by enhancing their physicochemical and ADMET properties or develop a sophisticated new algorithm to measure the potency and selectivity of new lead

candidate molecules. Therefore, in the first chapter, we will discuss the most recent advances and developments in CADD technology.¹⁸

Table 1.1 The list of drugs that came into existence with the help of CADD, its biological actions, and targets.

Drugs	Biological Actions	Targets
Captopril, ¹⁹ Valsartan ²⁰	Antihypertensive	Angiotensin Converting enzyme (ACE)
Norfloxacin ²¹	Urinary Tract infections	Topoisomerase II & IV
Epalrestat ²²	Diabetic neuropathy	Aldose Reductase
Dorzolamide ²³	Glaucoma	Carbonic anhydrase
Saquinavir, ²⁴ Indinavir, ²⁵ Ritonavir, ²⁶ Nelfinavir, ²⁷ Efavirenz, ²⁸ Darunavir, ²⁹ Raltegravir ³⁰	Human immunodeficiency virus (HIV)	HIV-1 protease
Imatinib ³¹	Acute lymphoblastic leukemia	Abi tyrosinase
Zolmitriptan ³²	Migraine	5-hydroxytryptamine (5HT)
Sunitinib ³³	kidney cancer	V EGF-R2 kinase
Aliskiren ³⁴	Human renin inhibitor	Angiotensinogen
Tomudex ³⁵	Colorectal cancer	Thymidylate synthase
Boceprevir, ³⁶ Telaprevir, ³⁷ Grazoprevir ³⁸	Hepatitis C virus (HCV) inhibitor	NS3/4A protease
Crizotinib, ³⁹ Dacomitinib ⁴⁰	Non-Small Cell Lung Cancer (NSCLC)	Anaplastic Lymphoma Kinase & ROS Proto-Oncogene 1
Saroglitazar ⁴¹	Diabetic Dyslipidemia	Peroxisome Proliferator-Activated (PPAR)
Rucaparib ⁴²	Prostate Cancer	Poly (ADP-ribose) polymerase
Lifitegrast ⁴³	Dry eye disease	LFA-1/ICAM-1
Vaborbactam ⁴⁴	Bacterial infections	beta-lactamase
Erdafitinib ⁴⁵	Urothelial carcinoma	Fibroblast growth factor receptor
Nolatrexed ⁴⁶	Liver cancer	EGFR, VEGF/PDGFR(receptors)

1.2 A Brief History of CADD

The last few decades witnessed significant developments in CADD methods, and the major milestone in the CADD can be classified into the following:

- 1900: Introduction of lock-and-key ideas by E. Fisher (1894) and P. Enrich (1909).

- 1970: Quantitative structure-activity relationships (QSAR) were established. However, it was limited to 2- Dimensional, retrospective analysis.
- 1980: The developments in the CADD field were significantly influenced by various crucial disciplines, including molecular biology, X-ray crystallography, multi-dimensional nuclear magnetic resonance (NMR), molecular modeling, and computer graphics.
- 1990: Combinatorial chemistry, high-throughput screening, and human genome bioinformatics are current approaches in creative medical science. Target identification and lead chemical discovery are aided by in silico techniques like homology modeling and virtual screening. Lead optimization benefits from cutting-edge strategies like scaffold hopping and free energy estimates. Cost-effective preclinical testing benefits from in silico ADMET prediction and pharmacokinetic modeling. There are three significant steps involved in drug design and drug discovery: (i) Target identification and validation, (ii) Lead generation and optimization, and finally, (iii) Clinical studies.

Table 1.2 lists commonly used molecular databases in CADD.

Dataset	Description	Examples
Protein Data Bank (PDB) ^{47,48}	Experimentally established 3D structures of protein, nucleic acids, and complex assemblies are stored in a database.	PDB ID: 6LU7 (Mpro), 1EVE (AChE), 2V5Z (MAO-B), 4NNI (Ribosomal protein S1 of M.TB), etc.
Chemical databases ^{49,50}	Small molecule databases and their biological activity such as interactions with specific proteins and possible medicinal usage.	Drug bank, PubChem, ChEMBL.
Genomic and proteomic datasets ^{51,52}	Genomic and protein sequences datasets from organisms.	Uniprot, Human Genome Project, Ensemble.

1. 3 The Impact of Technology on CADD

Developing medicinal chemicals that target specific molecules has become substantially more efficient with the emergence of computer techniques. CADD encompasses all preclinical or clinical trials, lead finding, optimization, and target identification. The speed and effectiveness of the drug development process have been significantly increased and enhanced due to considerable developments in CADD approaches in the post-genomic period.

1.3.1 Three Phases of the In-Silico Drug Design Methods

The method of computational drug discovery can be classified into three phases involving identifying, optimizing, and testing prospective drug candidates using computer-aided drug designing tools:

First phase: is target identification, in which potential pharmaceutical targets and the associated biological pathways are discovered using computational methods.

Second phase: is known as lead discovery. Several potential therapeutic compounds are screened and narrowed down to a select few with a high affinity for the target.

Third phase: The efficacy and safety properties of the selected drug candidates are greatly enhanced and optimized in the lead optimization phase. These enhanced prospects might then be the subject of preclinical and clinical research for future testing and development.

1.3.2 In Silico Validation of Targets and Identification Cavity

Identifying and validating targets is the initial and pivotal stage in the drug research pipeline. Selecting druggable targets from tens of thousands of potential macromolecules remains a challenge. A variety of technologies have recently developed to achieve these goals.

Table 1.3 Computational algorithms employed for binding-site prediction of proteins.

S.No.	Name of Packages	Methods
1	Fpocket ⁵³	Geometric clustering of ligand binding pockets
2	CASTp ⁵⁴	Identification of surface cavities
3	SURFNET ⁵⁵	Surface-based ligand binding site prediction
4	Meta Detector ⁵⁶	A consensus of multiple methods
5	DoGSiteScorer, ⁵⁷ Epock, ⁷⁰ MSpocket ⁷¹	Machine learning-based pocket prediction
6	MetaPocket ⁶³	A consensus of multiple methods
7	Bite Net ⁵⁸	The deep learning-based binding site prediction
8	Q-SiteFinder ⁶⁶	Voronoi tessellation-based pocket detection
9	LISE ⁶⁸	Ligand-induced surface expansion method
10	PoVME ⁶⁹	Volume-based pocket identification

A protein cavity is a crevice or pocket within a protein that can bind to specific ligands or small molecules. These cavities are typically necessary for protein function because they allow it to interact with other molecules and carry out biological tasks. The amino acid sequence and its three-dimensional structure

contribute to the size and shape of the protein cavities. Understanding the form and function of protein cavities is crucial for designing and developing new drugs. Some of the computational algorithms employed for the binding-site prediction of proteins are reported in Table 1.3.

1.3.3 In Silico ADMET Studies (Absorption, Distribution, Metabolism, Excretion, and Toxicity)

The primary factors contributing to the costly failures of pharmaceutical drugs in late-stage research, pharmacokinetics, and toxicity, are frequently seen as failed. As a result, these factors must be considered as shortly as possible in the drug identification process. Due to the development of combinatorial chemistry and high throughput screening, the number of compounds requiring preliminary data ADMET has considerably grown. The essential pharmacokinetics, metabolism, and toxicity endpoints may be modeled using in-silico technology, accelerating drug development.

1.3.4 In Silico Drug Safety Prediction

Early toxicity prediction is essential in drug discovery to reduce the financial risks and potential harm to the public. Significant losses and safety concerns might come from late-stage clinical studies or post-marketing discovery of unfavorable toxicological consequences. Several methods can be employed to predict toxicities, including Genotoxicity, liver toxicity, Inhibition of CYP450 enzymes, and Cardiovascular toxicity.

1.3.5 Advantages of CADD

1. We could reduce the quantity of biological and synthetic testing in this manner.
2. It produces the most effective medication candidate by filtering substances with undesirable qualities using in silico methods.
3. It is a rapid, automated, cost-effective, and time-saving technique.
4. It allows us to learn more about the patterns of drug-receptor interactions.
5. In silico studies may investigate enormous chemical libraries and uncover compounds with high hit rates, in contrast to traditional high throughput screening methods.
6. These strategies reduce the likelihood of failures in the latter stage.

1.4 Drug-Likeness Molecules

The Lipinski Rule of Five (L-Ro5) is frequently applied to determine how drug-like a molecule is. According to L-Ro5, molecules with the features that follow are predicted to have restricted absorption and permeability, a molecular weight (MW) of not more than 500 Da or g/mol, lipophilicity (LogP) of not

more than 5, and not more than 5 hydrogen-bond donors (HBD) and 10 not more than hydrogen-bond acceptors (HBA),⁵⁹ and the Ghose filter (GF) adds additional criteria to forecasts to make them better. The GF predicts that molecules with MW between 160 and 480 Da, a logP between -0.4 and 5.6, a molar refractivity (A) between 40 and 130, and a total number of atoms (TNA) between 20 and 70 are expected to exhibit considerable absorption.

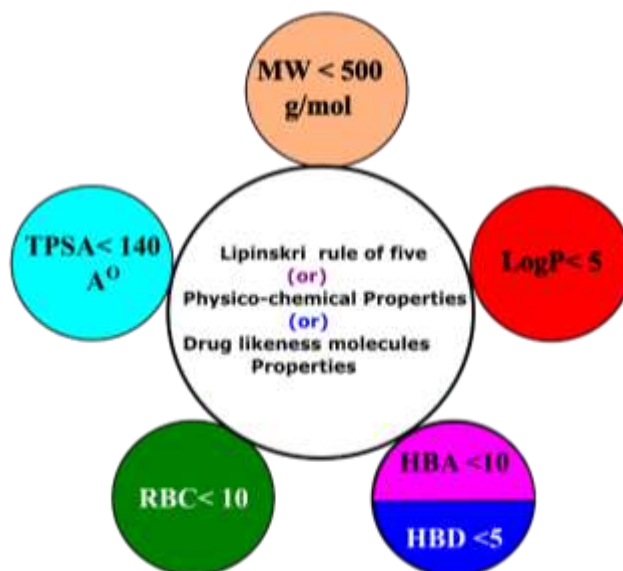


Figure 1.2 Physico-chemical properties of drug-likeness molecules.

Veber's rule (VR) suggests that molecules not more than 10 rotatable bonds (RB) and a polar surface area (PSA) of not more than 140 are more likely to be bioavailable, expanding the standard for bioavailability. Mugger's rule differs from L-Ro5, GF, and VR by incorporating to a further distinction between compounds that are drug-likeness and non-drug-likeness molecules. According to Mugger's rule, drug-like compounds are characterized by a MW of 200-600, a LogP range: -2 to 5, PSA limit: ≤ 150 , NR count: ≤ 7 , NC count: > 4 , NH count: > 1 , RB limit: ≤ 15 , HBD limit: ≤ 5 , HBA limit: ≤ 10 .

1.5 Main Categories of CADD Approaches:

CADD approaches are primarily classified into two categories for developing drugs or pharmaceuticals:

1.5.1 Structure-Based or Direct Drug Design (SBDD)

In contemporary pharmaceutical research, SBDD software finds promising chemical compounds with high binding free energies; calculating an empirical score from free energy in binding or affinity can evaluate the efficacy of Structure-Based Drug Design (SBDD). This score indicates the success rate of SBDD. Notably, SBDD is predicated on a prior structural understanding of the target protein, and docking

is used to determine the bio-affinity or interaction of tested chemical molecules using datasets. This technique may produce, a novel therapeutic molecule with enhanced interaction with the target protein.

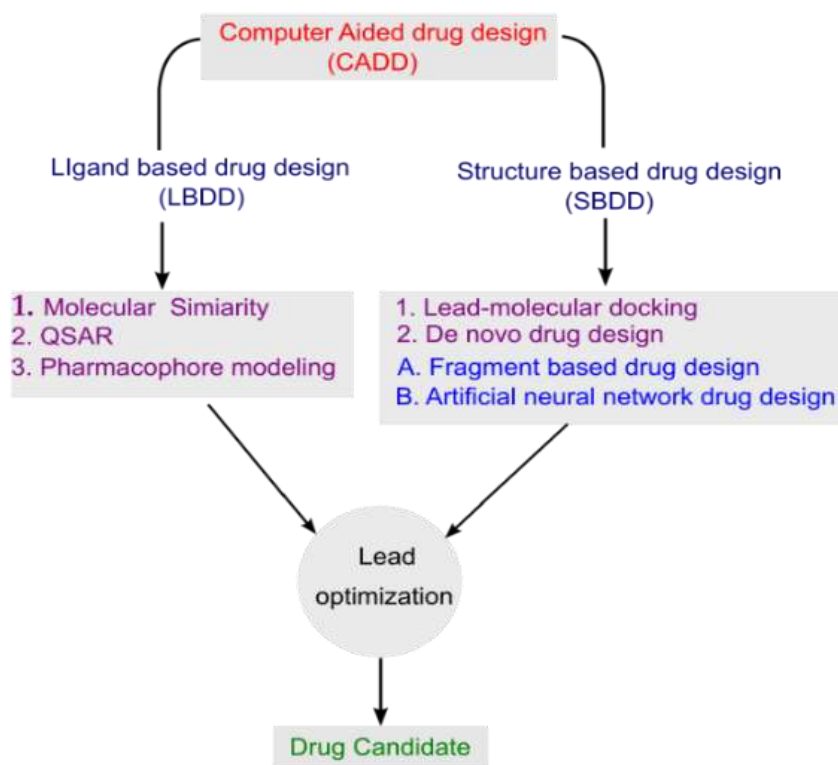


Figure 1.3 Overview of computer-aided drug design.

1.5.2. Ligand-based or Indirect Drug Design (LBDD)

LBDD, or ligand-based drug design, is used in drug discovery, but the ligands that bind to the protein are unknown. In this method, ligands are used to create a pharmacophore model or molecule with all the necessary elements to attach to the target active sites pharmacophore-based techniques and QSARs are primarily utilized in LBDD. In this method, compounds with similar structures are expected to interact with the target protein and have comparable biological functions. The exact atomic interactions of the protein-ligand complex are not necessary for LBDD, in contrast to SBDD.

1.6 Molecular Docking and Algorithms

Various features are used to design docking tools, such as rigid and flexible, and methodologies such as genetic algorithms, simulated annealing, Monte Carlo simulations, and iterative development algorithms

are employed using distinct approaches. The benchmark and an explanation of the molecular docking tools are provided.

Many different molecular docking software programmers were developed utilizing different algorithms.⁶⁰ This section highlights the benefits of commonly used algorithms and explores the diverse applications developed using each approach. It is important to note that docking software often utilizes multiple search algorithms simultaneously or independently.

1.6.1 Matching Algorithm

The Matching algorithm is widely recognized as a simple yet effective approach for consideration that must be given to the spatial alignment of two molecules to accommodate the geometric overlap between them.⁶¹ Different alignments between the receptor and the ligand are achieved in various ways. The DOCK program, for instance, heavily relies on the matching algorithm method as a part of its adaptable docking strategies. For example, in the first stages of the DOCK program, the locations of the potential ligand are determined, as well as the regions of the binding site known as sphere centers that are discovered using a Matching algorithm. Numerous other popular systems, such as DOCK, employ this tactic of Shape-matching algorithms.⁶²

1.6.2 Monte Carlo (MC) Algorithm

Many docking applications, including Auto Dock, MCDOCK, AutoDockVina, QXP, and ROSETTALIGAND, employ Monte Carlo (MC) algorithms.⁶³ Using the Metropolis criteria, the fundamental idea of MC is to accept or reject changes that occur randomly. Docking a ligand with a structure is utilized for determining the conformational structure of molecular affinity.⁶⁴ The Monte Carlo (MC) techniques, rooted in the Metropolis MC algorithm, are crucial in advancing docking investigations by providing acceptance criteria. In each iteration of the Algorithm, ligands undergo random alterations. If the binding position has an improved energy score. The alteration is permitted, while in cases where the energy score decreases, acceptance is determined based on the probability (P) defined in the equation below.

$$P = \exp[-(E_1 - E_0)/K_B T] \quad (Eq. 1.1)$$

K_B is the Boltzmann constant, T is the system temperature, and E_1 and E_0 are the energy scores before and after the system change, respectively. The majority of applications using MC simulation give a precise and accurate performance. However, time-dependent approaches like MD simulations do not suit the MC methodology.

1.6.3 Genetic Algorithm

The genetic Algorithm (GA) is a widely employed optimization technique for tackling docking problems. It has been extensively utilized to expedite calculations and identify the variables that significantly impact the activity of the investigated drug compound.⁶⁵ The GA approach draws inspiration from Darwin's theory of natural evolution, wherein two chromosomes (rates parents) combine genetically to generate a new chromosome that may exhibit superior characteristics compared to its parents.⁶⁶ This approach uses a variety of scoring functions (SFs) and several variables, such as crossover and mutation.⁶⁷

1.6.4 Particle Swarm Optimization

The Particle Swarm Optimization (PSO) technique approach was developed in 1995 by Kennedy and Eberhart. In a study conducted by Garcia-Godoy, Lopez-Camacho, Garcia-Nieto, Nebro, and Aldana-Montes in 2015. One metaheuristic employed in molecular docking is swarm optimization techniques. These approaches frame a docking problem as a parameter optimization problem and relate it to a clearly defined SF. The optimization approach looks for a ligand's docked conformation that uses the least energy. PSO is known for being a productive strategy for handling challenging search situations. Additionally, there are other docking algorithms based on swarm intelligence, such as SODOCK.⁶⁸ The PSO method has proven effective in addressing molecular docking challenges, particularly when minimizing the ligand score based on the scoring function. The developers have designed several docking techniques using these algorithms, each with its own advantages and disadvantages, shown in Table 1.4.

Table 1.4 The algorithms and programs for CADD docking tools.

Algorithms	Programs or Docking Tools
Matching algorithm	DOCK 4.0, ¹⁶ Ph4DOCK, ⁶⁹ GM-Dock ⁷⁰
Monte Carlo (MC)	MCDOC, ⁷¹ PRODOCK, ⁷² AutoDock ⁷³
Genetic Algorithm (GA)	GOLD3.1, ⁷⁴ GasDock, ⁷⁵ PSI-DOCK, ⁷⁶ Autodock4.0 ⁸²
Particle swarm optimization (PSO)	AutoDock, ClustMPSO ⁷⁷
Tabu search	SFDock and Proleads ⁷⁸
Incremental construction	eHiTS, ⁷⁹ DOCK 4.0, ⁸⁰ FlexX ⁸¹
Simulated annealing	MolDock, ⁸² AutoDock4.o and AutoDockVina ⁸³
Multiple Copy Simultaneous Search	HOOK, ⁸⁴ FlexX ³⁰
Evolutionary programming (EP)	MolDock, ³¹ GOLD, ⁸⁵ AutoDock, ²⁵ DIVALI, ⁸⁶ DARWIN, ⁸⁷ PSI- DOCK, ⁸⁸ FLIP Dock, ⁸⁹ Lead finder ⁹⁰
Fast Fourier transform Algorithm.	ZDOCK server, ⁹¹ Pathchdock, ⁹² HexServer ⁹³

1.7 Scoring Function (SF)

The design and definition of the scoring function (SF) is an important feature and critical for accurately rating docking candidates. The accuracy of the docking process relies on the precision of the scoring functions employed to describe the binding mode and ligand placement. These scoring functions are also instrumental in identifying potential drug candidates, thereby facilitating the discovery of protein targets. However, it remains a challenging endeavor to develop a rapid and highly accurate prognosis in molecular docking.⁹⁴

1.7.1 Force Field

A force field model is derived from atomic-scale information such as bond angles, lengths, and torsion angles. Mechanical calculations based on physical principles and experimental data are routinely used to construct force-field functions and associated parameters. Examples of force-field SFs are AutoDock, DOCK, GOLD, and D-score.⁹⁵

1.7.2 Empirical Scoring Function

Calculates binding affinity by adding the fundamental energy components of the protein ligand.⁹⁶ These empirical SFs were used in several investigations. The study compares POLSCORE⁹⁷ to the other SF programmers, Drug Score and X-SCORE. Overall, the analysis demonstrated that POLSCORE is more reliable in predicting the docked position. The empirical scoring function technique employs a training set comprising binding affinities of unknowns to determine the optimal weights for energy components. Various optimization strategies, such as linear regression analysis, can achieve this. Other empirical scoring functions, like F-Score, are also commonly utilized.⁹⁸

1.7.3 Knowledge-Based Scoring Functions

Knowledge-based SFs often leverage the structural details of known protein-ligand complexes to enhance their predictive capabilities.⁹⁹ Another name for it is statistical potential-based SFs; energy potentials obtained from empirically confirmed atomic structures are knowledge-based SFs. Researchers have successfully created and used predictive models for analyzing protein and ligand interactions and predicting protein structure, using scoring functions based on knowledge. Knowledge-based SFs and techniques have been compared in numerous research methodologies mentioned earlier, revealing that knowledge-based SFs can achieve a favorable trade-off between computational efficiency (speed) and predictive accuracy. It also demonstrates robustness when compared to the training set. Examples of knowledge-based SFs include PMF, Smog, Bleep, and Drug Score.¹⁰⁰

1.7.4 Machine Learning-Based Scoring Functions

Various machine learning (ML)-based scoring functions (MLSFs) may be used to create more reliable scoring functions, which can improve docking procedures.¹⁰¹ MLSFs outperform other traditional SFs in general, and they are used for grading and improving SF accuracy. When ML-SFs are of the supervised kind, they rely on training data sets.¹⁰² Support vector machines (SVM), random forests (RF), and convolutional neural networks (CNNs) are some of the ML methodologies utilized to create ML-SFs.¹⁰³

1.8 Molecular Dynamics (MD) Simulations and Free Energy Calculations

MD simulations of proteins were conducted for the first time in 1970, marking a significant milestone over 60 years after the initial discovery of the three-dimensional structure of proteins through X-ray crystallography. The Protein crystal structures archived in the Protein Data Bank (PDB) provide static representations, capturing a single conformation of the proteins. However, it is important to note that protein structures are inherently dynamic. Bonds within proteins undergo constant movement and fluctuations, leading to changes in conformation and occasionally in function. To gain a comprehensive understanding of the conformational space explored by proteins, molecular dynamics (MD) simulations are widely employed. MD simulations offer a means to simulate the motion of proteins, utilizing experimental structures obtained from the PDB and computational models derived from homology modeling or entirely constructed from MD simulations to analyze the spatial arrangement of atoms in three dimensions. This methodology replaces a static model with a dynamic one, where the atomic system is set into motion. By numerically solving the classical Newtonian dynamic equations, the simulation replicates the motion observed in the system. The technique of MD simulations is built upon Newton's second law, often referred to as the equation of motion, which serves as the fundamental principle governing the simulations

$$F = ma \quad (Eq. 1.2)$$

Here, m denotes the mass, a the acceleration of the particle, and F represents the force acting on it. It is possible to calculate their acceleration by simply determining the forces at work on each system component. An integrated set of motion equations provides a trajectory that captures particle positions, velocities, and accelerations over time. This trajectory can be utilized to calculate the average values of different properties.

The process of molecular dynamics (MD) simulation is deterministic, meaning that with the initial positions and velocities of each atom, the state of the system may be predicted at any point in time

ime. It is crucial to comprehend the significance of bonded and unbonded atom interactions, particularly Coulomb's law and the Lennard-Jones potential for van der Waals interactions. Both the system's bonded and unbonded components are covered by these interactions.¹⁰⁴ MD simulations can be computationally demanding and time-consuming, requiring substantial computational resources. Several widely used software packages, such as AMBER,¹⁰⁵ CHARMM,¹⁰⁶ and GROMACS,¹⁰⁷ have been developed to simulate protein flexibility through MD simulations in solvated protein systems.

1.8.1 Force Fields

A collection of variables and a mathematical formula referred to as the force field can be used to mathematically characterize the energy of a protein based on its atomic coordinates. This mathematical statement demonstrates the connection between the energy of the system and its three-dimensional coordinates. A force field employed to describe molecular systems is constructed from two terms: the equation for the bonded interactions, which encompasses bond length, bond angle, and dihedral parameters, and the equation for non-bonded interactions, which represent atoms with covalent bonds. The second term's equation, estimated using Coulomb's law and the Lennard-Jones potential, accounts for non-bonded interactions produced by van der Waals and electrostatic forces.

$$V(r) = \sum_{bonds} k_b (b - b_o)^2 + \sum_{angles} k_\theta (\theta - \theta_o)^2 + \sum_{torsions} k_\Phi [\cos(n\Phi + \delta) + 1] \\ + \sum_{\substack{non\ bond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \right] \quad (Eq. 1.3)$$

In a molecular system, the ensemble of atom coordinates is denoted as $V(r)$. Internal molecular parameters refer to the bonds (b), angles (θ), and torsions (Φ) within the molecule. Non-bonded terms encompass Lennard-Jones (LJ) parameters such as ϵ_{ij} (well-depth) and R_{min} (distance of minimum interaction energy). These LJ parameters define van der Waals interactions. Additionally, partial charges (q) are used in Coulomb's law to describe electrostatic contributions.

An analytical form in force fields represents the interatomic potential energy and other contributing factors. The force field parameters are typically obtained through classical methods like ab-initio or semi-empirical quantum mechanical calculations. Alternatively, they can be derived by fitting the force field to experimental data obtained from techniques.¹⁰⁸ The force field should be computationally efficient for rapid evaluation, while still capturing sufficient detail to replicate the properties of the systems under investigation accurately.

In scientific literature, many force fields are available to describe molecules of different complexities and address diverse systems. Examples include the Universal force fields (UFF), encompassing characteristics for all periodic table atoms and finding utility in various applications. These force fields have been developed in different iterations, such as CHARMM19, CHARMM22, CHARMM27, GROMOS96, GROMOS45A3, GROMOS53A5, GROMOS53A6, AMBER91, AMBER94, AMBER96, AMBER99, AMBER02, and more.¹⁰⁹ When investigating proteins and small molecules, it is crucial to ensure compatibility between the force field employed and the molecular components under investigation.

1.8.2 Root Mean Square Deviation (RMSD)

RMSD is a commonly employed quantitative metric for assessing the similarity between two sets of 3D atomic coordinates in an overlay. It can be calculated and displayed for various types and subsets of atoms, such as all carbon atoms in a protein, all atoms in a molecule, all residues in a protein, or all atoms when a protein is bound to a ligand. Studies have shown that folded regions exhibit more excellent stability, while loops tend to be more flexible, contributing to a reduction in the overall RMSD of the system.¹¹⁰

1.8.3 Binding Free Energy

Various methods, each with different levels of complexity, have been utilized to estimate binding free energy in biological macromolecular systems such as proteins.¹¹¹ Simplifying scoring methods is essential to achieving the requisite competency when searching enormous chemical databases of tiny compounds for a hit molecule that could one day serve as a lead or treatment candidate. The continuum solvent approximation may determine the binding free energy by assuming quadratic fluctuations around a particular configuration. It combines configurations obtained from molecular dynamics (MD) simulations conducted in explicit solvents with free energy estimators based on an implicit continuous solvent model. The MMPB-SA technique enables the determination of various types of free energies associated with biomolecules, including binding, polar, and non-polar free energies. The g-mmpbsa tool and several non-polar solvation models can be used to determine the binding free energy of protein-ligand complexes. One of these models, the repulsive model, considers how residue energy affects binding energy and the solvent-accessible surface area, volume, and other variables.

The MD trajectories that GROMACS generates are compatible with the program g-mmpbsa.¹¹² Use the following formulae to compute the binding free energy: The binding free energy is shown when an inhibitor forms a compound with a protein.

$$\Delta G_{bind} = G_{complex} - (G_{free-protein} + G_{free-inhibitor}) \quad (Eq. 1.4)$$

Here, $G_{free-protein}$ and $G_{free-inhibitor}$ represent the individual protein and inhibitor when they are individually present in the solvent. On the other hand, $G_{complex}$ refers to the combined free energy of the protein-inhibitor complex. As previously mentioned, the free energy of each separate entity is represented by the symbol G .

$$G = E_{MM} - TS + G_{solvation} \quad (Eq. 1.5)$$

The abbreviation TS , T , and S represents the contribution of entropy to the free energy in a vacuum, temperature, and entropy, respectively. The free energy connected to the solvation process is referred as $G_{solvation}$. Energy is needed to transport a solute from a vacuum into a solvent. The term refers to both the electrostatic and non-electrostatic parts of the free energy involved in solvation, denoted as G_{polar} and $G_{non-polar}$, respectively.

$$G_{solvation} = G_{polar} - (G_{free-protein} + G_{free-inhibitor}) \quad (Eq. 1.6)$$

The total potential energy of a molecule in a closed system is included in the E_{MM} , or Energy of Molecular Mechanics. Both bonded and non-bonded interactions are included, and the computation is based on the molecular mechanics (MM) force field parameters.

$$E_{MM} = E_{bonded} + E_{non-bonded} = E_{bonded} + (E_{vdW} + E_{elec}) \quad (Eq. 1.7)$$

The E_{bonded} is the interactions within the system, including bond, angle, dihedral, improper interactions, and non-bonded interactions. often called $E_{non-bonded}$. Particle interactions are influenced by several factors, including van der Waals, and electrostatic forces are typically simulated in molecular systems using the Coulomb and Lennard-Jones (LJ) potential functions.

1.9 Artificial Intelligence (AI)-Based Drug Design Methods

Artificial intelligence (AI) refers to computer algorithms that can learn from data and improve a task over time. These models were created using machine learning (ML), a branch of artificial intelligence. AI algorithms are available in a wide variety, including reinforcement learning, unsupervised learning, and supervised learning. In supervised learning, the Algorithm is improved by using labeled data, where both the input and output are known, to enhance its capacity to predict novel, unknown data. With reinforcement learning, an algorithm can take actions that maximize rewards in a particular environment. Table 1.5 briefly overviews four popular AI models, including their descriptions and some examples of common use cases.

Table 1.5 summary of the AI-based CADD models and their applications.

AI Model	Description	Use Cases
Convolutional Neural Network (CNN)	A type of neural network well-suited for image classification tasks due to its ability to identify patterns and features in images.	Image recognition, object detection, and medical image analysis.
Recurrent Neural Network (RNN)	A type of neural network well-suited for processing sequential data, such as text or speech. It uses a feedback loop to allow information to persist over time.	Natural language processing, speech recognition, and time series prediction.
Random Forest	An ensemble learning algorithm that combines multiple decision trees to make predictions. It is often used for classification and regression tasks.	Credit risk analysis, fraud detection, and predicting customer churn.
Support Vector Machine (SVM)	A type of machine learning algorithm well-suited for classification tasks, mainly when dealing with complex data. Finding the best hyperplane to separate the data into different classes works.	Image classification, text classification, and gene expression analysis.

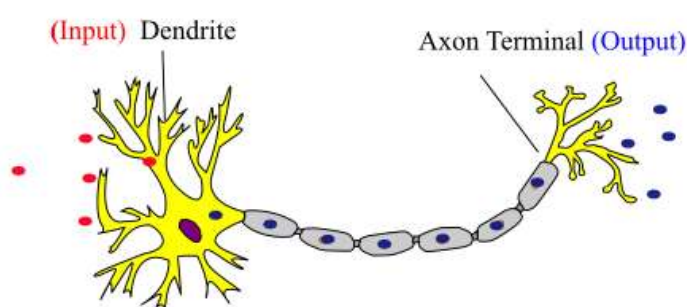


Figure 1.4 Biological neuron cell A comparison of the NNs.

Artificial neural networks (ANNs) are used in drug research and development. ANN may be able to structurally and functionally mimic the capacities of the human brain. Numerous neurons in the brain can store, retrieve, and link bits of information. It can recognize patterns based on previous training and knowledge. The human brain contains an estimated 100 billion or more neurons, and these neurons are interconnected through approximately 100 trillion synaptic connections. The capacity of the brain to communicate with its enormous number of neurons, which is required for a meaningful interpretation of the information, determines how it processes information¹¹³. The dendrites, cell body, axon, and synapses of a biological neuron collaborate to process information, transfer information, and receive signals in Figure 1.4.



Figure 1.5 Artificial neuron cell in ANN's.

1.9.1 AI in Drug Discovery and Applications

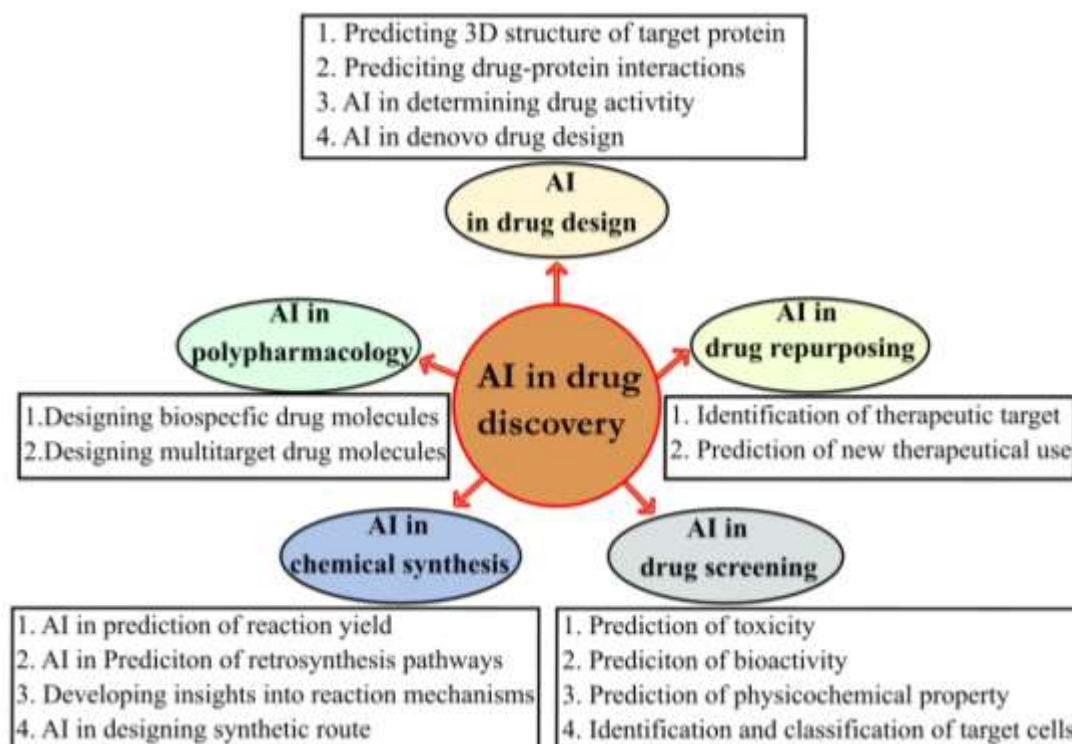


Figure 1.6 Various applications of AI in drug discovery.

The power of the brain lies in its ability to transmit information across vast networks of neurons rapidly within seconds. Both acquired skill sets and inherited characteristics significantly affect the ability of ones ability to receive and comprehend newly acquired information. The knowledge of neuroscience has improved with the mapping of the brain and the comprehension of the functions of different neurons. However, the exact mechanism of the brain is still unknown, and a computer has not entirely simulated it. With the advancements in computing power, there is ongoing development of networks that have the potential to function like the human brain.

The input, hidden, and output layers comprise the three artificial neuron layers.¹¹⁴ The input layer selects the input variable, which is later modified by the hidden layer before becoming calculated into the ultimate output values at the output layer in Figure 1.5.

Finding and developing novel drugs is challenging, expensive, and time-consuming. Typically, the R&D cycle lasts 10 to 15 years. Pharmaceutical companies invest significant money in drug research, yet the high attrition rate lowers R&D productivity. Only one in ten potential drug compounds advances beyond phase I clinical trials to receive regulatory approval. There is still a need for a new drug.^{115,116} The cost burden and time constraint of researching and creating novel pharmaceuticals requires significant financial resources. It is a time-intensive endeavor, and the pharmaceutical sector has increasingly resorted to AI to address these issues.

AI provides a range of tools and technologies that can speed up the validation of therapeutic targets, uncover possible hit and lead compounds, and improve drug design. AI thus has the potential to drastically cut the expense and time associated with finding new medicines, which would be advantageous for the healthcare sector.¹¹⁷

Table1.6 Some of the artificial intelligence tools employed in the drug discovery processes.

Tool	Details
DeepChem ¹¹⁸	A Python-based AI tool for various drug discovery task predictions
DeepNeuralNet-QSAR ¹¹⁹	Molecular activity predictions
DeepTox ¹²⁰	Toxicity predictions
Neural Graph Fingerprints ¹²¹	Property prediction of novel molecules
ORGANIC ¹²²	A practical approach for generating molecules with desired properties
Potential Net ¹²³	Ligand-binding affinity prediction based on a graph convolutional neural network

The chemical configuration required to trigger the desired reaction at the point of interest can be predicted using various computational techniques. These methods can also aid structural refinement to achieve multiple goals, including potency, safety, solubility, permeability, and synthesizability. The physical characteristics of the drug can also be predicted, and the synthesis procedure can be planned using computational approaches.^{124,125} One way to accelerate the elimination of non-lead compounds is to utilize all the available information, including structure- and ligand-based techniques. For instance, quantitative structure-activity relationship (QSAR) modeling has been implemented to select promising drug

candidates from a pool of up to a million options. Additionally, deep learning technology can now handle the vast amount of data produced during drug discovery and development, which has expanded due to the increase in big data in recent years.^{126,127} As seen in Figure 1.7, there are several ways that AI applications might speed up the drug development process.

A computational network consisting of one or multiple layers of artificial neurons (ANs) is commonly referred to as an artificial neural network (ANN) (see Figure 1.7). ANNs draw inspiration from the complex network of nerve cells in the human central nervous system. These networks aim to mimic the functioning of the human brain in performing specific tasks or functions of interest. In practice, ANNs can be implemented using electrical components or simulated through digital computer software, as Haykin et al. described.¹²⁸

Table1.7 Application of AI-based CADD methods in drug discovery.

Category	Subtitles	Description
1) AI in Drug Design	a) Predicting the 3D Structure of the Target Protein ¹²⁹	1) AI algorithms utilize protein sequence analysis and computational models to predict target protein structures, enhancing drug design efficiency. 2) By identifying potential binding sites and interactions, AI techniques revolutionize drug discovery, optimizing outcomes.
	b) Predicting Drug-Protein Interactions ¹³⁰	1) AI techniques, like machine learning and data analysis, predict drug-target interactions, improving therapy development. 2) Leveraging AI enables safer and more effective drugs by minimizing adverse effects through enhanced analysis.
	c) AI in Determining Drug Activity ¹³¹	1) AI techniques enhance drug compound evaluation and forecasting, optimizing research. 2) Efficient identification and prioritization of promising candidates expedite pharmaceutical development.
	d) AI in De Novo Drug Design ¹³²	1) AI algorithms are revolutionizing drug design by creating novel molecules. 2) Optimized characteristics of AI-designed drugs offer potential medical breakthroughs.
2) AI in Drug Repurposing	a) Identification of Therapeutic Use ¹³³	1) AI analysis identifies untapped therapeutic benefits in existing drugs. 2) Repurposing drugs with AI enhances patient care and treatment options.

	b) Predicting New Therapeutic Use ¹³⁴	1) AI analysis of molecular characteristics can unveil new therapeutic possibilities. 2) Harnessing AI's potential enables repurposing opportunities in medical research.
3) AI in Drug Screening	a) Prediction of Toxicity ¹³⁵	1) AI analysis of drug molecules enables the identification of novel therapeutic uses. 2) Harnessing the power of AI unveils repurposing opportunities in medicine.
	b) Prediction of Bio-Activity ¹³⁶	1) AI algorithms enhance drug discovery by predicting biological efficacy accurately. 2) Target identification and selection benefit from AI-powered forecasting technology.
	c) Prediction of Physicochemical Properties ¹³⁷	1) AI-driven predictive algorithms enhance drug-likeness molecules properties and optimization. 2) Pharmaceutical R&D experiences a transformative revolution through AI-powered technologies.
	d) Identification and Classification of Target Cells ¹³⁸	1) Leveraging advanced AI techniques to identify and categorize target cells accurately, revolutionizing drug targeting and treatment efficacy.
4) AI in Chemical Synthesis	a) AI Prediction of Reaction Yield ¹³⁹	1) AI techniques to predict the yield of chemical reactions, aiding in synthesis planning.
	b) AI Prediction of Retro Synthesis Pathways ¹⁴⁰	2) AI techniques to predict and suggest retro-synthesis pathways for efficiently synthesizing target molecules.
	c) Developing Reaction Mechanisms ¹⁴¹	1) AI algorithms to develop reaction mechanisms for chemical synthesis, improving efficiency.
	d) AI in Designing Synthetic Route ¹⁴²	1) AI methods to design optimal synthetic routes for efficiently producing target molecules.
5) AI in Polypharmacology	a) Designing Bio-Specific Drug Molecules ¹⁴³	1) AI algorithms to design drug molecules that target multiple biological pathways for improved efficacy.
	b) Designing Synthetic Route ¹⁴⁴	1) AI techniques to design synthetic routes for producing polypharmacological drugs.

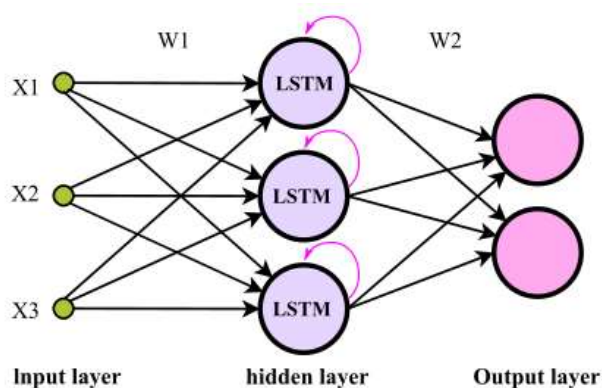


Figure 1.7 The architecture of an artificial neural network with input, hidden, and output layers.

The three layers commonly present in ANNs are input, output, and hidden layers. The initial layer and input point of the ANN are the input layer. The initial environment input is passed to the following layer in this layer for processing. The hidden layers come next after the input layer. ANNs in the buried layers modify it based on the information they receive. One of several factors that influence the transformation process is the nature of the activation function. However, ANNs can have several hidden layers. An ANN's output layer is its top layer. This layer determines the final result of the network.

1.9.2 Feed-Forward Neural Networks

Feed-forward neural network (FFNN) creates the output by transmitting the input signal from the outside environment through the network. The final output of an FFNN is computed for any input pattern using a single forward pass through the network, beginning at the input layer and ending at the output layer. There are no feedback loops between levels and the layers before them in FFNNs.¹¹⁷

1.9.2.1 Backpropagation Learning Algorithm

Backpropagation is one of the well-known techniques used in FNNs.¹⁴⁵ At the start of the learning process, the candidate weights and biases of the network are usually initialized with random values. These candidate weights and biases are updated or modified through a series of learning iterations. Epochs are employed to explain each learning iteration in backpropagation. An epoch typically has two phases:

1. Forward propagation: The input samples are fed into the FFNN, which then determines the actual output of the network, as explained previously.
2. Backward propagation: During this stage, the network error is computed by utilizing a loss function. The difference between the actual output and the output values desired for the input samples is an error of the network. The learning algorithm selects a set of weights and biases with the least error with the aid of the loss function, which incorporates a performance parameter. In the training phase of the study, the estimated error value is obtained by propagating from the output layer back to the input layer. This error value is then utilized to adjust the weights and biases of the neurons in the network. The mean squared error (MSE) was employed in this study as the loss function.

1.9.3 Recurrent Neural Networks (RNN)

RNN are a valuable network for processing temporal input and sequences. RNN repeats sequences while storing all object-specific data in a hidden state and keeping track of every item they have observed. The

compact form of a basic RNN is presented on the left side of Figure 1.8, while the form stretched across all time steps is displayed on the right side. The network's hidden state at time step t is determined using the activation function F to calculate the input X_t and the hidden state from the previous time step, h_{t-1} . The influence of the previous time step on the following one is preserved by the RNN's internal memory, which is made possible by this calculating method.

1.9.3.1 Types of RNN

RNN are more flexible than feed-forward networks, which only have a single input and output since the lengths of the inputs and outputs may be changed. RNNs are advantageous for modeling sequential data because of their flexibility.

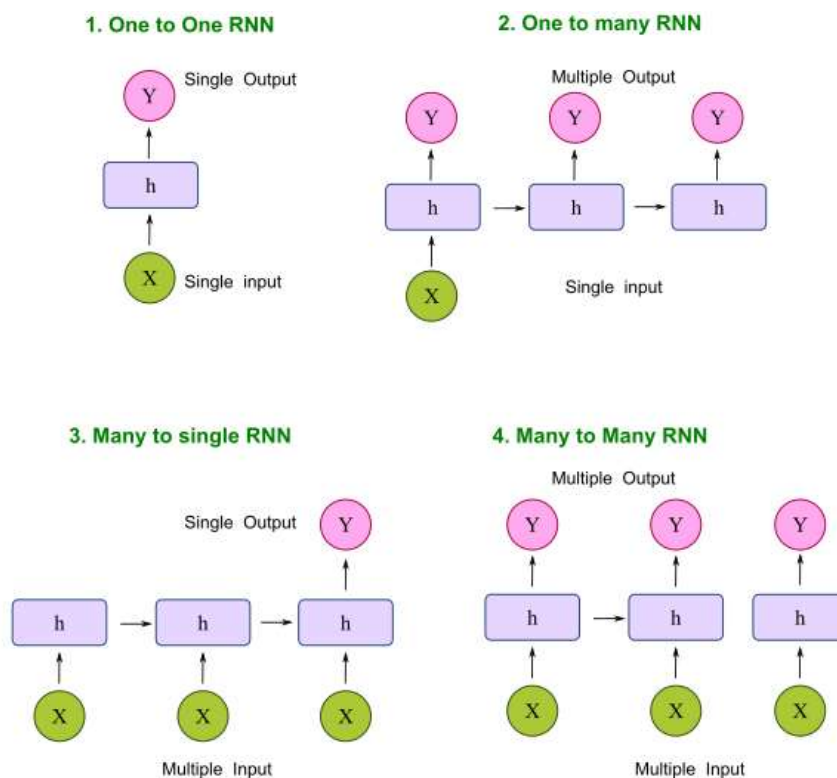


Figure 1.9 The inputs and outputs of four types of RNNs.

There are four main types of RNN, each with a different range of input and output lengths.

- One-to-one RNN is a primary neural network. It is widely applied to machine learning problems with only one input and output.
- One-to-many RNN one input and several outputs.
- Many-to-one RNN predicts one output from a variety of multiple inputs. It is a typical approach in sentiment classification, where text is the input and a category is the output.
- The term "many-to-many RNN" refers to RNN with many inputs and outputs.

1.9.3.2 Limitations of RNN

Simple RNN models often have two significant problems. The *gradient*, the slope of both the loss and error functions, relates to these issues.

1. Vanishing Gradient problems occur when the gradient becomes so tiny that altering the parameters is no longer advantageous; at this point, the Algorithm reaches a point of no return.
2. The *exploding gradient* problem arises when the gradient reaches an excessive size, rendering the model unstable. In this scenario, larger error gradients accumulate, resulting in too high model weights. This issue could lead to longer training times and lower model performance.

Reducing the number of hidden layers in the neural network is a straightforward way to address these problems, which will also simplify RNNs somewhat. Advanced RNN designs, such as the LSTM, can resolve these problems.

1.9.4 Long short-term memory (LSTM)

LSTM cells are particular building blocks employed by the long-short-term memory (LSTM) neural network, a type of RNN (Figure 1.9). The two hidden states calculated and stored by LSTM cells are cell state c and cell output h . Disappearing gradients is when networks lose the capacity to remember information from previous time steps. LSTM can deal with this issue. Conventional RNNs and extensive non-recurrent neural networks both frequently encounter this problem. The processing of older data remains relatively high when LSTM maintains the cell state c for many time steps (Figure 1.10).¹⁴⁶

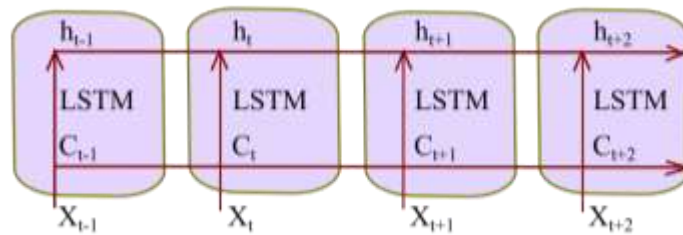


Figure 1.10 A cartoon picture of a simple LSTM neural network.

LSTM cells have an inbuilt computation system called gates to calculate the cell and output states. The input gate, forget gate, output gate, and input modulation gate are the four gates that make up an LSTM cell (Figure 1.11). These gates control O_w of data information that goes through the neural network and cell membrane.¹⁴⁷ It is computed using the input gate at time step t .

$$i_t = \sigma(w_i x_t + U_i h_{t-1} + b_i) \quad (\text{Eq. 1.8})$$

The forget gate f_t is determined by

$$f_t = \sigma(w_f x_t + U_f h_{t-1} + b_f) \quad (Eq. 1.9)$$

The output gate O_t is determined by

$$O_t = \sigma(w_o x_t + U_o h_{t-1} + b_o) \quad (Eq. 1.10)$$

To determine the cell state c , the input modulation gate ($\tilde{c}_t$) is determined by

$$\tilde{c}_t = \tanh(w_c x_t + U_c h_{t-1} + b_c) \quad (Eq. 1.11)$$

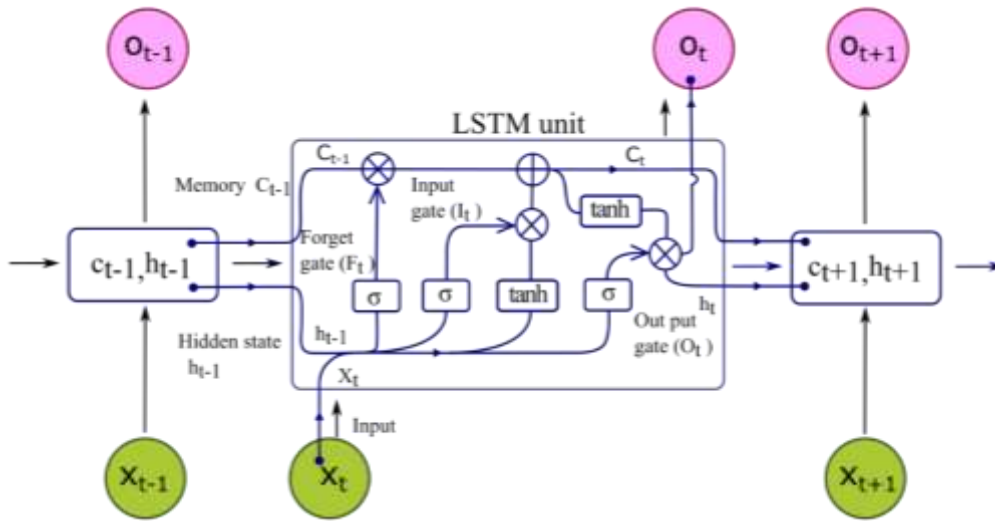


Figure 1.11 Detailed cartoon picture of the LSTM neural network.

Where x_t is the input at time t and h_{t-1} is the output state at the last time. The four gates' respective input x_t vectors are represented by the weight matrices step w_i , step, w_f step w_o , and step w_c . U_i, U_f, U_o and U_c make up the weight matrices for the four gates' previous output state, h_{t-1} , whereas b_i, b_f, b_o and b_c are the bias terms.

1.10 List of Diseases

1.10.1 Overview of Alzheimer's Disease (AD)

AD is a progressive neurological condition that deteriorates over time. It is characterized by notable alterations in the brain, resulting in the accumulation of specific proteins. As the disease advances, it leads to the shrinking of the brain and the eventual death of brain cells. The most common cause of dementia, a progressive decline in cognitive abilities, including memory, reasoning, behavior, and interpersonal skills, is AD. Particular changes significantly impact the ability of an individual to carry out daily activities. Approximately 6.5 million individuals aged 65 and above live with AD in the United States. Among them, over 70% are 75 years old or older. On a global scale, of the estimated 55 million people affected by dementia, around 60% to 70% are believed to have AD.¹⁴⁸ In the initial stages of AD, individuals may experience initial signs such as difficulty remembering recent events or conversations. As the condition progresses, it leads to more severe memory issues and the inability to carry out everyday tasks. While certain medications can potentially improve or slow down the advancement of symptoms, it's important to note that there is currently no known cure for AD.

Nevertheless, there are various programs and services available to provide support to individuals with the disease as well as their caregivers. In the advanced stages of the condition, the substantial loss of brain function can give rise to complications like dehydration, malnutrition, or infections, which, in some cases, can lead to death.¹⁴⁹ Acetylcholinesterase (AChE) plays a crucial role in the human body as an enzyme that breaks down the neurotransmitter acetylcholine. Acetylcholine is involved in transmitting signals between nerve cells and is particularly important for cognitive functions, including memory and learning. In drug design, AChE inhibitors play a significant role. This requires a thorough understanding of the structure and mechanism of enzymes. Scientists employ various strategies in Silico ADME properties, such as computer-aided drug design and molecular modeling, to design and optimize inhibitors with improved specificity and potency.

1.10.2 Overview of COVID-19

The Coronaviridae family of viruses, which includes the large group of viruses collectively known as coronaviruses, can infect a variety of animal species in addition to humans. While certain coronaviruses are known to cause mild respiratory illnesses that resemble the common cold, other coronaviruses have been associated with more deadly diseases, including SARS and Middle East Respiratory Syndrome (MERS). In December 2019, Wuhan, China, saw the emergence of a brand-new coronavirus that had never been observed in humans.¹⁵⁰ This coronavirus infection typically results in respiratory problems, fever, coughing, and breathing issues. It could result in pneumonia, severe acute respiratory syndrome, or even death in severe cases. The World Health Organisation (WHO) is aggressively collaborating with international scientists, governments, and partner organizations in order to fast expand our scientific

understanding of this unique virus. Their combined efforts are meant to provide quick guidance on precautionary measures to protect public health and halt the outbreak from spreading. The rights, responsibilities, and functions of health professionals must also be addressed. This includes workplace health and safety. Frontline healthcare workers play a crucial role in responding to outbreaks and are often exposed to various hazards that increase their susceptibility to infection with the pathogen responsible for the outbreak.¹⁵¹

1.10.3 Overview of Parkinson's Disease (PD)

PD is a progressive neurological condition distinguished by various motor symptoms, including slowed movement, tremors, muscle stiffness, difficulties with walking and balance, and involuntary movements. People with PD commonly experience a variety of non-motor problems in addition to motor abnormalities. These include discomfort, sensory issues, sleep disruptions, mental health conditions, and cognitive impairment. The motor symptoms associated with PD, such as dyskinesia (involuntary movements) and dystonia (painful involuntary muscle contractions), significantly impact speech, mobility, and overall functioning, leading to limitations in daily life activities. As the disease progresses, these symptoms worsen, resulting in a high incidence of disability and the need for increased care. Moreover, it is common for individuals with PD to develop dementia as the disease advances. Dementia in PD can manifest as cognitive decline, memory problems, and difficulties with thinking and reasoning.¹⁵² In addition to PD, other movement disorders such as multiple system atrophy, progressive supranuclear palsy, chorea, ataxia, and dystonia are also known to exist. Tremors, sluggish movement, and muscle rigidity are signs of several movement disorders that may resemble Parkinson's disease (PD). All movement disorders experience similar challenges in effectively identifying the ailment and providing appropriate treatment, much like PD.

Access to medication for these disorders is particularly limited in low- and middle-income countries (LMIC). While increasing age is a significant risk factor for developing PD, it is essential to note that younger individuals can also be affected. Furthermore, PD tends to affect men more frequently than women. Although the precise causes of Parkinson's disease (PD) are still not fully known, it is commonly accepted that a combination of hereditary factors and lifetime exposure to environmental variables contribute to the development of the disease. Some of these environmental factors include pesticides, solvents, and air pollution.¹⁵³

1.10.4 Overview of Tuberculosis Disease (TB)

TB is a bacterial infection caused by *Mycobacterium tuberculosis*, primarily affecting the lungs. When people with lung TB cough, sneeze, or discharge saliva, the disease is spread through the air. Infection can result from breathing in even a tiny number of these microorganisms. Annually, approximately 10 million people worldwide develop tuberculosis. Despite being a preventable and treatable condition, TB remains the top infectious cause of death, claiming the lives of around 1.5 million individuals annually. TB has a particularly devastating impact on individuals living with HIV, being the leading cause of death among this population.

Table 1.8 Disease, target proteins, its mechanisms, and PDB ID of the target protein.

Disease	Protein	Mechanism	PDB ID
Alzheimer's	Acetylcholinesterase (AChE)	Inhibition of AChE activity to increase ACh levels in the brain.	1EVE
COVID-19	Main protease (Mpro)	Inhibition of M ^{pro} activity to prevent viral replication.	6LU7
Parkinson's	Monoamine oxidase B (MAO-B)	Inhibition of MAO-B activity to increase dopamine levels in the brain.	2V5Z
Tuberculosis	Ribosomal RNA S1	Inhibition of ribosomal RNA S1, which is involved in the translation of mRNA into protein.	4NNI

Additionally, TB plays a significant role in contributing to the problem of antimicrobial resistance. While TB affects people globally, most cases occur in low- and middle-income countries.¹⁵⁴ Approximately 50% of TB cases are concentrated in eight countries: Bangladesh, China, India, Indonesia, Nigeria, Pakistan, the Philippines, and South Africa. These nations bear a significant burden of TB infections, contributing to a substantial portion of the global TB caseload. Roughly a quarter of the world's population is believed to have been exposed to TB bacteria. However, it is essential to note that not everyone infected will progress to active TB disease. Some individuals will naturally clear the infection, and those infected but not yet showing symptoms cannot transmit the disease to others. Individuals who have been infected with TB bacteria face a lifetime risk of approximately 5-10% of developing TB disease. However, certain factors can increase this risk. Those with compromised immune systems, such as individuals living with viral infections, malnutrition, diabetes, or those who engage in tobacco use, are more susceptible to falling ill with TB.¹⁵⁵

REFERENCES

1. Rang, H.; Hill, R. Drug Development: Introduction, Drug Discovery and Development-E-Book. *Technology in Transition* **2012**.
2. DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *J. Health Econ.* **2016**, *47*, 20–33.
3. Deore, A. B.; Dhumane, J. R.; Wagh, R.; Sonawane, R. The Stages of Drug Discovery & Development Process. *Asian J. Pharm. Res. Dev.* **2019**, *7* (6), 62–67.
4. Holien, J. K.; Coff, L.; Guy, A. J.; Boer, J. C. Drug Discovery in Real Life: An Online Learning Activity for Bioinformatics Students. *J. Chem. Educ.* **2023**.
5. Ragno, R.; Esposito, V.; Di Mario, M.; Masiello, S.; Viscovo, M.; Cramer, R. D. Teaching and Learning Computational Drug Design: Student Investigations of 3D Quantitative Structure-Activity Relationships through Web Applications. *J. Chem. Educ.* **2020**, *97* (7), 1922–1930.
6. Yang, X.; Wang, Y.; Byrne, R.; Schneider, G.; Yang, S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chem. Rev.* **2019**, *119* (18), 10520–10594.
7. Subbiah, V. The next Generation of Evidence-Based Medicine. *Nat. Med.* **2023**, *29* (1), 49–58.
8. Lutten, A.; Gullberg, H.; Abdurakhmanov, E.; Vo, D. D.; Akaberi, D.; Talibov, V. O.; Nekhotiaeva, N.; Vangeel, L.; De Jonghe, S.; Jochmans, D.; Krambrich, J.; Tas, A.; Lundgren, B.; Gravenfors, Y.; Craig, A. J.; Atilaw, Y.; Sandström, A.; Moodie, L. W. K.; Lundkvist, Å.; van Hemert, M. J.; Neyts, J.; Lennerstrand, J.; Kihlberg, J.; Sandberg, K.; Danielson, U. H.; Carlsson, J. Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J. Am. Chem. Soc.* **2022**, *144* (7), 2905–2920.
9. Maghsoudi, S.; Taghavi Shahraki, B.; Rameh, F.; Nazarabi, M.; Fatahi, Y.; Akhavan, O.; Rabiee, M.; Mostafavi, E.; Lima, E. C.; Saeb, M. R.; Rabiee, N. A Review on Computer-Aided Chemogenomics and Drug Repositioning for Rational COVID-19 Drug Discovery. *Chem. Biol. Drug Des.* **2022**, *100* (5), 699–721.
10. Khataniar, A.; Pathak, U.; Rajkhowa, S.; Jha, A. N. A Comprehensive Review of Drug Repurposing Strategies against Known Drug Targets of COVID-19. *COVID* **2022**, *2* (2), 148–167.
11. Muratov, E. N.; Amaro, R.; Andrade, C. H.; Brown, N.; Ekins, S.; Fourches, D.; Isayev, O.; Kozakov, D.; Medina-Franco, J. L.; Merz, K. M.; Oprea, T. I.; Poroikov, V.; Schneider, G.; Todd, M. H.; Varnek, A.; Winkler, D. A.; Zakharov, A. V.; Cherkasov, A.; Tropsha, A. A Critical Overview of Computational Approaches Employed for COVID-19 Drug Discovery. *Chem. Soc. Rev.* **2021**, *50* (16), 9121–9151.
12. Yamurai Bishi, L.; Chaitanya Vedithi, S.; L. Blundell, T.; Chitima Mugumbate, G. Computational Deorphaning of Mycobacterium Tuberculosis Targets. In *Drug Discovery and Development - New Advances*; IntechOpen, **2020**.
13. Mihović, N.; Tomašević, N.; Matić, S.; Mitrović, M. M.; Kostić, D. A.; Sabatino, M.; Antonini, L.; Ragno, R.; Mladenović, M. Human Estrogen Receptor α Antagonists. Part 1: 3-D QSAR-Driven Rational Design of Innovative Coumarin-Related Antiestrogens as Breast Cancer Suppressants through Structure-Based and Ligand-Based Studies. *J. Chem. Inf. Model.* **2021**, *61* (10), 5028–5053.
14. Ajiboye, B. O.; Iwaloye, O.; Owolabi, O. V.; Ejeje, J. N.; Okerewa, A.; Johnson, O. O.; Udebor, A. E.; Oyinloye, B. E. Screening of Potential Antidiabetic Phytochemicals from Gongronema Latifolium Leaf

- against Therapeutic Targets of Type 2 Diabetes Mellitus: Multi-Targets Drug Design. *SN Appl. Sci.* **2022**, 4 (1).
15. Elkouzi, A.; Vedam-Mai, V.; Eisinger, R. S.; Okun, M. S. Emerging Therapies in Parkinson Disease - Repurposed Drugs and New Approaches. *Nat. Rev. Neurol.* **2019**, 15 (4), 204–223.
 16. Vemula, D.; Jayasurya, P.; Sushmita, V.; Kumar, Y. N.; Bhandari, V. CADD, AI and ML in Drug Discovery: A Comprehensive Review. *Eur. J. Pharm. Sci.* **2023**, 181 (106324), 1–23.
 17. Baig, M. H.; Ahmad, K.; Roy, S.; Ashraf, J. M.; Adil, M.; Siddiqui, M. H.; Khan, S.; Kamal, M. A.; Provazník, Choi, I. Computer Aided Drug Design: Success and Limitations. *Curr. Pharm. Des.* **2016**, 22 (5), 572–581.
 18. Wagner, J. R.; Lee, C. T.; Durrant, J. D.; Malmstrom, R. D.; Feher, V. A.; Amaro, R. E. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **2016**, 116 (11), 6370–6390.
 19. Talele, T. T.; Khedkar, S. A.; Rigby, A. C. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem.* **2010**, 10 (1), 127–141.
 20. Aulakh, G. K.; Sodhi, R. K.; Singh, M. An Update on Non-Peptide Angiotensin Receptor Antagonists and Related RAAS Modulators. *Life Sci.* **2007**, 81 (8), 615–639.
 21. Takahashi H.; Hayakawa I.; Akimoto T. The history of the development and changes of quinolone antibacterial agents. *Yakushigaku Zasshi* **2003**, 38 (2), 161–179.
 22. Zhu, C. Aldose Reductase Inhibitors as Potential Therapeutic Drugs of Diabetic Complications. In *Diabetes Mellitus - Insights and Perspectives*; InTech, **2013**.
 23. Supuran, C. T. Advances in Structure-Based Drug Discovery of Carbonic Anhydrase Inhibitors. *Expert Opin. Drug Discov.* **2017**, 12 (1), 61–88.
 24. De Clercq, E. The History of Antiretrovirals: Key Discoveries over the Past 25 Years. *Rev. Med. Virol.* **2009**, 19 (5), 287–299.
 25. Van Drie, J. H. Computer-Aided Drug Design: The next 20 Years. *J. Comput. Aided Mol. Des.* **2007**, 21 (10–11), 591–601.
 26. Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal Structure of HIV-1 Protease in Complex with VX-478, a Potent and Orally Bioavailable Inhibitor of the Enzyme. *J. Am. Chem. Soc.* **1995**, 117 (3), 1181–1182.
 27. Aruksakunwong, O.; Promsri, S.; Wittayanarakul, K.; Nimmanpipug, P.; Lee, V.; Wijitkosoom, A.; Sompornpisut, P.; Hannongbua, S. Current Development on HIV-1 Protease Inhibitors. *Curr. Comput. Aided Drug Des.* **2007**, 3 (3), 201–213.
 28. Best, B. M.; Goicoechea, M. Efavirenz – Still First-Line King? *Expert Opin. Drug Metab. Toxicol.* **2008**, 4 (7), 965–972.
 29. Ghosh, A. K.; Dawson, Z. L.; Mitsuya, H. Darunavir, a Conceptually New HIV-1 Protease Inhibitor for the Treatment of Drug-Resistant HIV. *Bioorg. Med. Chem.* **2007**, 15 (24), 7576–7580.
 30. Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a Novel Binding Trench in HIV Integrase. *J. Med. Chem.* **2004**, 47 (8), 1879–1881.
 31. Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. *Science* **2000**, 289 (5486), 1938–1942.
 32. Fischer, J.; Robin Ganellin, C. *Analogue-Based Drug Discovery*; LibreDigital, 2006.

33. Sun, C. L.; Christensen, J. G.; McMahon, G. Discovery and Development of Sunitinib (SU11248): A Multitarget Tyrosine Kinase Inhibitor of Tumor Growth, Survival, and Angiogenesis. In *Kinase Inhibitor Drugs*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2009; pp 1–39.
34. Cohen, N. C. Structure-Based Drug Design and the Discovery of Aliskiren (Tekturna): Perseverance and Creativity to Overcome a R&D Pipeline Challenge: Discovery of Aliskiren. *Chem. Biol. Drug Des.* **2007**, *70* (6), 557–565.
35. Rutenber, E. E.; Stroud, R. M. Binding of the Anticancer Drug ZD1694 to E. Coli Thymidylate Synthase: Assessing Specificity and Affinity. *Structure* **1996**, *4* (11), 1317–1324.
36. Njoroge, F. G.; Chen, K. X.; Shih, N.-Y.; Piwinski, J. J. ChemInform Abstract: Challenges in Modern Drug Discovery: A Case Study of Boceprevir, an HCV Protease Inhibitor for the Treatment of Hepatitis C Virus Infection. *ChemInform* **2008**, *39* (17).
37. Rao, B. G.; Murcko, M.; Tebbe, M. J.; Kwong, A. D. Discovery and Development of Telaprevir (IncivekTM) - A Protease Inhibitor to Treat Hepatitis C Infection. In *Successful Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2015; pp 195–212.
38. Harper, S.; McCauley, J. A.; Rudd, M. T.; Ferrara, M.; DiFilippo, M.; Crescenzi, B.; Koch, U.; Petrocchi, A.; Holloway, M. K.; Butcher, J. W.; Romano, J. J.; Bush, K. J.; Gilbert, K. F.; McIntyre, C. J.; Nguyen, K. T.; Nizi, E.; Carroll, S. S.; Ludmerer, S. W.; Burlein, C.; DiMuzio, J. M.; Graham, D. J.; McHale, C. M.; Stahlhut, M. W.; Olsen, D. B.; Monteagudo, E.; Cianetti, S.; Giuliano, C.; Pucci, V.; Trainor, N.; Fandozzi, C. M.; Rowley, M.; Coleman, P. J.; Vacca, J. P.; Summa, V.; Liverton, N. J. Discovery of MK-5172, a Macrocyclic Hepatitis C Virus NS3/4a Protease Inhibitor. *ACS Med. Chem. Lett.* **2012**, *3* (4), 332–336.
39. Cui, J. J.; Tran-Dubé, M.; Shen, H.; Nambu, M.; Kung, P.-P.; Pairish, M.; Jia, L.; Meng, J.; Funk, L.; Botrous, I.; McTigue, M.; Grodsky, N.; Ryan, K.; Padrique, E.; Alton, G.; Timofeevski, S.; Yamazaki, S.; Li, Q.; Zou, H.; Christensen, J.; Mroczkowski, B.; Bender, S.; Kania, R. S.; Edwards, M. P. Structure Based Drug Design of Crizotinib (PF-02341066), a Potent and Selective Dual Inhibitor of Mesenchymal–Epithelial Transition Factor (c-MET) Kinase and Anaplastic Lymphoma Kinase (ALK). *J. Med. Chem.* **2011**, *54* (18), 6342–6363.
40. Reed, J. E.; Smaill, J. B. The Discovery of Dacomitinib, a Potent Irreversible EGFR Inhibitor. In *Comprehensive Accounts of Pharmaceutical Research and Development: From Discovery to Late-Stage Process Development Volume 1*; American Chemical Society: Washington, DC, 2016; pp 207–233.
41. Agrawal, R. The First Approved Agent in the Glitazar's Class: Saroglitazar. *Curr. Drug Targets* **2014**, *15* (2), 151–155.
42. White, A. W.; Almassy, R.; Calvert, A. H.; Curtin, N. J.; Griffin, R. J.; Hostomsky, Z.; Maegley, K.; Newell, D. R.; Srinivasan, S.; Golding, B. T. Resistance-Modifying Agents. 9. Synthesis and Biological Properties of Benzimidazole Inhibitors of the DNA Repair Enzyme Poly(ADP-Ribose) Polymerase. *J. Med. Chem.* **2000**, *43* (22), 4084–4097.
43. Abidi, A.; Shukla, P.; Ahmad, A. Lifitegrast: A Novel Drug for Treatment of Dry Eye Disease. *J. Pharmacol. Pharmacother.* **2016**, *7* (4), 194–198.
44. Hecker, S. J.; Reddy, K. R.; Totrov, M.; Hirst, G. C.; Lomovskaya, O.; Griffith, D. C.; King, P.; Tsvikovski, R.; Sun, D.; Sabet, M.; Tarazi, Z.; Clifton, M. C.; Atkins, K.; Raymond, A.; Potts, K. T.; Abendroth, J.; Boyer, S. H.; Loutit, J. S.; Morgan, E. E.; Durso, S.; Dudley, M. N. Discovery of a Cyclic

- Boronic Acid β -Lactamase Inhibitor (RPX7009) with Utility vs Class A Serine Carbapenemases. *J. Med. Chem.* **2015**, 58 (9), 3682–3692.
45. Murray, C. W.; Newell, D. R.; Angibaud, P. A Successful Collaboration between Academia, Biotech and Pharma Led to Discovery of Erdafitinib, a Selective FGFR Inhibitor Recently Approved by the FDA. *Medchem. Comm.* **2019**, 10 (9), 1509–1511.
 46. Gish, R. G.; Porta, C.; Lazar, L.; Ruff, P.; Feld, R.; Croitoru, A.; Feun, L.; Jeziorski, K.; Leighton, J.; Gallo, J.; Kennealey, G. T. Phase III Randomized Controlled Trial Comparing the Survival of Patients with Unresectable Hepatocellular Carcinoma Treated with Nilotrexed or Doxorubicin. *J. Clin. Oncol.* **2007**, 25 (21), 3069–3075.
 47. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, 28 (1), 235–242.
 48. Burley, S. K.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. *Methods Mol. Biol.* **2017**, 1607, 627–641.
 49. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Res.* **2021**, 49 (D1), D1388–D1395.
 50. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, 40 (Database issue), D1100–7.
 51. Hood, L.; Rowen, L. The Human Genome Project: Big Science Transforms Biology and medicine. *Genome Med.* **2013**, 5 (9), 79.
 52. UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, 49 (D1), D480–D489.
 53. Xu, D.; Zhang, Y. Ab Initio Protein Structure Assembly Using Continuous Structure Fragments and Optimized Knowledge-Based Force Field: QUARK Ab Initio Prediction Method. *Proteins* **2012**, 80 (7), 1715–1735.
 54. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **2018**, 46 (W1), W363–W367.
 55. Laskowski, R. A. SURFNET: A Program for Visualizing Molecular Surfaces, Cavities, and Intermolecular Interactions. *J. Mol. Graph.* **1995**, 13 (5), 323–330, 307–308.
 56. Lippi, M.; Passerini, A.; Punta, M.; Rost, B.; Frasconi, P. MetalDetector: A Web Server for Predicting Metal-Binding Sites and Disulfide Bridges in Proteins from Sequence. *Bioinformatics* **2008**, 24 (18), 2094–2095.
 57. Fährrolfes, R.; Bietz, S.; Flachsenberg, F.; Meyder, A.; Nittinger, E.; Otto, T.; Volkamer, A.; Rarey, M. ProteinsPlus: A Web Portal for Structure Analysis of Macromolecules. *Nucleic Acids Res.* **2017**, 45 (W1), W337–W343.
 58. Kozlovskii, I.; Popov, P. Spatiotemporal Identification of Druggable Binding Sites Using Deep Learning. *Commun. Biol.* **2020**, 3 (1), 618.
 59. Pathania, S.; Singh, P. K. Analyzing FDA-Approved Drugs for Compliance of Pharmacokinetic Principles: Should There Be a Critical Screening Parameter in Drug Designing Protocols? *Expert Opin. Drug Metab. Toxicol.* **2021**, 17 (4), 351–354.

60. Dias, R.; de Azevedo, W. F., Jr. Molecular Docking Algorithms. *Curr. Drug Targets* **2008**, 9 (12), 1040–1047.
61. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided Drug Des.* **2011**, 7 (2), 146–157.
62. Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. *J. Comput. Aided Mol. Des.* **2001**, 15 (5), 411–428.
63. Meiler, J.; Baker, D. ROSETTALIGAND: Protein-Small Molecule Docking with Full Side-Chain Flexibility. *Proteins* **2006**, 65 (3), 538–548.
64. Goodsell, D. S.; Olson, A. J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins* **1990**, 8 (3), 195–202.
65. Li, Z.; Gu, J.; Zhuang, H.; Kang, L.; Zhao, X.; Guo, Q. Adaptive Molecular Docking Method Based on Information Entropy Genetic Algorithm. *Appl. Soft Comput.* **2015**, 26, 299–302.
66. Katoch, S.; Chauhan, S. S.; Kumar, V. A Review on Genetic Algorithm: Past, Present, and Future. *Multimed. Tools Appl.* **2021**, 80 (5), 8091–8126.
67. Zheng, L.; Meng, J.; Jiang, K.; Lan, H.; Wang, Z.; Lin, M.; Li, W.; Guo, H.; Wei, Y.; Mu, Y. Improving Protein-Ligand Docking and Screening Accuracies by Incorporating a Scoring Function Correction Term. *Brief. Bioinform.* **2022**, 23 (3), 1–15.
68. Chen, H.-M.; Liu, B.-F.; Huang, H.-L.; Hwang, S.-F.; Ho, S.-Y. SODOCK: Swarm Optimization for Highly Flexible Protein-Ligand Docking. *J. Comput. Chem.* **2007**, 28 (2), 612–623.
69. Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: Pharmacophore-Based Protein-Ligand Docking. *J. Med. Chem.* **2004**, 47 (27), 6804–6811.
70. Wang, K.; Lyu, N.; Diao, H.; Jin, S.; Zeng, T.; Zhou, Y.; Wu, R. GM-DockZn: A Geometry Matching-Based Docking Algorithm for Zinc Proteins. *Bioinformatics* **2020**, 36 (13), 4004–4011.
71. Liu, M.; Wang, S. MCDOCK: A Monte Carlo Simulation Approach to the Molecular Docking Problem. *J. Comput. Aided Mol. Des.* **1999**, 13 (5), 435–451. <https://doi.org/10.1023/a:1008005918983>.
72. Trosset, J.-Y.; Scheraga, H. A. Prodock: Software Package for Protein Modeling and Docking. *J. Comput. Chem.* **1999**, 20 (4), 412–427.
73. Goodsell, D. S.; Sanner, M. F.; Olson, A. J.; Forli, S. The AutoDock Suite at 30. *Protein Sci.* **2021**, 30 (1), 31–43.
74. Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, 245 (1), 43–53.
75. Li, H.; Li, C.; Gui, C.; Luo, X.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. Gas Dock: A New Approach for Rapid Flexible Docking Based on an Improved Multi-Population Genetic Algorithm. *Bioorg. Med. Chem. Lett.* **2004**, 14 (18), 4671–4676.
76. Pei, J.; Wang, Q.; Liu, Z.; Li, Q.; Yang, K.; Lai, L. PSI-DOCK: Towards Highly Efficient and Accurate Flexible Ligand Docking. *Proteins* **2006**, 62 (4), 934–946.
77. Janson, S.; Merkle, D.; Middendorf, M. Molecular Docking with Multi-Objective Particle Swarm Optimization. *Appl. Soft Comput.* **2008**, 8 (1), 666–675.
78. Fogel, G. B.; Cheung, M.; Pittman, E.; Hecht, D. Modeling the Inhibition of Quadruple Mutant Plasmodium Falciparum Dihydrofolate Reductase by Pyrimethamine Derivatives. *J. Comput. Aided Mol. Des.* **2008**, 22 (1), 29–38.

79. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P. EHiTS: A New Fast, Exhaustive Flexible Ligand Docking System. *J. Mol. Graph. Model.* **2007**, *26* (1), 198–212.
80. Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for Molecular Docking: A Review. *Biophys. Rev.* **2017**, *9* (2), 91–102.
81. Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative Evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy. *Proteins* **2004**, *57* (2), 225–242. <https://doi.org/10.1002/prot.20149>.
82. Thomsen, R.; Christensen, M. H. MolDock: A New Technique for High-Accuracy Molecular Docking. *J. Med. Chem.* **2006**, *49* (11), 3315–3321.
83. Morris, G. M.; Lim-Wilby, M. Molecular Docking. Molecular Modeling of Proteins; *Humana Press*, 2008.
84. Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. HOOK: A Program for Finding Novel Molecular Architectures That Satisfy the Chemical and Steric Requirements of a Macromolecule Binding Site. *Proteins* **1994**, *19* (3), 199–221.
85. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
86. Clark, K. P.; Ajay. Flexible Ligand Docking without Parameter Adjustment across Four Ligand-Receptor Complexes. *J. Comput. Chem.* **1995**, *16* (10), 1210–1226.
87. Taylor, J. S.; Burnett, R. M. DARWIN: A Program for Docking Flexible Molecules. *Proteins* **2000**, *41* (2), 173–191.
88. Pei, J.; Wang, Q.; Liu, Z.; Li, Q.; Yang, K.; Lai, L. PSI-DOCK: Towards Highly Efficient and Accurate Flexible Ligand Docking. *Proteins* **2006**, *62* (4), 934–946.
89. Zhao, Y.; Sanner, M. F. FLIPDock: Docking Flexible Ligands into Flexible Receptors. *Proteins* **2007**, *68* (3), 726–737.
90. Stroganov, O. V. Lead Finder: An Approach to Improve Accuracy of Protein–Ligand Docking, Binding Energy Estimation, and Virtual Screening. *J. Chem. Inf. Model.* **2008**, *48* (12), 2371–2385.
91. Chen, R.; Li, L.; Weng, Z. ZDOCK: An Initial-Stage Protein-Docking Algorithm. *Proteins* **2003**, *52* (1), 80–87.
92. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: Servers for Rigid and Symmetric Docking. *Nucleic Acids Res.* **2005**, *33*, W363–367.
93. Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M.-D.; Ritchie, D. W. HexServer: An FFT-Based Protein Docking Server Powered by Graphics Processors. *Nucleic Acids Res.* **2010**, *38*, W445–9.
94. Xu, X.; Huang, M.; Zou, X. Docking-Based Inverse Virtual Screening: Methods, Applications, and Challenges. *Biophys. Rep.* **2018**, *4* (1), 1–16.
95. Vanommeslaeghe, K.; Yang, M.; MacKerell, A. D., Jr. Robustness in the Fitting of Molecular Mechanics Parameters. *J. Comput. Chem.* **2015**, *36* (14), 1083–1101. <https://doi.org/10.1002/jcc.23897>.
96. Guedes, I. A.; Pereira, F. S. S.; Dardenne, L. E. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges. *Front. Pharmacol.* **2018**, *9*, 1089.
97. Dias, R.; de Azevedo, W. F., Jr. Molecular Docking Algorithms. *Curr. Drug Targets* **2008**, *9* (12), 1040–1047.

98. Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. Monte Carlo Simulations for Proteins: Binding Affinities for Trypsin–Benzamidine Complexes via Free-Energy Perturbations. *J. Phys. Chem. B* **1997**, *101* (46), 9663–9669.
99. Li, H.; Sze, K.-H.; Lu, G.; Ballester, P. J. Machine-learning Scoring Functions for Structure-based Drug Lead Optimization. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10* (5).
100. Nguyen, D. D.; Wei, G.-W. AGL-Score: Algebraic Graph Learning Score for Protein-Ligand Binding Scoring, Ranking, Docking, and Screening. *J. Chem. Inf. Model.* **2019**, *59* (7), 3291–3304.
101. Maciej, P. J.; Ballester, P. Performance of Machine-Learning Scoring Functions in Structure-Based Virtual Screening. *Scientific Reports* **2017**, *7* (1), 1–10.
102. Li, H.; Sze, K.-H.; Lu, G.; Ballester, P. J. Machine-learning Scoring Functions for Structure-based Drug Lead Optimization. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10* (5).
103. Wang, H.; Zhou, Z.; Li, Y.; Chen, Z.; Lu, P.; Wang, W.; Liu, W.; Yu, L. Comparison of Machine Learning Methods for Classifying Mediastinal Lymph Node Metastasis of Non-Small Cell Lung Cancer from ¹⁸F-FDG PET/CT Images. *EJNMMI Research* **2017**, *7*.
104. Geng, H.; Chen, F.; Ye, J.; Jiang, F. Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 1162–1170.
105. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.
106. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.
107. Lindahl, E.; Hess, B. Van Der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *Molecular Modeling Annual.* **2001**, *7*, 306–317.
108. Chen, S. H. Spectroscopy in Biology and Chemistry: Neutron; X-Ray, Laser. A. Academic Press. 2017
109. Małolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. Symmetrization of the AMBER and CHARMM Force Fields. *J. Comput. Chem.* **2010**, *31* (7), 1402–1409.
110. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.
111. Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50* (7), 1625–1632.
112. Kumari, R.; Kumar, R. G_mmpbsa: GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54* (7), 1951–1962.
113. Subbulakshmi Radhakrishnan, S.; Sebastian, A.; Oberoi, A.; Das, S.; Das, S. A Biomimetic Neural Encoder for Spiking Neural Network. *Nat. Commun.* **2021**, *12* (1), 2143.
114. Akgün, E.; Demir, M. Modeling Course Achievements of Elementary Education Teacher Candidates with Artificial Neural Networks. *Int. J. Assess. Tools Educ.* **2018**, 491–509.
115. Martin, L.; Hutchens, M.; Hawkins, C.; Radnov, A. How Much Do Clinical Trials Cost? *Nat. Rev. Drug Discov.* **2017**, *16* (6), 381–382.
116. Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. An Analysis of the Attrition of Drug Candidates from Major Pharmaceutical Companies. *Nat. Rev. Drug Discov.* **2015**, *14* (7), 475–486.

117. Mak, K.-K.; Pichika, M. R. Artificial Intelligence in Drug Development: Present Status and Future Prospects. *Drug Discov. Today* **2019**, *24* (3), 773–780.
118. Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V. Deep Learning for the Life Sciences: Applying Deep Learning to Genomics; Microscopy, Drug Discovery, and More. *O'Reilly Media, Inc*, 2019.
119. Xu, Y.; Ma, J.; Liaw, A.; Sheridan, R. P.; Svetnik, V. Demystifying Multitask Deep Neural Networks for Quantitative Structure–Activity Relationships. *J. Chem. Inf. Model.* **2017**, *57* (10), 2490–2504..
120. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. DeepTox: Toxicity Prediction Using Deep Learning. *Toxicol. Lett.* **2017**, *280*, S69.
121. Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. arXiv [cs.LG], **2015**.
122. Sanchez-Lengeling, B.; Outeiral, C.; Guimaraes, G. L.; Aspuru-Guzik, A. Optimizing Distributions over Molecular Space. An Objective-Reinforced Generative Adversarial Network for Inverse-Design Chemistry (ORGANIC). *ChemRxiv*, **2017**.
123. Feinberg, E. N.; Sur, D.; Wu, Z.; Husic, B. E.; Mai, H.; Li, Y.; Sun, S.; Yang, J.; Ramsundar, B.; Pande, V. S. PotentialNet for Molecular Property Prediction. *ACS Cent. Sci.* **2018**, *4* (11), 1520–1530.
124. Sellwood, M. A.; Ahmed, M.; Segler, M. H.; Brown, N. Artificial Intelligence in Drug Discovery. *Future Med. Chem.* **2018**, *10* (17), 2025–2028.
125. Antonopoulos, I.; Robu, V.; Couraud, B.; Kirli, D.; Norbu, S.; Kiprakis, A.; Flynn, D.; Elizondo-Gonzalez, S.; Wattam, S. Artificial Intelligence and Machine Learning Approaches to Energy Demand-Side Response: A Systematic Review. *Renew. Sustain. Energy Rev.* **2020**, *130* (109899), 109899.
126. Zhang, N.; Yang, G.; Gao, Z.; Xu, C.; Zhang, Y.; Shi, R.; Keegan, J.; Xu, L.; Zhang, H.; Fan, Z.; Firmin, D. Deep Learning for Diagnosis of Chronic Myocardial Infarction on Nonenhanced Cardiac Cine MRI. *Radiology* **2019**, *291* (3), 606–617.
127. Zhang, Y.; Zhang, G.; Shang, Q. Computer-Aided Clinical Trial Recruitment Based on Domain-Specific Language Translation: A Case Study of Retinopathy of Prematurity. *J. Healthc. Eng.* **2017**, 1–9.
128. Pandey, A.; Schreurs, J.; Suykens, J. A. K. Generative Restricted Kernel Machines: A Framework for Multi-View Generation and Disentangled Feature Learning. *Neural Netw.* **2021**, *135*, 177–191.
129. Śledź, P.; Caflisch, A. Protein Structure-Based Drug Design: From Docking to Molecular Dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93–102.
130. Schneider, G.; Fechner, U. Computer-Based de Novo Design of Drug-like Molecules. *Nat. Rev. Drug Discov.* **2005**, *4* (8), 649–663.
131. Álvarez-Machancoses, Ó.; Fernández-Martínez, J. L. Using Artificial Intelligence Methods to Speed up Drug Discovery. *Expert Opin. Drug Discov.* **2019**, *14* (8), 769–777
132. Hessler, G.; Baringhaus, K.-H. Artificial Intelligence in Drug Design. *Molecules* **2018**, *23* (10).
133. Prasad, K.; Kumar, V. Artificial Intelligence-Driven Drug Repurposing and Structural Biology for SARS-CoV-2. *Curr. Res. Pharmacol. Drug Discov.* **2021**, *2* (100042), 100042.
134. Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R. K. Artificial Intelligence in Drug Discovery and Development. *Drug Discov. Today* **2021**, *26* (1), 80–93.
135. Basile, A. O.; Yahi, A.; Tatonetti, N. P. Artificial Intelligence for Drug Toxicity and Safety. *Trends Pharmacol. Sci.* **2019**, *40* (9), 624–635.

136. Tyrchan, C.; Evertsson, E. Matched Molecular Pair Analysis in Short: Algorithms, Applications and Limitations. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 86–90.
137. Zhang, H.; Xiang, M.-L.; Ma, C.-Y.; Huang, Q.; Li, W.; Xie, Y.; Wei, Y.-Q.; Yang, S.-Y. Three-Class Classification Models of LogS and LogP Derived by Using GA-CG-SVM Approach. *Mol. Divers.* **2009**, *13* (2), 261–268.
138. You, Y.; Lai, X.; Pan, Y.; Zheng, H.; Vera, J.; Liu, S.; Deng, S.; Zhang, L. Artificial Intelligence in Cancer Target Identification and Drug Discovery. *Signal Transduct. Target. Ther.* **2022**, *7* (1), 156.
139. Schwaller, P.; Vaucher, A. C.; Laino, T.; Reymond, J.-L. Prediction of Chemical Reaction Yields Using Deep Learning. *Mach. Learn. Sci. Technol.* **2021**, *2* (1), 015016.
140. Ishida, S.; Terayama, K.; Kojima, R.; Takasu, K.; Okuno, Y. AI-Driven Synthetic Route Design Incorporated with Retrosynthesis Knowledge. *J. Chem. Inf. Model.* **2022**, *62* (6), 1357–1367.
141. Burés, J.; Larrosa, I. Organic Reaction Mechanism Classification Using Machine-Learning. *Nature* **2023**, *613* (7945), 689–695.
142. Wang, Z.; Zhao, W.; Hao, G.; Song, B. Mapping the Resources and Approaches Facilitating Computer-Aided Synthesis Planning. *Org. Chem. Front.* **2021**, *8* (4), 812–824.
143. Capecchi, A.; Awale, M.; Probst, D.; Reymond, J.-L. PubChem and ChEMBL beyond Lipinski. *Mol. Inform.* **2019**, *38* (5), e1900016.
144. He, C.; Zhang, C.; Bian, T.; Jiao, K.; Su, W.; Wu, K.-J.; Su, A. A Review on Artificial Intelligence Enabled Design, Synthesis, and Process Optimization of Chemical Products for Industry 4.0. *Processes* **2023**, *11* (2), 330.
145. Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning Representations by Back-Propagating Errors. *Nature* **1986**, *323* (6088), 533–536.
146. Chollet, F. *Deep Learning with Python*; Manning Publications: New York, NY, 2017.
147. Chen, Q.; Zhu, X.; Ling, Z.-H.; Wei, S.; Jiang, H.; Inkpen, D. Enhanced LSTM for Natural Language Inference. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017.
148. Torre, J. C. Basics of Alzheimer's Disease Prevention. *J. Alzheimers. Dis.* **2010**, *20* (3), 687–688.
149. Grigorenko, A. P.; Rogaev, E. I. Molecular basics of Alzheimer's disease. *Mol. Biol.* **2007**, *41* (2), 331–345.
150. Shi, Y.; Wang, G.; Cai, X.-P.; Deng, J.-W.; Zheng, L.; Zhu, H.-H.; Zheng, M.; Yang, B.; Chen, Z. An Overview of COVID-19. *J. Zhejiang Univ. Sci. B* **2020**, *21* (5), 343–360.
151. Salahshoori, I.; Mobaraki-Asl, N.; Seyfaee, A.; Mirzaei Nasirabad, N.; Dehghan, Z.; Faraji, M.; Ganjkhani, M.; Babapoor, A.; Shadmehr, S. Z.; Hamrang, A. Overview of COVID-19 Disease: Virology, Epidemiology, Prevention Diagnosis, Treatment, and Vaccines. *Biologics* **2021**, *1* (1), 2–40.
152. Jankovic, J.; Tan, E. K. Parkinson's Disease: Etiopathogenesis and Treatment. *J. Neurol. Neurosurg. Psychiatry* **2020**, *91* (8), 795–808.
153. DeMaagd, G.; Philip, A. Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis. *P T* **2015**, *40* (8), 504–532.
154. Gill, C. M.; Dolan, L.; Piggott, L. M.; McLaughlin, A. M. New Developments in Tuberculosis Diagnosis and Treatment. *Breathe* **2022**, *18* (1), 210149.
155. World Health Organization, 2021. WHO global lists of high burden countries for tuberculosis (TB), TB/HIV and multidrug/rifampicin-resistant TB (MDR/RR-TB), 2021–2025: background document.

CHAPTER 2

Methodology of Recurrent Neural Network Method to Design Potential Drug-likeness Molecules

2.1 Computational Details

All the HF calculations are carried out using the Gaussian-09 package.¹ Autodock-Vina² is used for defining binding energy implemented in RDKit Package.³ Molecular dynamics (MD) simulations are done in GROMACS packages⁴ and VMD. Physicochemical or Lipinski rule of five⁵ and ADME properties are calculated by the Swiss-ADME website.⁶ The following sections detail the LORD molecular predictor in different steps. Firstly, it investigates the potential binding sites and then prepares the data for training. After training, it generates new drug-like molecules.

2.2 Potential Binding Site Analysis by Scanning MESP

To perform binding site analysis, MESP calculations are done for the acetylcholinesterase enzyme (AChE) with a 3D cubic box length of $40 \times 40 \times 40$. The MESP is evaluated by using Equation (1) in the Gaussian09 package.

$$V(r) = \sum \frac{Z_A}{|r - R_A|} - \int \frac{\rho(r')d^3r'}{|r - r'|} \quad (\text{Eq. 2.1})$$

Where Z_A is the charge of the nucleus located at R_A , MESP evaluation at each grid point is expensive for the entire protein. We calculated the acetylcholinesterase (AChE), main protease (M_{pro}), Monoamine oxidase-B (MAO-B), and ribosomal protein S1 of Mycobacterium tuberculosis (Mtb).

2.2.1 MESP Calculation for Identification of Binding Sites in Target Protein

The total protein of (AChE) is subdivided into 27 substrate structures sequentially, and each substrate structure contains 20 amino acids, as shown in Figure 2.1.

The uniform grid spacing of 0.3 Å is preserved in all MESP calculations. The MESP evaluations are done for all substrate structures with the same cubic box using cubegen as implemented in the g09 package. All the substrate cube files are loaded in VMD and transformed into one single cube file. Figure 2.1 shows two color regions of MESP distributions, such as dark blue color lobes inside, representing most negative regions with function value -0.20 of the protein. The grey color distribution shows low negative areas of the entire protein at -0.01 a.u. The total MESP distribution can be seen in Figure 2.1 with two layers. The most negative regions, such as inner blue color lobes, are further investigated by surface cavity analysis as highlighted by black circles with notations C_1 - C_4 . MESP function values are projected onto the plane to trace the exact potential of the binding site location. Scanning MESP function values within the plane resulted in surface cavities, as shown in Figure 2.1. The surface cavities were assumed to be the most potential binding sites. The surface cavities provide an understanding of the distribution of MESP function

values by following the gradients as color changes, as shown in Figure 2.1. The most negative MESP function values are spotted as inner and outer circles. The color distribution indicates that drug molecules can swallow into the cavity by following the gradient of charge distribution. This analysis provided the four best potential binding sites for further studies.

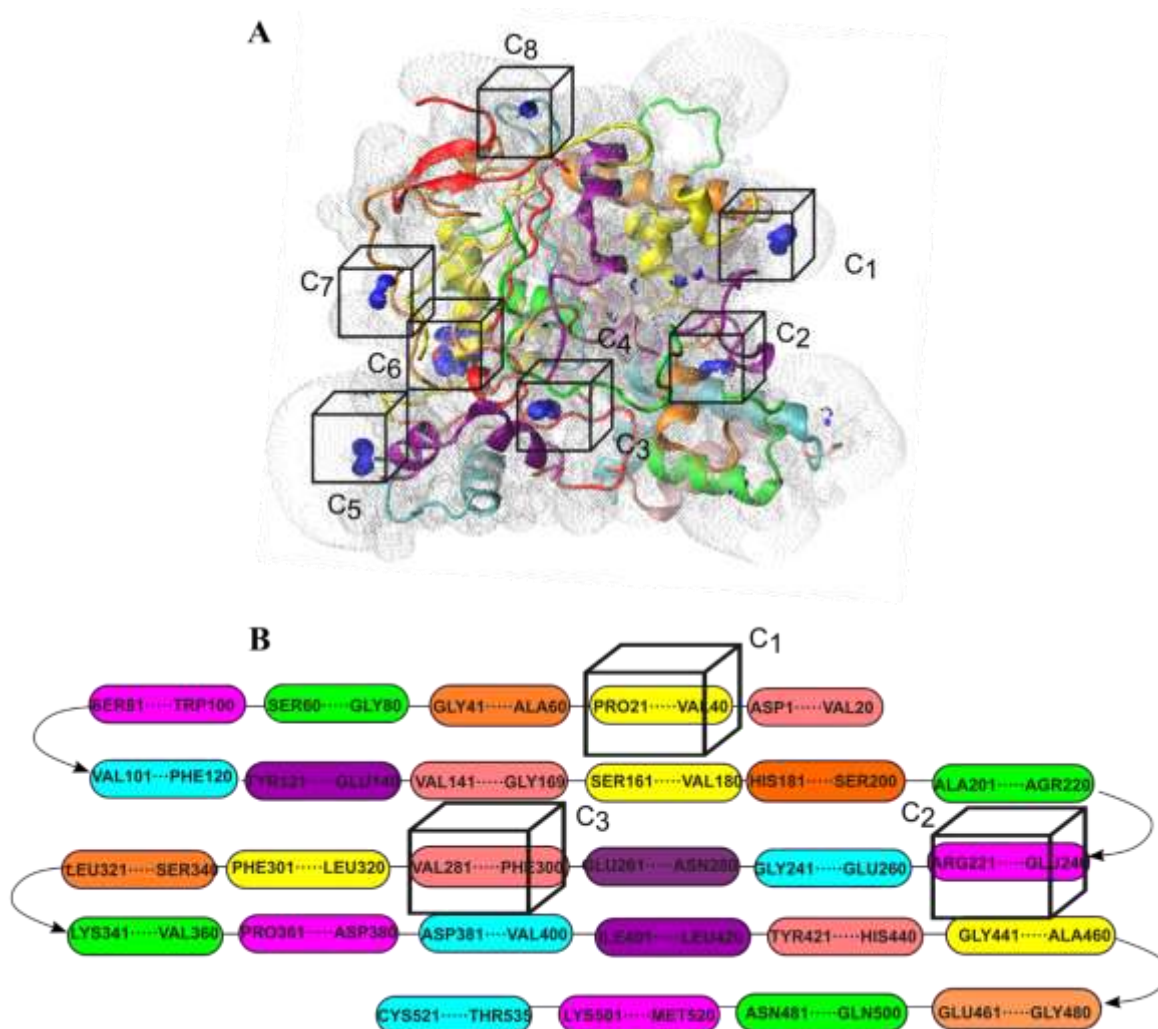


Figure 2.1 (A) MESP distribution of the AChE and (B) the procedure to construct substrate structures.

2.3 Data Preparation

Potential drug-likeness molecules were collected for this investigation from the ChEMBEL database. 1500 molecules in total were chosen, with a particular emphasis on substances that showed similarities to donepezil and literature-recommended molecules believed to be connected to Alzheimer's disease, as mentioned in Chapter 3, along with 1600 molecules that are antiviral drugs or show similarities to antiviral

compounds and literature-recommended molecules for the treatment of COVID-19 in Chapter 4. In Chapter 5, the focus of the investigation focused on Parkinson's disease, and 986 compounds related to safinamide drugs and their matching molecular similarities, as well as molecules proposed by the literature, were included. Finally, 786 compounds associated with pyrazinamide drugs and their molecular similarities, as well as molecules suggested in the literature for treating tuberculosis, were covered in Chapter 6 of the study.

The docking process involved evaluating all the molecules against multiple target binding sites, including the MESP-suggested and experimentally reported sites. The corresponding binding energies (BE) were determined and recorded for each molecule based on the specific type of binding site being considered. Consequently, the total binding energies obtained for each molecule across all the target binding sites were collected and stored in a file for further analysis.

2.4 Input Preparation for LORD Algorithm

Drug-likeness molecules are fragmented by the BRICS-BONDS⁷ fragmentation scheme as implemented in the RDKit package.⁸ Unique fragments are saved into the library by fingerprint similarity measure to avoid repetition of fragments as implemented in RDKit. In total fragments were saved in the library. Including all binding sites in the library helps identify their potential for binding interactions. The length of the library becomes 1000 is the sum of several drug fragments and several protein sites.

Symmetric functions were used to provide rotationally and translationally invariant coordinates to the LSTM networks as input vectors for each atom as inspired by Boehlert *al.*⁹ Symmetric functions such as radial (g_i^r) and angular (g_i^a) as shown in Equations (2) and (3). Symmetric functions cover the atomic environment accurately for each atom within the cut-off sphere.

$$g_i^r = \sum_{j \neq i} \exp(-\eta_r r_{ij}) f_c(r_{ij}) \quad (\text{Eq. 2.2})$$

$$g_i^a = \sum_{j \neq i} \sum_{k \neq i} \left(1 + \eta_a \cos(\theta_{ijk})\right)^\zeta \exp(-\eta_a (r_{ij} + r_{jk} + r_{ik})) f_c(r_{ij}) f_c(r_{jk}) f_c(r_{ik}) \quad (\text{Eq. 2.3})$$

$$f_c(r_{ij}) = \begin{cases} \frac{1}{2} \left[\cos\left(\frac{\pi r_{ij}}{r_c}\right) + 1 \right] & \text{if } r_{ij} \leq r_c \\ 0 & \text{if } r_{ij} > r_c \end{cases} \quad (\text{Eq. 2.4})$$

Here, r_{ij} , r_{jk} , and r_{ik} are the internuclear distance between i^{th} , j^{th} and k^{th} atoms in a molecule. $f_c(r_{ij})$, $f_c(r_{jk})$ and $f_c(r_{ik})$ are cut-off functions for respective r_{ij} , r_{jk} and r_{ik} . θ_{ijk} is the angle between j^{th} and k^{th} atoms

cantered at i^{th} atom. η_r and η_a are the width of the Gaussian function. The set of η_r, η_a, ζ , and λ are given in Table 2.1.

Table 2.1 Set of parameters η_r, η_a, ζ , and λ values for symmetric functions.

η_r	η_a	ζ	λ
0.001	0.001	1	1
0.010	0.003	2	-1
0.017	0.006	4	-
0.030	0.011	16	-
0.045	0.020	-	-
0.066	0.037	-	-
0.095	0.075	-	-
0.150	-	-	-
0.350	-	-	-

We have used a cut-off radius $r_c = 8\text{\AA}$. In the protein environment, the active site is covered within a 10 $\text{\AA}$ cut-off distance from the center of the cavity.

Symmetric function values are calculated for an entire library (drug fragments + protein sites(C_1, C_2, C_3, C_4)) using Equations (2) and (3). In total, 51 symmetric functions are used for each atom to describe the environment. To recognize each fragment uniquely, the sum of the symmetric function values for each atom within the fragment has been considered an input vector, as shown in Equation (5).

$$\text{Fragment feature} = \sum_{i=1}^N (D_i^{\text{rad}} + D_i^{\text{Ang}}) \quad (\text{Eq. 2.5})$$

Input is prepared so that the number of rows is the length of the molecule or the number of fragments and columns are 51 fragment features, i.e. (times steps $\times$ fragment features). The number of times steps is chosen based on the maximum molecular length in the database to maintain the uniform dimension while training by applying padding in Keras API.

2.5 LORD Output Preparation

LSTM simultaneously predicts two types of outputs, such as sequence and corrected binding energy (CBE) prediction. The sequence prediction output layer requires on-hot encoding of an entire library

(*times steps* $\times$ *Length of Library*) and CBE values for each molecule that involved in the training. CBE is calculated by Equation (2.6).

$$\text{CBE}(\bar{C}) = \frac{1}{2} \left(\text{BE} + \frac{1}{5} \left(\sum_{i=1,5} \frac{V_i}{\text{Max}(V_i)} \times \text{BE} \right) \right) \quad (\text{Eq. 2.6})$$

Where BE is the binding energy stored for each molecule concerning the protein site calculated with AutoDock Vina V_i is the variable consisting of the Lipinski rule of five, such as the rotational bonds, log(P) value, hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), and molecular weight. Actual binding energy is corrected according to the physicochemical properties calculated for each molecule in the database. LSTM networks are trained using CBE rather than BE to recognize the Lipinski rule of five. CBE enables LSTM networks to better prediction of the biologically active molecule.

2.6 Block Diagram of LORD Designer Algorithm

The block diagram for the LORD designer is shown in Figure 2.2. The research begins by retrieving the target protein structure from the Protein Data Bank (PDB) and accessing all available drug molecules stored in the ChEMBEL database. Once these resources are obtained, the subsequent step involves an analysis of the binding sites within the target protein, as previously discussed. Simultaneously, a drug fragment library is prepared for further investigations.

To facilitate the training of Long short-term memory (LSTM) networks, input data is prepared by employing the symmetric function procedure mentioned earlier. This input is then utilized for LSTM training, followed by testing using the prepared input. The network error is minimized through iterative training until reaching the desired minimum value. Upon achieving optimal network performance, the LORD (LSTM enable onsite recurrent molecular designer) algorithm comes into play, allowing the generation of novel drug molecules. The database is updated to incorporate the newly constructed molecules during this process.

Notably, the LORD algorithm incorporates a condition known as CBE (Chemical Bond Energy), which ensures that only molecules meeting the minimum cut-off criteria are saved, while the remaining ones are discarded to streamline the molecule-building procedure. The process of training and designing drug-likeness molecules is continued in generation mode. LSTM starts with the ChEMBEL database, called 1stgeneration LSTM, and LSTM is upgraded to better generations by adding newly generated data

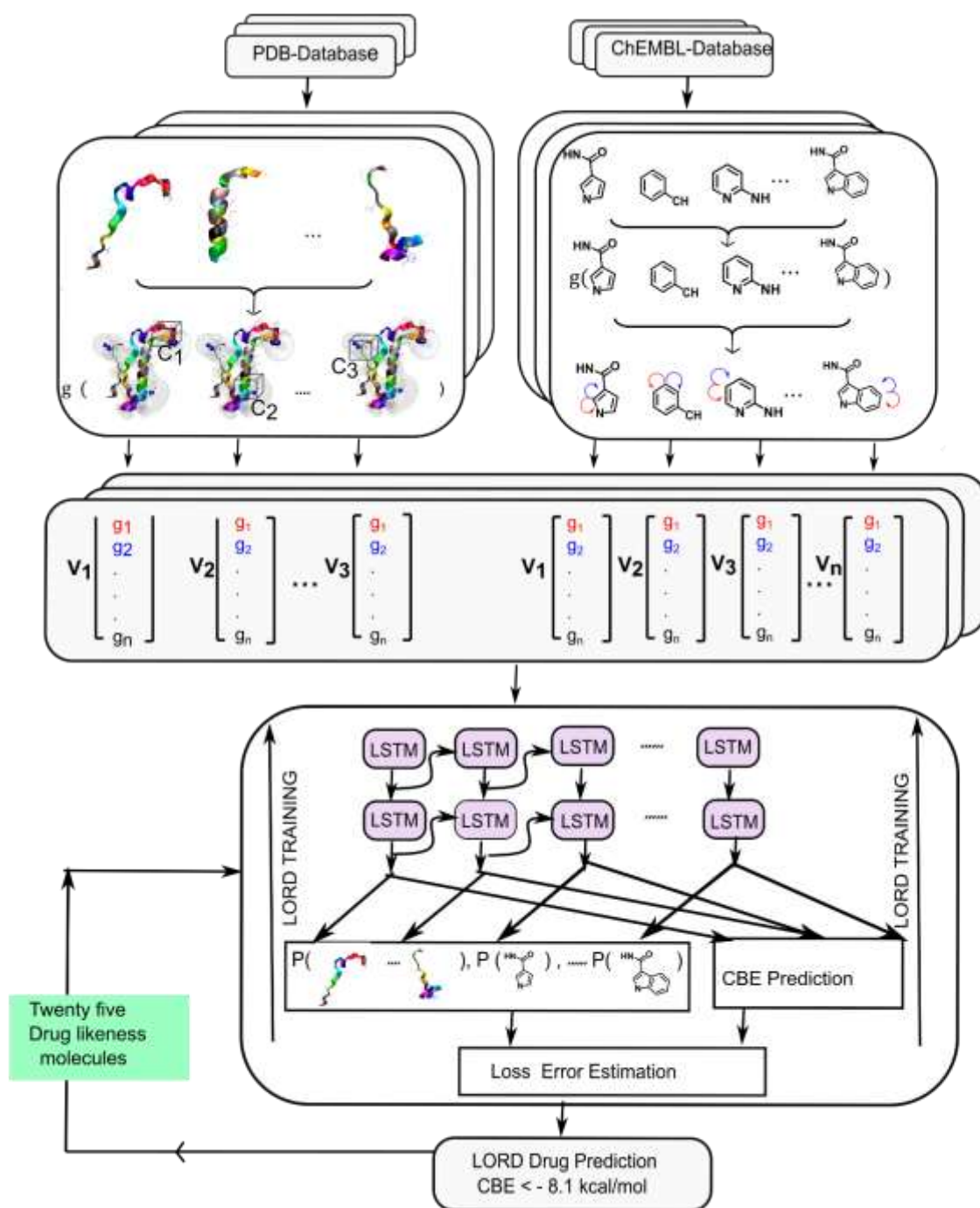


Figure 2.2 Block diagram for LORD designer algorithm.

2.7 LSTM Network Architecture and Training

LSTM networks consist of four layers, as shown in Figure 4. It begins with the Input layer, such as the symmetric function vector. Two LSTM layers and an output layer. The output layer consists of sequence

prediction with SoftMax function and $CBE(\bar{C})$ value prediction with dense layers simultaneously, as shown in Figure 4. Each LSTM layer consists of 256 hidden units, and the activation function is used. In the beginning network, the architecture is initialized with random weights. The total loss is the sum of the categorical loss and means squared loss, as shown in Figure 4. The categorical cross-entropy (CCE) and mean squared error (MSE) loss functions are defined by Equations (7) and (8) as implemented in Keras API.

$$MSE = \frac{1}{n} \sum_{i=1}^n (CBE - \overline{CBE})^2 \quad (\text{Eq. 2.7})$$

$$CCE = - \sum_{t=1}^{\text{length of library}} y_t \log(\bar{y}_t) \quad (\text{Eq. 2.8})$$

Where CBE is calculated by corrected binding and $\overline{CBE}$ is corrected binding from the network for the mean squared error function. And also y_t and $(\bar{y}_t)$ are real and network outputs of one-hot encoding for sequence generation. The loss function is back propagated to optimize the weights by Adam optimizer as implemented in Keras API.

2.8 LORD Molecular Drug Predictor

LORD utilizes the trained weights and generates molecules. To begin with, it requires a protein site as input, and then it generates a sequence of fragments as the time series goes on, as shown in Figure 4. At time series $(t - 1)$, it suggests the fragment output from the library, and in a subsequent step, it suggests the next fragment at t time series by providing the first fragment as input. The following time series indicates the number of fragments. Each fragment series is combined by using the RDKit package. At the same time, generating molecules number of fragments to generate is limited by maximum molecular weight as the cut-off is chosen as 500.

2.9 CHARMM Force Field in GROMACS

CHARMM (Chemistry at Harvard Macromolecular Mechanics) is a comprehensive collection of force fields and software for conducting molecular dynamics simulations and analysis. It offers united atom (CHARMM19) and all-atom (CHARMM22, CHARMM27, CHARMM36) force fields tailored to specific research needs. The CHARMM27 force field has been successfully adapted for GROMACS, a widely used molecular dynamics software. It is officially supported and provides accurate results within the GROMACS environment. On the other hand, the CHARMM36 force field files, which are continually

updated, can be obtained directly from the MacKerell lab website. The MacKerell lab is known for regularly producing the most current and up-to-date CHARMM force field files in the compatible GROMACS format.

To utilize the CHARMM force field parameters within GROMACS, researchers can employ the CHARMM36 force field. This force field is readily available in GROMACS as a precompiled binary file, encompassing all the necessary topology and parameter files required for simulations. By utilizing the CHARMM36 force field in GROMACS, researchers can benefit from a powerful combination of accurate force field parameters and the versatile simulation capabilities of GROMACS.¹⁰

To employ the force field in GROMACS, the following steps are followed:

- a) Begin by acquiring the CHARMM36 force field files from the official CHARMM website. These files contain the necessary parameters and specifications for the CHARMM36 force field.
- b) Once you have downloaded the CHARMM36 force field files, you will need to convert them into a format compatible with GROMACS. This can be achieved by utilizing the *gmx ffcharmm2gmx* command, specifically designed for this purpose. Executing this command will facilitate the conversion process and generate the essential topology and parameter files in the appropriate GROMACS format.
- c) Prepare your system for simulation using the standard GROMACS workflow.
- d) Use the generated topology and parameter files in your simulation using the *-f* and *-p* flags in the *grompp* command.

GROMACS is a popular molecular dynamics simulation package that can simulate protein-ligand complexes. Here is a general outline of the steps involved in running a GROMACS simulation for a protein-ligand complex:

1. Prepare the system: This process entails establishing the initial spatial arrangement of the protein-ligand complex, solvating the system in a water box, adding ions to neutralize the system, and assigning force field parameters to the protein and ligand molecules.
2. Energy minimization: Energy reduction aims to eliminate steric conflicts and unwanted interactions from the system. To do this, a series of reduction procedures are applied to the system until the energy converges to a minimum.
3. Equilibration: In this process, the system is allowed to adjust to the simulated conditions, such as temperature and pressure. The equilibration procedure consists of NVT (constant volume, constant temperature) and NPT (constant number of particles, pressure, and temperature).

4. Production run: This is an accurate simulation of the system; the simulation should be run for a significant length to portray the dynamics of the protein-ligand complex accurately. The simulation period typically ranges between nanoseconds to microseconds depending on the research issue.
5. Analysis: Many analyses may be run after the simulation to understand better how the protein-ligand complex functions. These investigations involve determining the root mean square deviation (RMSD) of the protein-ligand complex over time, figuring out the binding energy between the protein and ligand, and analyzing the protein-ligand complex's trajectory to search for essential interactions.
6. Analyze the trajectory of the protein-ligand complex to search for essential interactions.

2.10 Free Energy Calculation (g_mmpbsa)

The g_mmpbsa software employs the Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA)¹¹ approach to compute binding free energies in molecular systems. This technique commonly assesses the binding strength between multiple molecules, including protein-ligand complexes. By employing the MM/PBSA method, researchers can calculate the free energy of binding, a crucial parameter in understanding the stability and affinity of molecular interactions. The g_mmpbsa software facilitates the application of this approach, enabling efficient and accurate estimation of binding free energies in various molecular systems.

Kumari *et al.*¹² The g_mmpbsa (Molecular Mechanics/Poisson-Boltzmann Surface Area) method is a computational approach specifically designed to determine binding free energy in protein-ligand complexes. This method incorporates molecular dynamics simulations, where the protein-ligand complex is subjected to dynamic behavior, and the subsequent free energy changes are computed utilizing a specific formula. By employing the g_mmpbsa method, researchers can gain valuable insights into the energetics of protein-ligand interactions, enabling the evaluation of binding strengths and the prediction of potential binding affinities using followed formula.

$$\Delta G_{bind} = \Delta G_{complex} - (\Delta G_{protein} + \Delta G_{ligand}) \quad (\text{Eq. 2.9})$$

Where ΔG_{bind} is the binding free energy of the protein-ligand complex, $\Delta G_{complex}$ is the free energy of the complex, $\Delta G_{protein}$ represents the free energy of the protein when it is not bound to any ligand, and ΔG_{ligand} represents the free energy of the ligand when it is not bound to the protein. These values play a critical role in assessing the energetic contributions and stability of the protein-ligand complex.

The `g_mmpbsa` software package offers a comprehensive set of tools for conducting molecular dynamics simulations, computing electrostatic and van der Waals interactions within protein-ligand systems, and estimating solvation-free energy using the Poisson-Boltzmann Equation. By leveraging these calculations, the software facilitates the estimation of binding free energy for the complex, employing the formula mentioned earlier.

To utilize `g_mmpbsa` effectively, users must provide input files containing the coordinates of the protein and ligand and parameters necessary for the molecular dynamic simulation and Poisson-Boltzmann calculations. The software operates via a command-line interface, allowing users to execute commands and run the required calculations. Upon completion, the output files generated by `g_mmpbsa` will contain the estimated binding free energy and other pertinent information pertaining to the protein-ligand complex.

REFERENCES

1. Gaussian 09, Revision E.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.
2. Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings *J. Chem. Inf. Model.* **2021**, 61 (8), 3891–3898.
3. Bento, A. P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L. J.; De Veij, M.; Leach, A. R. An Open Source Chemical Structure Curation Pipeline Using RDKit *J. Chem. Inf. Model* **2020**, 12 (1), 51.
4. Kutzner, C.; Kniep, C.; Cherian, A.; Nordstrom, L.; Grubmüller, H.; de Groot, B. L.; Gapsys, V. GROMACS in the Cloud: A Global Supercomputer to Speed up Alchemical Drug Design *J. Chem. Inf. Model* **2022**, 62 (7), 1691–1711.
5. Shultz, M. D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs: Miniperspective *J. Med. Chem.* **2019**, 62 (4), 1701–1714.
6. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules *Sci. Rep.* **2017**, 7 (1).
7. Heikamp, K.; Zuccotto, F.; Kiczun, M.; Ray, P.; Gilbert, I.H. Exhaustive sampling of the fragment space associated to a molecule leading to the generation of conserved fragments *Chem. Biol. Drug Des.* **2018**, 91(3):655-667.
8. Liu, T.; Naderi, M.; Alvin, C.; Mukhopadhyay, S.; Brylinski, M. Break Down in Order to Build Up: Decomposing Small Molecules for Fragment-Based Drug Design with eMolFrag *J. Chem. Inf. Model.* **2017**, 57(4), 627-631.
9. Shao, Y.; Hellström, M.; Mitev, P. D.; Knijff, L.; Zhang, C. PiNN: A Python Library for Building Atomic Neural Networks of Molecules and Materials *J. Chem. Inf. Model.* **2020**, 60 (3), 1184–1193.
10. Vermaas, J. V.; Hardy, D. J.; Stone, J. E.; Tajkhorshid, E.; Kohlmeyer, A. TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD *J. Chem. Inf. Model.* **2016**, 56 (6), 1112–1116.
11. Poli, G.; Granchi, C.; Rizzolio, F.; Tuccinardi, T. Application of MM-PBSA Methods in Virtual Screening. *Molecules* **2020**, 25 (8), 1971.
12. Kumari, R.; Kumar, R.; Open Source Drug Discovery Consortium; Lynn, A. G_mmpbsa--a GROMACS Tool for High-Throughput MM-PBSA Calculations *J. Chem. Inf. Model.* **2014**, 54 (7), 1951–1962.

CHAPTER 3

Design of Potential Druglikeness Molecules for Alzheimer's Disease

3.1 Introduction to Alzheimer's Disease

Alzheimer's Disease (AD) is a prevalent neurological condition characterized by the progressive loss of memory and cognitive function due to the degeneration of neurons in the brain. AD is widely recognized as the most prevalent form of dementia, and it arises from disruptions in the transmission of signals between nerve cells and muscle cells in the brain, typically caused by the inhibition of an enzyme called cholinesterase.¹ This leads to the degeneration of neurons in the cerebral cortex and the formation of neurofibrillary tangles and plaques containing a beta-amyloid protein. These pathological changes profoundly impact the experience of individuals living with AD.² AI/ML can help discover the complexity necessary in discovering new intervention targets for neurodegenerative disorder by developing novel approaches, models, and algorithms to enable research of the complex non-linear dynamics between biology, environment, sickness, and public health. This method can improve our understanding of neurodegeneration and provide novel treatment for patients suffering from catastrophic neurodegenerative disease.³ In 2015, 46.8 million people lived with AD or a related form of dementia globally, with 1.9 million fatalities. In 2017, there were over 50 million Alzheimer's sufferers; in 2030, this figure will be 74.7 million; in 2050, it will be 131 million. The countries witnessing the most rapid growth in the population of AD patients include Finland, the United States, Canada, Iceland, Sweden, Switzerland, Norway, Denmark, the Netherlands, Belgium, India, Cambodia, Georgia, and Singapore, among others are among the countries with the fastest-growing patient populations.⁴

Acetylcholine (ACh) is a neurotransmitter that plays a role in the normal functioning of the brain. It involves many important cognitive processes, such as attention, learning, and memory formation. ACh is produced by cholinergic neurons, which are concentrated in the basal forebrain and brainstem. In AD, there is a significant reduction in cholinergic function due to the loss of cholinergic neurons in the brain. This loss of function is thought to contribute to the cognitive and memory impairments characteristic of AD. Specifically, the breakdown of ACh is accelerated due to an increase in the activity of an enzyme called Acetylcholinesterase (AChE).^{5,6} It's an enzyme that breaks down ACh into acetyl and choline as a result, there is a decrease in the level of ACh in the brain, leading to cognitive and memory deficits.

Acetylcholinesterase inhibitors (AChEIs) are a class of drugs that are commonly used to treat the cognitive and memory impairments associated with AD. These drugs are Donepezil, Rivastigmine,

Galantamine, Tacrine, and memantine, which are market available, ondansetron is a highly effective cholinesterase inhibitor and other novel related AChEIs,^{7,8,9} including Eptastigmine, Phenserine, Huperzine A, Dimebon, Pozanicline (ABT-089), RG3487, GSK239512, Varenicline, ABT-288, Nelonicline (ABT-126), Encenicline (EVP-6124), S 38093 LadostigilHemitartrate, GLN-1062, and SUVN-G3031 work by inhibiting the activity of AChE, which allows on the controlled breakdown of Ach and therefore increases the ACh concentrations in the brain.^{10,11}

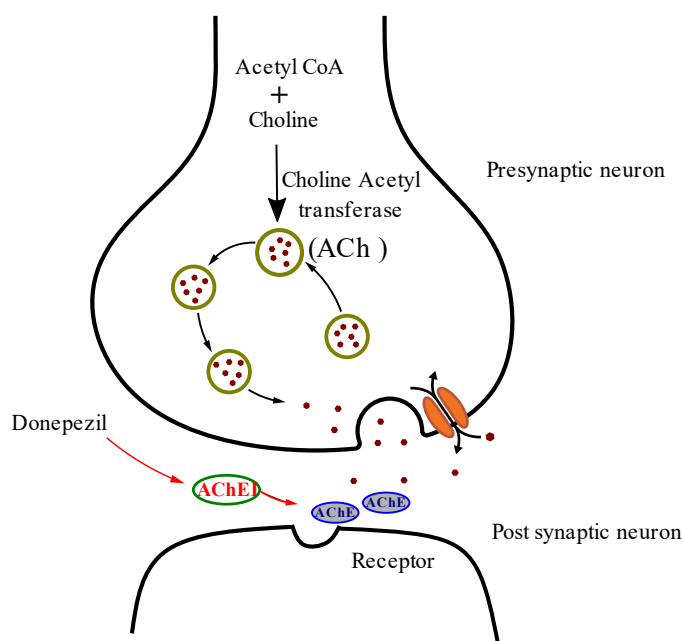


Figure 3.1 Mechanisms of action Acetylcholinesterase inhibitors in AD.

The X-ray crystal structure of AChE of the protein id: 1EVE^{12,13} was crystallized with Donepezil and was retrieved from the protein data bank (PDB). In which the active site of amino acids are Tyrosine 70 (Try70), Asparticacid 72 (Asp72), Tryptophan 84(Trp84), Glycine 118(Gly118), Glycine 119(Gly119), Tyrosine121(Try121), Tyrosine 130(Try130), Serine 200(Ser200), Alanine 201(Ala201), Tryptophan 279(Trp279), Phenylalanine 288(Phe288), Phenylalanine290(Phe290), Phenylalanine 330(Phe 330), Phenylalanine 331(Phe 331), Tyrosine 334(Try334), Histidine(His440) in the torpedo californica acetylcholinesterase (TcAhE).

Numerous computational research methods have been documented regarding AChEIs, focusing on the substances involved. Through molecular docking studies, it has been predicted that the interactions between the protein and ligand complex exhibit robust binding activity. The protein and ligand complex suggest high binding activity of the compounds such as Flurbiprofen, Isoniazide,¹⁴

Spirooxindole,¹⁵ Thiazole-related compounds,¹⁶ Pyrimidine, Oxadiazole¹⁷ and Pyridine derivatives¹⁸ were employed for the AChE inhibitory activity, and Table 3.1 includes to some of the previous literature, we focused on discusses the methodology and providing a summary.

Table 3.1 Overview of previous literature and focus on methodology and summary.

Methodology	Summary
Small-molecule, therapeutics, biologics	This article explores various aspects of AD and drug design, focusing on dementia, the blood-brain barrier, small-molecule therapeutics, biologics, neurodegenerative disease, and neuroinflammation. It provides valuable insights into the challenges and advancements in developing drugs for AD and sheds light on the underlying mechanisms of this neurodegenerative condition. ¹⁹
Molecular docking, MD simulations, and QSAR approaches	Computational modeling has emerged as a powerful tool for designing multi-target-directed inhibitors. This approach simultaneously targets multiple pathological factors in the disease, potentially enhancing therapeutic efficacy. Researchers can use computational techniques to develop innovative drug candidates with improved potency and selectivity, offering promising prospects in the fight against AD. ²⁰
Crystal structure, molecular docking and simulation, tacrine.	The newly determined crystal structure of AChE in complex with inhibitors offers valuable insights, aiding the development of novel drugs with enhanced efficacy. The findings open doors for more targeted and effective drug design strategies, potentially leading to improved treatments for various conditions. ²¹
Donepezil, computational Studies, and biological Validation	Donepezil-like compound(D ₁) being investigated for AD treatment shares similarities with the FDA-approved drug donepezil. D ₁ has shown promising effects in elevating acetylcholine levels and reducing A β plaques in <i>Caenorhabditis elegans</i> , surpassing the functionality of Donepezil at similar doses. Moreover, D ₁ exhibits an agonistic effect on the α 7 nicotinic acetylcholine receptor, making it a potentially beneficial AD treatment with its dual-binding site characteristics and additional impact on neurotransmission. ²²
E-pharmacophore-based virtual screening, molecular docking, and MD simulations	New therapeutic candidates with potential value in treating conditions related to AChE dysfunction. A combination of virtual screening and molecular dynamics simulations is employed. Using computational methods, scientists aim to pinpoint promising compounds that can be further investigated and developed, ultimately advancing the field of drug discovery and paving the way for potential treatments targeting an AChE. ²³
Molecular docking, MD simulations	Fragment-based design and virtual screening to identify a promising new acetylcholinesterase inhibitor. This innovative approach combines computational methods and molecular modeling to efficiently discover potential drug candidates, paving the way for developing more effective treatments for AD. ²⁴
Fragment-based drug design,	Innovative research focuses on creating activity rules and designing chemical fragments to facilitate the virtual discovery of novel dual inhibitors targeting

molecular docking and dual-target inhibitor	both AChE and BACE1 enzymes, with the aim of combatting AD. This study aims to contribute to developing effective therapies against this debilitating neurodegenerative disorder. ²⁵
Peptides and proteins, Inhibitors	A novel approach utilizing virtual screening and binding free energy calculations is employed to identify potential AChE inhibitors. This study aims to discover new compounds that can effectively inhibit AChE, a key enzyme involved in neurodegenerative disorders, by using computational techniques. ²⁶
Docking studies, MD simulations, and MM-GBSA calculations	This study utilizes molecular dynamics and structure-based virtual screening to identify potential natural compounds that can modulate the signaling pathway. By targeting this pathway, the identified compounds hold promise as therapeutics for AD, providing new avenues for developing effective treatments. ²⁷
Virtual screening, molecular docking, and MD simulations	New acetylcholinesterase inhibitors are achieved through virtual screening, in vitro experiments, and molecular dynamics simulations. This study presents a novel approach to discovering potential therapeutic agents that target acetylcholinesterase, offering promising avenues for further exploration in the field of drug discovery. ²⁸
Molecular docking, MD simulations, and DFT calculations	This study aims to identify highly effective acetylcholinesterase inhibitors derived from plants with the potential to be used in treating Alzheimer's disease. By employing advanced algorithms, the research seeks to discover promising candidates for further investigation and development of therapeutic strategies. ²⁹
Natural compounds, molecular docking, MD simulations, and pharmacokinetic predictions	Exploring the potential of natural compounds: computational screening for acetylcholinesterase inhibition. This study employs biocomputational techniques to evaluate the effectiveness of various natural compounds in targeting acetylcholinesterase, a key enzyme implicated in neurological disorders. Screening these compounds can identify potential candidates for further investigation and drug development. ³⁰
Pharmacophore studies, molecular docking, MD simulations	A novel approach combining pharmacophore modeling, virtual screening, docking simulation, and bioassay has identified potent acetylcholinesterase inhibitors. This discovery presents new candidates with promising potential in treating neurological disorders. ³¹
Synthesis, molecular docking, and biological Evaluations	In this study, researchers aimed to develop new thiazole-based derivatives as potential AChE inhibitors. A series of compounds were synthesized and evaluated for their inhibitory activity against AChE, an enzyme implicated in AD. Promising results were obtained, indicating that these thiazole derivatives have the potential to be effective AChE inhibitors, which could contribute to the development of novel treatments for AD. ³²

The present chapter shows a novel design strategy for designing drug-like molecules by applying an AChE target. The methodology section explained the binding site analysis using MESP and LORD designer theory and implementation. The results section gives a detailed discussion about

physicochemical and ADME properties and MD simulation results for LORD 25 potential drug-like molecules by LORD with concluding remarks.

3.2 Result and Discussion

3.2.1 MESP Calculation for Identification of Binding Sites in Alzheimer's Target Protein

The whole protein of AChE is subdivided into 27 substrate structures sequentially, and each substrate structure contains 20 amino acids, as shown in Figure 3.2. A uniform grid spacing of 0.3 Å is maintained in all MESP calculations. The MESP evaluations are done for all substrate structures with the same cubic box using *cubegen* as implemented in the g09 package.

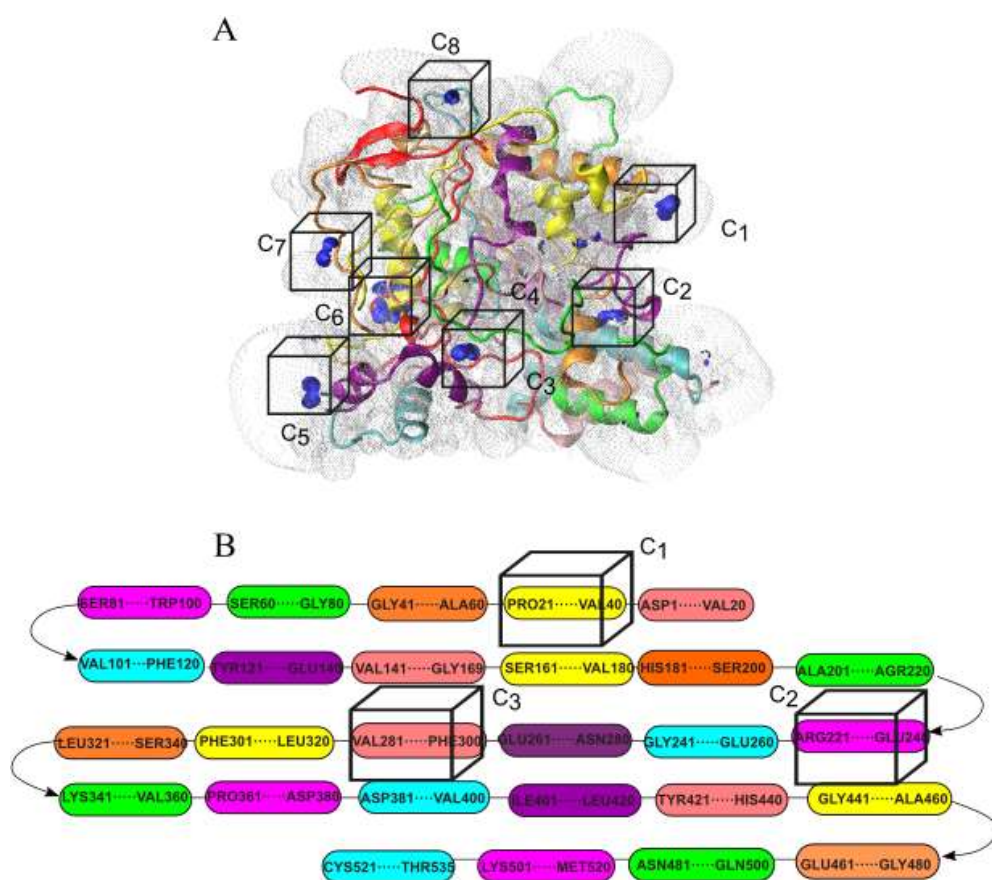
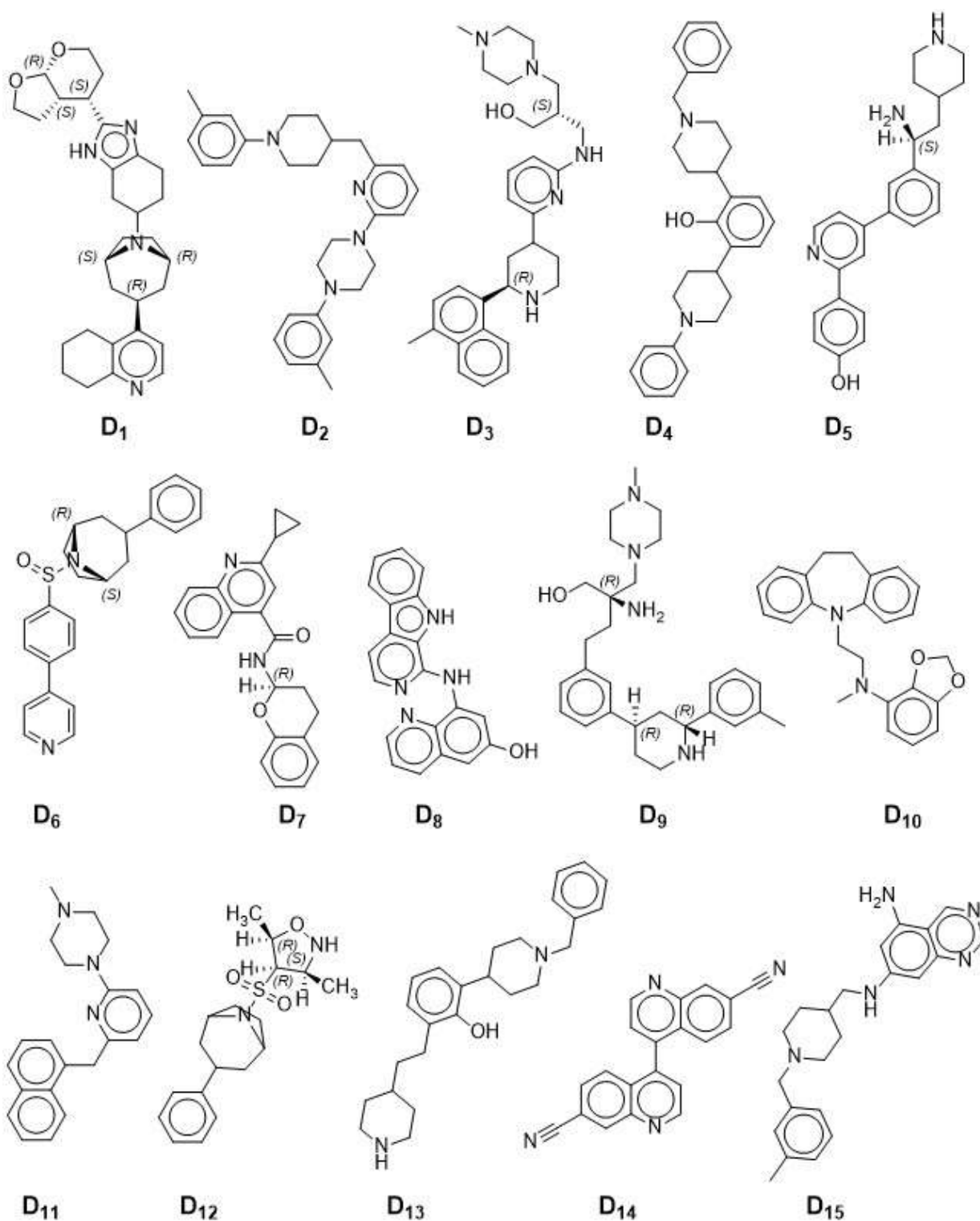


Figure 3.2 (A) MESP distribution of the AChE and **(B)** Substrate structures.

Figure 3.2 depicts MESP isosurfaces, such as dark blue lobes, representing most negative regions with function value -0.20 a.u. of the protein. The red color distribution shows another MESP isosurface with -0.01 a.u. for the complete protein. The total MESP distribution can be seen in Figure

3.2.2 Design of Potential Drug-Likeness Molecules

LORD molecular designer designed 25 Potential drug-likeness molecules and compared them to the market-available donepezil drug.



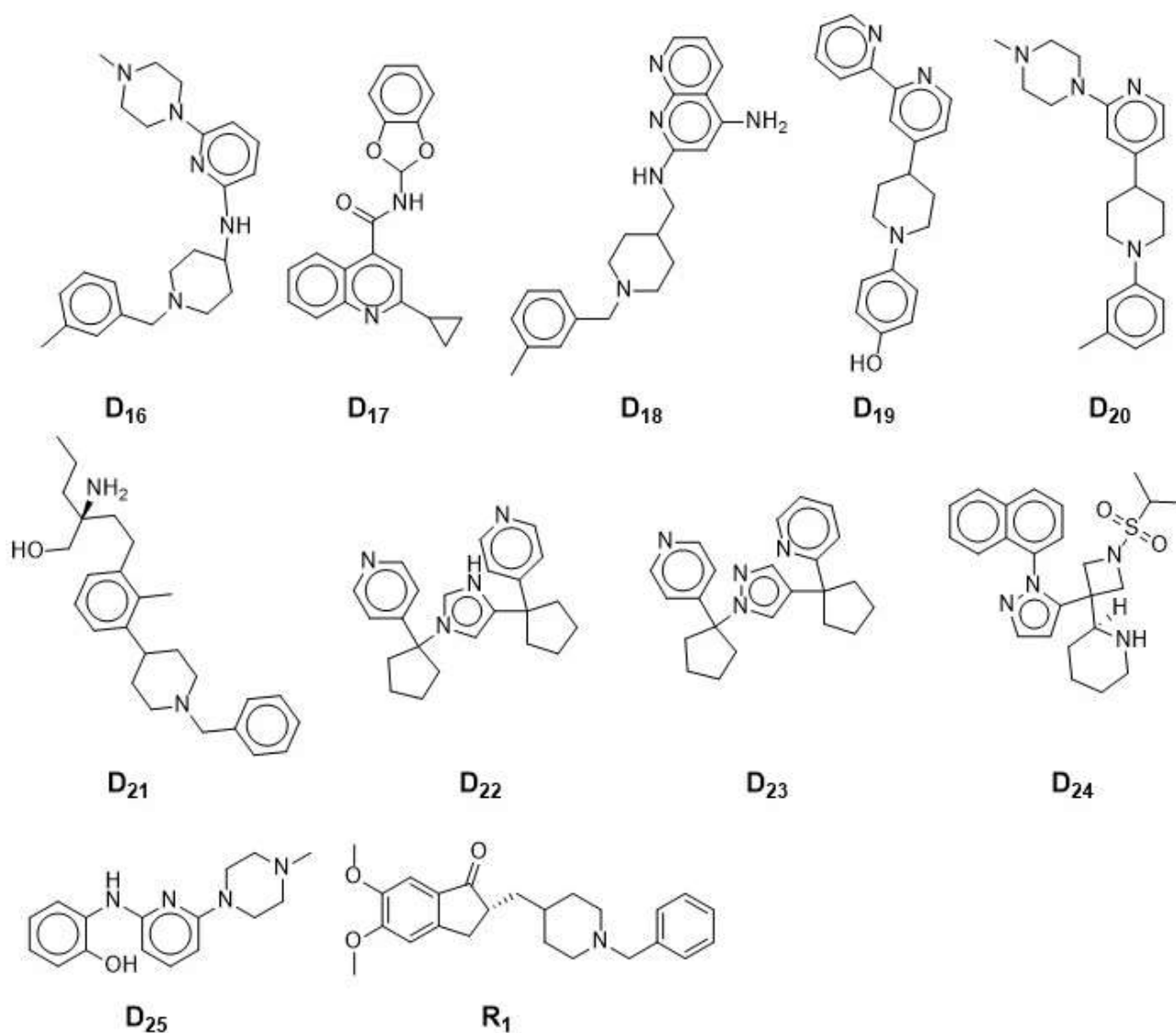


Figure 3.3 The structure of the twenty-five drug-likeness predicted molecules and the reference Donepezil (R₁) molecules.

3.2 with two layers. The most negative regions, such as inner blue color lobes, are further investigated by surface cavity analysis as highlighted by black circles with notations C₁-C₄. MESP function values are projected onto the plane to trace the location of the potential binding site. Scanning MESP function values within the plane resulted in surface cavities, as shown in Figure 3.2. The surface cavities were assumed to be the most potential binding sites. The surface cavities provide an understanding of the distribution of MESP function values by following the gradients as color changes, as shown in Figure 3.2. The color distribution indicates that drug molecules can swallow into the cavity by following the

gradient of charge distribution. This analysis provided the four best potential binding sites for further studies.

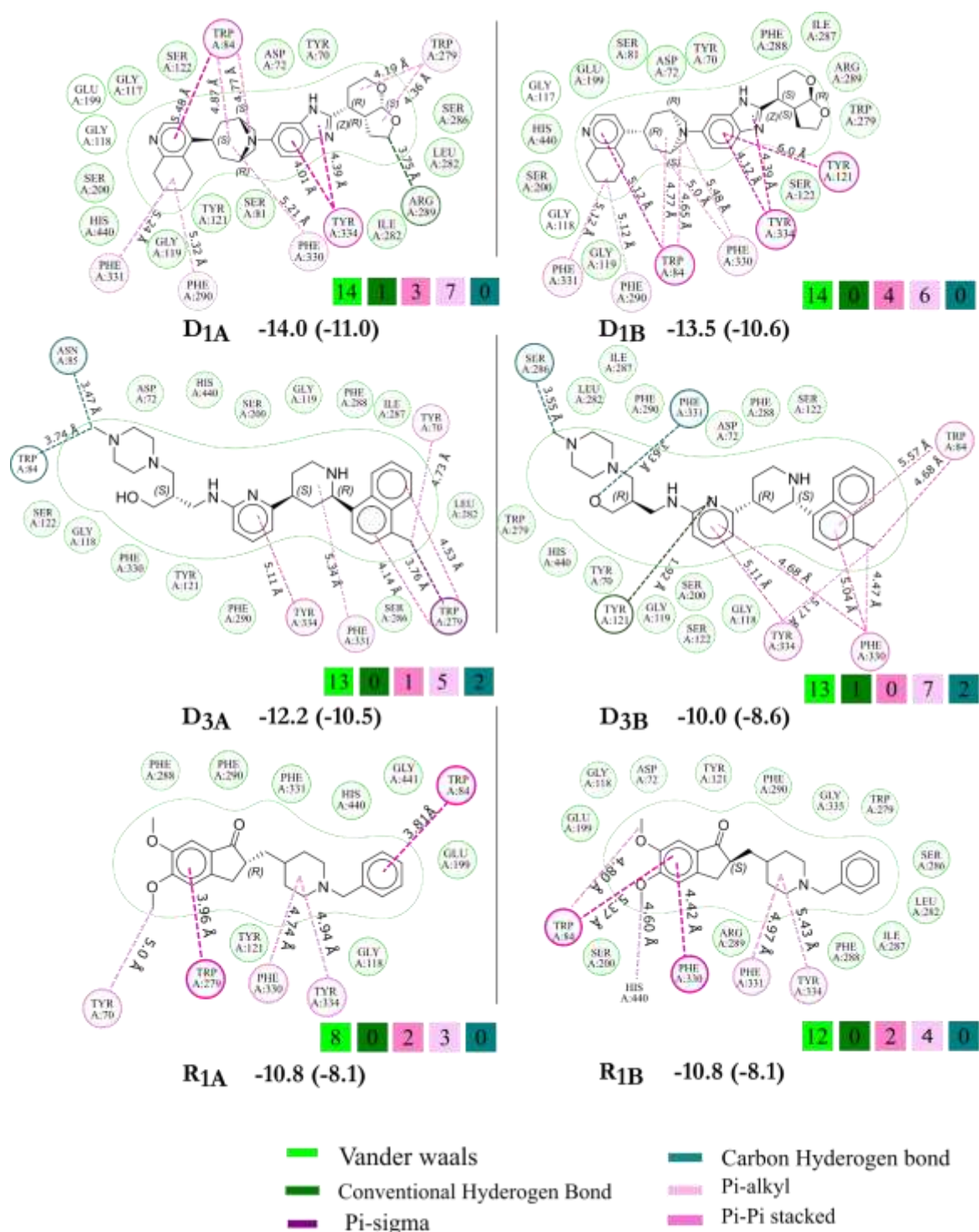


Figure 3.4 Stereo-isomers of D₁, D₃, and R₁ drug-likeness molecules and their interactions with the target protein.

3.2.3 Interactions of Stereo-Isomers with Target Molecules

Stereochemistry is crucial to the development of pharmaceuticals, given that many drugs actions depend on their three-dimensional structure. Drug designers can choose the most active and secure stereoisomer or create a drug that can specifically target a certain stereoisomer by studying the stereochemistry of a drug molecule and the mechanism of its action. D_{1A}, D_{1B}, D_{3A}, D_{3B}, and R_{1A}, R_{1B} are the enantiomers that reflect the absolute stereochemistry of the stereocentres in 3D structures using ChemDraw 3D software. Stereoisomers were investigated for their interaction with Site-4 (C₄) of the acetylcholinesterase of the target protein.

During docking, the protein's active amino acids interact with the ligand creating hydrogen bonds, electrostatic interactions, and van der Waals interactions. These interactions stabilize the protein-ligand combination. These findings provided evidence for the impact of the stereo centers on the molecular docking energy (BE), corrected binding energy (CBE), protein-ligand complex and the nature of interactions.

In this D_{1A} BE of -14.0 kcal/mol, CBE was -11.0 kcal/mol, and hexahydro-4H-furo[2,3-b] pyran fragments were more significant than D_{1B}. The protein and ligand complex can develop two P-alkyl bonding interactions with TRP A: 279 and one hydrogen bond with ARG A: 289 amino acid residues. Given that D_{1B} have BE and CBE are lower than those of D_{1A} (-13.5 and -10.6 kcal/mol, respectively), it must have a lower binding energy. In the complex, D_{3A} interacts with the Pi-sigma bonding of TRP A:279 and contains fragments of 1-methylnaphthalene with BE -12.5 kcal/mol and CBE -10.5 kcal/mol. D_{3B} has a lower binding energy than D_{3A}, with a BE of -10.0 kcal/mol and a CBE of -8.6 kcal/mol, respectively, and both R_{1A}, R_{1B} have BE -10.8 kcal/mol, CBE -8.1 kcal/mol.

In Figure 6, D_{1A}, D_{3A}, and R_{1A} are stereoisomers of drug-likeness molecules with more incredible binding energy than D_{1B}, D_{3B}, and R_{1B} stereoisomers. We were seen to be stabilized by the minimal distance between amino acid residues and the protein-ligand nature of their interaction. The molecular docking results are significantly influenced by stereochemistry, and accurate ligand binding process and affinity predictions depend on accurate ligand and protein structure modeling. In this study, we found that the interaction between D₁, and D₃ stereoisomers is better than that of the drug Donepezil (R₁) due to the higher binding energies and shorter minimal distance between amino acid residues.

3.2.4 Protein-Ligand Complex and their Interaction Studies of the Top Four Drug-Likeness Molecules at Selected Four Binding Sites

To visualize the target binding site, we have provided the protein-ligand complex for all four target binding sites in Figure 3.5 (left column). On the other side, we have shown the BE and CBE of the top four (D_1, D_2, D_3, D_4) LORD-generated molecules and reference molecules of the Donepezil (R_1) for each site (C_1 to C_4) from top to bottom in the right column of Figure 3.5 CBE values for LORD's top four molecules are much lower in energy than the reference molecules shown in the green and blue color bar plot in all the sites. LORD molecules dominate the present existing market drugs in all four sites.

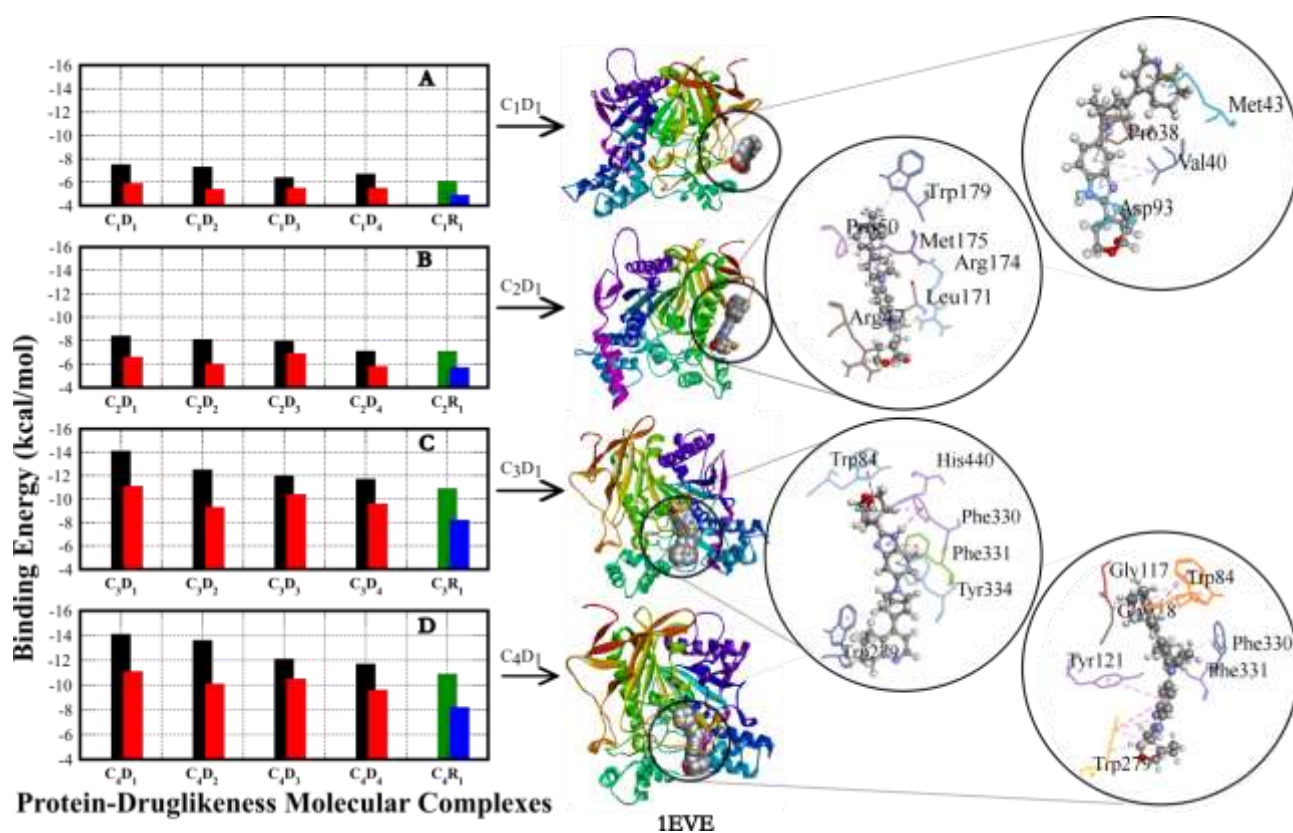


Figure 3.5 Protein-ligand complex and their interaction studies of the top five drug-likeness molecules at selected five binding sites. (A) MESP suggested site-1(C_1), (B) MESP suggested site-2 (C_2), (C) MESP suggested cavity site-3 (C_3), and (D) Experimentally reported site-4 (C_4).

3.2.5 Binding sites analysis from protein-ligand complex

MESP cavity analysis has explored the interactions between proteins and their ligands at each of the four sites (C₁-C₄). The 25 potential drug-likeness molecules can bind to three different protein cavities based on the MESP strategy conducted on it and compared with the experimentally suggested active binding cavity or site. An active cavity is a particular type of cavity present in the protein structure that is crucial to the activity of the protein.

The common active amino acid residues are Glycine A 32 (GLY A:32), Glutamic acid A:37 (GLU A:37), Proline A:38 (PRO A:38), Valine A:40 (VAL A:40), Glycine A 41 (GLY A:41), Lysine A:51 (LSY A:51), Lysine A:52 (LSY A: 52), Proline (PRO 53), Trptophan A:54 (TRP A:54), Serine A:55 (SER A:55), Tryptophan A:58 (TRP A: 58), Proline A:64 (PRO A:64), Asparagine A:65 (ASN A: 65), Serine (SER A:91), Asparagine (ASP A: 93), Tryptophan A:96 (TYR A:96) found in Site-1(C₁)

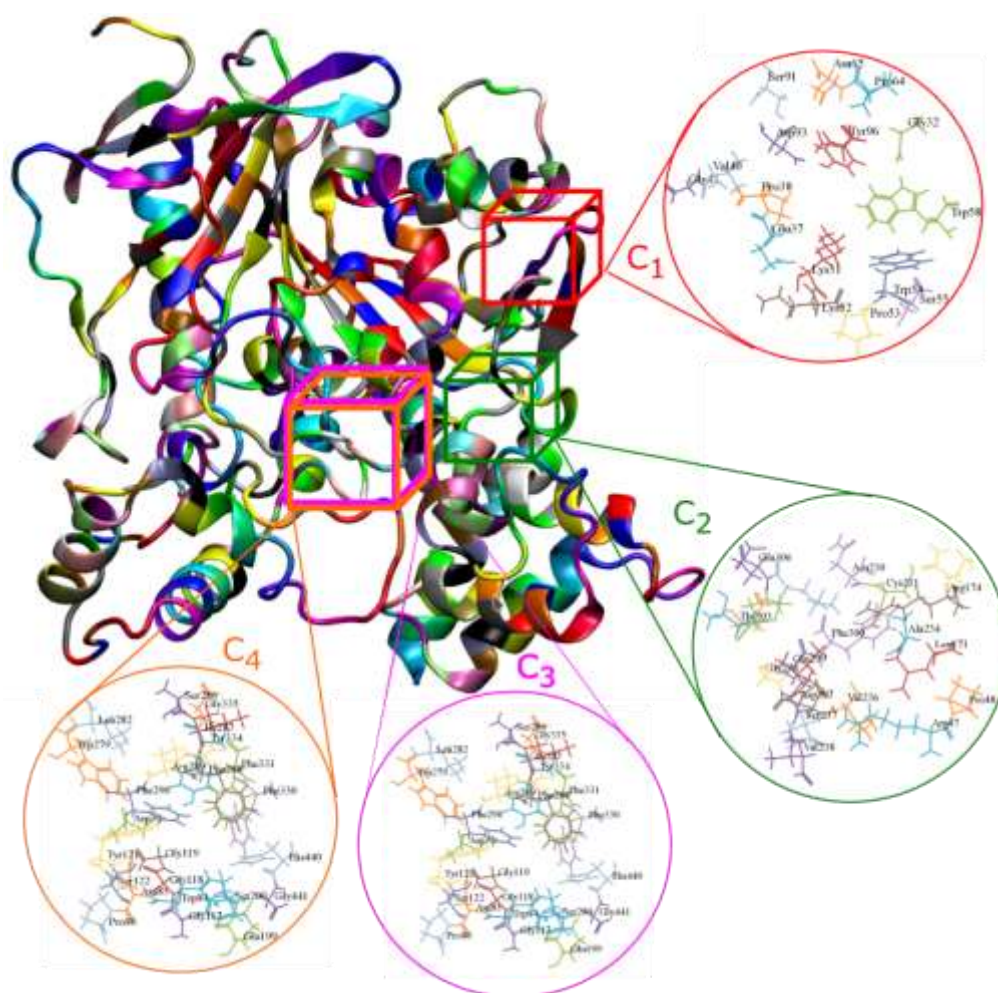


Figure 3.6 Acetylcholinesterase enzyme and four cavities, C₁, C₂, C₃, C₄

It demonstrated decreased binding energy compared to the other three sites. The residues of amino acids were found at Site-2 (C₂). It is shown in Serine A:12,(SER A:12), Phenylalanine A:35 (PHE A:35), arginine (ARG A:47), Proline A:48 (PRO A:48), Lysine A:51 (LYS A:51), Proline A:53 (PRO A:53), Leucine A:171 (LEU A:171), Arginine A:174 (ARG A:174), Glutamic acid A:178 (GLN A:178), Trptophan (TRP A:179), Aparagine A:230 (ASN A: 230), Cysteine A:231 (CYS A: 231), Alanine A:234 (ALA A:234), Valine A:236 (VAL A:236), Serine A:237 (SER A: 237), Valine A:238 (VAL A:238), Isolecuine A:296 (ILE A:296), Asparagine ASP:297 (ASP 297), Glycine A:298 (GLY A: 298), Proline A:302 (PRO A: 302), Threonine A:303 (THR A: 303), Serine A:304 (SER A: 304), Leucine A:305 (LEU A: 305), Glutamine A:306 (GLU A: 306), Histidine A:362 (HIS A: 362), Histidine A:398 (HIS A: 398), Proline A:403 (PRO A: 403) that site-4(C₄) and site-3(C₃) have more incredible binding energy than site-2 (C₂), while site-1(C₁) has less binding energy than site-2(C₂).

We noticed that the active amino acids at the site-3 (C₃) and experimentally reported site-4 (C₄) correspond, including Aspartic acid A:72, (ASP A:72), Trptophan A:84 (TRP A:84), Asparagine A:85 (ASN A:85), Proline A:86 (PRO A:86), Glycine A:117 (GLY A:117), Glycine A:118 (GLY A:118), Glycine A:119 (GLY A:119), Typtophan A:121 (TYR A:121), Serine A:122 (SER A:122), Tyrptophan A:130 (TYR A:130), Glutamine A:199 (GLU A:199), Serine A:200 (SER A: 200), Trptophan A:279 (TRP A:279), Leucine A:282 (LEU A:282), Serine A:286 (SER A:286), Isolecuine A:287 (ILE A:287), Phenylalanine A:288 (PHE A:288), Arginine A:289 (ARG A:289), Phenylalanine A:290 (PHE A:290), Phenylalanine A:330, (PHE A: 330), Phenylalanine A:331 (PHE A:331), Tyrosine A:334 (TYR A:334), Glycine A:335 (GLY A:335), Histidine A:440 (HIS A: 440), Glycine A:441 (GLY A:441) are amino acid interaction and shown better binding energy.

To understand the concept of a specific cavity in a protein, one starts by describing the three-dimensional structure of proteins. The lengthy complex structures known as proteins develop from chains of amino acids. The specialized folding and twisting of these amino acid indicators create a particular protein substrate. Suppose the amino acid residues are arranged in certain nooks or cavities within this structure. In that case, they may interact with the substrate in a certain way or go through conformational changes with the help of MESP studies.

3.2.6 Cavity-Drug-Likeness Molecules Interaction Matrix

Drug design depends on how proteins and ligands interact, given that numerous drugs modify the function or signaling of specific proteins by attaching to receptors. Understanding the protein's three-dimensional structure and locating possible ligand-binding sites are important steps in drug discovery.

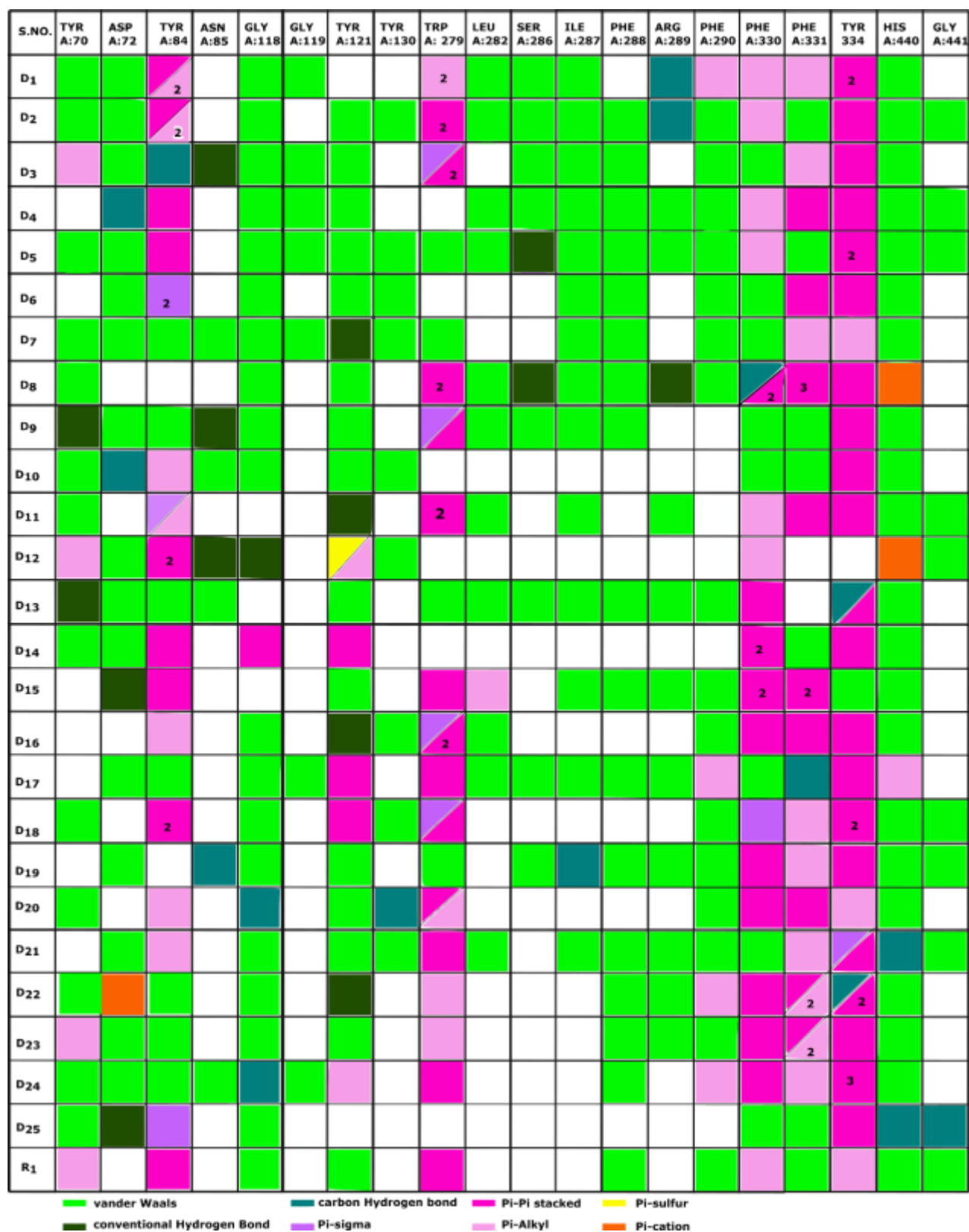


Figure 3.7 Interaction matrix of main protease with 25 drug-likeness molecules at the experimentally reported site-4 (C₄).

Computer modeling and simulation approaches are frequently utilized to forecast how ligands will interact with the protein and find ligands that are likely to have powerful and precise binding interactions. The Donepezil drug and 25 additional compounds with drug-likeness characteristics were put to docking simulations against the target protein acetylcholinesterase in the study. Here, Figure 3.7, displays the results of these interactions between the proteins and their associated ligands after various types of interactions have been recorded. Although van der Waals interactions and other weak interactions (WI) are also addressed, hydrogen bonds, conventional hydrogen, Pi-Pi, Pi-alkyl, Pi-sulphur, and Pi-cation are interactions considered with strong interactions (SI). Hydrogen bonds are shown as dark green, conventional hydrogen interactions as pine green, non-covalent interactions are depicted as pink for Pi-Pi interactions, light pink for Pi-alkyl, yellow for Pi-sulphur, red for Pi-cation, and light green for weak van der Waals interactions respectively.

We analyzed the interactions between the donepezil complex and the active amino acid residues in different complexes. In R1 complex, we observed 5 strong interactions (SI) and 7 weak interactions (WI) of active amino acids. Similarly, the D₁ complex showed interactions with 11 SI and 8 WI active amino acid residues. The D₂ complex exhibited interactions with 8 SI and 13 WI active amino acids, while the D₃ complex showed interactions with 8 SI and 10 WI active amino acids. D₄ complex exhibited interaction with 5 SI and 11 WI active amino acids. In the D₅ complex, interactions were observed with 5 SI and 15 WI active amino acids, while the D₆ complex showed interactions with 4 SI and 9 WI active amino acids. In the D₇ complex, we found interactions with 3 SI and 13 WI active amino acids. The D₈ complex displayed interactions with 12 SI and 7 WI active amino acids, while the D₉ complex showed interactions with 5 SI and 11 WI active amino acids. In the D₁₀ complex, we observed interactions with 3 SI and 8 WI active amino acids, and in the D₁₁ complex, there were interactions with 8 SI and 6 WI active amino acids. The D₁₂ complex exhibited interactions with 7 SI and 3 WI active amino acids, while the D₁₃ complex showed interactions with 4 SI and 12 WI active amino acids. In the D₁₄ complex, interactions were observed with 6 SI and 4 WI active amino acids, and in the D₁₅ complex, we found interactions with 8 SI and 7 WI active amino acids. The D₁₆ complex showed interactions with 8 SI and 5 WI active amino acids, while the D₁₇ complex exhibited interactions with 10 SI and 6 WI active amino acids. In the D₁₈ complex, we observed interactions with 9 SI and 6 WI active amino acids, while the D₁₉ complex showed interactions with 5 SI and 10 WI active amino acids. The D₂₀ complex displayed interactions with 8 SI and 4 WI active amino acids. In the D₂₁ complex, interactions were observed with 6 SI and 11 WI active amino acids, while the D₂₂

complex showed interactions with 11 SI and 6 WI active amino acids. The D₂₃ complex exhibited interactions with 7 SI and 8 WI active amino acids, and in the D₂₄ complex, we found interactions with 9 SI and 7 WI active amino acids. Finally, the D₂₅ complex displayed interactions with 5 SI and 4 WI active amino acids.

3.2.7 Physico-Chemical Properties and ADME Properties

To validate twenty-five Potential drug-likeness molecules, we have provided the physic-chemical properties, such as the Lipinski rule of five in Figure 3.8, for all twenty-five molecules. All the candidate molecules show molecular weight ranging from 284.23 to 484.63; hence, less than 500 Daltons or g/molten Partition coefficient ranges between 2.64-4.97 which is less than 5, and No. of HBD is between 0-3, and No. of HBA is between 1-5. The molecular polar surface area ranges between 26.17-73.36, is less than 140 angstroms, and the total number of the rotatable bond are ranged between 1-8, is less than 10. Figure 3.8 shows that most molecules followed the Lipinski rule of five. Hence, they can be studied for further analysis.

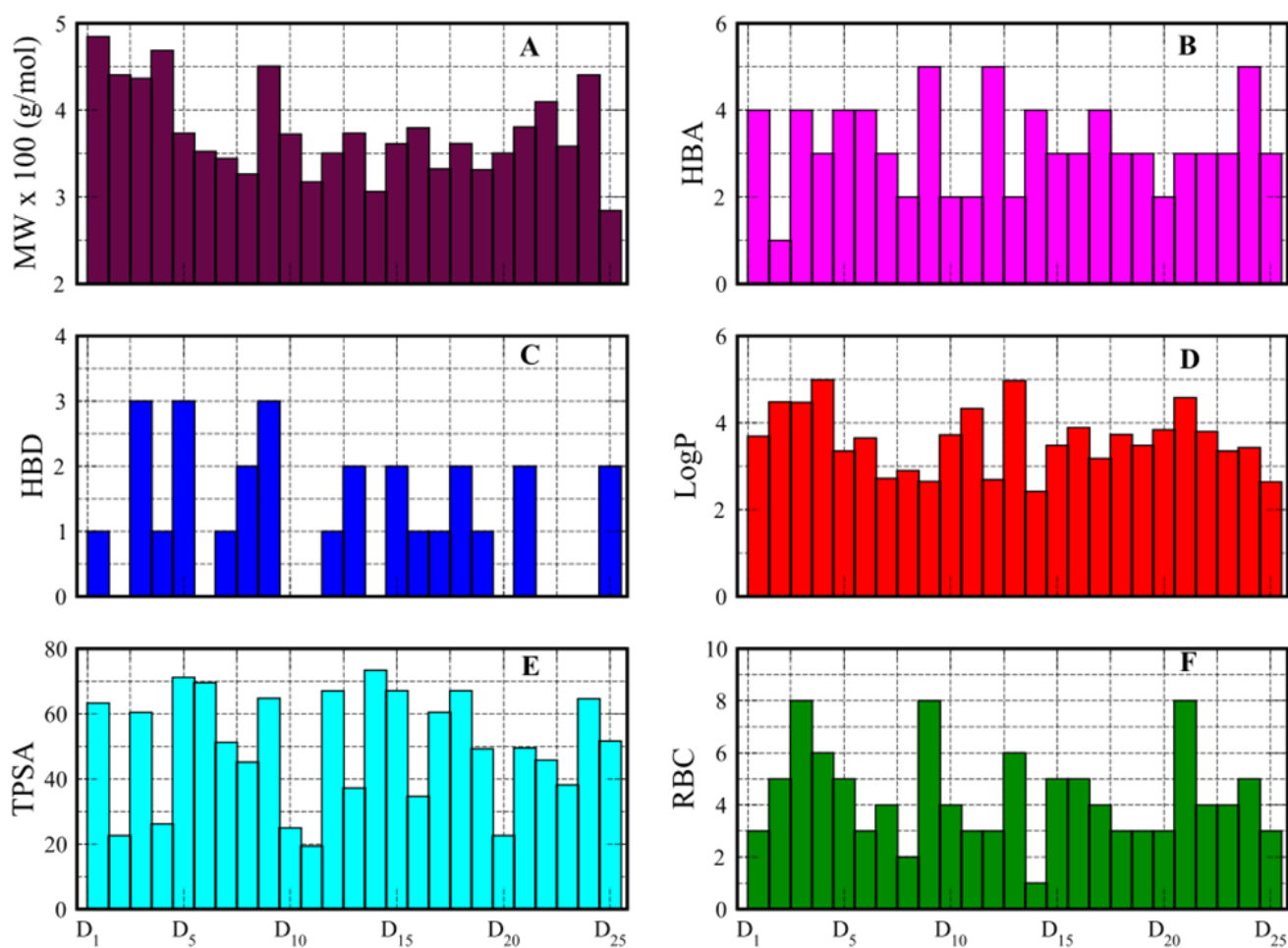


Figure 3.8 Physico-chemical properties of 25 drug-likeness molecules (A) Molecular weight (g/mol), (B) Hydrogen bond acceptor count, (C) Hydrogen bond donor count, (D) Partition coefficient (Logp), (E) Total polar surface area, and (F) Rotatable bond count.

ADME properties are generated using the SWISS ADME website to succeed with twenty-five molecules. ADME properties are most effective in predicting human pharmacokinetic properties to be successful in clinical trials. The top five potential molecules and reference molecule(R₁) have shown in Table 3.2, and ADME offers high gastrointestinal absorption (GI). Four of them have a permeability of the blood-brain barrier (BBB). The molecules determined for the Cytochrome P450 isomers are represented in Table 3.2.

Table 3.2 ADME / Pharmacokinetic proprieties of the site-wise selected molecules in gastrointestinal absorption. ^bBlood brain barrier permeant. ^cP-gp substrate. ^dCytochrome P450 family 1 subfamily A member2. ^eCytochrome P450 family 2 subfamily c member19. ^fCytochrome P450 family 2 subfamily C member 9. ^gCytochrome P450 family 2 subfamily D member 6. ^hCytochrome P450 family 3 subfamily A member4. ⁱSkin permeation in cm.

S.NO	GI abs ^a	BBB permeant ^b	P- gpsubstrate ^c	CYP1A2 Inhibitor ^d	CYP2C19 Inhibitor ^e	CYP2C9 Inhibitor ^f	CYP2D6 Inhibitor ^g	CYP3A4 Inhibitor ^h	Log K _p ⁱ
D1	High	Yes	No	No	No	No	No	Yes	-5.64
D2	High	Yes	Yes	No	No	No	Yes	Yes	-4.38
D3	High	Yes	Yes	No	No	No	Yes	No	-5.79
D4	High	Yes	Yes	No	No	No	Yes	No	-4.5
R1	High	Yes	Yes	No	No	No	Yes	Yes	-5.58

3.2.8 Correlation of Potential Drug-Likeness Molecules vs Site Wise Binding Energy

The CBE values for twenty-five molecules are displayed at each location, as indicated by Figure 3.9. The range of CBE is -4.0 to -10.6 kcal /mol for all four sites of the acetylcholinesterase. In this figure 3.9A, the X-axis represents the number of drug-like molecules, while the Y-axis represents the binding energy. In this Figure 3.9B, the X-axis represents the number of binding sites or cavities, while the Y-axis represents the binding energy. Figures 3.9 A and 3.9 B show that blue is the primary interaction site with the experimental site for all the drug-like molecules. CBE analysis showed that C₃ and C₄

contained more potential for protein and ligand complex binding energy than C₁ and C₂, as shown in Figure 3.9. The blue line of C₄ (Experimental reported active site) represents binding energy value -6.8 to 10.6 kcal/mol, and the black line C₁ was shown -3.5 to -6.2 kcal/mol binding energy. The red line for C₂ showed -4.1 to -7.0 kcal/mol, and the green line for C₃ was -6.7 to -10.6 kcal/mol. Two sites (C₃, C₄) are better than the others (C₁, C₂) based on their binding energies, as shown in Figure 3.9.

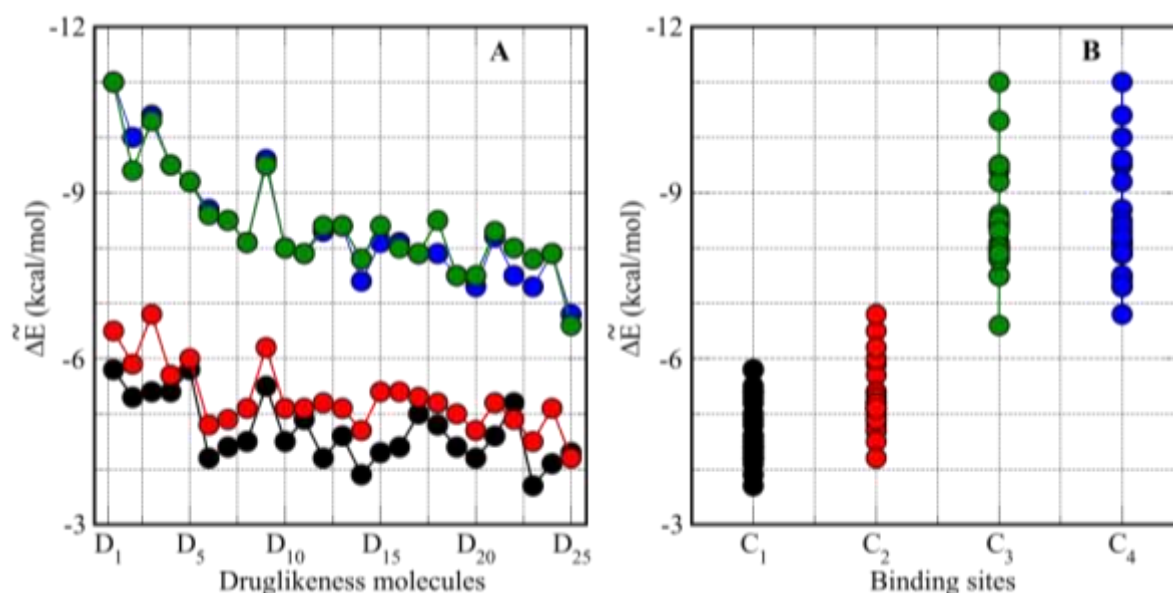


Figure 3.9 Corrected binding energy (CBE) trend curves for LORD 25 molecules at all four sites. (A) The X-axis is 25 drug-likeness molecules (D₁ to D₂₅), and Y-axis is CBE values and (B) The X-axis is MESP-suggested sites (C₁, C₂, C₃), and Y-axis is CBE values.

3.2.9 Molecular Dynamics (MD) Simulations

Acetylcholinesterase enzyme receptor with top four drug-like molecules and donepezil drug molecule is further studied by MD simulations using GROMACS 5.1.2 package. To compare the stability, we have also conducted MD simulations for Donepezil. The MD simulations were performed using the CHARMM force field to assess the energy evaluations of the protein. The ligand topology was also determined using the CHARMM General Force Field (CGenFF). This combination of force fields allowed for comprehensive and accurate evaluations of both the protein and ligand components during the MD simulations.⁵⁸ After the protein-ligand complex was successfully docked, it was immersed in a cubic box using the TIP3 water model for solvation. An appropriate number of chlorine ions (CL⁻) were added to achieve a neutral charge, followed by energy minimization using the steepest descent

approach with 10,000 steps. The NVT ensemble was utilized to maintain a temperature of 298 K throughout the simulation, while the NPT ensemble, with a pressure range of 1.0 bar to 250 bars, was employed to simulate the ligand-protein complex. The Particle Mesh Ewald and LINCS algorithms were utilized for short-range electrostatic interactions. A Van der Waals distance cut-off of 10 Å was applied to constrain all bonds. The stability of the system was maintained through simulations lasting 100 ns. The RMSD plots are shown in Figure 3.10 for four top drug molecules with Donepezil.

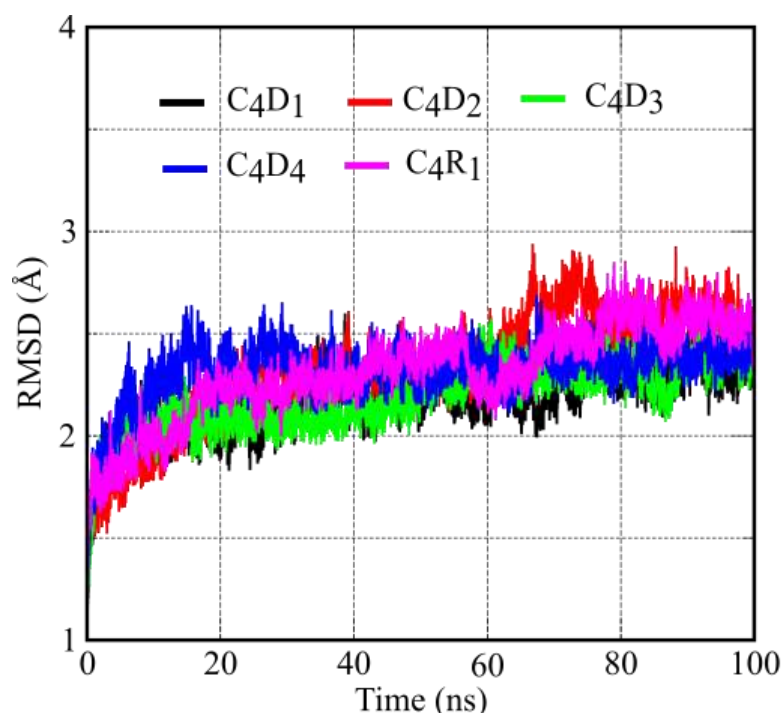


Figure 3.10 MD simulation studies of the RMSD for the top four drug-likeness molecules (D₁, D₂, D₃, D₄) and Reference (R₁).

The interaction between the AChE target and four drug-like molecules, along with Donepezil, was compared. To analyze the dynamic behavior, stability of interaction, and structural changes in the complexes, RMSD calculations were employed. Molecular dynamics (MD) simulations of AChE were performed for three ligand molecules over a duration of 100 ns. Molecular alignment was assessed by superimposing the average structure at the start of the simulation (0 ns) with the output structures at 100 ns. The stability of the protein-ligand complexes for the four-candidate drug-like molecules was evaluated based on the RMSD and RMSF values obtained from the MD simulations. Figure 3.10 illustrates the tight binding of these four candidate ligands within site 1 (C₁) of AChE, as indicated by their RMSD values.

3.2.10 Free Energy Calculations

The protein-ligand complexes binding energy, or binding energy (BE) as it is more often known, was determined using the *g-mmpbsa* software. The molecular mechanics (MM) force field is used to calculate the vacuum potential energy resulting from different interactions, such as bound and unbound forces. The MM force field uses the Lennard-Jones potential function to determine the van der Waals (Evdw) energy contribution and the Coulomb potential function to determine the electrostatic energy contribution. To determine the polar contribution and calculate the free energy of solvation, the Poisson-Boltzmann equation's solution is also used. In the case of non-polar contribution, it is typically believed that the solvent-accessible surface area (SASA) and the non-electrostatic solvation energy have a direct correlation. The non-polar energy term includes the van der Waals interaction and the attractive and repellent forces produced by cavities created between the solute and solvent. Snapshots of the equilibrated area from the molecular dynamics (MD) trajectory are taken during the binding energy (BE) computations. The computations are done using the default settings supplied by Kumari et al. using the *MmPbsaDecomp.py* script, which is part of the *g-mmpbsa* package.

For the five protein-ligand complexes that were chosen, calculations of binding energy (BE) using the MM-PBSA program were made. The top four drug-likeness as D₁, D₂, D₃, D₄, and Donepezil's reference (R₁) molecule all had BE values between 0-10 ns, 30-40 ns, 60-70 ns, and 90-100 ns. The BE distribution calculation revealed that it significantly contributed to the overall BE, as shown in Table 3.3.

Table 3.3 Free energy calculation of the top four drug-likeness (D₁-D₄) molecules and Donepezil (R₁).

Time intervals (ns)	Complex free energies (kJ/mol)				
	D ₁	D ₂	D ₃	D ₄	R ₁
0-25	-134.4 ±/-15.0	-144.3 +/- 17.3	-113.4 +/- 15.8	-110.9 +/- 19.2	-111.1 +/- 16.2
25-50	-134.5 +/- 13.5	-120.7 +/-12.8	-115.2 +/- 16.6	-112.7 +/- 14.7	-122.6 +/- 11.5
50-75	-136.2 +/- 12.3	-145.8 +/- 17.3	-114.1 +/- 16.2	-117.8 +/- 12.5	-116.7 +/- 16.1
75-100	-143.5 +/- 12.7	-119.0 +/- 10.7	-124.7 +/- 12.9	-118.9 +/- 14.2	-121.5 +/- 11.7

3.3 Conclusions

The present research introduces an innovative approach to designing drug-like molecules that target proteins at their binding sites. The current study provides a novel method for designing drug-like compounds targeting target proteins on-site. The molecular structure is built through LSTM-suggested fragment addition. Due to LORD's ability to include physicochemical features through corrected binding energy training, the probability of unnecessary molecule generations is extremely rare. The LORD uses a reverse engineering technique, which involves learning about the environment of the target binding site and then constructing molecules. Hence, the produced molecule has an excellent possibility of becoming a drug-like molecule. LORD has the advantage of being computationally cheap due to the platform technique used while scanning the MESP cavities, and it will automatically find target binding sites. LORD is smart enough to build potential molecules based on the target site's environment. Thus, the molecules generated will be appropriate for that site. This work investigates LORD extensively in this study on the Acetylcholinesterase (AChE), which causes AD. We discovered four putative binding sites, one of which was an experimentally active site. Four potential binding sites are utilized to build molecules by LORD, producing 25 potential drug candidates. The experimentally active site, when compared to the other sites in CBE values, is the most dominant of the four binding sites. LORD 25 molecules have good Corrected binding energy (CBE) (kcal/mol), AutoDockvina binding energy (BE), physicochemical, and ADME qualities when compared to commercially available medications such as Donepezil. In addition, MD simulations of the top-four potential drug-like compounds were investigated, and it was discovered that these are more stable in protein during long simulation runs than donepezil drugs.

REFERENCES

1. Gupta, M.; Weaver, D. F. Alzheimer's: The ABCDE Paradigm. *ACS Chem. Neurosci.* **2022**, *13* (9), 1355–1357.
2. Sehar, U.; Rawat, P.; Reddy, A. P.; Kopel, J.; Reddy, P. H. Amyloid Beta in Aging and Alzheimer's Disease. *Int. J. Mol. Sci.* **2022**, *23* (21), 12924.
3. Khachaturian, A. S.; Dengel, A.; Dočkal, V.; Hroboň, P.; Tolar, M. Editorial: Accelerating Innovations for Enhanced Brain Health. Can Artificial Intelligence Advance New Pathways for Drug Discovery for Alzheimer's and Other Neurodegenerative Disorders? *J. Prev. Alzheimers Dis.* **2023**, *10* (1), 1–4.
4. Prince, M.; Wimo, A.; Guerchet, M.; Ali, G.-C.; Wu, Y.-T.; Prina, M.; Kit, Y.; Xia, Z. *Alz.co.uk*. <https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf> (accessed 2023-03-06)
5. Wang, X.; Li, P.; Ding, Q.; Wu, C.; Zhang, W.; Tang, B. Observation of Acetylcholinesterase in Stress-Induced Depression Phenotypes by Two-Photon Fluorescence Imaging in the Mouse Brain. *J. Am. Chem. Soc.* **2019**, *141* (5), 2061–2068.
6. van der Westhuizen, C. J.; Stander, A.; Riley, D. L.; Panayides, J.-L. Discovery of Novel Acetylcholinesterase Inhibitors by Virtual Screening, in Vitro Screening, and Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2022**, *62* (6), 1550–1572.
7. Akhoun, B. A.; Choudhary, S.; Tiwari, H.; Kumar, A.; Barik, M. R.; Rathor, L.; Pandey, R.; Nargotra, A. Discovery of a New Donepezil-like Acetylcholinesterase Inhibitor for Targeting Alzheimer's Disease: Computational Studies with Biological Validation. *J. Chem. Inf. Model.* **2020**, *60* (10), 4717–4729.
8. Brewster, J. T., 2nd; Dell'Acqua, S.; Thach, D. Q.; Sessler, J. L. Classics in Chemical Neuroscience: Donepezil. *ACS Chem. Neurosci.* **2019**, *10* (1), 155–167.
9. Dighe, S. N.; Tippana, M.; van Akker, S.; Collet, T. A. Structure-Based Scaffold Repurposing toward the Discovery of Novel Cholinesterase Inhibitors. *ACS Omega* **2020**, *5* (48), 30971–30979.
10. Galimberti, D.; Scarpini, E. Old and New Acetylcholinesterase Inhibitors for Alzheimer's disease. *Expert Opin. Investig. Drugs* **2016**, *25* (10), 1181–1187.
11. Siopa, F.; Frade, R. F. M.; Diniz, A.; Andrade, J. M.; Nicolai, M.; Meirinhos, A.; Lucas, S. D.; Marcelo, F.; Afonso, C. A. M.; Rijo, P. Acetylcholinesterase Choline-Based Ionic Liquid Inhibitors: In Vitro and in Silico Molecular Docking Studies. *ACS Omega* **2018**, *3* (12), 17145–17154.
12. Kryger, G.; Silman, I.; Sussman, J. L. Structure of Acetylcholinesterase Complexed with E2020 (Aricept): Implications for the Design of New Anti-Alzheimer Drugs. *Structure* **1999**, *7* (3), 297–307.
13. Soyer, Z.; Uysal, S.; Parlar, S.; Tarikogullari Dogan, A. H.; Alptuzun, V. Synthesis and Molecular Docking Studies of Some 4-Phthalimidobenzenesulfonamide Derivatives as Acetylcholinesterase and Butyrylcholinesterase Inhibitors. *J. Enzyme Inhib. Med. Chem.* **2017**, *32* (1), 13–19.
14. Asghar, A.; Yousuf, M.; Fareed, G.; Nazir, R.; Hassan, A.; Maalik, A.; Noor, T.; Iqbal, N.; Rasheed, L. Synthesis, Acetylcholinesterase (AChE) and Butyrylcholinesterase (BuChE) Activities, and Molecular Docking Studies of a Novel Compound Based on Combination of Flurbiprofen and Isoniazide. *RSC Adv.* **2020**, *10* (33), 19346–19352.
15. Islam, M. S.; Al-Majid, A. M.; Azam, M.; Verma, V. P.; Barakat, A.; Haukka, M.; Elgazar, A. A.; Mira, A.; Badria, F. A. Construction of Spirooxindole Analogues Engrafted with Indole and Pyrazole Scaffolds as Acetylcholinesterase Inhibitors. *ACS Omega* **2021**, *6* (47), 31539–31556.
16. Hemaïda, A. Y.; Hassan, G. S.; Maarouf, A. R.; Joubert, J.; El-Emam, A. A. Synthesis and Biological Evaluation of Thiazole-Based Derivatives as Potential Acetylcholinesterase Inhibitors. *ACS Omega* **2021**, *6* (29), 19202–19211.

17. Elghazawy, N. H.; Zaafar, D.; Hassan, R. R.; Mahmoud, M. Y.; Bedda, L.; Bakr, A. F.; Arafa, R. K. Discovery of New 1,3,4-Oxadiazoles with Dual Activity Targeting the Cholinergic Pathway as Effective Anti-Alzheimer Agents. *ACS Chem. Neurosci.* **2022**, *13* (8), 1187–1205.
18. Elghazawy, N. H.; Zaafar, D.; Hassan, R. R.; Mahmoud, M. Y.; Bedda, L.; Bakr, A. F.; Arafa, R. K. Discovery of New 1,3,4-Oxadiazoles with Dual Activity Targeting the Cholinergic Pathway as Effective Anti-Alzheimer Agents. *ACS Chem. Neurosci.* **2022**, *13* (8), 1187–1205.
19. Gupta, M.; Weaver, D. F. Alzheimer's: The ABCDE Paradigm. *ACS Chem. Neurosci.* **2022**, *13* (9), 1355–1357.
20. Kumar, A.; Sharma, A. Computational Modeling of Multi-Target-Directed Inhibitors against Alzheimer's Disease. In *Neuromethods*; Springer New York: New York, NY, **2018**; pp 533–571.
21. Dileep, K. V.; Ihara, K.; Mishima-Tsumagari, C.; Kukimoto-Niino, M.; Yonemochi, M.; Hanada, K.; Shirouzu, M.; Zhang, K. Y. J. Crystal Structure of Human Acetylcholinesterase in Complex with Tacrine: Implications for Drug Discovery. *Int. J. Biol. Macromol.* **2022**, *210*, 172–181.
22. Akhoun, B. A.; Choudhary, S.; Tiwari, H.; Kumar, A.; Barik, M. R.; Rathor, L.; Pandey, R.; Nargotra, A. Discovery of a New Donepezil-like Acetylcholinesterase Inhibitor for Targeting Alzheimer's Disease: Computational Studies with Biological Validation. *J. Chem. Inf. Model.* **2020**, *60* (10), 4717–4729.
23. Malik, R.; Choudhary, B. S.; Srivastava, S.; Mehta, P.; Sharma, M. Identification of Novel Acetylcholinesterase Inhibitors through E-Pharmacophore-Based Virtual Screening and Molecular Dynamics Simulations. *J. Biomol. Struct. Dyn.* **2017**, *35* (15), 3268–3284.
24. Stavrakov, G.; Philipova, I.; Lukarski, A.; Atanasova, M.; Georgiev, B.; Atanasova, T.; Konstantinov, S.; Doytchinova, I. Discovery of a Novel Acetylcholinesterase Inhibitor by Fragment-Based Design and Virtual Screening. *Molecules* **2021**, *26* (7), 2058.
25. Bao, L.-Q.; Baecker, D.; Mai Dung, D. T.; Phuong Nhung, N.; Thi Thuan, N.; Nguyen, P. L.; Phuong Dung, P. T.; Huong, T. T. L.; Rasulev, B.; Casanola-Martin, G. M.; Nam, N.-H.; Pham-The, H. Development of Activity Rules and Chemical Fragment Design for in Silico Discovery of AChE and BACE1 Dual Inhibitors against Alzheimer's Disease. *Molecules* **2023**, *28* (8).
26. Nguyen, T. H.; Tran, P.-T.; Pham, N. Q. A.; Hoang, V.-H.; Hiep, D. M.; Ngo, S. T. Identifying Possible AChE Inhibitors from Drug-like Molecules via Machine Learning and Experimental Studies. *ACS Omega* **2022**, *7* (24), 20673–20682.
27. Manandhar, S.; Sankhe, R.; Priya, K.; Hari, G.; Kumar B, H.; Mehta, C. H.; Nayak, U. Y.; Pai, K. S. R. Molecular Dynamics and Structure-Based Virtual Screening and Identification of Natural Compounds as Wnt Signaling Modulators: Possible Therapeutics for Alzheimer's Disease. *Mol. Divers.* **2022**, *26* (5), 2793–2811.
28. Van der Westhuizen, C. J.; Stander, A.; Riley, D. L.; Panayides, J.-L. Discovery of Novel Acetylcholinesterase Inhibitors by Virtual Screening, in Vitro Screening, and Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2022**, *62* (6), 1550–1572.
29. Sarkar, B.; Alam, S.; Rajib, T. K.; Islam, S. S.; Araf, Y.; Ullah, M. A. Identification of the Most Potent Acetylcholinesterase Inhibitors from Plants for Possible Treatment of Alzheimer's Disease: A Computational Approach. *Egypt. J. Med. Hum. Genet.* **2021**, *22* (1).
30. Ahmad, S. S.; Khan, M. B.; Ahmad, K.; Lim, J.-H.; Shaikh, S.; Lee, E.-J.; Choi, I. Biocomputational Screening of Natural Compounds against Acetylcholinesterase. *Molecules* **2021**, *26* (9).
31. Zhang, Y.; Zhang, S.; Xu, G.; Yan, H.; Pu, Y.; Zuo, Z. The Discovery of New Acetylcholinesterase Inhibitors Derived from Pharmacophore Modeling, Virtual Screening, Docking Simulation and Bioassays. *Mol. Biosyst.* **2016**, *12* (12), 3734–3742.

32. Hemaida, A. Y.; Hassan, G. S.; Maarouf, A. R.; Joubert, J.; El-Emam, A. A. Synthesis and Biological Evaluation of Thiazole-Based Derivatives as Potential Acetylcholinesterase Inhibitors. *ACS Omega* **2021**, 6 (29), 19202–19211.

CHAPTER 4

Designing of Potential Drug-Likeness Molecules for COVID-19

4.1 Introduction to COVID-19

The recently identified RNA virus coronavirus disease-2019 (COVID-19) or severe acute respiratory syndrome coronavirus II (SARS-CoV-2) represents a significant health risk to the general public.¹ The most common symptoms of the coronavirus are fever, dry cough, sore throat, headache, loss of taste, and severe symptoms such as difficulty breathing and chest pain, which sometimes leads to pneumonia.² COVID-19-infected people show mild to moderate illness. SARS-CoV-2 belongs to the Coronaviridae family of Coronaviruses are enveloped, positive-sense, single-stranded ribonucleic acids (RNA) genomes like MERS-CoV and SARS-CoV.³ The novel coronavirus genome consists of many essential proteins such as nucleocapsid protein, Spike protein (S), and Envelope protein (E), Membrane protein (M). These proteins showed crucial roles in the gene expression and replication process of the coronavirus that will lead to the breaking down of the polyproteins in the human genome.⁴

The 3C-like protease (3CL-pro) is a different designation of the SARS-CoV-2 virus's main protease (M^{pro}). By using replicase enzymes, this protease is in charge of breaking down polyproteins. M^{pro} is a dimeric protein comprising the protomers A and B, which are two identical parts. Each protomer is made up of three different domains: domain I, which is made up of residues 8 to 101, domain II (residues 102 to 184) and has an antiparallel beta-barrel structure, and domain III (residues 201 to 303) and contains a lot of alpha helices. A flexible loop region spanning residues 185 to 200 connects domain II to M^{pro} . Catalytic dyads, made up of Cys145 and His41, are essential to the activity of enzyme. In the space between domains I and II, M^{pro} substrate-binding site is found. The crystal structure of the main protease with inhibitor N3 complex is PDB ID: 6LU7.⁵ Fragment molecular orbitals (FMO) method-based interaction analysis on Moro by splitting into five fragments indicates that the essential amino acid residues are His41, His163, His164, and Glu166 due to hydrogen bond interactions.⁶ M^{pro} is an essential drug target due to its indispensable role in viral replication of the life cycle and transcription virus inside the host. And hence, one can develop effective antiviral drugs for treating COVID-19 infection.^{7,8} CADD plays a significant role in the drug discovery journey and holds particular significance in combating the COVID-19 pandemic.^{9,10}

Recent literature studies indicate that drug repurposing processes are an essential strategy to identify drugs by applying three types of proteins such as main protease (M^{pro}), Papain-like-protease (PL^{pro}), RNA-dependent polymerase (RdRp) or Nucleotide inhibitor to combat COVID-19 pandemic.^{11, 12, 13} The computational studies on existing antiviral medicines¹⁴ such as Remdesivir, Hydroxyl ethylamine derivative Favipiravir, Ritonavir, Lopinavir, Oseltamivir,

Ribavir, Galidesivir, and Riamilovir are promising to combat COVID-19.¹⁵ Also, there are other potential medicines such as Chloroquine, hydroxyl Chloroquine (HCQ), Mycopheacidsc acids (MPA), Premirrolast, isoniazid, and Eriodictyol, Azithromycin. Telaprevir and paritaprevir are studied by covalent interactions, and asunaprevir, simeprevir, and paritaprevir are studied by noncovalent interactions, and Hydroxyl ethylamine derivatives in the treatment for COVID-19.¹⁶ Previous a selection of relevant literature, while our focus in this section is to discuss the methodology and offer a concise summary in Table 4.1.

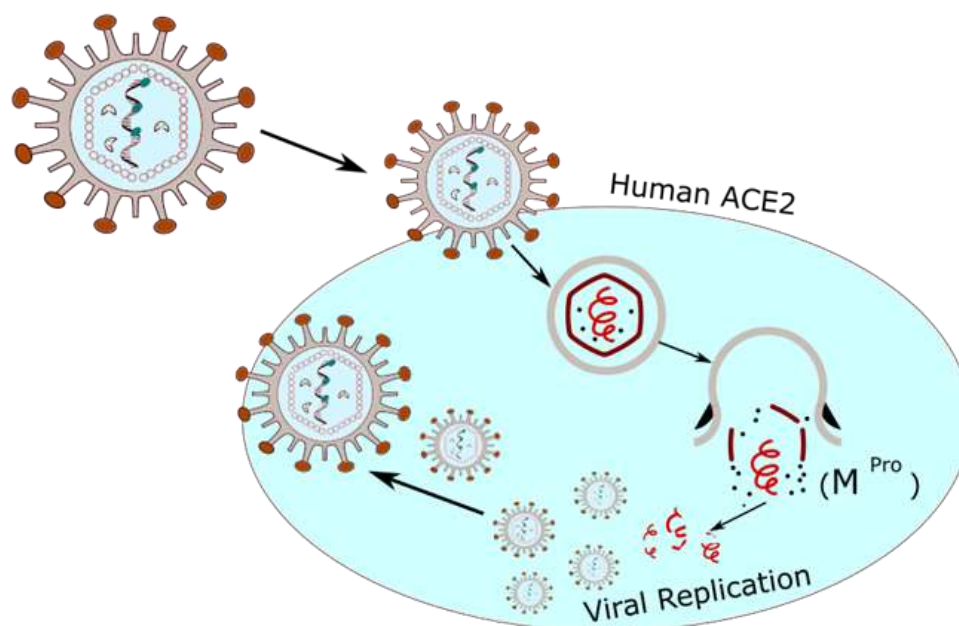


Figure 4.1 Mechanism action of viral replication.

Table 4.1 Summary of literature on methodology.

Methodology	Summary
X-ray crystallography	The crystal structure of SARS-CoV-2 M ^{Pro} provides crucial insights into the design of potent α -ketoamide inhibitors, facilitating the development of enhanced antiviral medications to combat COVID-19. By leveraging these important details, researchers can devise more effective strategies to target the viral protease, potentially leading to improved treatment options against the disease. ¹⁷
COVID-19, molecular docking, virtual screening, machine learning, and molecular dynamics simulations.	This critical overview assesses the computational approaches employed for COVID-19 drug discovery. The study identifies key techniques such as molecular docking, virtual screening, machine learning, and molecular dynamics simulations through an extensive literature review. The strengths and limitations of these approaches are evaluated, shedding light on their contributions to the search for effective treatments for COVID-19. ¹⁸
X-ray crystallography,	The recent breakthrough in COVID-19 research has unveiled the structure of the M ^{Pro} enzyme, a key target for potential drug development against the virus.

virtual screening and biochemical assays	Alongside this discovery, scientists have also identified inhibitors that could effectively target and block the activity of the M ^{pro} enzyme. ¹⁹
SARS-CoV-2 virus, machine learning, Pharmacology	This review of 17,000 studies highlighted the significant role of computational approaches in combating COVID-19. The findings demonstrated that AI-driven methodologies have contributed to various aspects, including epidemiological modeling, drug discovery, vaccine development, and patient care. By leveraging computational tools and techniques, researchers have been able to accelerate the understanding of the virus, facilitate data-driven decision-making, and enhance public health strategies to mitigate the impact of the pandemic effectively. ²⁰
Molecular docking and MD simulations	Virtual screening and structural optimization techniques have successfully pinpointed potential inhibitors for the M ^{pro} enzyme. This breakthrough discovery offers a promising pathway toward creating highly effective antiviral medications. Researchers have identified compounds that exhibit potent inhibitory effects on M ^{pro} by utilizing computational methods and refining the molecular structure. ²¹
X-ray crystallography and molecular modeling	Establishing the structural foundation for developing inhibitors against the M ^{pro} enzyme of SARS-CoV has provided valuable insights, facilitating the progress in drug development. This advancement lays the groundwork for designing effective inhibitors to target the M ^{pro} enzyme, potentially contributing to developing therapeutic interventions against SARS-CoV. ²²
Virtual screening and molecular docking	Promising novel small-molecule inhibitors have been identified to target the SARS-CoV-2 M ^{pro} enzyme, providing potential candidates for developing effective antiviral therapies against COVID-19. These inhibitors hold promise in combating the virus by inhibiting the activity of the M ^{pro} enzyme, which plays a crucial role in viral replication. ²³
Virtual screening, molecular docking and MD simulations	Virtual screening and molecular dynamics simulations have been employed to identify potential inhibitors for the M ^{pro} enzyme in SARS-CoV-2. This breakthrough offers hope for developing powerful antiviral treatments, instilling optimism in the ongoing battle against the COVID-19 pandemic. ²⁴
Structure-based virtual screening and MD simulations	Promising antiviral drug candidates have emerged by identifying potent M ^{pro} inhibitors possessing favorable drug-likeness characteristics. These findings showcase their potential as effective agents in developing antiviral medications. Their favorable properties make them worthy contenders for further investigation and potential use in combating viral infections. ²⁵
Computational and experimental studies	Boceprevir, GC-376, and calpain inhibitors II and XII exhibit promising antiviral activity by specifically targeting the main protease of SARS-CoV-2, thereby impeding viral replication. These compounds hold significant potential as therapeutic options for combating the virus, offering valuable strategies for antiviral intervention. By inhibiting the main viral protease, they disrupt essential viral processes and may contribute to the development of effective antiviral treatments. ²⁶
In silico screening and MD simulations	The identification of FDA-approved drugs as potential inhibitors for the SARS-CoV-2 M ^{pro} enzyme opens up possibilities for repurposing them as treatments for COVID-19. These drugs exhibit inhibitory properties, indicating their potential to target and hinder the activity of the M ^{pro} enzyme associated with the virus. This discovery provides an avenue for exploring existing medications in the fight against COVID-19. ²⁷
Molecular docking and MD simulations	New possibilities for COVID-19 treatment have emerged with the identification of potential inhibitors that target the M ^{pro} enzyme. These inhibitors have been found within FDA-approved drugs, providing opportunities to repurpose existing

medications. This discovery paves the way for exploring alternative therapeutic options for COVID-19 by leveraging the safety profiles and availability of already approved drugs.²⁸

Fragment
molecular orbital
calculations and
MD simulations

Quantum mechanics/molecular mechanics simulations were employed to assess the effectiveness of M^{pro} inhibitors, yielding crucial insights into their binding mechanisms. These simulations offered valuable contributions to developing new drugs, as they shed light on the interactions between the inhibitors and their target, paving the way for enhanced drug design and optimization.²⁹

X-ray
crystallography
and SBDD

The constant evolution and diversity of SARS-CoV-2 M^{pro} inhibitors pose significant challenges in antiviral drug design. Furthermore, the emergence of drug resistance further complicates the development of effective treatments. It is imperative to devise new strategies and approaches to combat these challenges and create novel antiviral drugs to effectively target the virus and overcome resistance mechanisms.³⁰

Virtual
screening and
MD simulations

Novel phytochemicals derived from medicinal plants have been discovered through molecular docking and deep learning techniques, showcasing their promising inhibitory effects against the SARS-CoV-2 M^{pro} protein.³¹

X-ray
crystallography,
fragment
screening, and
medicinal
chemistry

Researchers are making significant strides in developing potent noncovalent inhibitors targeting the SARS-CoV-2 M^{pro} protein. These inhibitors have demonstrated promising efficacy in preclinical studies, effectively blocking the activity of the viral protease and impeding viral replication. This approach offers a potential avenue for developing effective antiviral therapies against COVID-19 without relying on covalent binding mechanisms.³²

N-heterocyclic
compounds,
DFT
calculations, and
molecular
docking

The study employed molecular docking and DFT calculations to evaluate the binding affinity and interaction between the N-heterocycles and viral proteins. The aim was to predict their antiviral activity and identify key structural features contributing to their effectiveness.³³

Insilico antiviral
screening,
molecular
docking, and
MD simulations

The study utilized computational techniques to examine the anti-SARS-CoV-2 activity of chloroquine and its analogy. Additionally, the researchers performed a virtual screening of main protease inhibitors to identify potential compounds with inhibitory effects on the virus and highlight promising compounds that could be further studied for their efficacy against SARS-CoV-2.³⁴

4.2 Results and Discussion

4. .2.1 MESP Calculation for Identification Binding Sites M^{pro} in COVID-19 Target

Figure 4.2 shows the whole MESP distribution with two layers. Surface cavity analysis explores the most negative areas, such as the inner blue color lobes marked in black circles with notations C1-C5. MESP function values are projected onto the plane to pinpoint the precise location of the possible binding site. The scanning MESP function values inside the plane produced surface cavities.

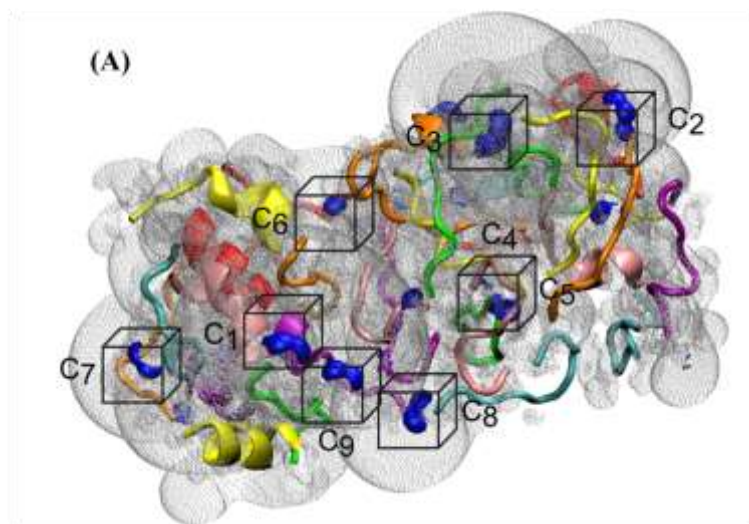
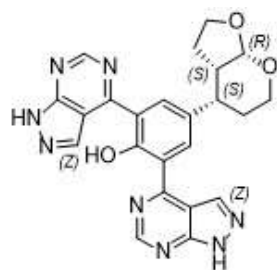


Figure 4.2 Substrate structures (A) and (B) MESP distribution of the M^{pro} target protein.

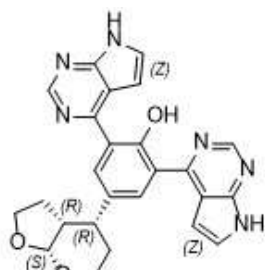
The surface cavities were expected to have the most significant number of possible binding sites. As illustrated in Figure 4.2, the surface cavities help to comprehend the distribution of MESP function values by tracking the gradients as the color changes. The inner and outer circles represent the two negative MESP function isosurfaces. The color distribution suggests that drug molecules can be ingested into the cavity by following the charge distribution gradient. This analysis identified the top five probable binding sites for future investigation.

4.2.2 Design of Potential Candidate Molecules

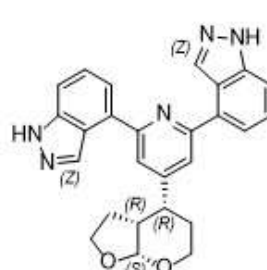
The chapter contains 50 potential candidates developed utilizing the LORD method and compared to Remdesivir (R₁), a presently available drug in the market.



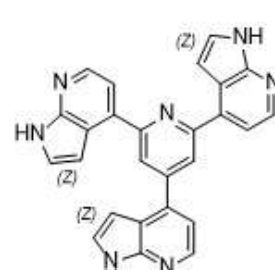
D₁



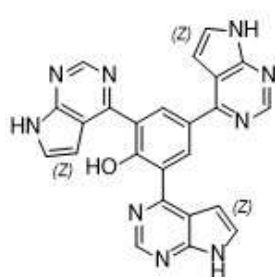
D₂



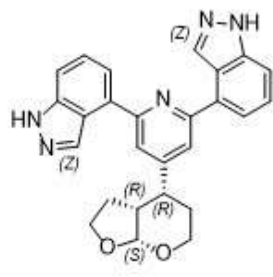
D₃



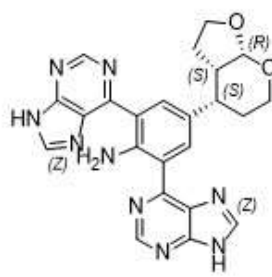
D₄



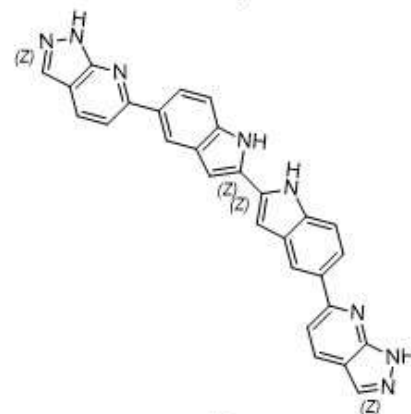
D₅



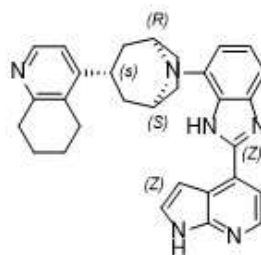
D₆



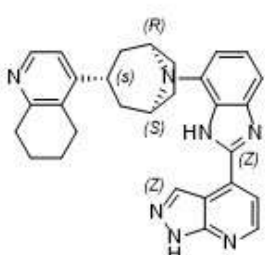
D₇



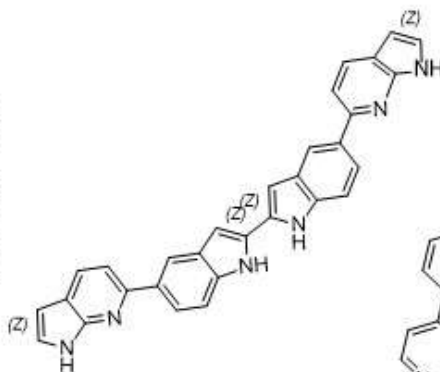
D₈



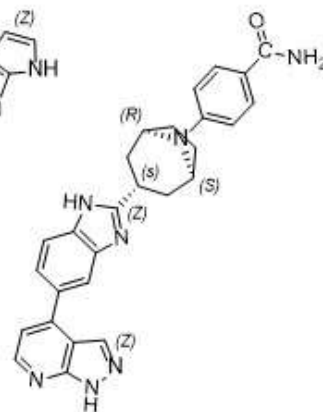
D₉



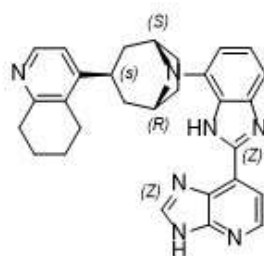
D₁₀



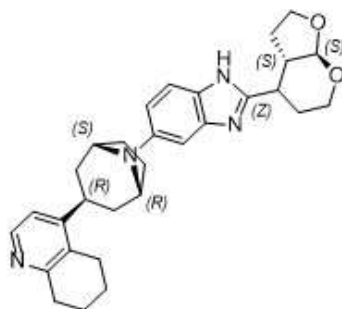
D₁₁



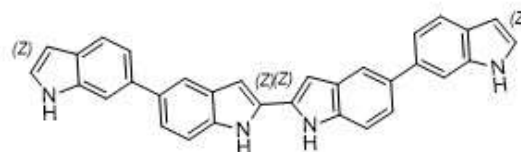
D₁₂



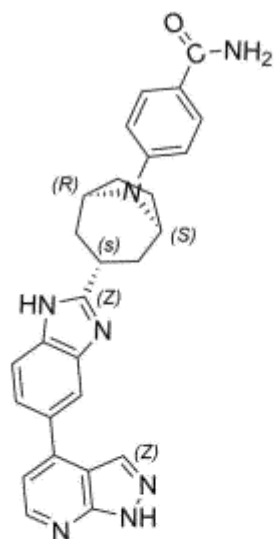
D₁₃



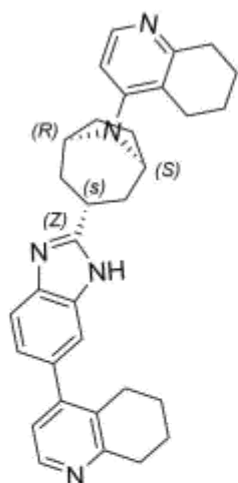
D₁₄



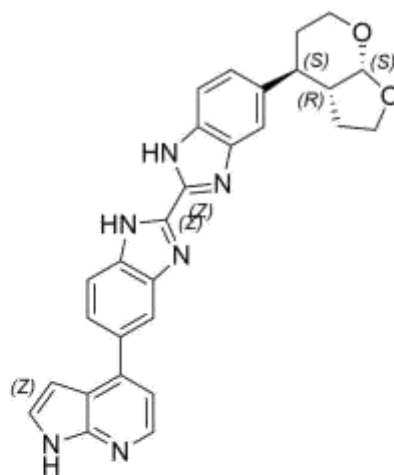
D₁₅



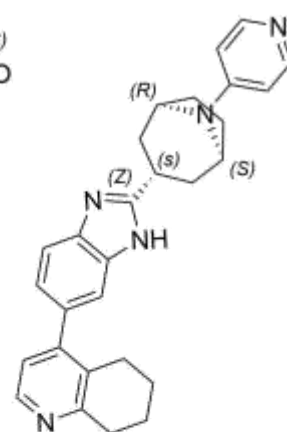
D16



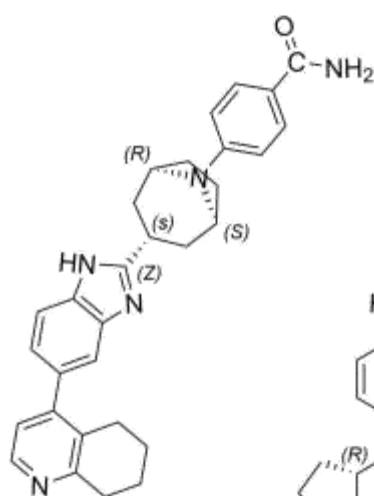
D17



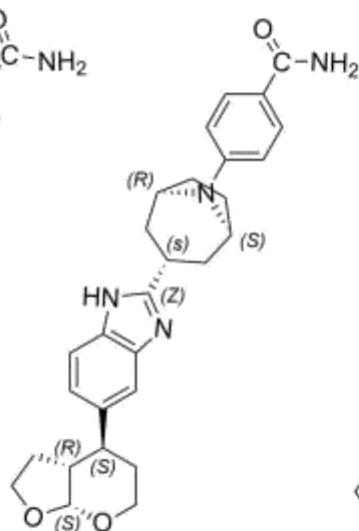
D18



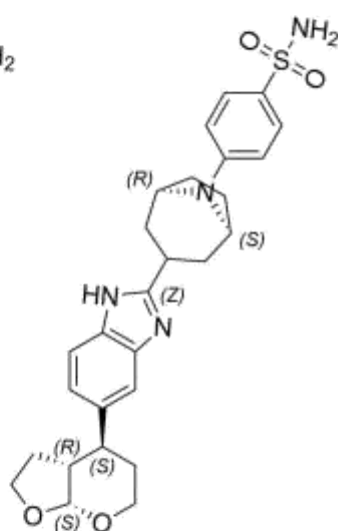
D19



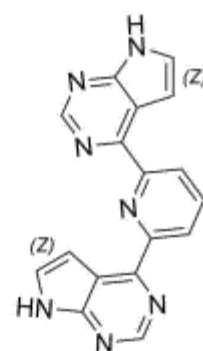
D20



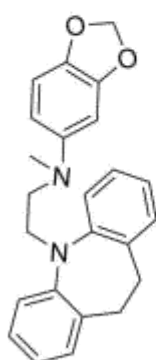
D21



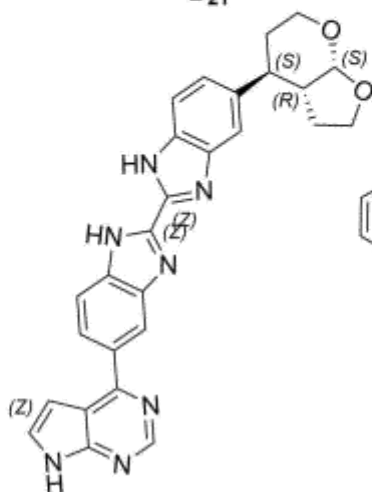
D22



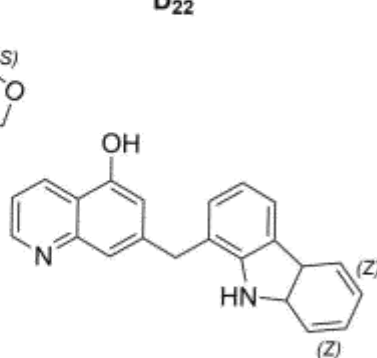
D23



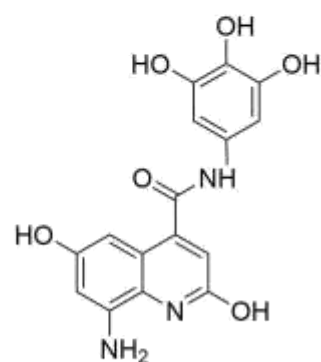
D24



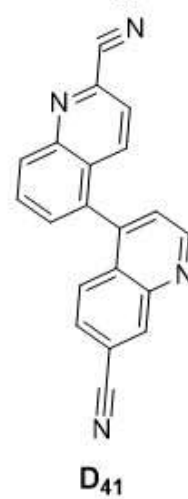
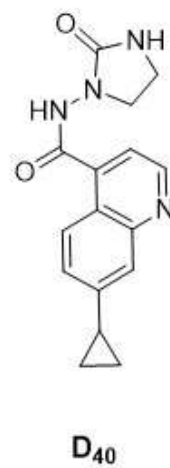
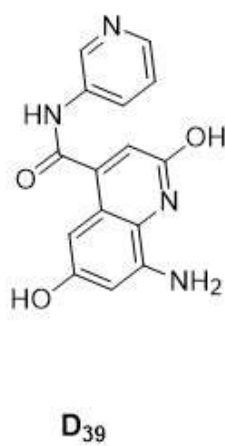
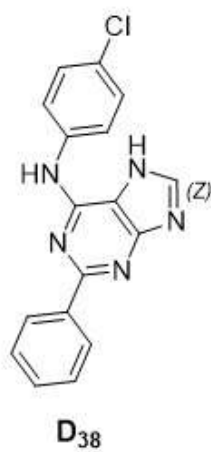
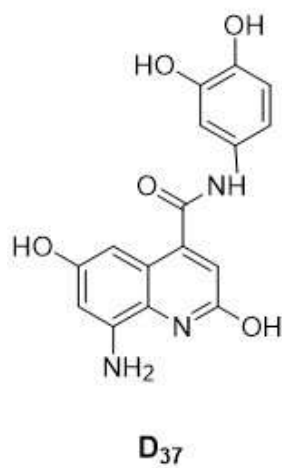
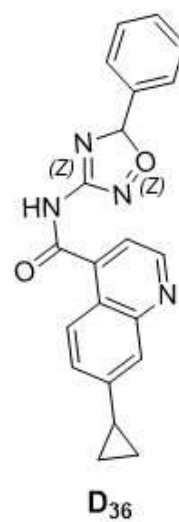
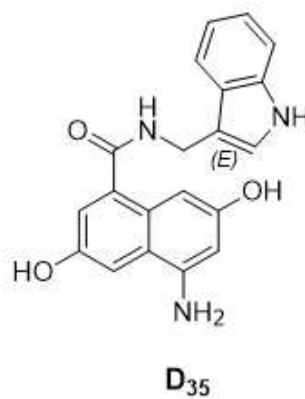
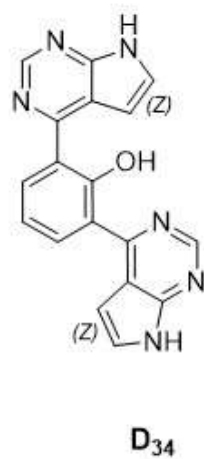
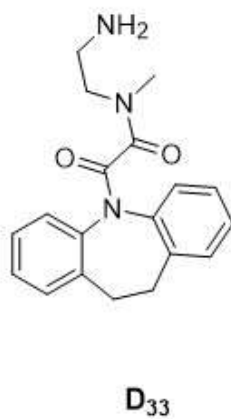
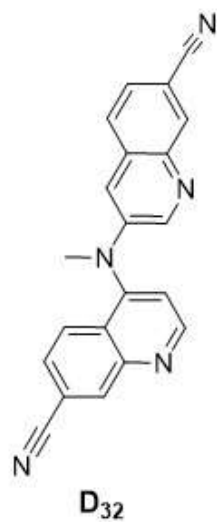
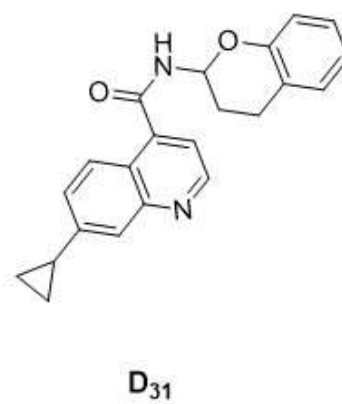
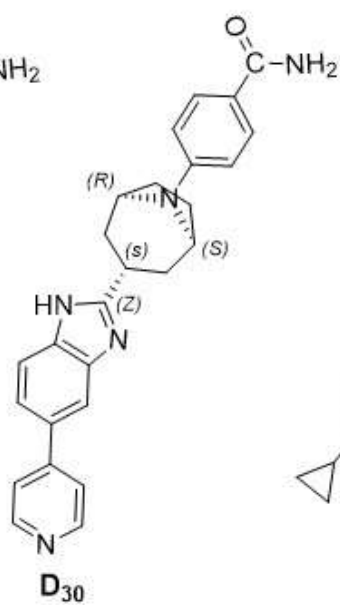
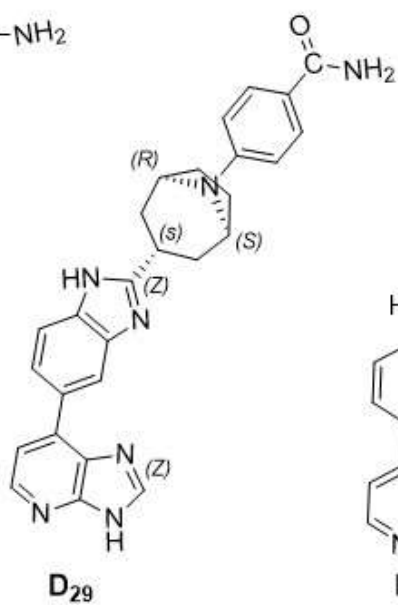
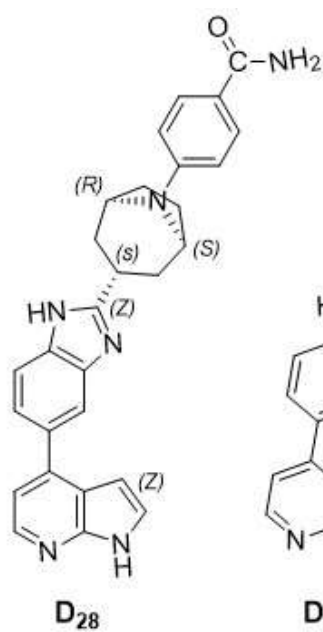
D25



D26



D27



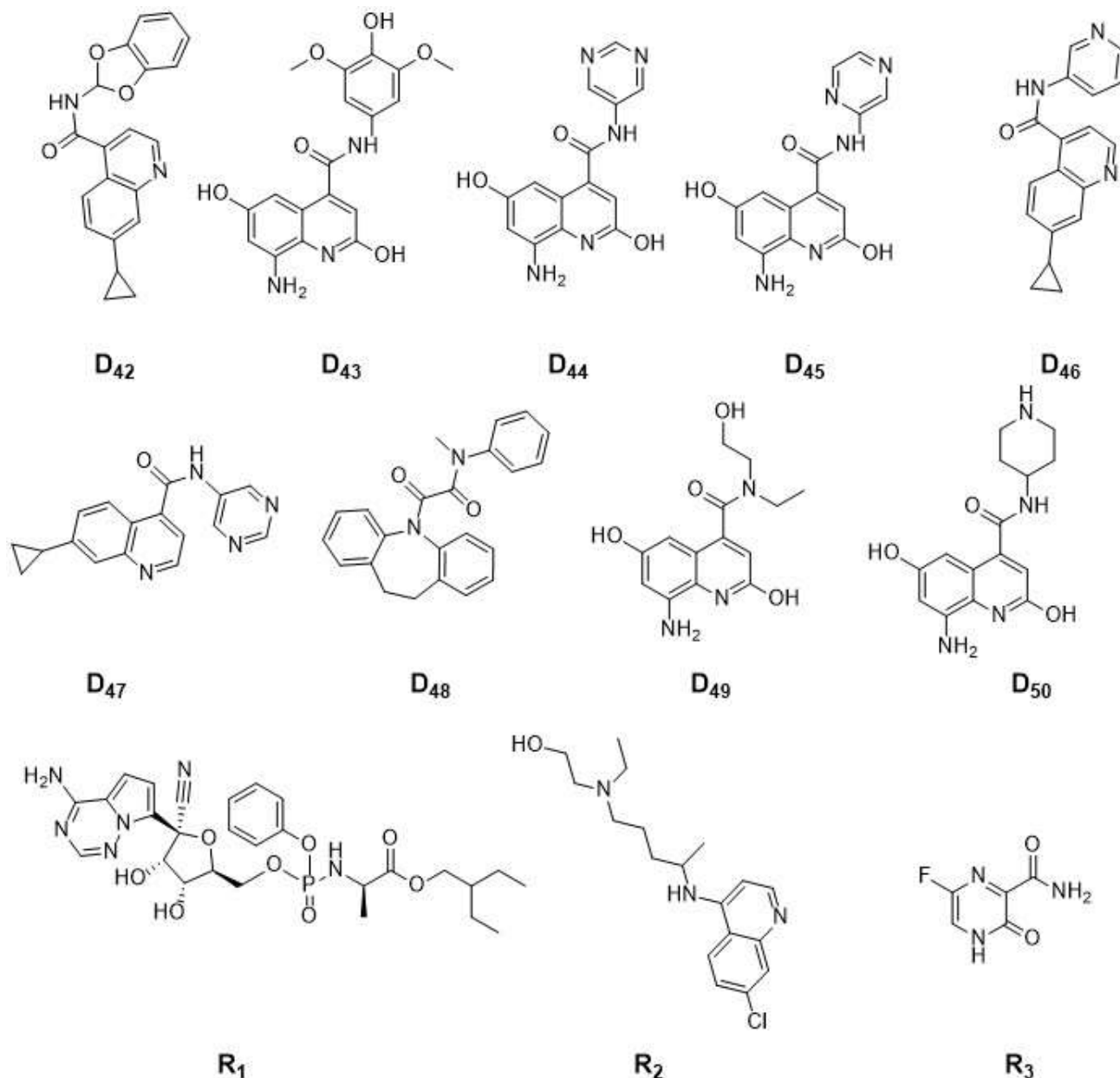


Figure 4.3 Fifty potential candidate drug-likeness molecule structures.

4.2.3 Stereochemistry and Protein-Ligand Complex Interaction Studies

The stereochemistry of six stereoisomers, D_{1A}, D_{1B}, D_{2A}, D_{2B}, and R_{1A}, R_{1B}, as well as their interactions with Site-5 (C₅) of the M^{pro} target protein. The enantiomers were visualized in 3D structures using Chemdraw 3D software to demonstrate their absolute stereochemistry. The investigation examined the impact of stereochemistry on the docking energy between stereoisomers and the target protein. Specifically, the study focused on analyzing how different stereo centers influenced the docking process. The protein's active amino acids create a variety of interactions with the ligand, including hydrogen bonds, electrostatic interactions, and van der Waals contacts, the protein-ligand combinations are stabilized by these interactions.

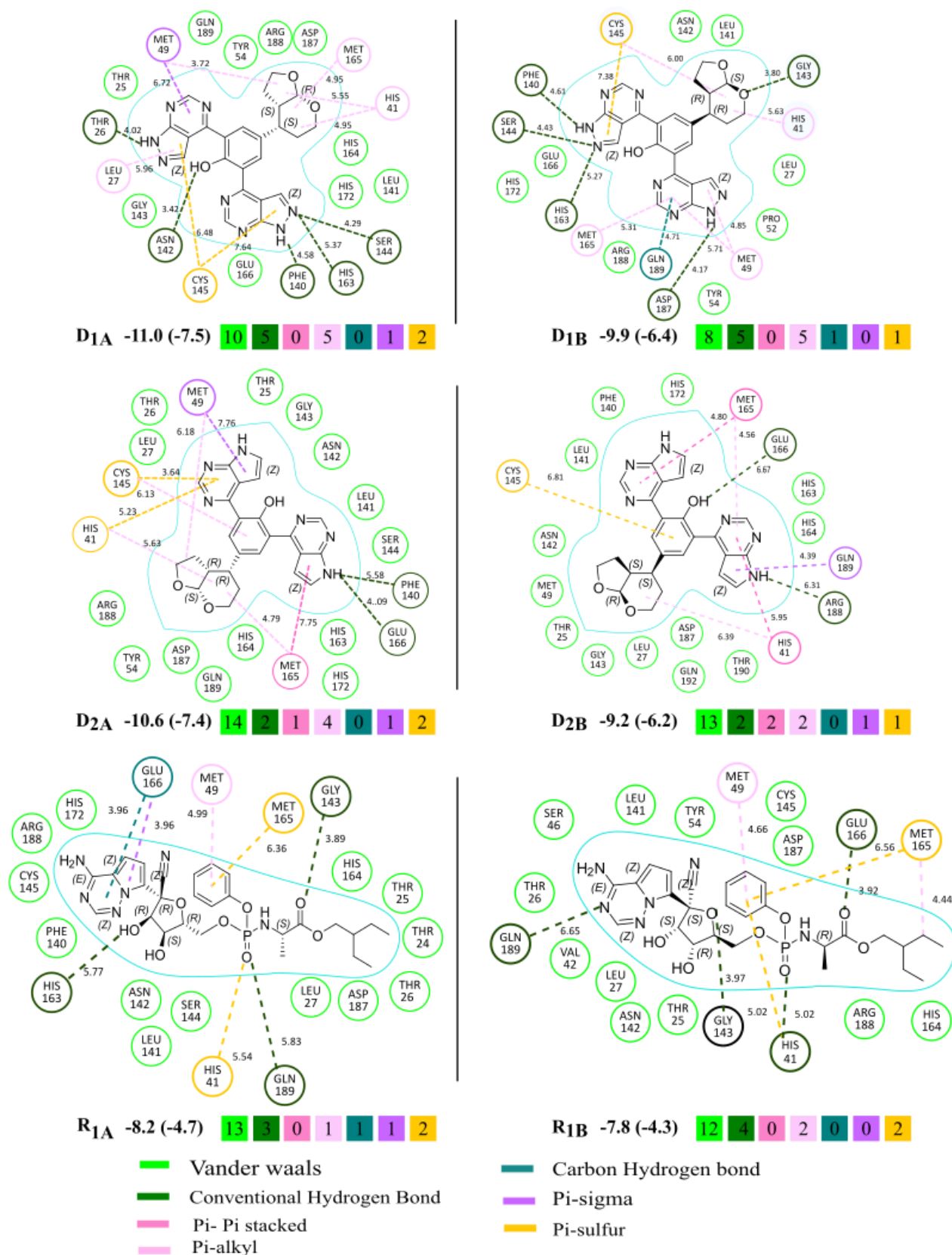


Figure 4.4 Stereochemistry and its Role in D₁, D₂, R₁ drug-likeness molecules and their interactions.

D_{1A} showed a higher BE of -11.0 kcal/mol and a higher CBE of -7.5 kcal/mol than D_{1B} (BE=-9.9 kcal/mol and CBE=-6.4 kcal/mol). D_{1A} and the protein complex generated one P-sigma bonding

contact with MET 49 and one hydrogen bonding connection with ASN 142 amino acid residues with phenol. D_{1A} also has more substantial 1H-pyrazolo[3,4-d] pyrimidine fragments than D_{1B}. D_{2A} also had a greater BE = -10.6 kcal/mol and CBE = -7.4 kcal/mol than D_{2B} (BE = -9.2 kcal/mol and CBE = -6.2 kcal/mol), and it interacted with the Pi-sigma bonding of MET 49. Additionally, R_{1A} had a BE of -10.8 kcal/mol, a CBE of -8.1 kcal/mol, and fragments of pyrrolo[2,1-f] [1,2,4] triazine-4-amine. R_{1A} and protein complex interacted with GLU 166 amino acid residue via one Pi-sigma bonding and one carbon-hydrogen bonding interaction, making it more significant than R_{1B} (BE = -7.8 kcal/mol and CBE = -4.3 kcal/mol).

The 3D chemdraw software to visualize the stereoisomers and observed that D_{1A}, D_{2A}, and R_{1A} stereoisomers had higher BE than D_{1B}, D_{2B}, and R_{1B} stereoisomers. The minimal distance between amino acid residues and the protein-ligand nature of their interaction influenced the complexes' stability. The study concluded that the stereochemistry of the ligand influences the kind and strength of the interaction with the target protein, and stereoisomers with higher BE and lower minimum distance between amino acid residues interacted effectively with the target protein. D₁ and D₂ stereoisomers better interacted with the target protein than Remdesivir (R₁).

4.2.4 Protein-Ligand Complex and their Interaction Studies of the Top Four Drug-Likeness Molecules at Selected Four Binding Sites

We have shown the protein-ligand complex for each of the five target binding sites in Figure 4.5. These illustrations may be seen in the Figure left column. A Black circle in each picture denotes the precise position between the protein that have particular interactions with the ligand complex at each site. Additionally, we have presented the BE and CBE values for the top five LORD-generated molecules (D₁-D₅) as well as three reference molecules, Remdesivir (R₁), Hydroxy chloroquinoline (R₂), and Favipiravir (R₃) - for each site, labeled C₁ to C₅ from top to bottom, in the right column of Figure 4.5. The CBE values for the top five molecules generated by LORD were observed to be much lower than the three reference molecules shown in the green and blue color-coded bar plots in all five sites. This indicates that the LORD-generated molecules have stronger binding affinities to the target protein, an essential property for an effective drug.

Furthermore, the LORD-generated molecules were found to outperform the currently available drugs in the market for all five binding sites. This suggests that the LORD-generated molecules have the potential to be more effective in treating diseases or conditions that target these specific binding sites. Our results demonstrate that the LORD-generated molecules have strong binding affinities to the target protein and could potentially serve as better drug candidates than currently available drugs in the market.

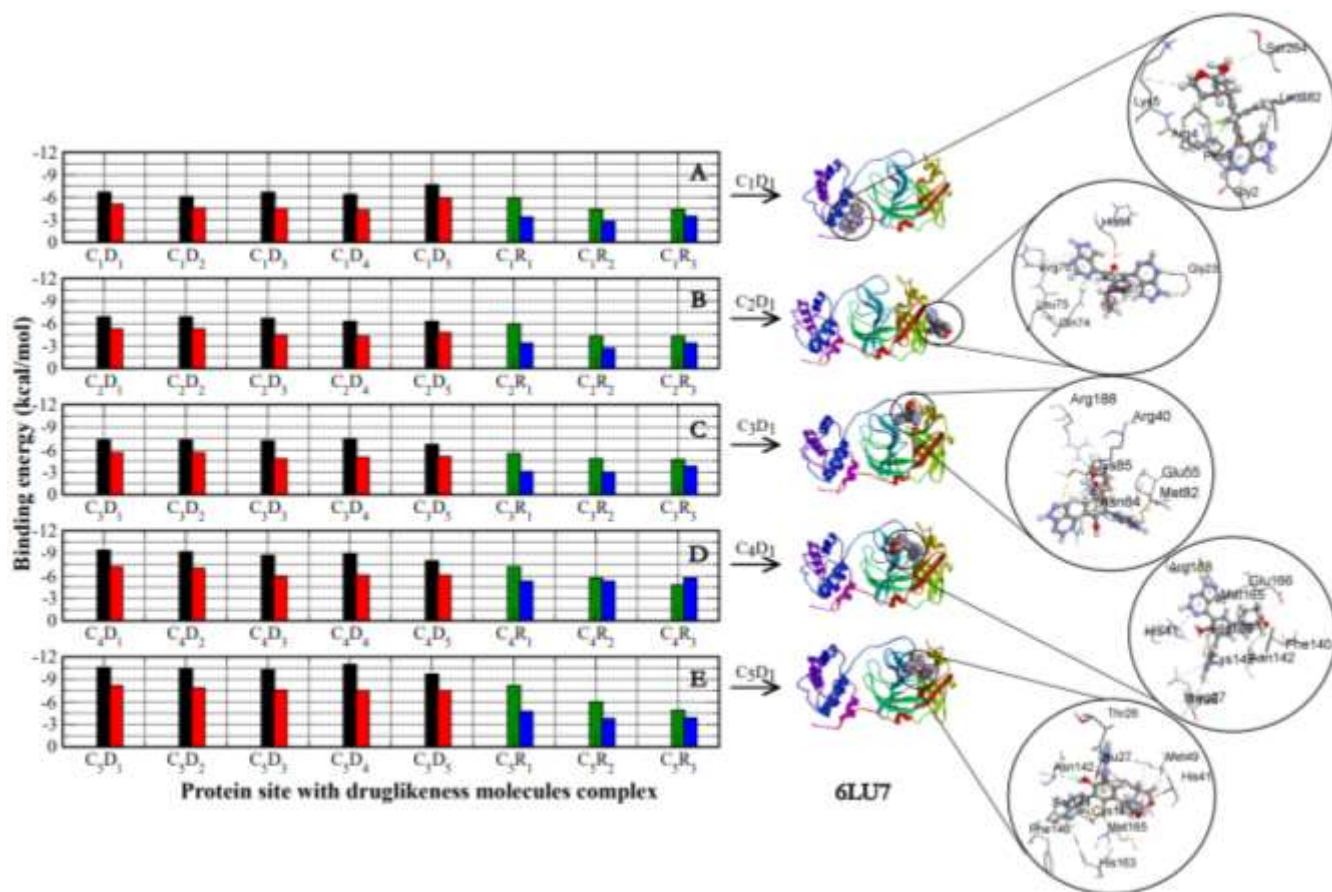


Figure 4.5 The Protein Site with drug-likeness molecules complex and their interaction studies of the top five drug-likeness molecules and three reference molecules at Selected five binding sites. (A) MESP suggested site-1(C₁), (B) MESP suggested site-2 (C₂), (C) MESP suggested cavity site-3 (C₃), and (D) Experimentally reported site-4 (C₄).

4.2.5 Binding Sites Analysis from Protein-Ligand Complex

We identified specific amino acid residues active in the five target binding sites (C₁ to C₅). In Site-1 (C₁), we found the active residues to be Glycine 12(Gyl2), Phenylalanine 3 (Phe3), Arginine 4 (Arg4), Lysine 5 (Lys5), Tyrosine126 (Tyr126), Glutamine 127 (Gln127), Tryptophan 207 (Trp207), Asparagine 214 (Asn214), Leucine 282 (Leu282), Serine 284 (Ser284), Alanine 285 (Ala285), Leucine 286 (Leu286), and Threonine 292 (Thr292). In comparison to the other sites, Site-1 exhibited decreased binding energy. In Site-2 (C₂), we observed active residues including Threonine 21 (Thr21), Cysteine 22 (Cys22), Glycine 23 (Gly23), threonine 24 (Thr24), Threonine 25 (Thr25), Lysine 61 (Lys61), Serine (Ser62), Asparagine 63 (Asn63), Histidine 64 (His64), Asparagine 65 (Asn65), Phenylalanine 66 (Phe66), leucine 67 (Leu67), Glutamine 74 (Gln74), Leucine 75 (Leu75), Arginine 76 (Arg76), Valine 77 (Val77), Isoleucine 78 (Ile78), and Glycine 79 (Gly79). Site-4 (C₄) and Site-3 (C₃) had higher BE than Site-2, whereas Site-1 had lower binding energy. Active amino acid residues in Site-3 (C₃) included Proline 52 (Pro52), Asparagine 53 (Asn53), Tyrosine (Tyr54), Glutamic acid 55 (Glu55), Methionine 82 (Met82),

Glutamine 83 (Gln83), Valine 86 (Val86), Leucine 87 (Leu87), Glycine 179 (Gly179), Asparagine 180 (Asn180), Phenylalanine 181 (Phe181), Tyrosine 182 (Tyr182), Phenylalanine 185 (Phe185), Valine 186 (Val186), Aspartic acid 187 (Asp187) and Arginine (Arg188). Interestingly, we noticed that these active residues at Site-4 corresponded to the experimentally reported active residues in Site-5, including Threonine 26 (Thr26), Leucine 27 (Leu27), Phenylalanine 140 (Phe140), Leucine 141 (Leu141), Asparagine 142 (Asn142), Glycine 143 (Gly143), Serine 144 (Ser144), Cysteine (Cys145), Histidine 163 (His163), Histidine 164 (His164), Methionine 165 (Met165), glutamic acid 166 (Glu166), Aspartic acid 187 (Asp187), Arginine 188 (Arg188), Glutamine189(Gln189), Threonine 190 (Thr190), and Alanine 191(Ala191) These amino acids interacted with each other and resulted in better binding energy.

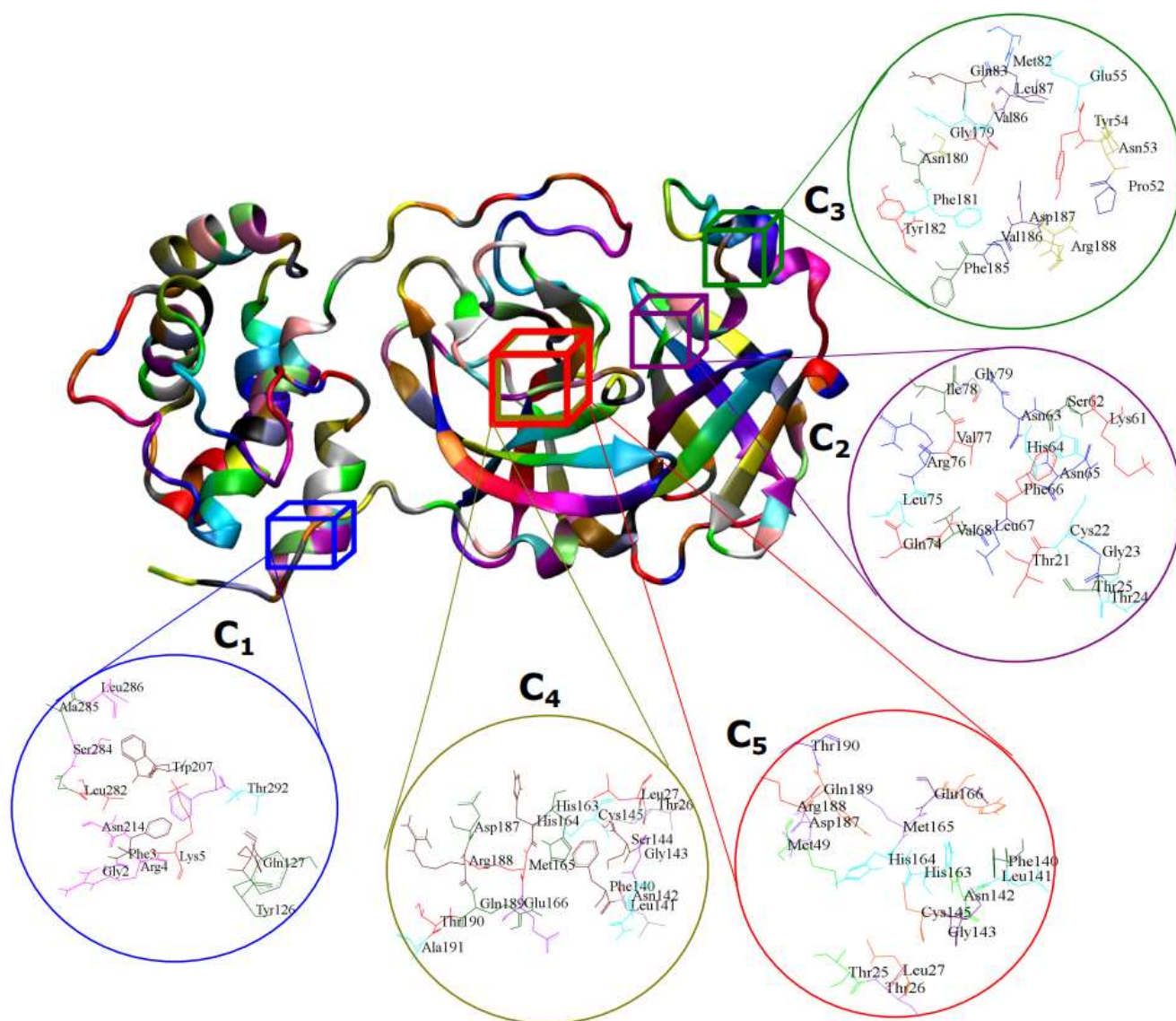


Figure 4.6 The four active sites, C₁, C₂, C₃, and C₄, of the target, main protease.

4.2.6 Cavity Drug-Likeness Interaction Matrix

In this study, we docked fifty potential drug-likeness molecular interactions against the target protein. Figure 4.7 depicts interactions between proteins and ligands and summarises our findings, using different colors for different types of interactions. We investigated the interactions between the Remdesivir complex and various active amino acid residues.

S.NO.	THR 25	THR 26	LEU 27	HIS 41	MET 49	PHE 140	LEU 141	ASN 142	GLY 143	SER 144	CYS 145	HIS 163	HIS 164	MET 165	GLU 166	HIS 172	ASP 187	ARG 188	GLN 189	THR 190	SI
D1				2																	12
D2																					10
D3																					11
D4																					8
D5																					8
D6																					7
D7																					8
D8																					9
D9																					7
D10																					7
D11																					7
D12																					6
D13																					8
D14																					5
D15																					8
D16																					4
D17																					8
D18																					6
D19																					10
D20																					6
D21																					5
D22																					3
D23																					6
D24																					6
D25																					7
R1																					8

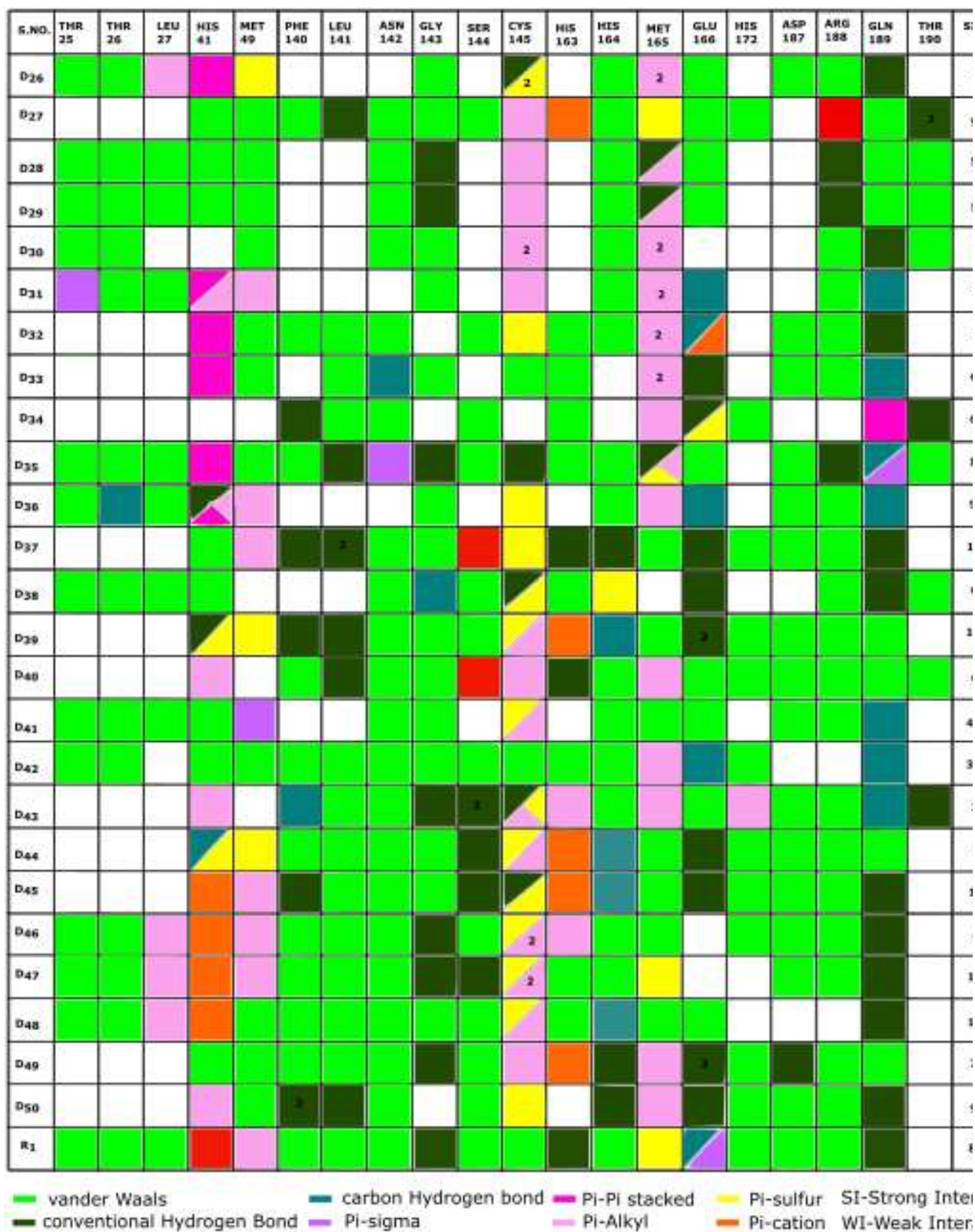


Figure 4.7 Experimentally reported Site5 (C₅) of the main protease with 50 drug-likeness molecules.

We investigated the Remdesivir-protein interactions complex with 20 active amino acid residues, 20 active amino acid residue interactions in the D₁, 20 active amino acids in the D₂, 22 active amino acids in the D₃, 22 active amino acids in the D₄, 23 active amino acids in the D₅, 21 active amino acids in the D₆, 21 active amino acids in the D₇, 17 active amino acids in the D₈, 13 active amino acids in the D₉, 17 active amino acids in the D₁₀, 15 active amino acid in the D₁₁, 15 active amino acids in the D₁₂, 14 active amino acids in the D₁₃, 16 active amino acids in the D₁₄, 16 active amino acids in the D₁₅, 13 active amino acids in the D₁₆, 16 active amino acids in the D₁₇, 17 active amino acids in the D₁₈, 20 active amino acids in the D₁₉, 15 active amino acids in the D₂₀, 14 active amino acids in the D₂₁, 14 active amino acids in the D₂₂, 15 active amino acids in the D₂₃, 17 active amino acids in the D₂₄ and 14 active amino acids in the D₂₅, 16 active amino acid residue interactions in the D₂₆, 19 active amino acids in the D₂₇, 15 active amino acids in the D₂₈, 21 active amino acids in the D₂₉, 14 active amino acids in the D₃₀, 14 active amino acids in the D₃₁, 16 active amino acids in the D₃₂, 13 active amino acids in the D₃₃, 11 active amino acids in the D₃₄, 22 active amino acids in the D₃₅, 15 active amino acids in the D₃₆, 17 active amino acids in the D₃₇, 15 active amino acids in the D₃₈, 19 active amino acids in the D₃₉, 16 active amino acids in the D₄₀, 15 active amino acids in the D₄₁, 16 active amino acids in the D₄₂, 18 active amino acids in the D₄₃, 18 active amino acids in the D₄₄, 17 active amino acids in the D₄₅, 19 active amino acids in the D₄₇, 17 active amino acids in the D₄₈, 16 active amino acids in the D₄₉, and 15 active amino acids in the D₅₀. Essentially, the interaction between proteins and ligands is important for drug design because it makes it possible to create substances that can precisely target and control the activity of disease-related proteins, leading to the development of effective and secure therapeutics.

4.2.7 Physicochemical and ADME Properties

To validate LORD algorithm on 50 molecules, we have provided the physicochemical properties, such as the Lipinski rule of five in Figure 4.8 for all the fifty molecules. All the candidate molecules show molecular weight ranging from 456.46 to 474.60, less than 500 Daltons. The Partition coefficient ranges between 1.65-3.42, less than 5; HBD is between 2-4, and HBA is between 3-9. The molecular polar surface area ranges between 73.49-147.61, is less than 140Å², and the total number of the rotatable bond ranges between 3-4 is less than 10. Figure 4.8 represents that most of the molecules followed the Lipinski rule of five; hence, they can be studied for further analysis. LORD once again proved that it incorporates Lipinski's rule of five by learning CBE while training data.

ADME properties are generated by using the SWISS ADME website for the 50 molecules predicted using the LORD algorithm. ADME properties are most effective in predicting human pharmacokinetic properties to be successful in clinical trials. The top five potential molecules have shown in Table 4.2.

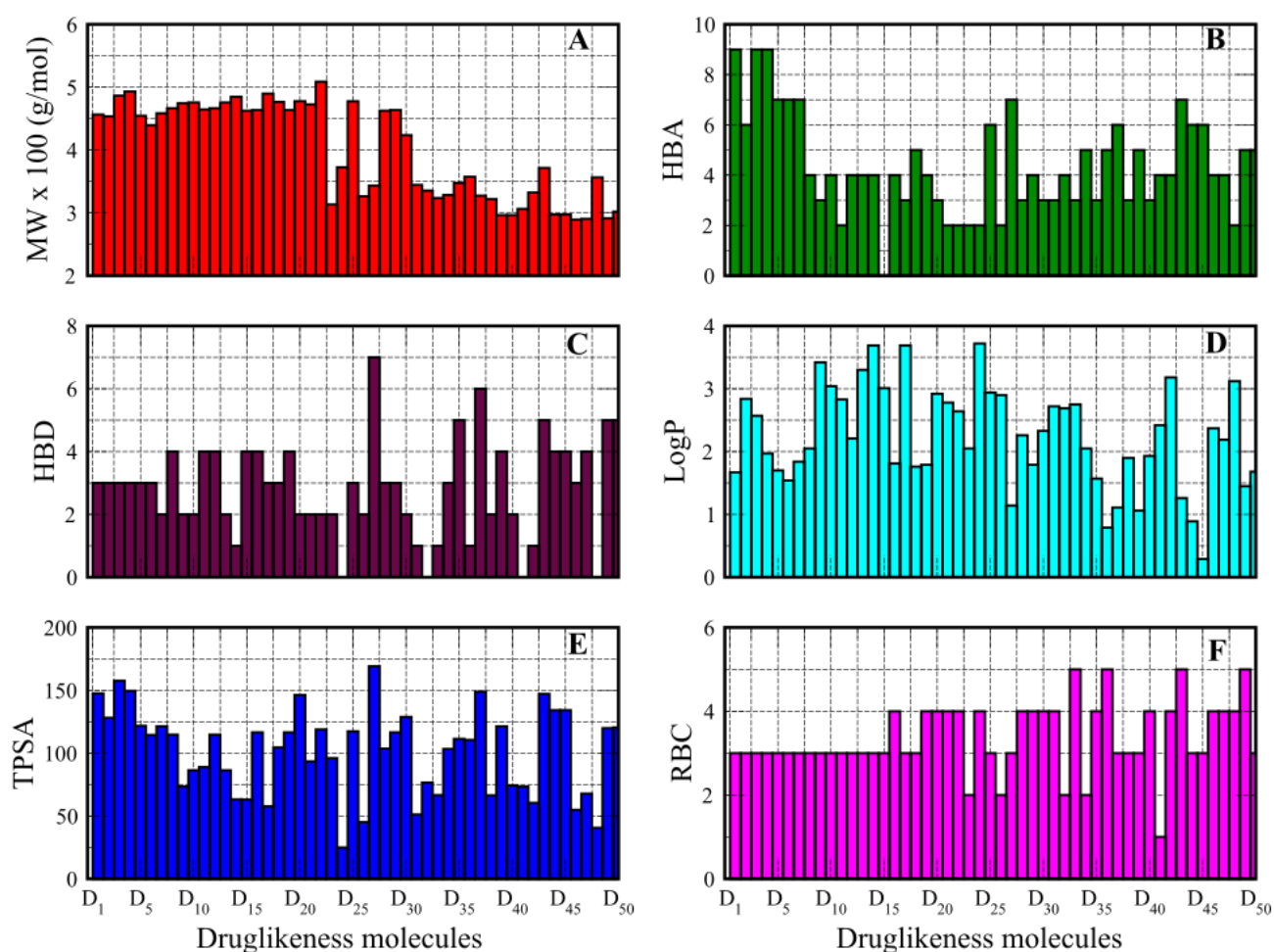


Figure 4.8 Physico-chemical properties of LORD 50 molecules. (A) Molecular weight (g/mol), (B) Hydrogen bond acceptor, (C) Hydrogen bond donor, (D) Partition Coefficient (Logp), (E) Total polar surface area, and (F) Rotatable bond count.

Table 4.2 ADME or Pharmacokinetic proprieties of the site-wise selected molecules in ^aGastrointestinal absorption, ^bBlood brain barrier permeant, ^cP-gp substrate, ^dCytochrome P450 family 1 subfamily A member2, ^eCytochrome P450 family 2 subfamily C member19, ^fCytochrome P450 family 2 subfamily C member 9, ^gCytochrome P450 family 2 subfamily D member 6, ^hCytochrome P450 family 3 subfamily A member4, and ⁱSkin permeation in cm.

S.No	GI abs ^a	BBB permnt	P-gp substrate ^c	CYP1A2 Inhibitor ^d	CYP2C19 Inhibitor ^e	CYP2C9 Inhibitor ^f	CYP2D6 Inhibitor ^g	CYP3A4 Inhibitor ^h	Log Kp ⁱ
D ₁	Low	No	No	No	No	No	No	Yes	-7.88
D ₂	High	No	Yes	No	No	Yes	Yes	Yes	-7.08
D ₃	High	No	No	No	No	No	Yes	Yes	-5.16
D ₄	Low	No	No	No	No	No	No	No	-5.70
D ₅	High	No	Yes	No	No	Yes	Yes	Yes	-6.41
R ₁	Low	No	Yes	No	No	No	No	Yes	-8.62

4.2.8 Correlation of Potential Drug-Likeness Molecules vs. Site-Wise Binding Energy

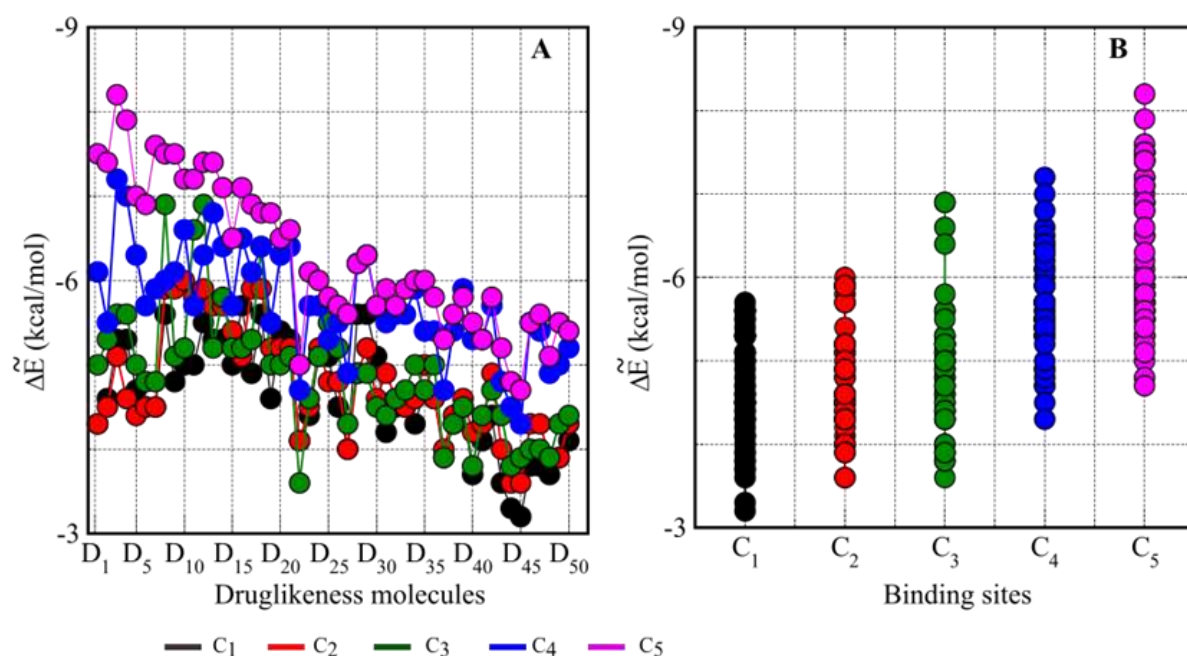


Figure 4.9 Corrected binding energy trend curves for LORD 50 molecules at all five sites. (A) The X-axis is 50 drug-likeness molecules (D₁ to D₅₀), and Y-axis is CBE values, (B) The X-axis is ESP-suggested sites (C₁, C₂, C₃, C₄), and Y-axis is CBE values.

The CBE values for LORD 50 molecules at each site are depicted in Figure 4.9. The range of CBE is -3.0 to -8.0 kcal/mol for all five sites of the main protease of the Sars-Cov-2 virus. In Figure 4.9 A, the X-axis represents 50 potential drug-likeness molecules, and Y-axis corresponds to CBE Figure 4.9 B. The X-axis represents the number of cavities or binding sites, while the Y-axis displays the CBE.

Figure 4.9 shows that Pink is a significant interaction and experimental site for all the drug-likeness molecules. And also, CBE analysis showed that Site 4 and Site 5 are more potential for protein-ligand complex interactions compared to sites 1, 2, and 3, as indicated in Figure 4.9.

In which the black line of C_1 (Experimental reported active site) represents binding energy value -3.2 to -5.6 kcal/mol, and the red line C_2 was shown -3.6 to -6.0 kcal/mol binding energy. The green line for C_3 was shown at -3.9 to -6.9 kcal/mol, the blue line for C_4 was shown at -4.3 to -7.2 kcal/mol, and the pink line for C_5 showed -4.8 to -8.2 kcal/mol. Two sites (C_4 , C_5) are better than the other three sites (C_1 , C_2 , and C_3) based on their CBE value shown in Figure 4.9.

4.2.9 Molecular Dynamics (MD) Simulations

The main protease of SARS-CoV-2 was extensively investigated to identify potential molecules employing MD simulations using the GROMACS 5.1.2 program. Three top drug-likeness compounds were discovered and compared to stability of reference molecule, Remdesivir using MD simulations. In the MD simulations, the protein energy evaluations in this study were conducted using the CHARMM force field, while the ligand topology was generated using the CHARMM General Force Field (CGenFF). The TIP3 water model was employed within a cubical box to facilitate the dissolution of the protein-ligand complex. To ensure system neutrality, chloride ions were introduced. The energy of the system was minimized using the steepest descent method. To maintain a constant temperature of 298 K, an NVT ensemble was utilized.

Additionally, the NPT ensemble was employed to simulate the behavior of the ligand-protein combination. The Particle Mesh Ewald and LINCS algorithms were used to examine short-range electrostatics, and a Van der Waals distance threshold of 10 was utilized to limit all bonds. To assure stability, the simulations were run for 100 ns, and the RMSD plots were compared to compare the stability of the three possible drug-like molecules with Remdesivir.

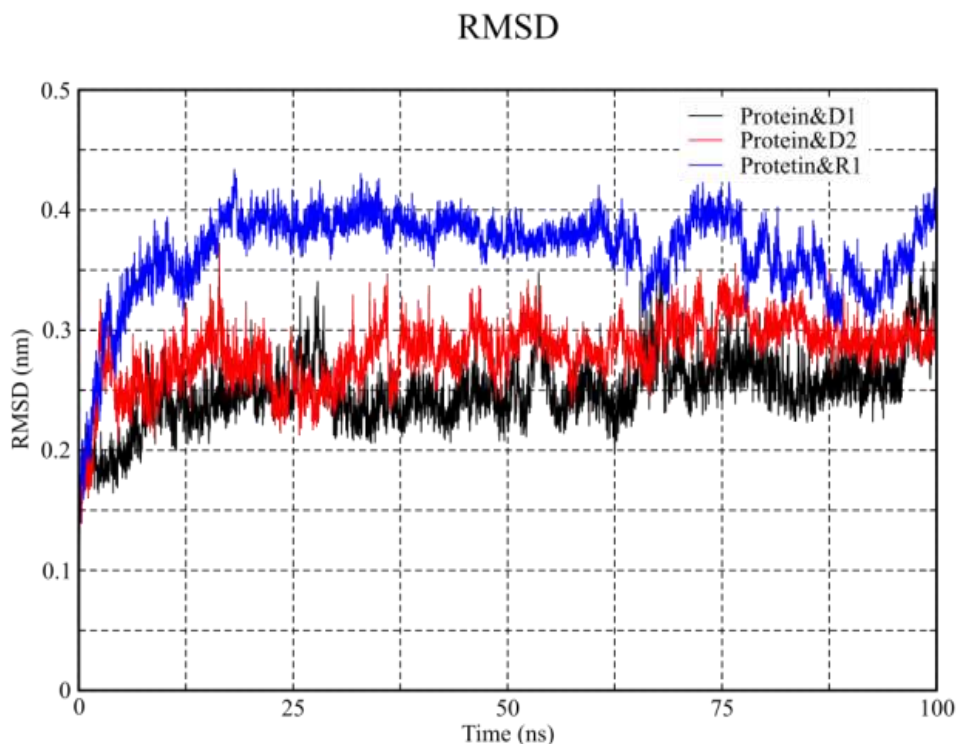


Figure 4.10. RMSD deviations of Protein-Ligand complex.

The molecular interactions of the two potential drug-like compounds and Remdesivir with the SARS-CoV-2 virus were compared using the RMSD calculation. The stability of the association between the ligands and the protein and the structural changes in the complex were investigated. The RMSD result suggested that the three potential ligands were strongly bound inside Site 5 (C₅) of the M^{pro}. RMSD values range from 0.1-0.3.2 nm to 0.2-0.3.5 nm. The MD simulation of the SARS-CoV-2 virus's major protease revealed good molecular contact of the three ligand molecules, and the stability of the connection between drugs and ligands was investigated. Molecular superposition was used to compare the average of the input (0 ns) and output (100 ns) structures. The findings revealed that the two-candidate potential drug-likeness compounds were stable and had effective molecular interactions with the M^{pro}.

4.2.10 Free Energy Calculations

The protein-ligand complexes binding energy (BE) as it is more often known, was determined using the *g-mmpbsa* software. Calculating the vacuum potential energy resulting from various interactions, including both bound and unbound states, requires the application of the molecular mechanics (MM) force field. The computation of the electrostatic and van der Waals (Evdw) energy contributions

involves the use of the Coulomb potential function and the Lennard-Jones potential function, respectively. The solution to the Poisson-Boltzmann equation allowed for identifying the polar contribution while evaluating the free energy of solvation. It has traditionally been assumed that the solvent-accessible surface area (SASA) and the non-electrostatic solvation energy are closely related when evaluating the non-polar contribution. The non-polar energy term (G_{nonpolar}) incorporates van der Waals interactions and the attractive and repulsive forces that arise between the solute and solvent due to cavities. During the Free energy calculations, snapshots of the equilibrated area from the molecular dynamics (MD) trajectory were captured. All computations were performed using the MmPbsaDecomp.py script, which is part of the *g-mmpbsa* package, with the default parameters provided by Kumari et al.

Table 4.3 Free energy calculation of the top four drug-likeness (D₁-D₄) molecules and Remdesivir (R₁).

Time intervals (ns)	Complex free energies (kJ/mol)		
	D ₁	D ₂	R ₁
0-25	-17.6 +/- 11.6	-49.5 +/- 13.7	-76.2 +/- 19.3
25-50	-20.8 +/- 12.4	-43.6 +/- 14.1	-59.6 +/- 15.5
50-75	-29.2 +/- 12.9	-44.0 +/- 11.0	-64.9 +/- 13.3
75-100	-31.7 +/- 8.8	-38.2 +/- 12.2	-64.9 +/- 13.3

For the two protein-ligand complexes that were chosen, calculations of binding energy (BE) using the MM-PBSA program were made. The top four drug-likeness as, D₁, D₂, and Remdesivir as reference (R₁) molecules, had BE values at 0-25ns, 25-50ns, 50-75ns, and 75-100ns. The BE distribution calculation revealed that it significantly contributed to the overall BE, as shown in Table 4.3.

4.3 Conclusions

The present work demonstrates a novel approach to designing on-site drug-likeness molecules for target proteins. On-site drug design strategy allows the algorithm better in building molecules by

exactly locating the target binding site environment by MESP cavity studies. LORD method can be utilized for any target protein to design drug-likeness molecules. LORD can incorporate the physicochemical properties by corrected binding energy training. Hence, the possibility of irrelevant molecular generations is very much infrequent. LORD follows a reverse engineering approach, such as knowing the environment of the target binding site and building molecules. Hence, it is a very high chance that the constructed molecule can become a drug-likeness molecule. The advantages of LORD are computationally cheap due to the substrate approach while scanning the MESP cavities, and it will locate target binding sites automatically. LORD algorithm is intelligent towards building molecules based on the target site environment. Hence, the suggestion of molecules will be relevant to that site.

In this work, we found five potential binding sites, including an experimentally active site. Five potential binding sites are utilized to build molecules by LORD, producing 50 potential drug candidates. Out of all five binding sites, the experimentally active site is more dominant than the other sites in CBE values. LORD 50 molecules possess excellent BE (kcal/mol), and Physico-chemical properties and have shown better ADME properties than market-available drugs such as Remdesivir, Favipiravir, and Hydroxychloroquine. Further, MD simulations of top-four candidate drug-like molecules are studied and found to be stable in protein for long simulation runs, unlike less stable Remdesivir drug molecules.

REFERENCES

1. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M.-L.; Zhang, Y.-L.; Dai, F.-H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E. C.; Zhang, Y.-Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, 579 (7798), 265–269.
2. Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P.; Carter, L. J.; Smoot, J.; Gregg, A. C.; Daniels, A. D.; Jervey, S.; Albaiu, D. Research and Develop-Ment on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Dis-Eases. *ACS Central Science* **2020** (3), 315–331.
3. Prasad, A.; Prasad, M. SARS-CoV-2: The Emergence of a Viral Pathogen Causing Havoc on Human Existence. *J. Genet.* **2020**, 99 (1).
4. Cui, J.; Li, F.; Shi, Z.-L. Origin and Evolution of Pathogenic Coronaviruses. *Nat. Rev. Microbiol.* **2019**, 17 (3), 181–192.
5. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, 582 (7811), 289–293.
6. Hatada, R.; Okuwaki, K.; Mochizuki, Y.; Handa, Y.; Fukuzawa, K.; Komeiji, Y.; Okiyama, Y.; Tanaka, S. Fragment Molecular Orbital Based In-Teraction Analyses on COVID-19 Main Protease - Inhibitor N3 Complex (PDB ID: 6LU7). *J. Chem. Inf. Model.* **2020** (7), 3593–3602.
7. Khater, I.; Nassar, A. In Silico Molecular Docking Analysis for Repurposing Approved Antiviral Drugs against SARS-CoV-2 Main Protease. *Biochem. Biophys. Rep.* **2021**, 27 (101032), 101032.
8. Li, G.; Hilgenfeld, R.; Whitley, R.; De Clercq, E. Therapeutic Strategies for COVID-19: Progress and Lessons Learned. *Nat. Rev. Drug Discov.* **2023**.
9. Muratov, E. N.; Amaro, R.; Andrade, C. H.; Brown, N.; Ekins, S.; Fourches, D.; Isayev, O.; Kozakov, D.; Medina-Franco, J. L.; Merz, K. M.; Oprea, T. I.; Poroikov, V.; Schneider, G.; Todd, M. H.; Varnek, A.; Winkler, D. A.; Zakharov, A. V.; Cherkasov, A.; Tropsha, A. A Critical Overview of Computational Approaches Employed for COVID-19 Drug Discovery. *Chem. Soc. Rev.* **2021**, 50 (16), 9121–9151.
10. Napolitano, F.; Xu, X.; Gao, X. Impact of Computational Approaches in the Fight against COVID-19: An AI Guided Review of 17 000 Studies. *Brief. Bioinform.* **2022**, 23 (1).
11. Maiti, B. K. Can Papain-like Protease Inhibitors Halt SARS-CoV-2 Replication? *ACS Pharmacol. Transl. Sci.* **2020**, 3, 1017–1019.
12. Ghahremanpour, M. M.; Tirado-Rives, J.; Deshmukh, M.; Ippolito, J. A.; Zhang, C.-H.; Cabeza de Vaca, I.; Liosi, M.-E.; Anderson, K. S.; Jorgensen, W. L. Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.* **2020**, 11 (12), 2526–2533.
13. Achutha, A. S.; Pushpa, V. L.; Suchitra, S. Theoret-Ical Insights into the Anti-SARS-CoV-2 Activity of Chloroquine and Its Analogs and in Silico Screen-Ing of Main Protease Inhibitors. *J. Proteome Res.* **2020**, 11, 4706–4717.
14. Gao, K.; Nguyen, D. D.; Chen, J.; Wang, R.; Wei, G.-W. Repositioning of 8565 Existing Drugs for COVID-19. *J. Phys. Chem. Lett.* **2020**, 11 (13), 5373–5382.

15. Analogs against 3CLproProtein Target of SARS-CoV-2: Molecular Docking, Molecular Dynamics Simulation, and Structure-Activity Relationship Studies. *J.Chem.inf.Model* **2020** 12,5754–5770.
16. Manandhar, A.; Blass, B. E.; Colussi, D. J.; Almi, I.; Abou-Gharbia, M.; Klein, M. L.; Elokely, K. M. Targeting SARS-CoV-2 M3CLpro by HCV NS3/4a Inhibitors: In Silico Modeling and in Vitro Screening. *J Chem. Inf. Model.* **2021** (12) 1020-1032
17. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α -Ketoamide Inhibitors. *Science* **2020**, 368 (6489), 409–412.
18. Muratov, E. N.; Amaro, R.; Andrade, C. H.; Brown, N.; Ekins, S.; Fourches, D.; Isayev, O.; Kozakov, D.; Medina-Franco, J. L.; Merz, K. M.; Oprea, T. I.; Poroikov, V.; Schneider, G.; Todd, M. H.; Varnek, A.; Winkler, D. A.; Zakharov, A. V.; Cherkasov, A.; Tropsha, A. A Critical Overview of Computational Approaches Employed for COVID-19 Drug Discovery. *Chem. Soc. Rev.* **2021**, 50 (16), 9121–9151.
19. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, 582 (7811), 289–293.
20. Napolitano, F.; Xu, X.; Gao, X. Impact of Computational Approaches in the Fight against COVID-19: An AI Guided Review of 17 000 Studies. *Brief. Bioinform.* **2022**, 23 (1).
21. Kanhed, A. M.; Patel, D. V.; Teli, D. M.; Patel, N. R.; Chhabria, M. T.; Yadav, M. R. Identification of Potential Mpro Inhibitors for the Treatment of COVID-19 by Using Systematic Virtual Screening Approach. *Mol. Divers.* **2021**, 25 (1), 383–401.
22. Mengist, H. M.; Dilnessa, T.; Jin, T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.* **2021**, 9, 622898.
23. Mercorelli, B.; Desantis, J.; Celegato, M.; Bazzacco, A.; Siragusa, L.; Benedetti, P.; Eleuteri, M.; Croci, F.; Cruciani, G.; Goracci, L.; Loregian, A. Discovery of Novel SARS-CoV-2 Inhibitors Targeting the Main Protease Mpro by Virtual Screenings and Hit Optimization. *Antiviral Res.* **2022**, 204, 105350.
24. Sobhia, M. E.; Ghosh, K.; Sivangula, S.; Kumar, S.; Singh, H. Identification of Potential SARS-CoV-2 Mpro Inhibitors Integrating Molecular Docking and Water Thermodynamics. *J. Biomol. Struct. Dyn.* **2022**, 40 (11), 5079–5089.
25. Mehmood, A.; Nawab, S.; Wang, Y.; Chandra Kaushik, A.; Wei, D.-Q. Discovering Potent Inhibitors against the Mpro of the SARS-CoV-2. A Medicinal Chemistry Approach. *Comput. Biol. Med.* **2022**, 143, 105235.
26. Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y.; Wang, J. Boceprevir, GC-376, and Calpain Inhibitors II, XII Inhibit SARS-CoV-2 Viral Replication by Targeting the Viral Main Protease. *Cell Res.* **2020**, 30 (8), 678–692.
27. Kanhed, A. M.; Patel, D. V.; Teli, D. M.; Patel, N. R.; Chhabria, M. T.; Yadav, M. R. Identification of Potential Mpro Inhibitors for the Treatment of COVID-19 by Using Systematic Virtual Screening Approach. *Mol. Divers.* **2021**, 25 (1), 383–401.

28. Ray, A. K.; Sen Gupta, P. S.; Panda, S. K.; Biswal, S.; Bhattacharya, U.; Rana, M. K. Repurposing of FDA-Approved Drugs as Potential Inhibitors of the SARS-CoV-2 Main Protease: Molecular Insights into Improved Therapeutic Discovery. *Comput. Biol. Med.* **2022**, *142*, 105183.
29. Wang, Y.; Murlidaran, S.; Pearlman, D. A. Quantum Simulations of SARS-CoV-2 Main Protease Enable High-Quality Scoring of Diverse Ligands. *J. Comput. Aided Mol. Des.* **2021**, *35* (9), 963–971.
30. Fischer, C.; Feys, J. R. SARS-CoV-2 Mpro Inhibitors: Achieved Diversity, Developing Resistance and Future Strategies. *Future Pharmacology* **2023**, *3* (1), 80–107.
31. Hossain, A.; Rahman, M. E.; Rahman, M. S.; Nasirujjaman, K.; Matin, M. N.; Faruqe, M. O.; Rabbee, M. F. Identification of Medicinal Plant-Based Phytochemicals as a Potential Inhibitor for SARS-CoV-2 Main Protease (Mpro) Using Molecular Docking and Deep Learning Methods. *Comput. Biol. Med.* **2023**, *157* (106785), 106785.
32. Hou, N.; Shuai, L.; Zhang, L.; Xie, X.; Tang, K.; Zhu, Y.; Yu, Y.; Zhang, W.; Tan, Q.; Zhong, G.; Wen, Z.; Wang, C.; He, X.; Huo, H.; Gao, H.; Xu, Y.; Xue, J.; Peng, C.; Zou, J.; Schindewolf, C.; Menachery, V.; Su, W.; Yuan, Y.; Shen, Z.; Zhang, R.; Yuan, S.; Yu, H.; Shi, P.-Y.; Bu, Z.; Huang, J.; Hu, Q. Development of Highly Potent Noncovalent Inhibitors of SARS-CoV-2 3CLpro. *ACS Cent. Sci.* **2023**, *9* (2), 217–227.
33. Hagar, M.; Ahmed, H. A.; Aljohani, G.; Alhaddad, O. A. Investigation of Some Antiviral N-Heterocycles as COVID 19 Drug: Molecular Dock-Ing and DFT Calculations *Int. J. Mol. Sci.* **2020**, *21*(11), 3922.
34. Achutha, A. S.; Pushpa, V. L.; Suchitra, S. Theoret-Ical Insights into the Anti-SARS-CoV-2 Activity of Chloroquine and Its Analogs and in Silico Screen-Ing of Main Protease Inhibitors. *J. Proteome Res.* **2020**, *11*, 4706–4717.

CHAPTER 5

Design of Potential Drug-likeness Molecules for Parkinson's Disease

5.1 Introduction to Parkinson's Disease

Parkinson's Disease (PD) is a common neurological disorder wherein the midbrain's substantia nigra loses dopaminergic neurons. With its rising prevalence, PD has become a major public health concern, affecting approximately 7-10 million individuals aged 60 and above globally. Second, only to Alzheimer's disease (AD), PD requires increased research endeavors and interventions to enhance the well-being of those afflicted, underscoring the need to address its escalating challenges effectively.^{1,2} Because dopamine replacement therapy is still the most effective symptomatic medication for Parkinson's disease, most contemporary treatments focus on symptomatic alleviation to enhance patient quality of life. For a more in-depth examination of contemporary pharmacotherapies and their efficacy, consider,³ Flavin adenine dinucleotide (FAD)-containing enzymes that are connected to the mitochondrial outer membrane and responsible for catalyzing the oxidative deamination of monoamine neurotransmitters and dietary amines are found in specific human monoamine oxidase (MAO) enzymes.⁴ The two main types of MAO isoforms are monoamine oxidase type A (MAO-A) and monoamine oxidase type B (MAO-B). MAO isoforms are distinguished based on their genetic makeup, distribution within tissues, and the specific substances they act upon. MAO-A, an enzyme, is crucial in breaking down norepinephrine, serotonin, and tyramine. Selective inhibitors targeting MAO-A are commonly used in treating depression. By inhibiting this enzyme, the levels of these neurotransmitters are increased, which helps alleviate depressive symptoms, and the MAO-B type is employed for selectivity of the metabolized dopamine, so selective MAO-B inhibitors to the treatment of the PD.⁵

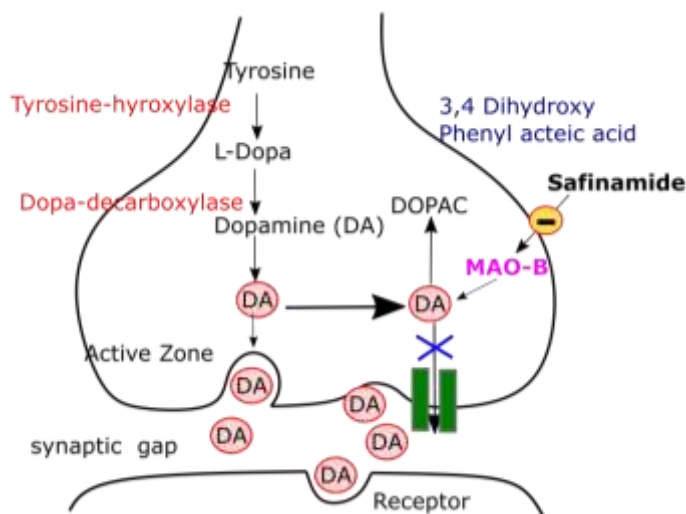


Figure 5. 1 Cartoon illustrating the Monoamine Oxidase-B enzyme role in PD.

The current studies on the crystal structure of the human MAO-B type with safinamide complex (PDB ID= 2V5Z, Resolution =1.6Å) are available in the protein databank.⁶ In which complex of human MAO-B with safinamide have hydrogen bond interaction of the Tyrosine-435(Tyr-435), Leucine (Leu-171), amino acids, and Pi-interaction of the Leucine (Leu-171), Cysteine-172(Cys-172), Tyrosine-326(Tyr-326), Phenylalanine (Phe-168), Leucine-164(Leu-164), Isoleucine-199(Ile-199) amino acid residues.⁷

The recent literature is reported as chemical scaffolds of the N-heterocyclic compounds⁸ and their related compounds⁹ 3-Arylcoumarin-tetracyclic tacrine,¹⁰ coumarin derivatives,¹¹ chacolone derivatives,¹² and their related compounds are used as novel MAO-B inhibitors are employed as Anti Parkinson's agent. And Table 5.1 present a selection of relevant previous literature, while our focus in this section is to discuss the methodology and summary.

Table 5.1 Overview of previous literature and focus on methodology and summary.

Methodology	Summary
X-ray crystallography and molecular docking	In a comprehensive study, researchers extensively investigated the structural aspects of the binding mode demonstrated by selective MAO-B inhibitors, which paved the way for rational drug design strategies. Through meticulous analysis, the study yielded valuable insights into the specific interactions and molecular features that contribute to the inhibitor's selectivity. This newfound knowledge offers researchers a deeper understanding of the critical binding mechanism involved, thus enabling the development of more potent and customized MAO-B inhibitors. ¹³
X-ray crystallography and computational analysis	A study on the crystal structure analysis of MAO-B and its reversible inhibitors has been conducted to facilitate the development of new propargylaminoindan compounds with enhanced selectivity and pharmacological properties. By analyzing the complex structure, researchers aim to design novel compounds that exhibit improved therapeutic benefits. ¹⁴
Fragment-based drug design, synthesis, MAO inhibitory in Vitro and Vivo assay.	In this study, we employed a computational fragment-based design to identify a novel, potent, and selective inhibitor of the MAO-B. Using a combination of virtual screening, docking simulations, and MD simulations, we identified a promising fragment that exhibited strong binding affinity and selectivity for MAO-B. ¹⁵
Molecular docking, MD simulations, and QSAR studies	Novel MAO-B hit inhibitors using multidimensional molecular modeling approaches and application of binary QSAR models for prediction of their therapeutic activity, pharmacokinetic, and toxicity properties, the authors explore the development of new MAO-B inhibitors using molecular modeling techniques. ¹⁶
Drug Repurposing, Molecular docking, and molecular MD simulations	Virtual screening and molecular dynamics simulations to discover new MAO-B inhibitors that could have neuroprotective properties. The study aims to identify potential compounds that could be further investigated for treating neurodegenerative disorders such as AD by employing computational techniques. ¹⁷
Molecular docking, MD simulations, and	The mechanism of irreversible inhibition of MAO enzymes by the antiparkinsonian propargylamine inhibitors, rasagiline, and selegiline. Using computational methods,

quantum chemical analysis	the study provides valuable insights into the binding interactions and structural changes involved, enhancing our understanding of these inhibitors' therapeutic effects. ¹⁸
pharmacophore modeling, QSAR analysis, Molecular docking, MD simulation, and free energy calculations	New insights on the activity and selectivity of MAO-B inhibitors through in silico methods explore the use of computational methods to analyze the effectiveness and specificity of MAO-B inhibitors. The study provides valuable insights into the design and development of potential drugs targeting MAO-B, aiding in treating neurodegenerative disorders such as PD. ¹⁹
Natural product, molecular docking and MD simulations	In silico study to identify new MAO-A selective inhibitors from natural sources by virtual screening and molecular dynamics simulation" explores the use of computational methods to discover potential inhibitors of the enzyme MAO-A. The study focuses on natural sources and employs virtual screening and MD simulations to identify promising compounds. ²⁰
Molecular docking and MD simulations	Identification of new small molecule MAO-B inhibitors through pharmacophore-based virtual screening, molecular docking, and MD simulation studies; they utilized computational methods to identify potential inhibitors for MAO-B, by employing pharmacophore-based virtual screening, molecular docking, and MD simulation techniques. ²¹
Pharmacophore modeling, molecular docking, MD simulations, and machine learning	This study proposes a novel approach to developing treatments for AD and PD by targeting MAO-B. Combining machine learning and computer-aided drug design, the researchers identified potential inhibitors for MAO-B. These findings hold promise for the development of new therapies that could help mitigate the progression of these neurodegenerative diseases. ²²
Flavonoids, molecular docking, and synthesis approaches	The use of computational methods to investigate flavonoid derivatives as potential inhibitors of MAO-B. By studying the structural properties and interactions of these compounds, researchers aim to design novel compounds that can effectively target MAO-B and potentially have therapeutic effects. ²³
QSAR modeling, Molecular docking, and MD simulations	Chemical library design, QSAR modeling, and MD simulations of naturally occurring coumarins as dual inhibitors of MAO-B and AChE, explore the potential of naturally occurring coumarins as dual inhibitors of MAO-B and AChE. ²⁴
Benzofuran, Safinamide compounds and Molecular docking	This study focuses on designing, synthesizing, and evaluating new compounds targeting MAO-B for treating Parkinson's disease. The researchers aimed to improve the pharmacokinetic properties of these compounds. The findings of the study may contribute to the development of novel MAO-B inhibitors with enhanced therapeutic potential for PD. ²⁵
MAO inhibitors	MAO-B inhibitors for the treatment of PD: past, present, and future" explores the historical development, current usage, and prospects of MAO-B inhibitors in the treatment of PD. ²⁶
Selegiline, rasagiline, safinamide	
Unsaturated Ketone, ADME prediction, Docking studies, MD simulations, and binding free energy	New MAO-B inhibitors using a multi-faceted approach. The study incorporates docking-based alignment, 3D-QSAR, ADMET prediction, molecular dynamics simulation, and MM_GBSA binding free energy to design potential inhibitors. The findings could contribute to the development of new therapeutic strategies for diseases associated with MAO-B activity. ²⁷
QSAR modeling, molecular docking, in silico ADME	The use of computational methods to study a group of compounds that target MAO-B for the treatment of Alzheimer's disease. The article discusses the application of quantitative structure-activity relationship (QSAR) models, simulation techniques,

and ADMET/pharmacokinetics assessment to evaluate the potential effectiveness of these compounds.²⁸

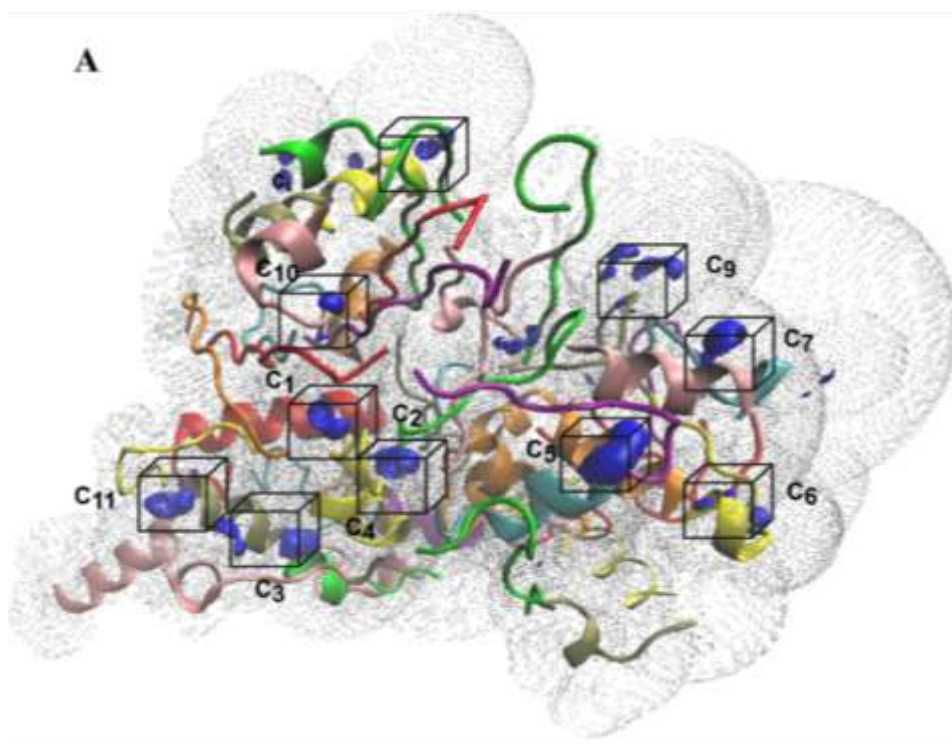
2H-chromen-2-one core,
3D-QSAR, and
molecular docking.

The study utilized both structure-based and ligand-based approaches to derive 3D-QSAR predictive models. By analyzing the structures of the inhibitors and their binding interactions with MAO-B, the researchers developed models that could predict the activity of new inhibitors based on their molecular features.²⁹

Therefore, in the current study, the successful application of ESP studies was employed for finding the potential binding site in the protein target, and LORD generator for designing twenty-five candidate potential drug-likeness for employs to inhibit MAO-B in PD.

5.2 Results and Discussion

5.2.1 MESP Calculation for Identification Binding Sites in Target Protein



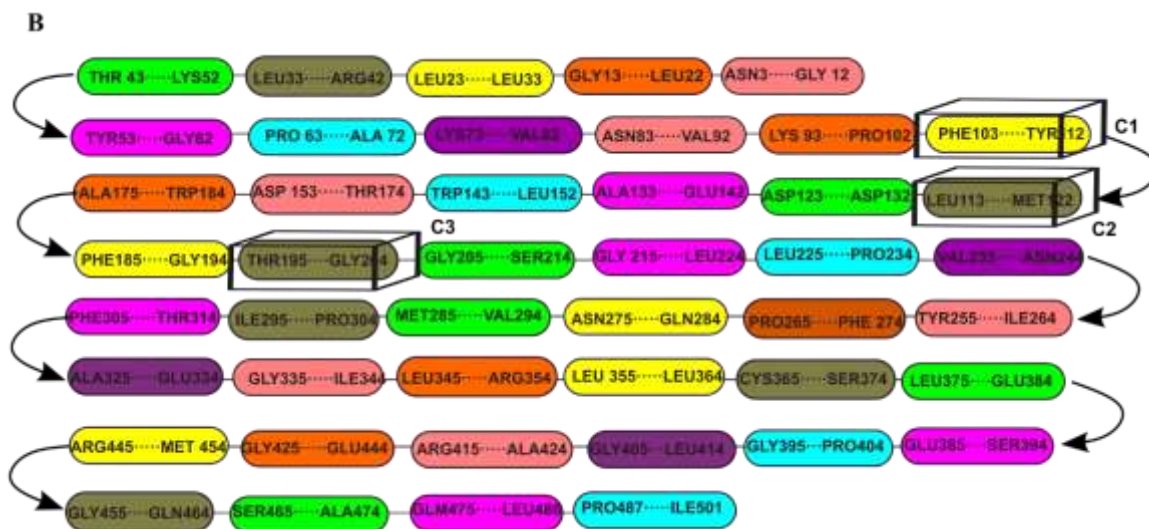
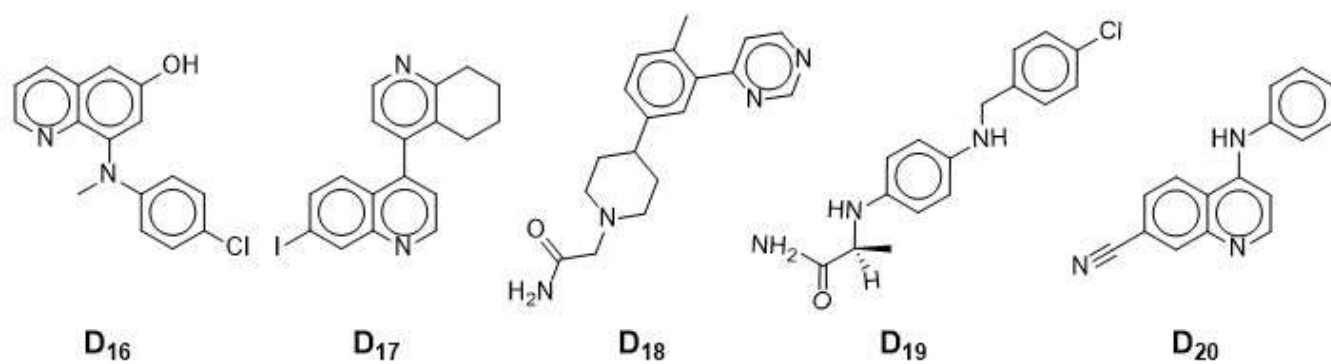
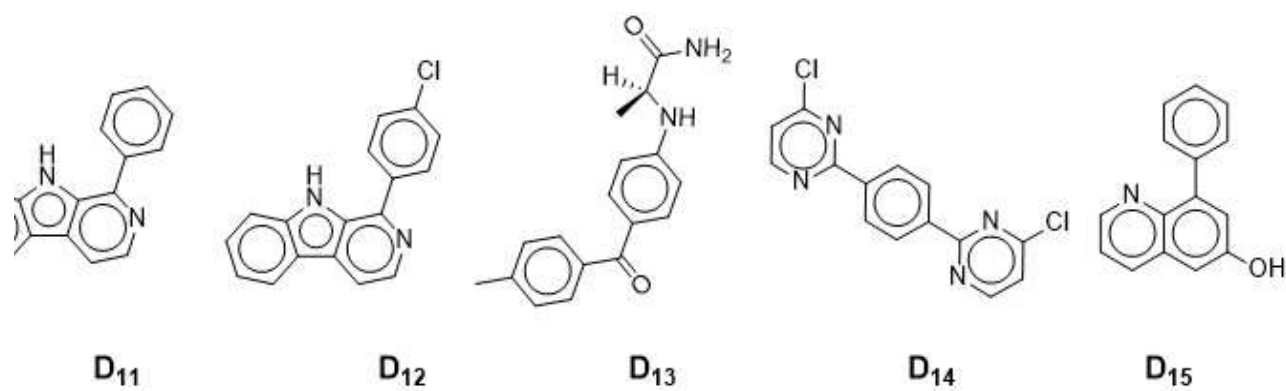
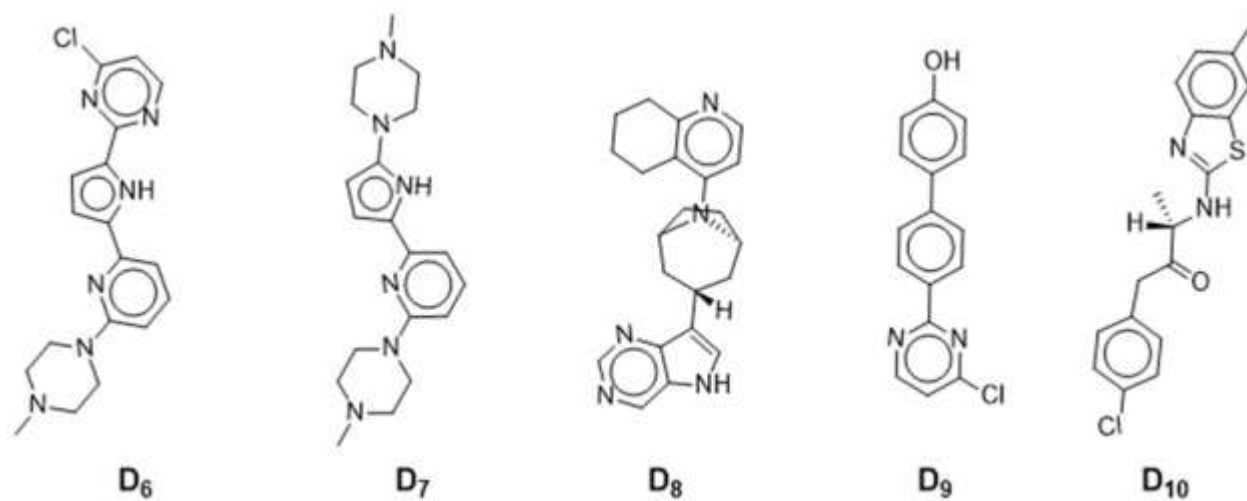
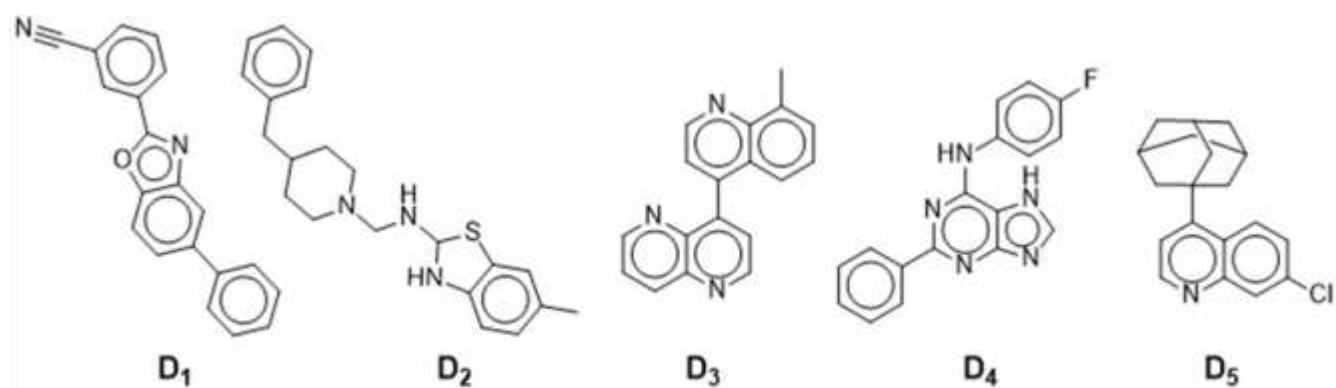


Figure 5.2 (A) MESP distribution of the MAO-B and (B) Substrate structures.

All the substrate cube files are loaded in VMD and transformed into one cube file. Figure 5.2 shows two color regions of MESP distributions, such as dark blue lobes inside red, representing most negative regions with function value -0.20 a.u. of the protein. The red color distribution shows low negative regions with a function value of -0.01 a.u. of the whole protein. The total MESP distribution can be seen in Figure 5.2 with two layers. The most negative regions, such as inner blue color lobes, are further investigated by surface cavity analysis as highlighted by black circles with notations C_1 - C_4 . To trace the exact location of the potential binding site, MESP function values are projected onto the plane. Scanning MESP function values within the plane resulted in surface cavities, as shown in Figure 5.2. The surface cavities were assumed to be the most potential binding sites. The two most negative MESP function values are spotted as inner and outer circles. This analysis provided the four best possible binding sites for further studies.

5.2.2 Design Potential Candidate Molecules

25 LORD-generated molecules are designed; the twenty-five with reference market drug Safinamide (R_1) are selected based on AutoDock vina binding energy, and corrected binding energy are reported in Figure 5.3.



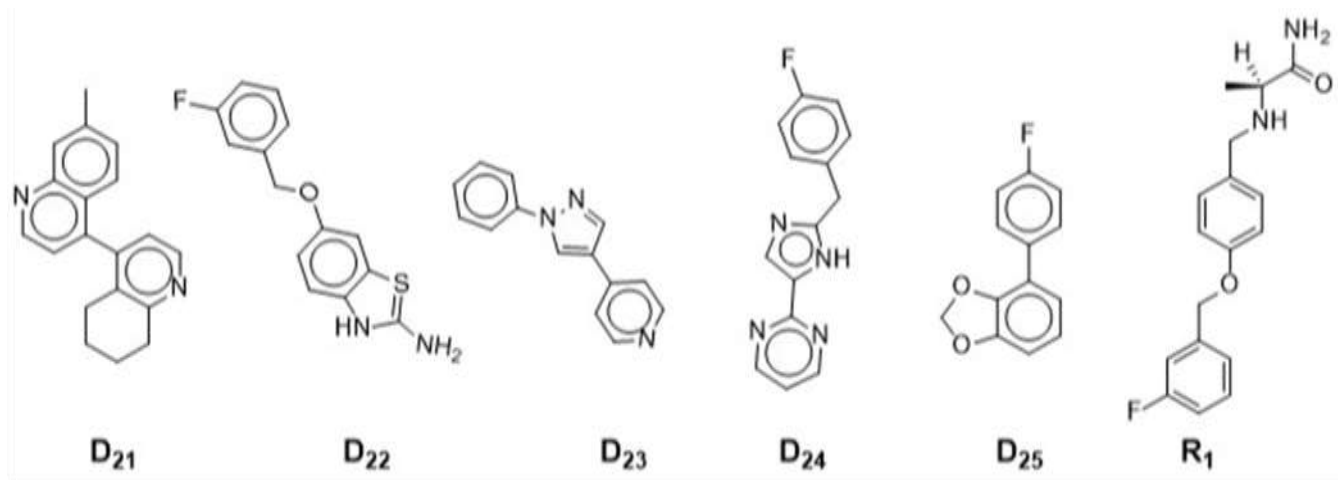


Figure 5.3 Potential 25 molecules predicted using LORD drug designer.

5.2.3 Stereochemistry and Protein-Ligand Complex Interaction Studies

Stereochemistry is vital for the generation of drugs, given that many therapies' activities depend on their three-dimensional configuration. Consideration should be given to stereoselectivity when designing small molecules that interact with their objectives. Drug developers can choose the most effective and secure stereoisomer or create a treatment that can specifically target a certain stereoisomer by assessing the stereochemistry of a therapeutic molecule and the mechanism of action.

The interaction of stereoisomers D_{1A}, D_{1B}, R_{1A}, and R_{1B} with the Site-4 (C₄) and chemical structure of a target protein was investigated using ChemDraw 3D software. Among the stereoisomers, D_{2A} showed a higher BE of -11.1 kcal/mol and a higher CBE of -8.3 kcal/mol than D_{2B}. The molecule also created two Pi-sulfur bonding connections with PHE A: 168, one hydrogen bond, and two Pi-sigma contacts with ARG A: 289 amino acid residues, indicating a more significant engagement with the target protein. D_{2B}, on the other hand, had better BE and CBE values than D_{2B}, with BE of -10.9 kcal/mol and CBE of -8.1 kcal/mol. Both R_{1A} and R_{1B} BE values are shown at -9.2 kcal/mol and -7.0 kcal/mol, respectively.

Figure 5.4 presented the stereoisomers D_{2A} and R_{1A} as drug-likeness molecules with much greater binding energy than the D_{2B} and R_{1B} stereoisomers. The minimal distance between amino acid residues and the kind of interaction with the protein was demonstrated to maintain the stereoisomer relationship. The interaction between D₂ stereoisomers has been shown to be better than that of the Sildenafil (R₁) drug due to their higher binding energies and shorter minimum distance between amino acid residues. The findings provide important insights into the impact of stereochemistry on ligand-protein interactions, which may aid in medicine design and development by optimizing stereochemistry for improved efficacy.

potential to be more effective treatments than those already available. LORD chemicals have lower CBE values than reference molecules, suggesting they interact favorably with the target protein.

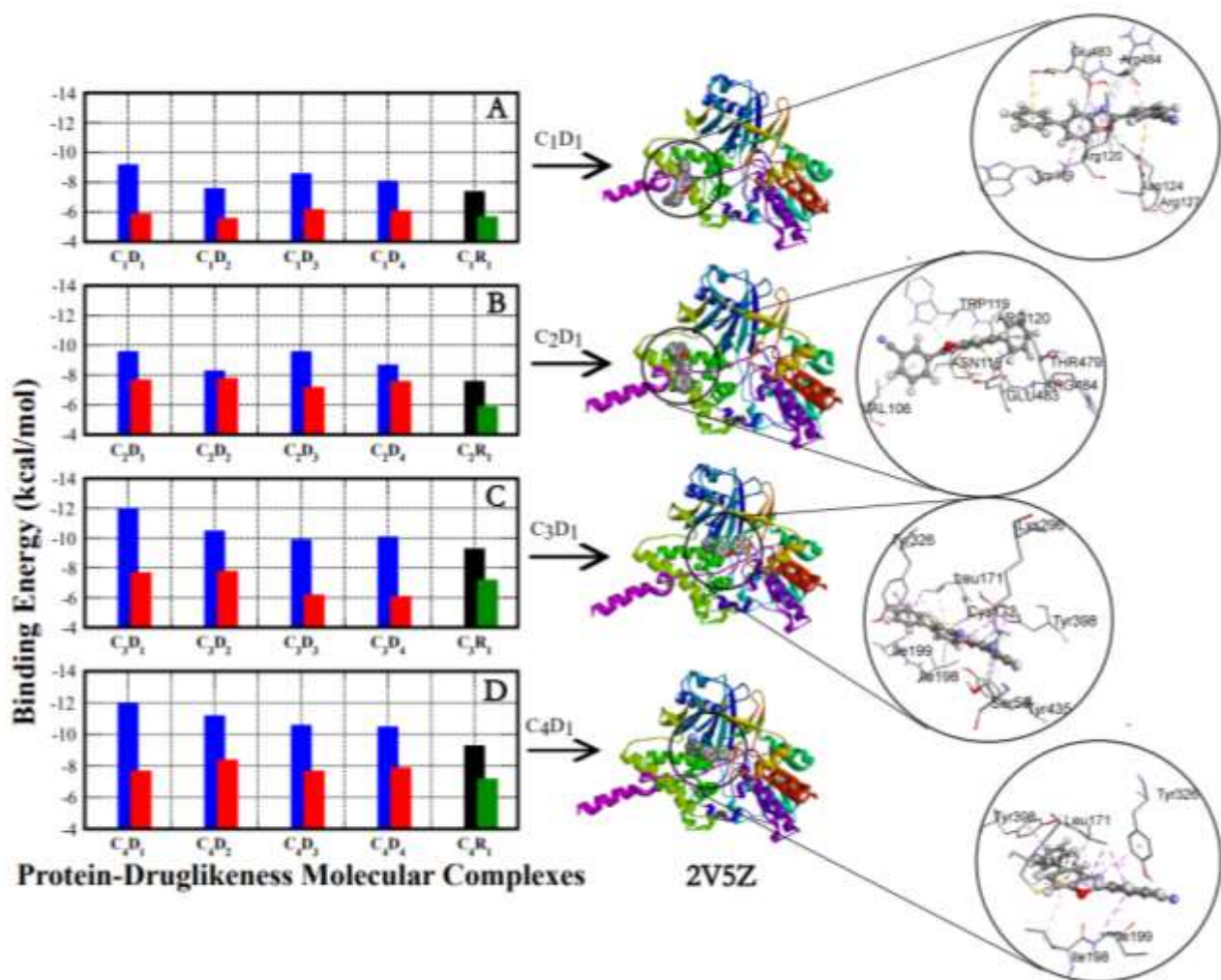


Figure 5.5 The binding site analysis of the specific drug molecule with each site, (A) MESP suggested site-1(C₁), (B) MESP suggested site-2 (C₂), (C) MESP suggested cavity site-3 (C₃), and (D) Experimentally reported site (C₄).

Figure 5.5 depicts the protein-ligand combination for all four target binding sites in the left column. C₁, C₂, C₃, and C₄ are the target binding sites. BE and CBE of the top four molecules generated by LORD (D₁, D₂, D₃, D₄) and reference molecules of safinamide (R₁) for each site (C₁ to C₄) are shown in a color bar plot on the right side of Figure 5.5. The black and green color bars represent the BE and CBE values of the reference molecules, respectively. The CBE values for LORD's top four molecules are much lower in energy than the reference compounds displayed in the black and green color bar plot for all

four locations. The findings give information on the design and development of new medicinal compounds with improved binding properties. The use of LORD in synthesizing innovative pharmaceutical compounds has the potential to produce more effective drugs with fewer side effects and better therapeutic outcomes.

5.2.5 Binding Sites Analysis from Protein-Ligand Complex

The ESP cavity analysis was performed to explore the interactions between proteins and their ligands at four unique sites labeled C₁ through C₄.

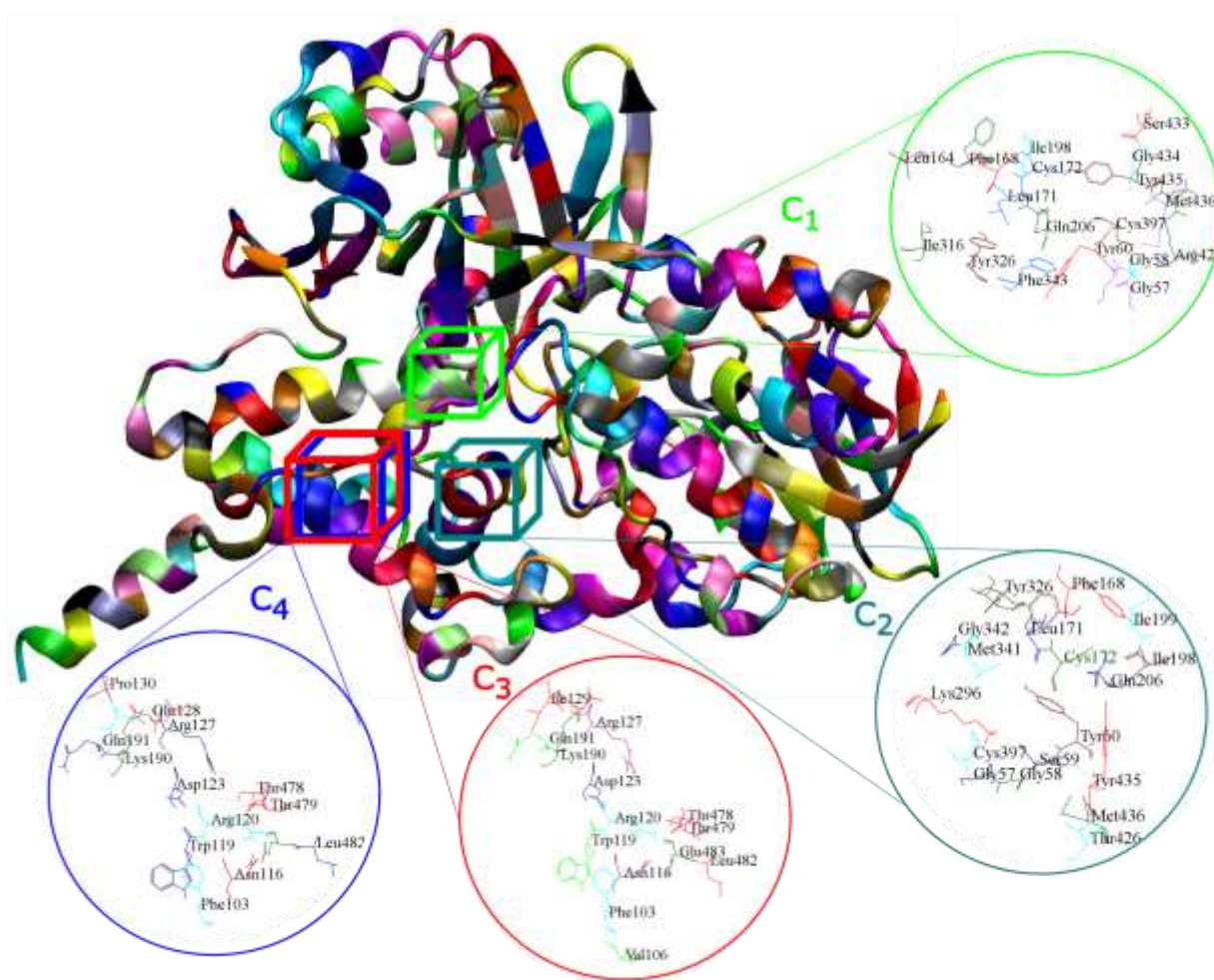


Figure 5.6 Monoamine Oxidase-B enzyme Protein active sites (C₁, C₂, C₃, and C₄).

Active amino acid residues at C₁ with lower binding energies were identified to be Arginine42 (ARG42), Glycine 57 (GLY57), Glycine 58 (GLY58), Tyrosine 60 (TYR60), Leucine 164 (LEU164),

Phenylalanine 168 (PHE168), Leucine 171 (LEU171), Cysteine172 (CYS172), Isoleucine 198 (ILE198), Glutamine 206 (GLN206), Isoleucine 316 (ILE316), Tyrosine 326 (TYR326), Phenylalanine 343 (PHE343), Cysteine 397 (CYS397), Serine 33 (SER33), Glycine (GLY434), Tyrosine 435 (TYR435), and Methionine 436 (MET436). At C₂, it was discovered that Glycine 57 (GLY57), Glycine 58 (GLY58), Serine 59 (SER59), Tyrosine 60 (TYR60), Phenylalanine 168 (PHE168), Leucine 171 (LEU171), Cysteine 172 (CYS172), Isoleucine198 (ILE198), Isoleucine 199 (ILE199), Glutamine 206 (GLN206), Lysine 296 (LYS296), Tyrosine 326 (TYR326), Methionine (MET341), Glycine 342 (GLY342), Cysteine 397 (CYS397), Threonine 426 (THR426), Tyrosine 435 (TYR435), and Methionine 436 (MET436) occurred. When compared to sites C₁ and C₂, sites C₃ and C₄ have higher binding energies with various amino acid residues.

Further analysis indicated a link between the experimentally reported site C₄ and the active amino acids at site C₃. Phenylalanine 103 (PHE103), Valine 106 (VAL106), Asparagine 116 (ASN116), Tryptophan 119 (TRP119), Arginine 120 (ARG120), Aspartic acid (ASP123), Arginine 127 (ARG127), Glutamic acid 128 (GLU128), Proline 130 (PRO130), Lysine 191 (LYS191), Glutamine 191 (GLN191), Threonine 478 (THR478), Threonine 479 (THR479), Leucine (LEU482), and Glutamic acid (GLU483) had the most significant binding energies at C₃. These findings suggest that amino acid residues at different sites have an essential role in protein-ligand interactions and that knowing the binding energies at each location is critical for predicting and changing protein-ligand interactions.

5.2.6 Cavity-Drug-Likeness Molecules Interaction Matrix

Drug design is based on how proteins and ligands interact, even though several drugs alter the function or signaling of specific proteins by attaching to receptors. A key step in the drug development process is identifying possible ligand-binding sites to forecast how ligands will interact with the protein and find ligands likely to have robust and precise binding interactions; computer modeling and simulation methods are frequently utilized. The MAO-B target protein was used to screen 25 potential drug-likeness molecules results are reported in Table 5.7, where interactions between proteins and their ligands are marked as hydrogen bonds, conventional hydrogen interactions, Pi-Pi contacts, Pi-alkyl interactions, Pi-sulphur interactions, Pi-cation interactions, and van der Waals interactions. Hydrogen bonds, traditional hydrogen contacts, Pi-Pi interactions, Pi-alkyl interactions, Pi-sulphur interactions, and Pi-cation interactions represent strong interactions (SI). In contrast, van der Waals interactions illustrate weak interactions (WI).

S.NO.	ARG 42	GLY 57	GLY 58	SER 59	TYR 60	TRP 119	LEU 164	PHE 168	LEU 171	CYS 172	IIL 198	ILE 199	GLN 206	ILE 316	TYR 326	PHE 343	CYS 397	TYR 398	TYR 435	MET 436	SI	WI
D1									2												9	6
D2									2			2									12	7
D3																		4			8	8
D4																	2				8	3
D5									2									2			8	6
D6																					9	7
D7	2																				5	6
D8																		2			7	7
D9									2									2			8	5
D10																					11	6
D11																		3	2		8	5
D12									3												7	6
D13																					6	4
D14									2									2			7	8
D15																		2			5	6
D16																					4	6
D17																					11	9
D18																					8	4
D19																					4	4
D20																		3	2		7	7
D21																		2			10	4
D22									2												7	7
D23																					6	6
D24																					3	8
D25																					4	4
R1																					8	8

vander Waals

conventional Hydrogen Bond

carbon Hydrogen bond

Pi-sigma

Pi-Pi stacked

Pi-Alkyl

Pi-sulfur

Pi-cation

SI-Strong Interaction

WI-Weak Interaction

Table 5.7 The protein target-Drug interaction matrix of 25 drug-likeness molecules with the MAO-B target protein within the experimentally reported site4 (C₄).

The safinamide complex was investigated for its interactions with 16 active amino acid residues. The interactions were studied for each of the 25 complexes, referred to as D₁ to D₂₅. The active amino acid residues included 15 in the D₁ complex, 19 in the D₂ complex, 16 in the D₃ complex, 12 in the D₄ complex, 14 in the D₅ complex, 16 in the D₆ complex, 11 in the D₇ complex, 14 in the D₈ complex, 13 in the D₉ complex, 17 in the D₁₀ complex, 13 in the D₁₁ complex, 13 in the D₁₂ complex, 10 in the D₁₃ complex, 15 in the D₁₄ complex, 11 in the D₁₅ complex, 10 in the D₁₆ complex, 20 in the D₁₇ complex, 22 in the D₁₈ complex, 8 in the D₁₉ complex, 14 in the D₂₀ complex, 14 in the D₂₁ complex, 14 in the D₂₂ complex, 12 in the D₂₃ complex, 11 in the D₂₄ complex, and 8 in the D₂₅ complex.

5.2.7 Physicochemical Properties and ADME Properties

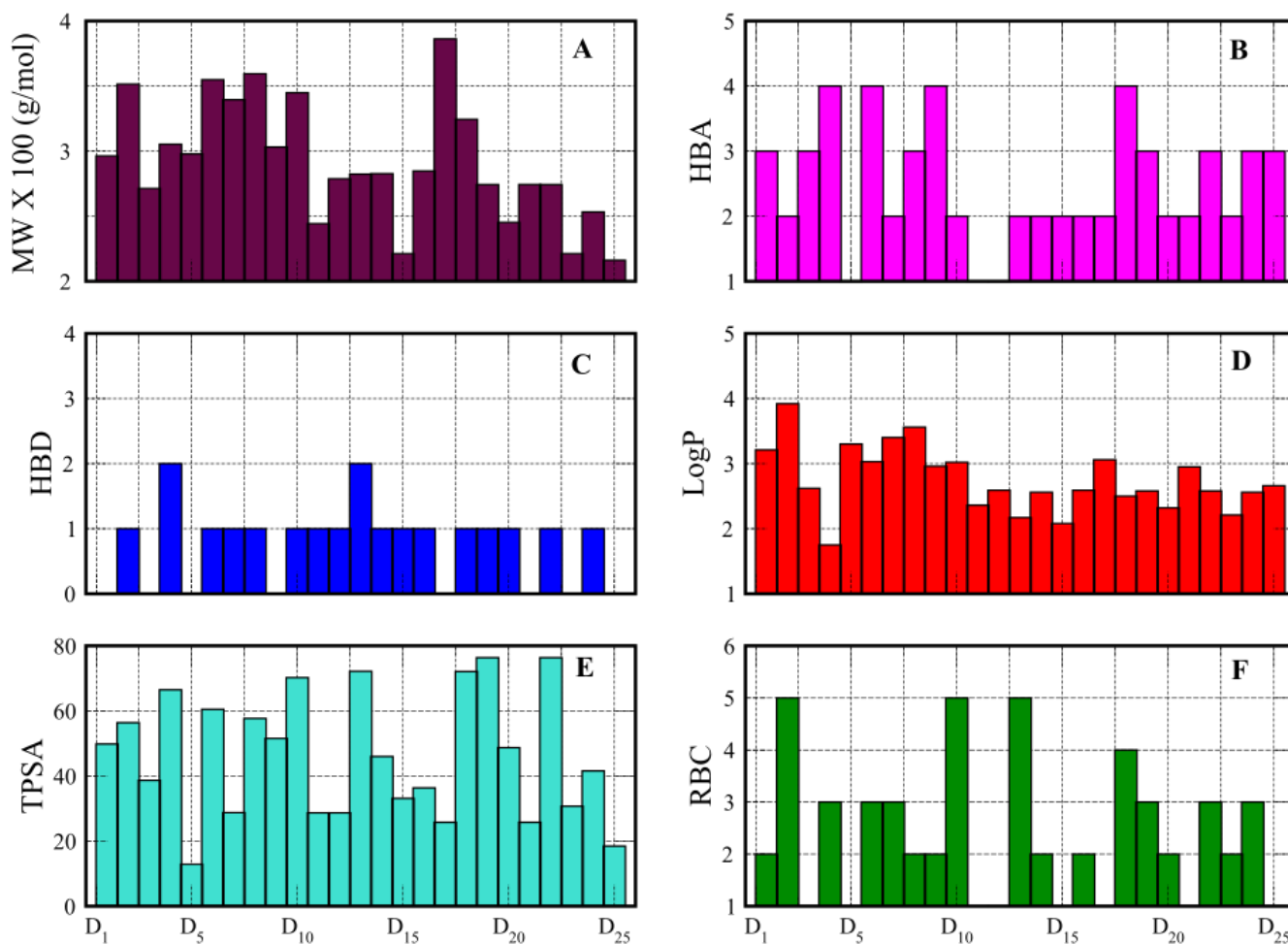


Figure 5.8 Physico-chemical properties of LORD 25 molecules. (A) Molecular weight (g/mol). (B) Hydrogen bond acceptor, (C) Hydrogen bond donor, (D) Partition Coefficient (Logp), (E) Total polar surface area and (F) Rotatable bond count.

Essentially, the ability to precisely target and modulate the activity of disease-related proteins is a vital component of the interaction between proteins and ligands, which is crucial for the creation of reliable and efficient therapies.

We used Lipinski's rule of five to evaluate the physicochemical characteristics of 25 compounds. These potential compounds have molecular weights between 216.21 and 386.23 g/mol, below the 500 g/mol threshold. Their lipophilicity-indicating partition coefficients are within the desirable range of 1.75-3.92, which denotes acceptable membrane permeability. The compounds have between 0 and 2 hydrogen bond donors (HBD), whereas there are between 1 and 5 hydrogen bond acceptors (HBA). We also measured the total polar surface area (TPSA) of the molecules. Which is a measure of the ability of molecules to interact with polar surfaces, ranges between 12.89-76.38 angstroms, and is less than 140 angstroms. The total number of rotatable bonds in the molecules ranges between 1-5, less than 10. Figure 5.8 shows that most molecules fit Lipinski's rule of five requirements, indicating that they are likely to have strong pharmacokinetic characteristics and can be researched further for medication development. This discovery highlights LORD's efficacy, taught using a machine learning method that employs Lipinski's rule of five.

We utilized the SWISS-ADME website to determine the suitability of the 25 potential compounds found by LORD. Based on the ADME study, we found the top five candidate compounds shown in Table 5.2.

Table 5.2 ADME / Pharmacokinetic proprieties of the site-wise selected molecules in ^aGastrointestinal absorption, ^bBlood brain barrier permeant, ^cP-gp substrate, ^dCytochrome P450 family 1 subfamily A member2, ^eCytochrome P450 family 2 subfamily c member19, ^fCytochrome P450 family 2 subfamily C member 9, ^gCytochrome P450 family 2 subfamily D member 6. ^hCytochrome P450 family 3 subfamily A member4, ⁱSkin permeation in cm.

S.No.	GI abs ^a	BBB permeant ^b	P-gp substrate ^c	CYP1A2 Inhibitor ^d	CYP2C19 Inhibitor ^e	CYP2C9 Inhibitor ^f	CYP2D6 Inhibitor ^g	CYP3A4 Inhibitor ^h	Log Kp ⁱ
D ₁	High	Yes	No	No	No	No	No	Yes	-5.64
D ₂	High	Yes	Yes	No	No	No	Yes	Yes	-4.38
D ₃	High	Yes	Yes	No	No	No	Yes	No	-5.79
D ₄	High	Yes	Yes	No	No	No	Yes	No	-4.5
R ₁	High	Yes	Yes	No	No	No	Yes	Yes	-5.58

5.2.8 Correlation of Potential Drug-Likeness Molecules vs. Site-Wise Binding Energy

CBE values for LORD 25 molecules are shown at each site, as shown in Figure 5.8. The range of CBE is -4.0 to -10.6 kcal/mol for all four sites of the MAO-B. Figure 5.9A illustrates the correlation between the number of drug-like molecules (X-axis) and their respective binding energies (Y-axis). On the other hand, Figure 5.9B depicts the relationship between the number of binding sites or cavities (X-axis) and the corresponding binding energies (Y-axis).

Figure 5.9 shows that the green color is the primary interaction site with the experimental site for all the drug-like molecules. CBE analysis showed that Site 4 and Site 3 maintained more potential for protein-ligand complex interactions than Site 1, 2, as indicated by Figure 5.9. We performed computational binding energy (CBE) analysis for four locations of the MAO-B enzyme to evaluate the binding affinity of the LORD 25 molecules, as shown in Figure 5.8. All four locations had CBE values of -3.8 to -8.5 kcal/mol. In Figures 5.9A and 5.9B, we also plotted the number of drug-likeness molecules against binding energy and the number of binding sites against binding energy. As shown by the green line in Figure 5.9, our study revealed that the green interaction site was the most suitable for all drug similarity compounds. Figure 5.9 shows that Sites 3 and 4 had a higher potential for protein-ligand complex interactions than Sites 1 and 2.

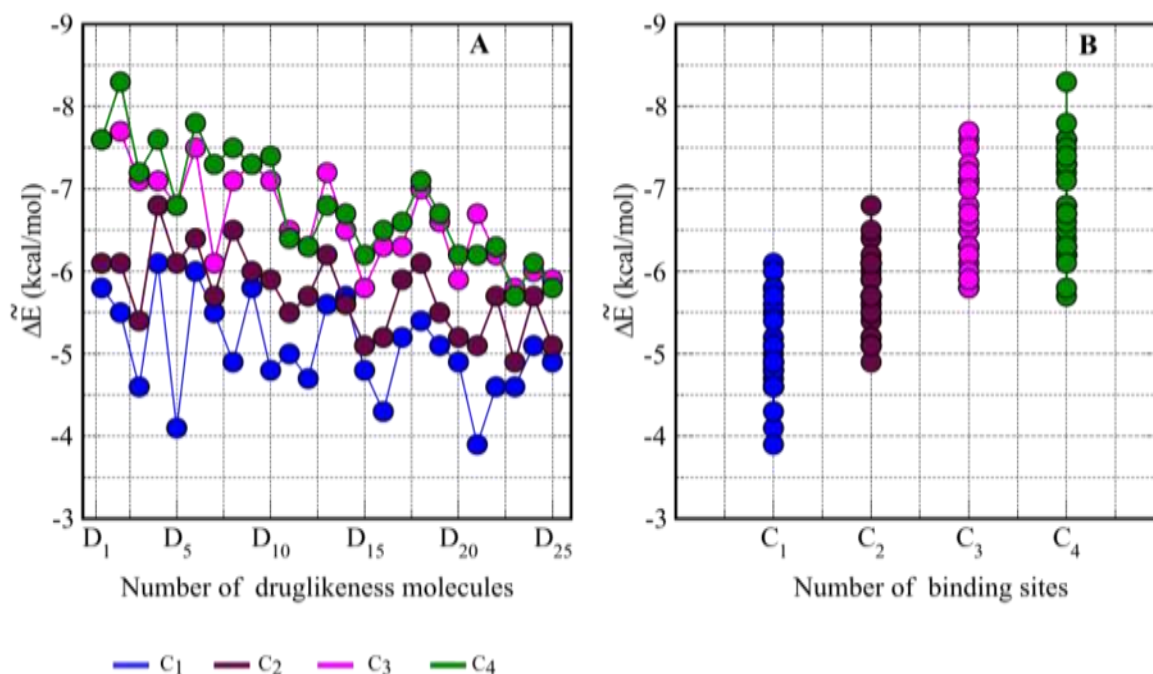


Figure 5.9 Corrected binding energy (CBE) trend curves for LORD 25 molecules at all four sites. (A) The X-axis is 25 drug-likeness molecules (D₁ to D₂₅), and Y-axis is CBE values and (B) The x-axis is ESP-suggested sites (C₁, C₂, C₃) and y-axis is CBE values.

Moreover, we identified the binding energies of each site-based on the CBE analysis. The blue line in Site 1 showed a CBE of -3.9 to -6.1 kcal/mol. The maroon color line in Site 2 exhibited a CBE of -4.9 to -6.8 kcal/mol, and the pink color line in Site 3 showed a BE of -5.8 to -7.7 kcal/mol, while the green color line in Site 4, which represents the experimental active site, showed CBE value of -4.9 to -8.3 kcal/mol. Our analysis suggests that Site 3 and Site 4 are more promising binding sites than Site 1 and Site 2, as their binding energies are stronger. This information can guide the selection of the most promising molecules for further experimental validation.

5.2.9 Molecular Dynamic (MD) Simulations

MD simulations investigated the stability of the top three D₁, D₂, and D₃ molecules, and safinamide drug molecules bind to the MAO-B enzyme receptor. For the MD simulations, the GROMACS 5.1.2 software was employed.

The CHARMM force field was employed to evaluate the energy of the protein, while the CHARMM General Force Field (CGenFF) was utilized to generate ligand topologies. After docking, the protein-ligand complex was solvated in a cubic box employing the TIP3 water model. To neutralize the system, chlorine (CL⁻) ions were added, and the energy was minimized using the steepest descent method with 10,000 steps. The NVT ensemble maintained the system's temperature at 298 K, while the NPT ensemble, ranged from 1.0 bar to 250 bars, was employed to simulate the ligand-protein complex.

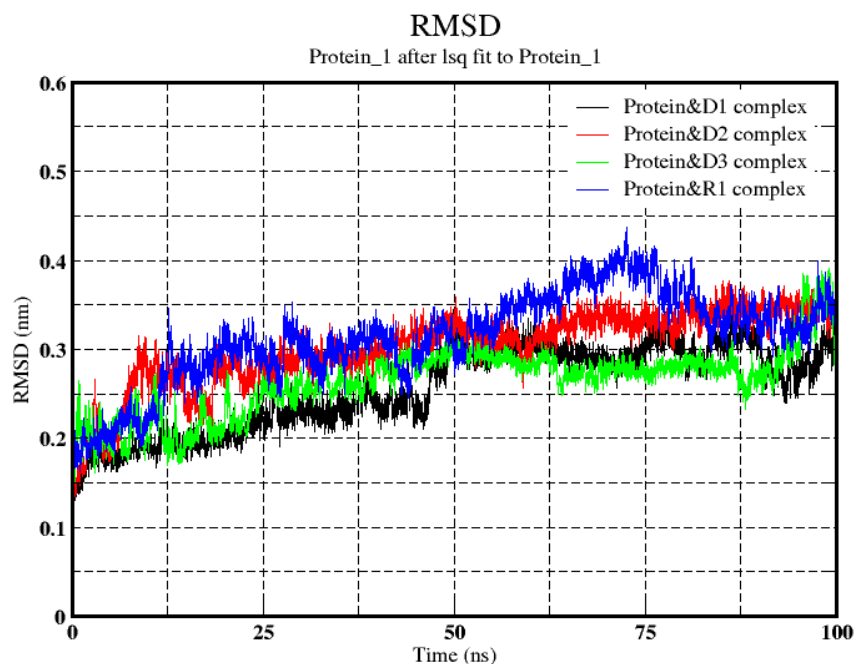


Figure 5.10 RMSD of the MD simulations.

The Particle Mesh Ewald and LINCS algorithms were utilized for short-range electrostatics, and a Van der Waals distance cut-off of 10 was used to limit all bonds. To preserve stability, the simulations were run for 100 ns. The strength of the top four drug-like compounds was compared using safinamide as the reference molecule. The RMSD graphs show how much the protein-ligand complex structure has deviated from its beginning structure. We can tell which molecule has a more stable protein-ligand complex structure by analyzing the RMSD graphs.

5.2.10 Free Energy Calculations

The *g-mmpbsa* program determines the free binding energy for protein-ligand complexes. This algorithm assesses the vacuum potential energy resulting from interactions between the bound and unbound protein and ligand using a molecular mechanics (MM) force field. The MM force field uses the Lennard-Jones potential function to compute the van der Waals (Evdw) energy contribution and the Coulomb potential function to determine the electrostatic energy contribution.

Table 5.3 Free energy calculation of the top four drug-likeness (D₁-D₄) molecules and Safinamide (R₁).

Time interval s (ns)	Complex Free Energies (kJ/mol)				
	D ₁	D ₂	D ₃	D ₄	R ₁
0-25	-93.3 +/- 8.9	-126.0 +/- 13.6	-63.9 +/- 17.0	-35.7 +/- 10.4	-101.3 +/- 8.7
25-50	-97.0 +/- 11.9	-130.1 +/- 11.0	-74.8 +/- 14.2	-62.0 +/- 24.7	-104.5 +/- 9.9
50-75	-103.4 +/- 7.7	-126.6 +/- 9.7	-71.4 +/- 13.5	-85.4 +/- 7.7	-101.4 +/- 7.4
75-100	-101.2 +/- 7.7	-124.5 +/- 10.0	-68.7 +/- 7.9	-81.0 +/- 9.4	-98.6 +/- 10.1

To assess the polar contribution, the solvation energy is evaluated using the Poisson-Boltzmann equation. The non-polar contribution is computed based on the assumption that it is proportional to the solvent-accessible surface area (SASA). The non-polar energy term (G_{nonpolar}) incorporates the van der Waals interactions and the attractive and repulsive forces arising from cavities between the solute and solvent. In order to calculate the binding energy, snapshots from the equilibrated region of the molecular dynamics (MD) trajectory are selected. The MmPbsaDecomp.py script provided by Kumari et al, with default settings, is employed to carry out all calculations using the *g-mmpbsa* package.

For the four protein-ligand complexes (D₁-D₄) and one protein-drug molecule(R₁) complex that were chosen, calculations of binding energy (BE) using the MM-PBSA program were made. The top four drug-likeness as D₁, D₂, D₃, D₄, and Safinamide (R₁) molecules, all had BE values between 0-10 ns, 30-40 ns, 60-70 ns, and 90-100 ns. The BE distribution calculation revealed that it significantly contributed to the overall BE, as shown in Table 5.3.

5.3 Conclusions

LORD is smart enough to build potential molecules based on the target site's environment. Thus, the molecules generated will be appropriate for that site. In this work, LORD is investigated extensively in this study on the MAO-B, which causes PD. We discovered four putative binding sites, one of which was an experimentally active site. Four potential binding sites are utilized to build molecules by LORD, producing 25 potential drug candidates. LORD 25 molecules have good corrected binding energy (CBE), AutoDockvina binding energy (BE), physicochemical, and ADME qualities compared to commercially available drugs such as Safinamide. In addition, MD simulations of the top-four potential drug-like compounds were investigated, and it was discovered that these are more stable in protein during long simulation runs than the Safinamide drug.

REFERENCES

1. Parkinson Disease-Associated Cognitive Impairment. *Nat. Rev. Dis. Primers* **2021**, 7 (1), 46.
2. Cui, L.; Hou, N.-N.; Wu, H.-M.; Zuo, X.; Lian, Y.-Z.; Zhang, C.-N.; Wang, Z.-F.; Zhang, X.; Zhu, J.-H. Prevalence of Alzheimer's Disease and Parkinson's Disease in China: An Updated Systematical Analysis. *Front. Aging Neurosci.* **2020**, 12, 603854.
3. Mao, Q.; Qin, W.-Z.; Zhang, A.; Ye, N. Recent Advances in Dopaminergic Strategies for the Treatment of Parkinson's Disease. *Acta Pharmacol. Sin.* **2020**, 41 (4), 471–482.
4. Carradori, S.; Silvestri, R. New Frontiers in Selective Human MAO-B Inhibitors: Miniperspective. *J. Med. Chem.* **2015**, 58 (17), 6717–6732.
5. Gealageas, R.; Devineau, A.; So, P. P. L.; Kim, C. M. J.; Surendraddoss, J.; Buchwalder, C.; Heller, M.; Goebeler, V.; Dullaghan, E. M.; Grierson, D. S.; Putnins, E. E. Development of Novel Monoamine Oxidase-B (MAO-B) Inhibitors with Reduced Blood–Brain Barrier Permeability for the Potential Management of Noncentral Nervous System Diseases. *J. Med. Chem.* **2018**, 61 (16), 7043–7064.
6. Binda, C.; Wang, J.; Pisani, L.; Caccia, C.; Carotti, A.; Salvati, P.; Edmondson, D. E.; Mattevi, A. Structures of Human Monoamine Oxidase B Complexes with Selective Noncovalent Inhibitors: Saffinamide and Coumarin Analogs. *J. Med. Chem.* **2007**, 50 (23), 5848–5852.
7. Azam, F.; Abodabos, H. S.; Taban, I. M.; Rfieda, A. R.; Mahmood, D.; Anwar, M. J.; Khan, S.; Sizochenko, N.; Poli, G.; Tuccinardi, T.; Ali, H. I. Rutin as Promising Drug for the Treatment of Parkinson's Disease: An Assessment of MAO-B Inhibitory Potential by Docking, Molecular Dynamics and DFT Studies. *Mol. Simul.* **2019**, 45 (18), 1563–1571.
8. Agrawal, N.; Mishra, P. Novel Isoxazole Derivatives as Potential Antiparkinson Agents: Synthesis, Evaluation of Monoamine Oxidase Inhibitory Activity and Docking Studies. *Med. Chem. Res.* **2019**, 28 (9), 1488–1501.
9. Takao, K.; Endo, S.; Nagai, J.; Kamauchi, H.; Takemura, Y.; Uesawa, Y.; Sugita, Y. 2-Styrylchromone Derivatives as Potent and Selective Monoamine Oxidase B Inhibitors. *Bioorg. Chem.* **2019**, 92 (103285), 103285.
10. Dai, Y.; Xu, Y.; Guo, C.; Xue, X.; Lin, D.; Lin, K. Discovery and Evaluation of New Compounds Targeting Ribosomal Protein S1 in Antibiotic-Resistant Mycobacterium Tuberculosis. *Eur. J. Med. Chem.* **2020**, 196 (112317), 112317.
11. Costas-Lago, M. C.; Besada, P.; Rodríguez-Enríquez, F.; Viña, D.; Vilar, S.; Uriarte, E.; Borges, F.; Terán, C. Synthesis and Structure-Activity Relationship Study of Novel 3-Heteroarylcoumarins Based on Pyridazine Scaffold as Selective MAO-B Inhibitors. *Eur. J. Med. Chem.* **2017**, 139, 1–11.
12. Zhang, C.; Yang, K.; Yu, S.; Su, J.; Yuan, S.; Han, J.; Chen, Y.; Gu, J.; Zhou, T.; Bai, R.; Xie, Y. Design, Synthesis and Biological Evaluation of Hydroxypyridinone-Coumarin Hybrids as Multimodal Monoamine Oxidase B Inhibitors and Iron Chelates against Alzheimer's Disease. *Eur. J. Med. Chem.* **2019**, 180, 367–382.

13. Binda, C.; Li, M.; Hubalek, F.; Restelli, N.; Edmondson, D. E.; Mattevi, A. Insights into the Mode of Inhibition of Human Mitochondrial Monoamine Oxidase B from High-Resolution Crystal Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (17), 9750–9755.
14. Binda, C.; Hubálek, F.; Li, M.; Herzig, Y.; Sterling, J.; Edmondson, D. E.; Mattevi, A. Crystal Structures of Monoamine Oxidase B in Complex with Four Inhibitors of the N-Propargylaminoindan Class. *J. Med. Chem.* **2004**, *47* (7), 1767–1774.
15. Jin, C.-F.; Wang, Z.-Z.; Chen, K.-Z.; Xu, T.-F.; Hao, G.-F. Computational Fragment-Based Design Facilitates Discovery of Potent and Selective Monoamine Oxidase-B (MAO-B) Inhibitor. *J. Med. Chem.* **2020**, *63* (23), 15021–15036.
16. Is, Y. S.; Durdagi, S.; Aksoydan, B.; Yurtsever, M. Proposing Novel MAO-B Hit Inhibitors Using Multidimensional Molecular Modeling Approaches and Application of Binary QSAR Models for Prediction of Their Therapeutic Activity, Pharmacokinetic and Toxicity Properties. *ACS Chem. Neurosci.* **2018**, *9* (7), 1768–1782.
17. Crisan, L.; Istrate, D.; Bora, A.; Pacureanu, L. Virtual Screening and Drug Repurposing Experiments to Identify Potential Novel Selective MAO-B Inhibitors for Parkinson's Disease Treatment. *Mol. Divers.* **2021**, *25* (3), 1775–1794.
18. Tandarić, T.; Vianello, R. Computational Insight into the Mechanism of the Irreversible Inhibition of Monoamine Oxidase Enzymes by the Antiparkinsonian Propargylamine Inhibitors Rasagiline and Selegiline. *ACS Chem. Neurosci.* **2019**, *10* (8), 3532–3542.
19. Liliana, P.; Alina, B.; Luminita, C. New Insights on the Activity and Selectivity of MAO-B Inhibitors through In Silico Methods. *Int. J. Mol. Sci.* **2023**, *24*(11), 9583.
20. Jalal, K.; Khan, K.; Haleem, D. J.; Uddin, R. In Silico Study to Identify New Monoamine Oxidase Type a (MAO-A) Selective Inhibitors from Natural Source by Virtual Screening and Molecular Dynamics Simulation. *J. Mol. Struct.* **2022**, *1254*, 132244.
21. Kumar, S.; Ayyannan, S. R. Identification of New Small Molecule Monoamine Oxidase-B Inhibitors through Pharmacophore-Based Virtual Screening, Molecular Docking and Molecular Dynamics Simulation Studies. *J. Biomol. Struct. Dyn.* **2022**, 1–22.
22. Iddiqui, A. J.; Jahan, S.; Siddiqui, M. A.; Khan, A.; Alshahrani, M. M.; Badraoui, R.; Adnan, M. Targeting Monoamine Oxidase B for the Treatment of Alzheimer's and Parkinson's Diseases Using Novel Inhibitors Identified Using an Integrated Approach of Machine Learning and Computer-Aided Drug Design. *Mathematics* **2023**, *11* (6), 1464.
23. Gogineni, V.; Nael, M. A.; Chaurasiya, N. D.; Elokely, K. M.; McCurdy, C. R.; Rimoldi, J. M.; Cutler, S. J.; Tekwani, B. L.; León, F. Computationally Assisted Lead Optimization of Novel Potent and Selective MAO-B Inhibitors. *Biomedicines* **2021**, *9* (10), 1304.
24. Boulaamane, Y.; Kandpal, P.; Chandra, A.; Britel, M. R.; Maurady, A. Chemical Library Design, QSAR Modeling and Molecular Dynamics Simulations of Naturally Occurring Coumarins as Dual Inhibitors of MAO-B and AChE. *J. Biomol. Struct. Dyn.* **2023**, 1–18.
25. Yi, C.; Liu, X.; Chen, K.; Liang, H.; Jin, C. Design, Synthesis and Evaluation of Novel Monoamine Oxidase B (MAO-B) Inhibitors with Improved Pharmacokinetic Properties for Parkinson's Disease. *Eur. J. Med. Chem.* **2023**, *252*, 115308.

26. Tan, Y.-Y.; Jenner, P.; Chen, S.-D. Monoamine Oxidase-B Inhibitors for the Treatment of Parkinson's Disease: Past, Present, and Future. *J. Parkinsons. Dis.* **2022**, *12* (2), 477–493.
27. El Aissouq, A.; Chedadi, O.; Bouachrine, M.; Ouammou, A.; Khalil, F. Development of Novel Monoamine Oxidase B (MAO-B) Inhibitors by Combined Application of Docking-Based Alignment, 3D-QSAR, ADMET Prediction, Molecular Dynamics Simulation, and MM_GBSA Binding Free Energy. *J. Biomol. Struct. Dyn.* **2022**, 1–14.
28. Ajala, A.; Uzairu, A.; Shallangwa, G. A.; Abechi, S. E. QSAR, Simulation Techniques, and ADMET/Pharmacokinetics Assessment of a Set of Compounds That Target MAO-B as Anti-Alzheimer Agent. *Futur. J. Pharm. Sci.* **2023**, *9*, 4.
29. Mladenović, M.; Patsilnakos, A.; Pirolli, A.; Sabatino, M.; Ragno, R. Understanding the Molecular Determinant of Reversible Human Monoamine Oxidase B Inhibitors Containing 2H-Chromen-2-One Core: Structure-Based and Ligand-Based Derived Three-Dimensional Quantitative Structure-Activity Relationships Predictive Models. *J. Chem. Inf. Model.* **2017**, *57* (4), 787–814.

CHAPTER 6

Designing of Potential Drug-Likeness Molecules for Tuberculosis

6.1 Introduction to Tuberculosis:

Tuberculosis (TB) is a bacterial infection caused by *Mycobacterium tuberculosis*, primarily affects the lungs but can also affect other human organs. In 2019, the World Health Organisation (WHO) projected that 1.7 million people died due to an estimated 10 million new cases of TB being identified.¹

By extending the range of crucial proteins used in TB chemotherapy, discovering ligands for most of the Mtb proteins may result in new TB drugs and targets. Only roughly 10 Mtb targets have historically been the focus of drug development efforts, leaving many other crucial proteins untapped. This strategy addresses medication resistance, toxicity, and the length of TB treatment by finding ligands for Mtb proteins.²

The four initial lines of TB treatment main drugs are isoniazid (INH), rifampicin (RIF), Pyrazinamide (PZA), and ethambutol (EMB). The discovery of effective TB drugs is a complicated and difficult procedure. One of the significant issues is the high attrition rate of candidate molecules throughout clinical development. Researchers are looking into new targets and discovering new pharmaceuticals or drugs for Anti TB therapy to address this. Another difficulty is drug resistance, which has become a serious public health concern worldwide. Researchers aim to treat drug-resistant tuberculosis by finding and conforming novel therapeutic targets.³

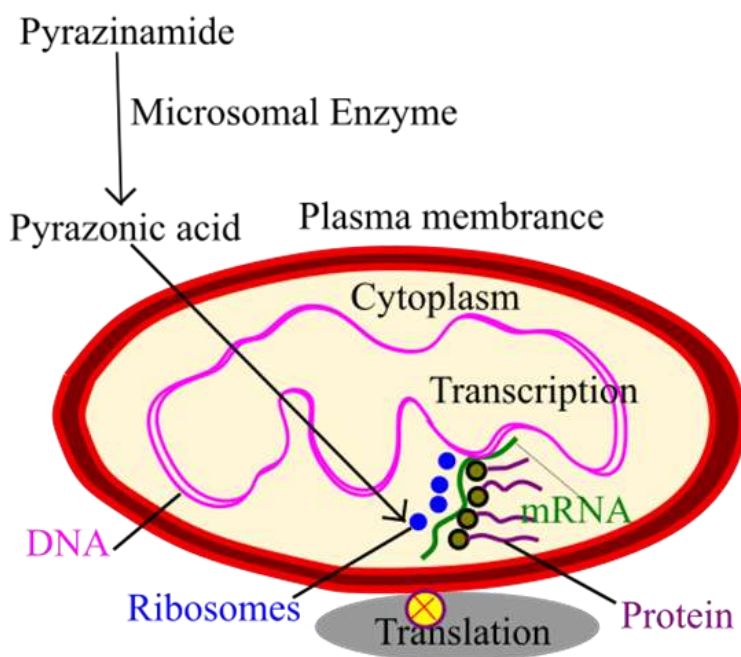


Figure 6.1 Pyrazinamide drug inhibits the role of ribosomal protein S1 of mycobacterium tuberculosis.

The main role of ribosomal protein S1 (RpsA) in relation to PZA resistance is crucial, as highlighted in the previous study. Notably, several compounds, namely PXYC1, PXYC2, PXYC12, PXYC13, and PXYD3, have demonstrated antibiotic properties, and the multiple dynamic simulations of PZA were conducted action on MTB mutations such as T370P and W403G are employed for the vital role of affecting protein activity.⁴ The other three mutants, N11k, P69T, and D126N, are, interact with PZA drug molecules are controlled or inhibited growth in Ribosomal protein S1 of MTB.⁵

The in-silico generation of novel inhibitors and their potential as an oral administration strategy. Due to their binding free energy, ADMET characteristics, and drug-likeness, benzimidazole compounds are an excellent choice for FtsZ inhibitors.^{6,7} We presented Table 6.1, which highlights key the previous literature, while our primary emphasis in this section centers on discussing methodology and providing a summary.

Table 6.1 Overview of previous literature on methodology and summary.

Methodology	Summary
Mycobacterium tuberculosis, proteome modeling, and virtual screening.	This study uses a computational approach to deorphanize targets in mycobacterium tuberculosis, the bacterium responsible for tuberculosis. By integrating genomic and proteomic data, we identified potential drug targets among the uncharacterized proteins in the M. tuberculosis genome. Our methodology combines protein-protein interaction networks, sequence homology analysis, and functional annotation to prioritize the most promising candidates for further experimental validation. ⁸
X-ray crystallography, pyrazinamide-S1 complex in Mycobacterium tuberculosis	This research elucidates the mechanism of action of Pyrazinamide, an important drug used to treat Mycobacterium tuberculosis infections. By determining the crystal structure of the pyrazinamide-S1 complex, the study reveals the specific binding interactions and highlights the key amino acid residues involved in the binding process. These findings provide valuable insights for developing more effective therapeutic strategies against tuberculosis. ⁹
Molecular docking, structure-based drug design	The application of structure-based computer-aided drug design (SB-CADD) in discovering potential drugs against Mycobacterium tuberculosis, the causative agent of tuberculosis. The study highlights the importance of computational methods in identifying compounds that target specific protein structures involved in the pathogenesis of tuberculosis. The findings provide valuable insights for developing novel therapeutics to combat this infectious disease. ¹⁰
Structure-activity relationship (SAR) studies, Molecular docking, and Synthesis approaches	This study focuses on identifying novel compounds that can effectively target the ribosomal protein S1 in antibiotic-resistant strains of Mycobacterium tuberculosis. Through a comprehensive evaluation process, several promising compounds have been discovered and assessed for their potential antimicrobial activity. The findings of this research provide valuable insights for developing new therapies against drug-resistant tuberculosis, addressing a critical global health concern. ¹¹
Molecular docking, pharmacophore, molecular dynamics	This study focuses on in silico drug design for ribosomal protein S1 (RpsA) with an Ala438 deletion, using the active compound Zrl15 as a basis. The researchers employed computational techniques to design potential drug candidates that can specifically target

simulations, and Free energy calculations	RpsA and potentially restore its functionality. This approach holds promise for the development of novel therapeutics against RpsA-related disorders. ¹²
Pyrazinamide, multidrug-resistant tb (mdr-tb), molecular docking, and molecular dynamic simulation.	In this study, we employed computational methods to investigate the binding mechanism between pyrazinoic acid (PA) and the RpsA protein. By utilizing molecular docking and molecular dynamics simulations, we revealed the key interactions involved in the PA-RpsA binding process. Our findings suggest that PA forms hydrogen bonds and hydrophobic interactions with specific residues in the binding pocket of RpsA, providing valuable insights for the Design of novel anti-tuberculosis drugs targeting this protein. ¹³
Benzimidazole, molecular docking and molecular dynamic simulations.	In This research focused on exploring the potential of benzimidazole congeners as innovative drug candidates. Through in-silico investigations, the study evaluated their drug likeliness by performing molecular docking simulations and analyzing the interactions with target proteins. Additionally, physicochemical properties were assessed to determine their suitability as drugs. The findings from this study provide valuable insights into the promising applications of benzimidazole congeners in drug discovery and development. ¹⁴
In silico structure-based drug screening, molecular modelling, pharmacophore modelling, QSAR.	Computational medicinal chemistry techniques to identify new chemical structures with promising anti-tuberculosis activity. They analyze various compounds through a rational drug design approach and propose potential candidates for further investigation. This study highlights the importance of computational methods in accelerating the discovery of effective anti-tuberculosis drugs, potentially aiding in the global fight against tuberculosis. ¹⁵
and in vitro translation reaction.	New compounds can inhibit the peptidyl transferase center in Mycobacterium tuberculosis. These small molecules have the potential to be developed into drugs for treating tuberculosis, providing a promising avenue for combating this infectious disease. ¹⁶
Virtual screening, Trans-translation, Drug-resistant	Potential therapeutic targets against drug-resistant strains of Mycobacterium tuberculosis. By investigating the lead compounds and identifying critical residues within the ribosomal protein S1, the researchers aim to develop novel drugs that can effectively combat the challenges posed by drug resistance in tuberculosis treatment. This research holds promise for advancing the development of new treatment options for tuberculosis. ¹⁷
Beilschmie compound, Multi-drug resistant tuberculosis and molecular docking, and virtual screening	Structure-based computational study to explore beilschmie compounds and their potential as targeted therapies against multidrug-resistant mycobacterium tuberculosis. The study aimed to identify promising compounds for further investigation by analyzing the molecular structures and interactions. The findings contribute to the development of novel treatments for tuberculosis, especially against drug-resistant strains. ¹⁸
Mutation Identification, Molecular docking, MD simulations, and Free energy calculations	Newly identified mutations in the ribosomal protein S1 (rpsA) are linked to resistance against Pyrazinamide. The researchers employed advanced computational methodologies, including molecular dynamics simulations and free energy calculations, to explore the effects of these mutations on the protein's stability and binding affinity to Pyrazinamide. ¹⁹
Structured-based drug design,	This study aimed to identify potential inhibitors for tuberculosis using a combination of structure-based drug design, molecular docking, and molecular dynamics simulation. The crystal structure of a relevant protein was obtained and used for the virtual screening of a compound library. Promising compounds were selected based on their

molecular docking and MD simulations	binding affinities, and their stability and dynamics were further evaluated through molecular dynamics simulations. These findings provide valuable insights for the development of new anti-tuberculosis drugs. ²⁰
Natural compounds, molecular docking, and MD simulations	This study uses a structure-based strategy to discover a natural compound inhibitor that can effectively target <i>M. tuberculosis</i> thioredoxin reductase. The binding affinity, stability, and dynamics of protein-ligand complexes were evaluated by employing molecular docking and MD simulation techniques. The findings from this study offer important insights into potential natural compounds that could be utilized as inhibitors to combat <i>M. tuberculosis</i> infection. ²¹
Computational techniques, virtual screening, drug resistance studies, in-vitro and in vivo studies	This study aimed to explore opportunities for overcoming tuberculosis by targeting emerging targets and developing inhibitors. A literature review identified potential targets, and computational techniques were used to screen and identify inhibitors against these targets. In vitro and in vivo experiments were conducted to evaluate the efficacy of the identified inhibitors against <i>Mycobacterium tuberculosis</i> . ²²

In this, we present a study on the successful application of MESP studies employed for finding the potential binding site in the protein target and LSTM generator for designing candidate potential drug-likeness molecules from starting fragment molecules. Presently, important molecular properties based on the physicochemical and ADMET properties using RDKit and Swiss-ADMET, Density function theory DFT calculation using the Gaussian software package, and molecular dynamics simulations studies were employed for dynamic behavior, stability of the protein-ligand complex by using Gromacs package. The free binding energies calculations and active amino acid residues of protein with ligand binding site interactions were calculated by the MM-PBSA method.

6.2 Result and Discussion

6.2.1 MESP Calculation for Identification of Binding Site in Target Protein

We have collected 16 protein fragments of amino acid residues from the ribosomal protein S1 of *Mycobacterium tuberculosis* (TB). There are 278-287, 288-297, 298-307, 308-317, 318-327, 328-337, 338-347, 348-357, 358-367, 368-377, 378-387, 388-397, 398-407, 408-417, 418-427, 428-438 amino acid residues and then calculated the Molecular electrostatic potential surface area (MESP) using by gaussian package. In comparison, MESP was employed at the possible binding site on the protein fragment surface of the amino acid.

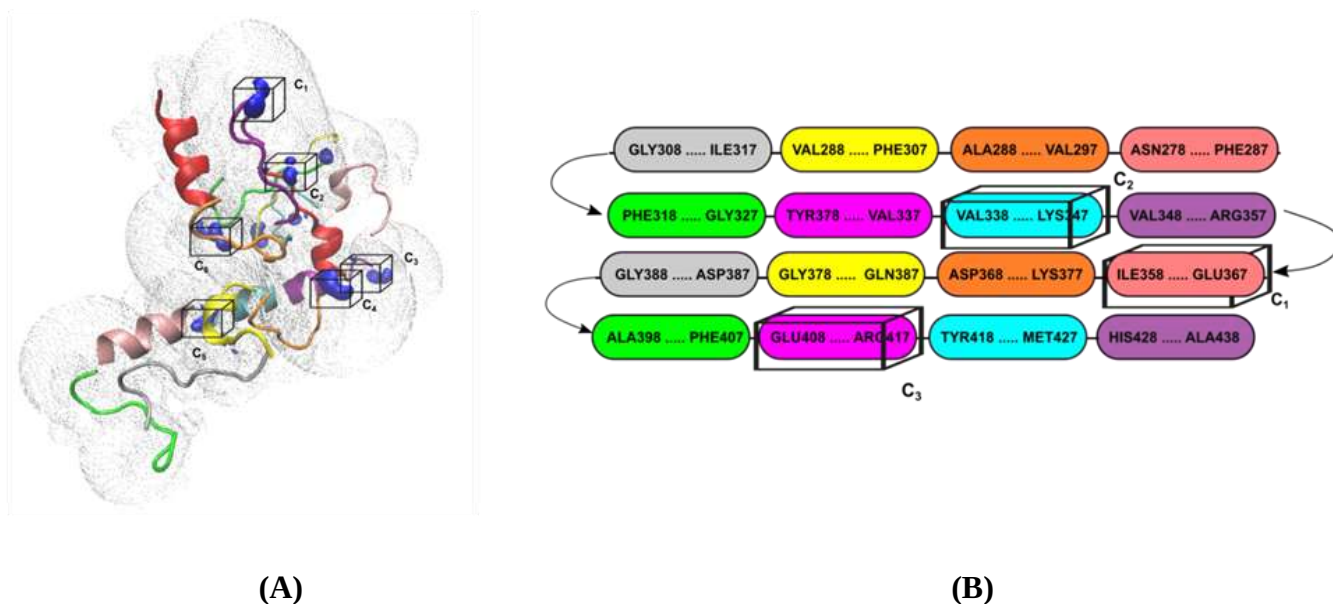
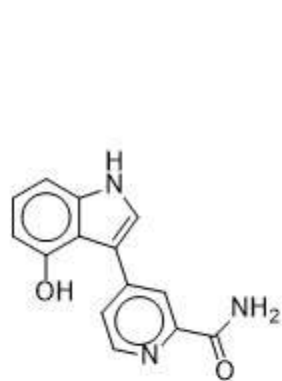


Figure 6.2 ESP studies of the ribosomal protein S1 of *Mycobacterium tuberculosis*.

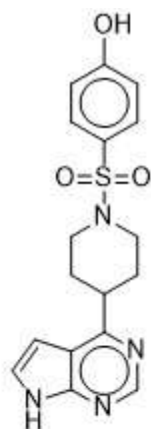
All the substrate cube files are loaded in VMD and transformed into one single cube file. Figure 6.2 shows two color regions of MESP distributions, such as dark blue lobes inside red, representing most negative regions with function value -0.20 of the protein. The red color distribution shows low negative regions with a function value of -0.01 a.u. of the whole protein. The total MESP distribution can be seen in Figure 6.2 with two layers. The most negative regions, such as inner blue color lobes, are further investigated by surface cavity analysis as highlighted by black circles with notations C₁-C₄. To trace the exact location of the potential binding site, MESP function values are projected onto the plane. Scanning ESP function values within the plane resulted in surface cavities, as shown in Figure 6.2. This analysis provided the four best potential binding sites for further studies.

6.2.2 Design of Potential Drug-Likeness Molecules

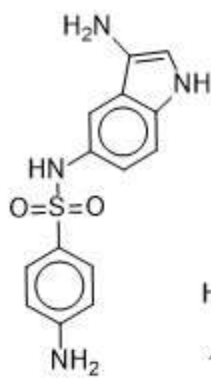
In the present manuscript, 25 LORD-generated molecules are provided. Twenty-five with reference market drug pyrazinamide (R₁) are selected based on AutoDock vina binding energy, and corrected binding energy are shown in Table Figure 6.3.



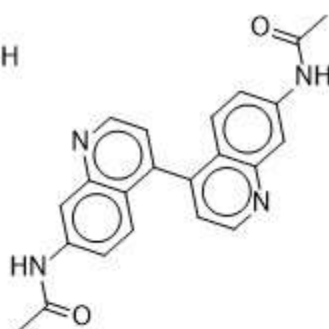
D1



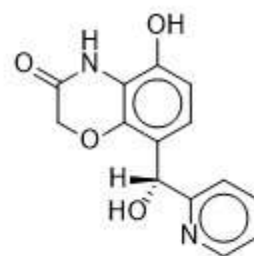
D2



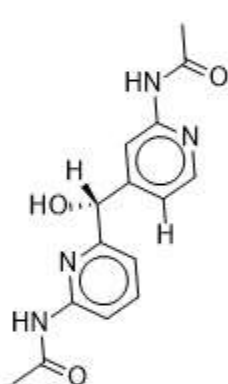
D3



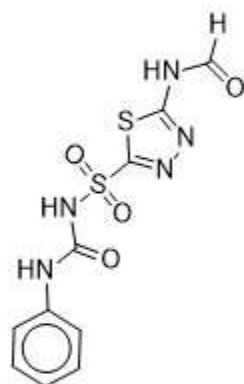
D4



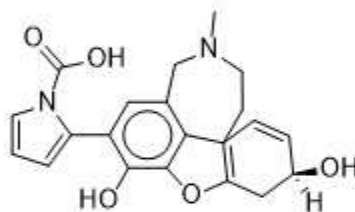
D5



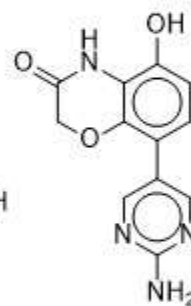
D6



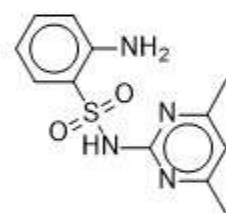
D7



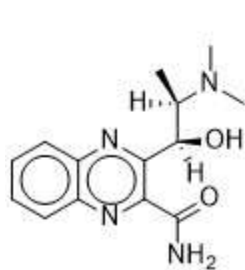
D8



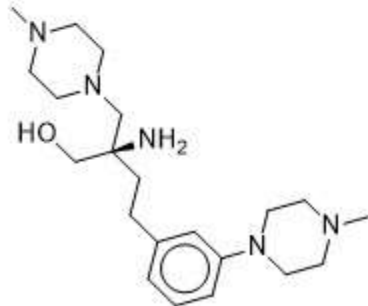
D9



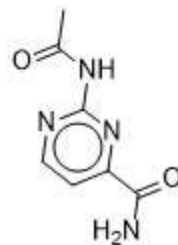
D10



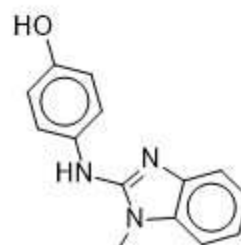
D11



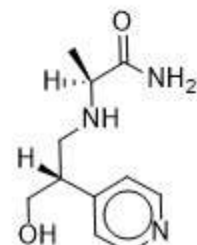
D12



D13



D14



D15

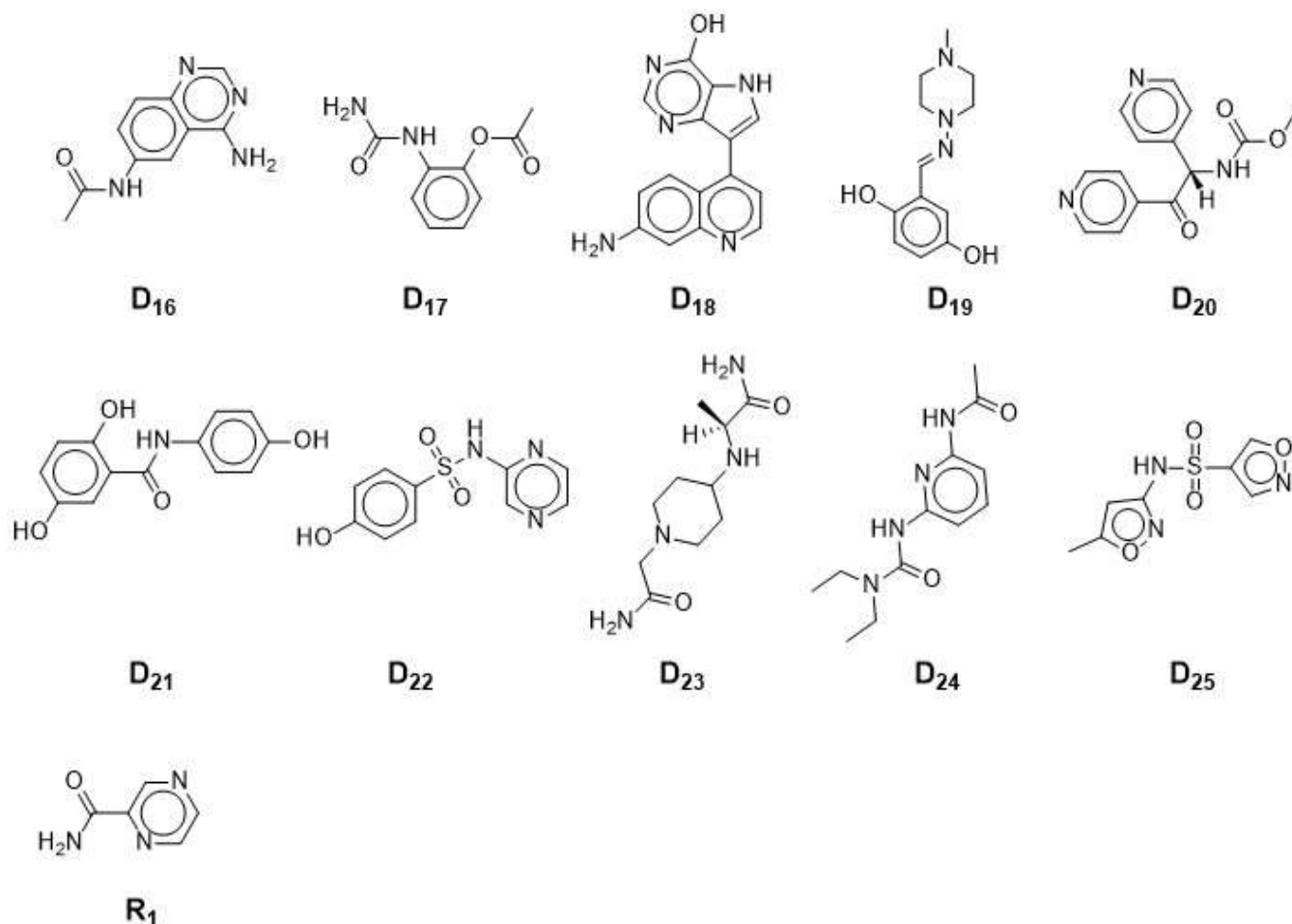


Figure 6.3 Twenty-five potential candidate drug-likeness molecules and Reference molecule geometric structures.

6.2.3 Stereochemistry and Protein-Ligand Complex Interaction Studies

Stereochemistry is important in drug discovery because the biological activity of many drugs is dependent on their three-dimensional configuration. Stereoselectivity must be considered when designing small molecules that interact with biological targets. Stereoisomers can be synthesized separately or isolated using chiral chromatography to produce new drugs. The activity of these particular stereoisomers can then be determined. Drug developers can pick the most efficient and secure stereoisomer or design a therapy targeting a specific stereoisomer by analyzing the stereochemistry of a medicinal molecule and its mechanism of action. In this way, we investigated the stereochemistry of a potential candidate molecule and its interactions with experimental Site-4 (C₄) of the target protein from *Mycobacterium tuberculosis* (TB). We employed ChemDraw 3D software to design 3D structures that accurately matched the stereochemistry of the stereocenters and studied the interactions of four

stereoisomers (D_{5A}, D_{5B}, D_{6A}, and D_{6B}) with the target protein, and determined molecular docking energy (BE) and corrected binding energy (CBE).

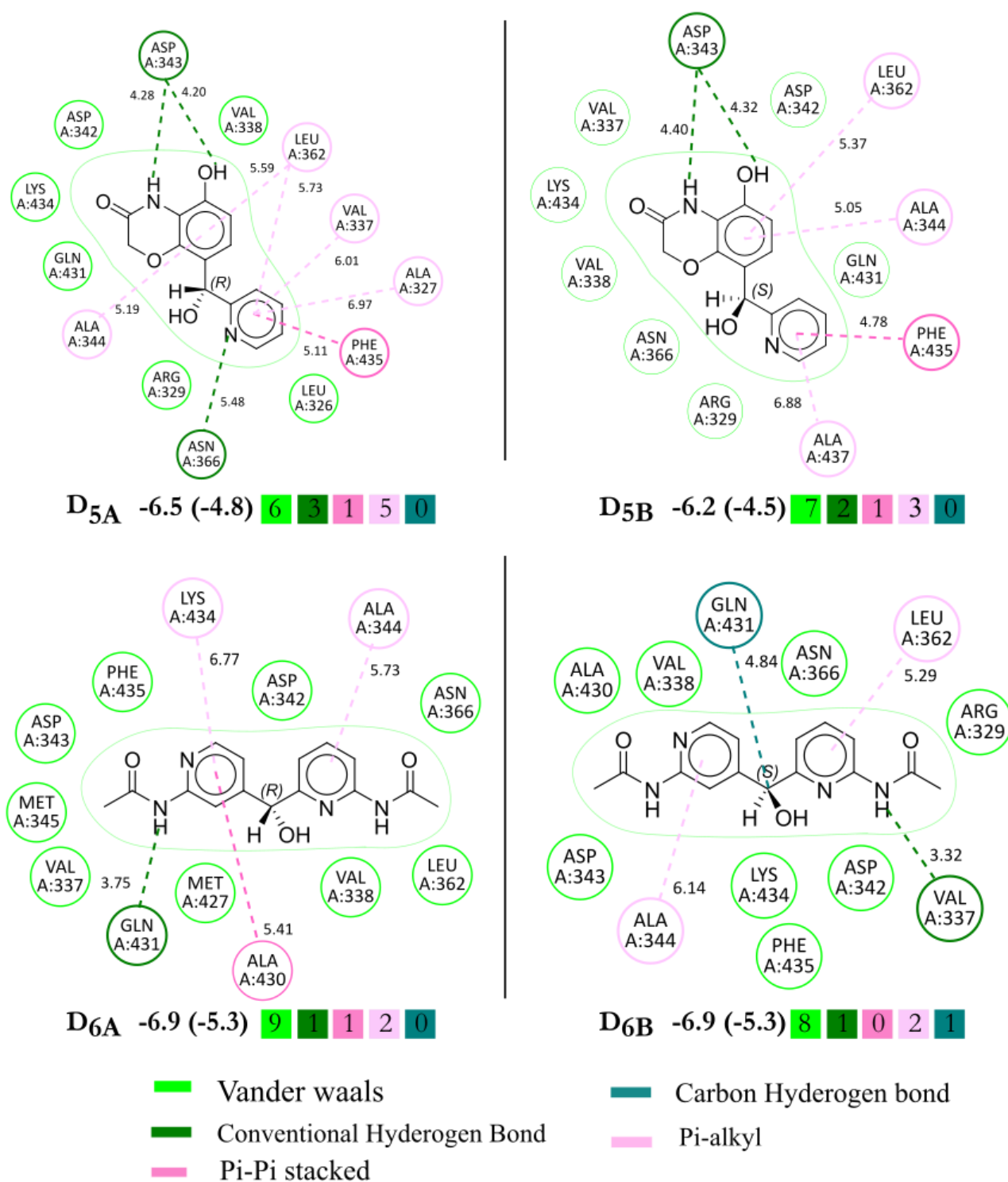
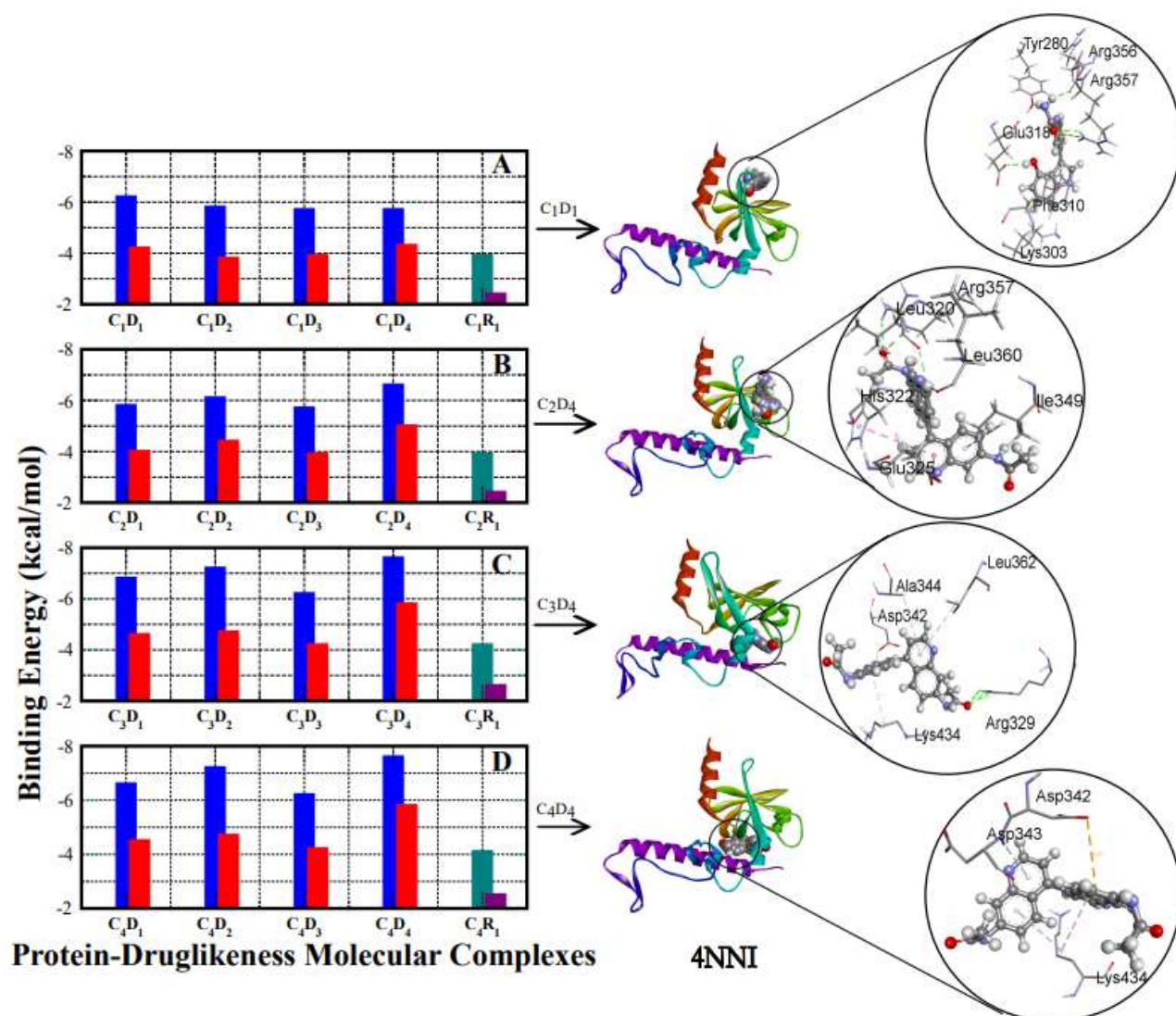


Figure 6.4 Stereo-isomer of D₅ and D₆ drug molecules and their interactions with target protein.

D_{5A} exhibited a BE of -6.5 kcal/mol and a CBE of -4.8 kcal/mol, with pyridine fragments being more important than D_{5B}. D_{5A} has one hydrogen bond contact with ASN A:366 and three Pi-alkyl interactions with ALA A:327. D_{5B}, on the other hand, had a BE shown at -6.2 kcal/mol and a CBE of -4.5 kcal/mol. D_{6A} and D_{6B} have a BE of 6.9 kcal/mol and a CBE of -5.3 kcal/mol. These results explain how stereo centers affect molecular docking energy and the complexity of interactions between proteins and ligands. Drug developers can design more effective and safe therapeutics by understanding the stereochemistry of chemical compounds and their interactions with biological targets.

6.2.4 Protein-Ligand Complex and Their Interaction Studies of the Top Four Drug-Likeness Molecules at Selected Four Binding Sites



We presented a diagram of the protein-ligand complex in Figure 6.5 (left column) for all four binding sites to help you understand the binding site for the target. As seen in Figures 6 and 5, the black circle on each picture depicts the precise position of the ligand complex interactions. Furthermore, in the right column of Figure 6.5, we compared the BE and CBE of the top five compounds created by LORD (D₁, D₂, D₃, D₄) for each site (C₁ to C₄) with the reference molecule, Pyrazinamide (R₁).

The BE and CBE values for the top five potential drug-likeness molecules designed by LORD are better binding energy than the reference molecules, as seen in the maroon and green4 color bar plots for all four binding sites. Finally, our LORD-generated potential molecules beat traditional market drugs in all four binding sites.

6.2.5 Binding Sites Analysis from Protein-Ligand Complex

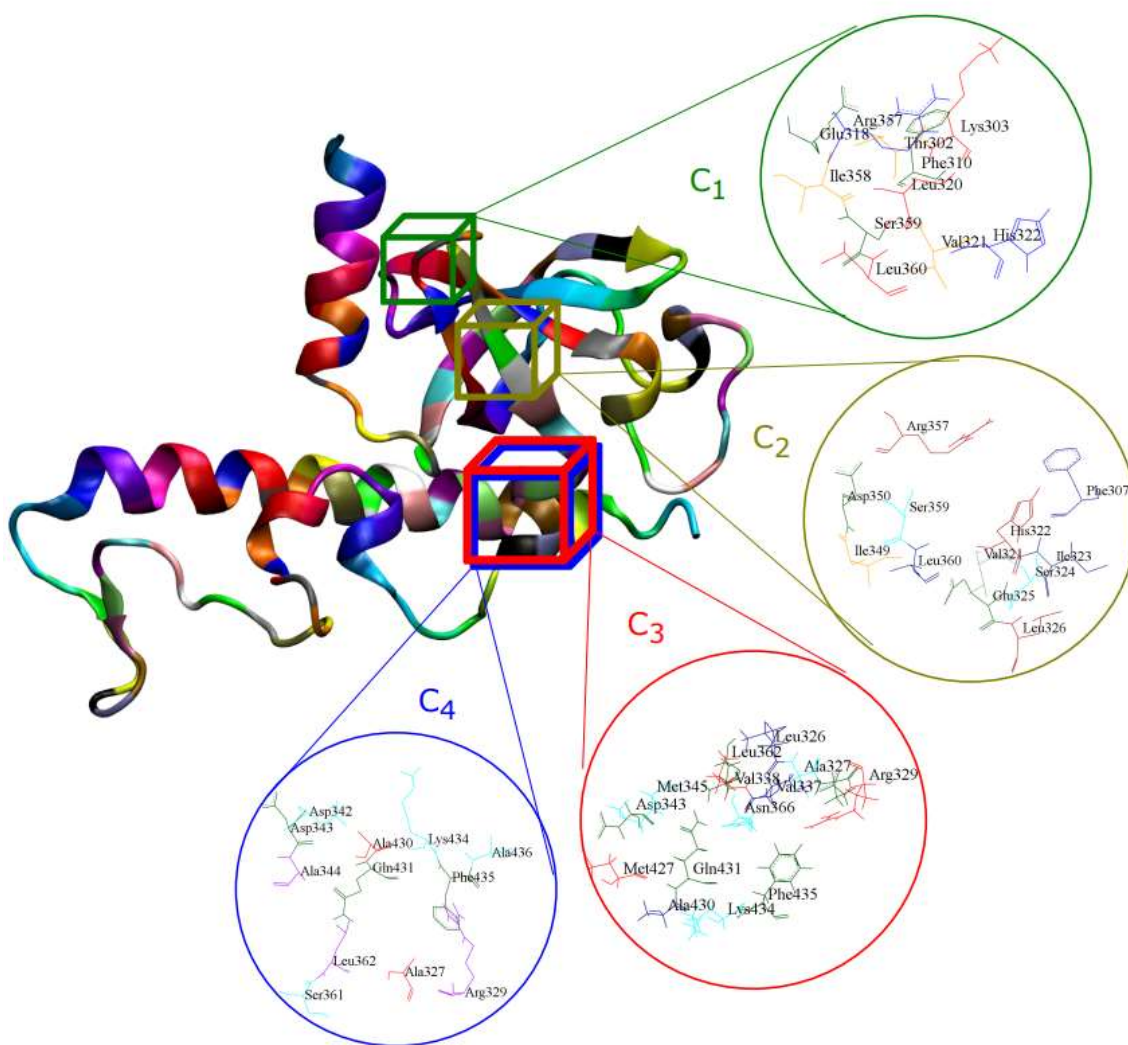


Figure 6.6 Ribosomal protein S1 of mycobacterium tuberculosis active sites (C₁, C₂, C₃, C₄).

The ESP cavity analysis was employed to explore the interactions between proteins and their ligands at four different binding sets (C₁-C₄). Our findings demonstrated that the amino acid residues at Site-1 (C₁) have high binding energies than the other three binding sites. These residues include Thr 302, Lys 303, Phe 310, Glu 318, Leu 320, Val 321, His 322, Arg 357, Ile 358, Ser 359, and Leu 360. At Site-2 (C₂), Phe 307, Val 321, His 322, Ile 323, Ser 324, Glu 325, Leu 326, Ile 349, Asp 350, Arg 357, Ser 359, and Leu 360 were identified as active amino acid residues. Additionally, it was found that sites 4 and 3 exhibit higher binding energies than C₁ and C₂, respectively, while C₁ and C₂ exhibit lower energies. Our investigation also revealed that the active amino acids at site-3 (C₃) and site-4 (C₄) correspond with the experimentally confirmed site-4 (C₄). These active amino acids include Leu 326, Ala 327, Val329, Val337, Val338, Asp 343, Met 345, Leu 362, Asn 366, Met 427, Ala 430, Gln 431, Lys 434, and Phe 435.

6.2.6 Cavity-Drug-Likeness Molecules Interaction Matrix

Understanding how proteins and ligands interact is crucial in potential molecule development. It is essential to identify potential ligand-binding sites and precisely determine the three-dimensional structure of the protein to facilitate the development of effective drugs. Computer modeling and simulation methods are widely used to predict how ligands interact with the protein and identify ligands likely to have robust and precise binding interactions. For this, 25 potential drug-likeness molecules were docked to the target protein, and the docked compounds were graded using a strict filter that considered four parameters. Strong contacts were represented by dark green for hydrogen bonds, pine green for typical hydrogen interactions, and pink for non-covalent interactions such as Pi-Pi, Pi-alkyl, Pi-sulphur, and Pi-cation. Weak van der Waals interactions were represented by light green. The pyrazinamide complex was investigated for interactions with four active amino acid residues. In contrast, the D₁ complex had 11 active amino acid residue interactions, D₂ had 13 active amino acids, D₃ had 16 active amino acids, D₄ had 12 active amino acids, D₅ had 9 active amino acids, D₆ had 12 active amino acids, D₇ had 9 active amino acids, D₈ had 12 active amino acids, D₉ had 12 active amino acids, D₁₀ had 12 active amino acids, D₁₁ had 13 active amino acids, D₁₂ had 11 active amino acids, D₁₃ had 11 active amino acids, D₁₄ had 12 active amino acids, D₁₅ had 13 active amino acids, D₁₆ had 10 active amino acids, D₁₇ had 11 active amino acids, D₁₈ had 9 active amino acids, D₁₉ had 9 active amino acids, D₂₀ had 9 active amino acids, D₂₁ had 11 active amino acids, D₂₂ had 10 active amino acids, D₂₃ had 9 active amino acids, D₂₄ had 10 active amino acids, and D₂₅ had 11 active amino acids. The binding energy was employed to calculate and analyze all of these interactions, and the results are shown in Figure 6.7.

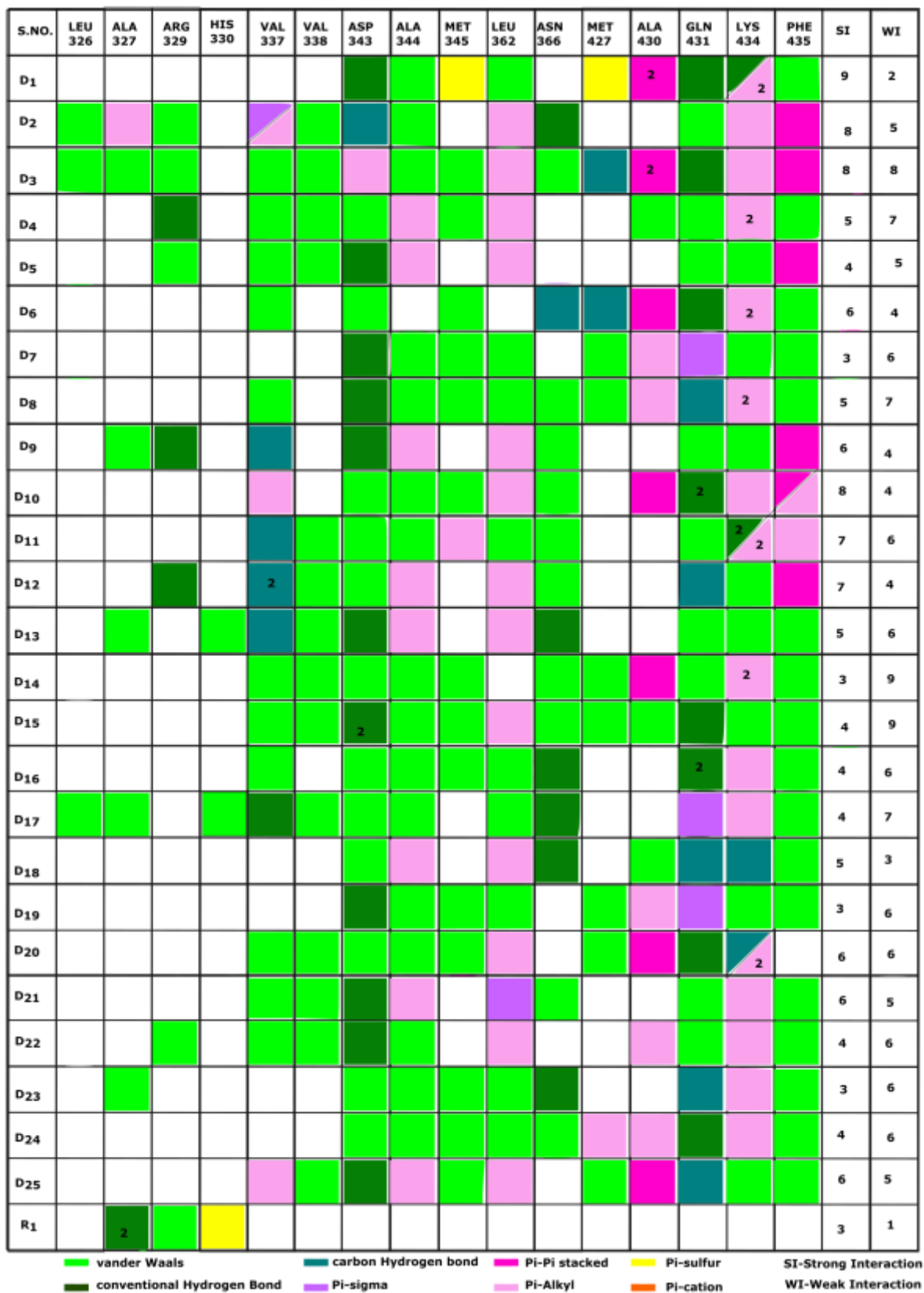


Figure 6.7 Target-Drug Interaction matrix for 25 drug candidates.

Determining the interactions between proteins and ligands is critical for developing reliable and effective potential drug-likeness molecules. A way to precisely target and regulate the activity of disease-related proteins is essential for developing potential drug-likeness molecules.

6.2.7 Physicochemical and ADME properties

The physicochemical features of twenty-five potential drug-likeness molecules have been analyzed using the Lipinski rule of five to determine their applicability. The Lipinski rule of five is a series of principles used to assess a molecule's drug-likeness based on its physicochemical features.

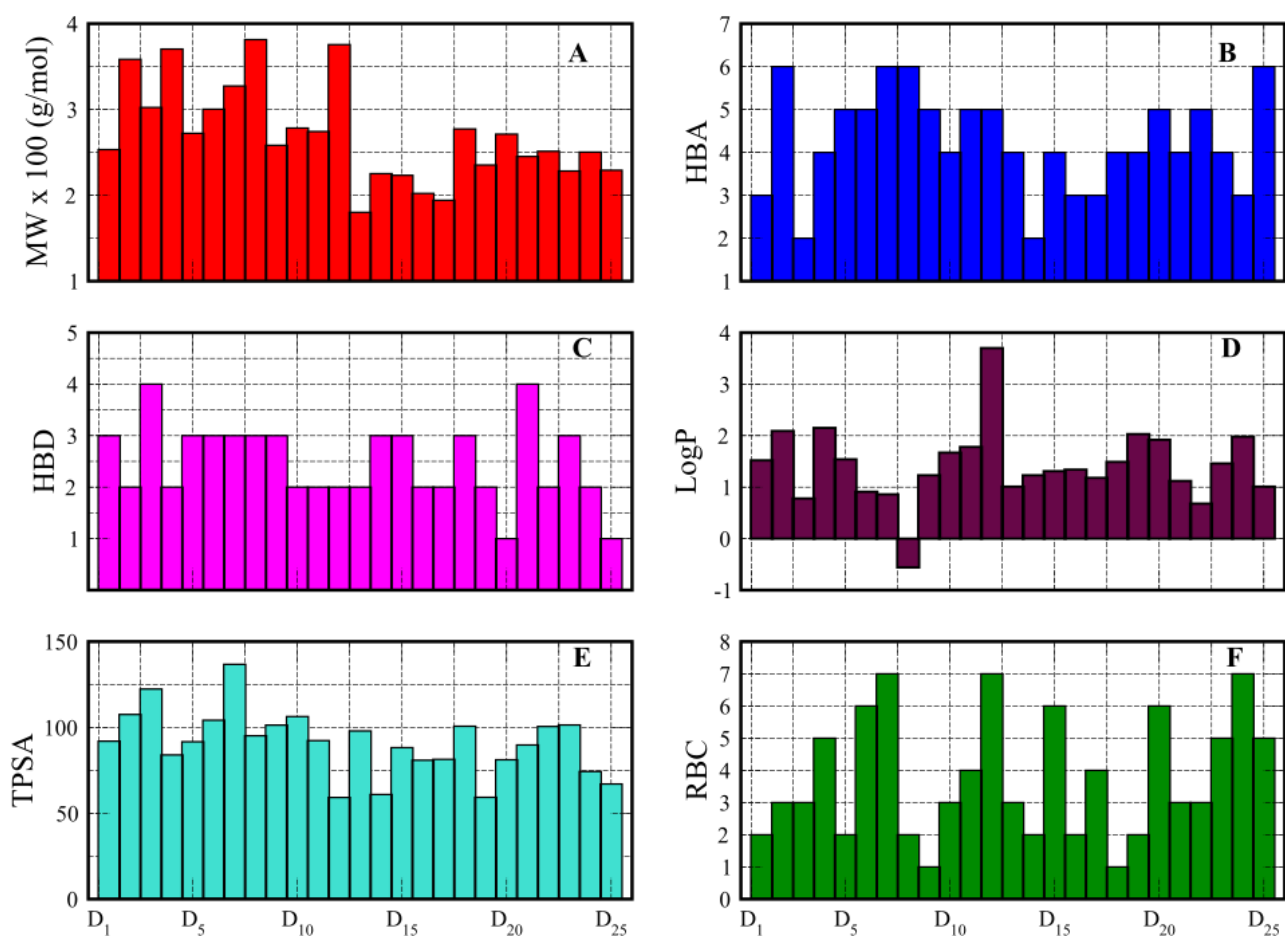


Figure 6.8 Physico-chemical properties of 25 drug-likeness molecules. (A) Molecular weight (g/mol), (B) Hydrogen bond acceptor, (C) Hydrogen bond donor, (D) Partition Coefficient (LogP), (E) Total polar surface area, and (F) Rotatable bond count.

The analysis showed that all twenty-five molecules had an MW ranging from 306.32 to 484.63 g/mol, which is less than the upper limit of 500 g/mol set by the Lipinski rule of five. The logP values of these molecules ranged between 2.64 and 4.97, which is less than the upper limit of 5 set by the Lipinski rule of five. The number of HBD was found to be between 0-3, while the number of HBA was between 1-5. The molecular polar surface area of the molecules ranged from 26.17 to 73.36 angstroms, which is less than the upper limit of 140 angstroms. Additionally, the TPSA in these molecules ranged from 1-8, less than the upper limit of 10 set by the Lipinski rule of five.

Table 6.2 ADME / Pharmacokinetic proprieties of the selected molecules in ^aGastrointestinal absorption, ^bBlood brain barrier permeant, ^cP-gp substrate, ^dCytochrome P450 family 1 subfamily A member2, ^eCytochrome P450 family 2 subfamily c member19, ^fCytochrome P450 family 2 subfamily C member 9, ^gCytochrome P450 family 2 subfamily D member 6, ^hCytochrome P450 family 3 subfamily A member4, ⁱSkin permeation in cm.

S.No.	GI abs ^a	BBB permeant ^b	P-gp substrate ^c	CYP1A2 Inhibitor ^d	CYP2C19 Inhibitor ^e	CYP2C9 Inhibitor ^f	CYP2D6 Inhibitor ^g	CYP3A4 Inhibitor ^h	Log K _p ⁱ
D ₁	High	No	No	Yes	No	No	Yes	Yes	-6.81
D ₂	High	No	Yes	Yes	No	No	No	No	-7.26
D ₃	High	No	Yes	Yes	No	No	No	No	-7.29
D ₄	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-6.83
R ₁	High	No	No	No	No	No	No	No	-7.48

To succeed with twenty-five molecules, ADME (Absorption, Distribution, Metabolism, and Excretion) qualities are critical for determining the effectiveness, safety, and pharmacokinetic aspects of potential drug-likeness molecules during drug development. The SWISS ADME website is a tool for predicting these attributes based on the molecule's chemical structure. Prediction of ADME features is critical in finding possible drug candidates with a high probability of success in clinical trials. The top five candidate potent drug-likeness molecules with promising ADME features are listed in Table 6.2.

6.2.8 Correlation of Potential Drug-Likeness Molecules vs. Site-wise binding Energy

The binding energies of twenty-five drug-like compounds were determined for four distinct target protein binding sites. The results are shown in Figures 6.8(A) and 6.8(B), with binding energies ranging from -2.5 to -5.8 kcal/mol. Figure 6.8(A) plots the number of drug-likeness molecules against their respective binding energies, while Figure 6.8(B) plots the number of binding sites against their binding energies. In both pictures, the green color denotes sites that interact considerably with the experimental site. As shown in Figure 6.8, the study indicated that binding sites C₃ and C₄ had more potential for binding energies of protein and ligand complexes than C₁ and C₂.

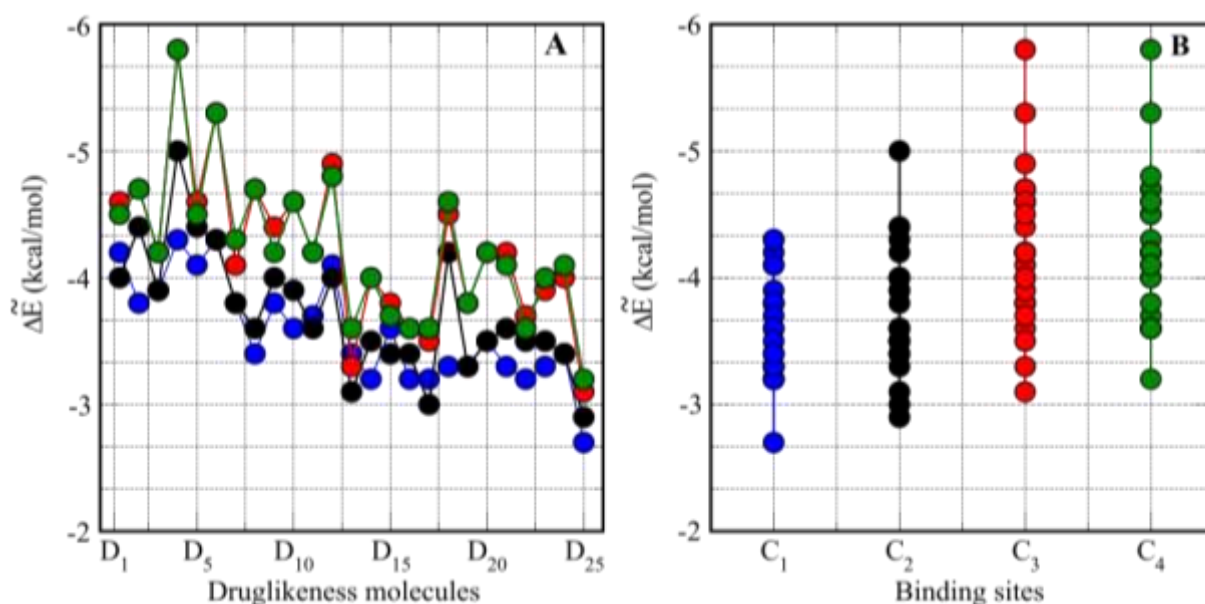


Figure 6.9 Corrected binding energy (CBE) trend curves for LORD 25 molecules at all four sites. (A) The X-axis is 25 drug-likeness molecules (D₁ to D₂₅), and Y-axis is CBE values and (B) The X-axis is ESP-suggested sites (C₁,C₂,C₃,C₄), and Y-axis is CBE values.

6.2.9 Molecular Dynamics (MD) Simulations

MD simulations of the pyrazinamide drug and four drug-like molecules (D₁, D₂, D₃, and D₄) were carried out using the GROMACS 5.1.2 software. The CHARMM force field was employed to calculate the protein's energy and molecular dynamics (MD) simulations, and the CHARMM General Force Field (CGenFF) was used to create the ligand's topology. The protein-ligand combination was put inside a container and submerged once the docking process was finished, thereby solvating the system using the TIP3 water model. Chloride (Cl⁻) ions were appropriately added to the complex to keep it neutral. The

steepest descent approach was used for a total of 10,000 steps to conserve energy. The simulations employed the NVT ensemble while maintaining the temperature at 298 K and the NPT ensemble ranging from 1.0 bar to 250 bar to simulate the ligand-protein interaction. The Particle Mesh Ewald and LINCS algorithms were used to address short-range electrostatic interactions, and all bonds were restricted using a Van Der Waals distance threshold of 10. To assure stability, the simulations were carried out for 100 ns, and the RMSD graphs for the four top drug molecules with Pyrazinamide are presented in Figure 6.10. The researchers were most likely comparing the stability of the four drug-likeness molecules with Pyrazinamide to evaluate their potential as therapeutic candidates.

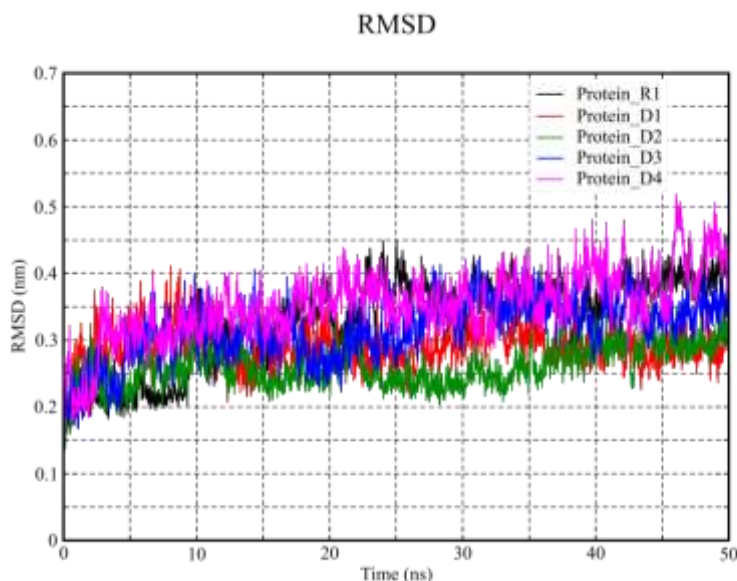


Figure 6.10 MD simulation studies of the RMSD, top four drug-likeness molecules (D₁, D₂, D₃, D₄) and reference drug molecule (R₁).

6.2.10 Free Energy Calculations

The protein-ligand complexes binding energy, or binding energy (BE) as it is more often known, was determined using the *g-mmpbsa* software. The molecular mechanics (MM) force field is used to model both bound and unbound states to determine the vacuum potential energy resulting from various interactions. This involves utilizing the Coulomb and Lennard-Jones potential functions to compute the electrostatic and van der Waals (Evdw) energy contributions. Additionally, the Poisson- Boltzmann equation is a significant advance that was made in precisely identifying the polar component during the evaluation of the solvation-free energy.

In general, considering that non-electrostatic solvation energy correlates precisely with the solvent-accessible surface area (SASA) when calculating the non-polar contribution. The non-polar energy term (G_{nonpolar}) includes the van der Waals interaction, such as the attracting and repelling forces that emerge between the solute and solvent due to cavities. During the Free energy calculations, snapshots of the equilibrated area from the molecular dynamics (MD) trajectory are captured. To carry out these computations, the default parameters outlined by Kumari et al. are utilized, and the MmPbsaDecomp.py script included with the *g-mmpbsa* package is employed.

The D₁, D₂, D₃ have better complex free energy than the reference of Pyrazinamide (R₁) are shown in Table 6.3.

Time intervals (ns)	Complex free energy (kJ/mol)				
	D ₁	D ₂	D ₃	D ₄	R ₁
0-25	-62.8 +/- 8.5	-41.5 +/- 13.8	-61.0 +/- 16.4	-51.9 +/- 15.3	-16.0 +/- 6.4
25-50	-60.6 +/- 13.7	-61.7 +/- 14.4	-61.0 +/- 16.4	-54.2 +/- 34.2	-20.1 +/- 14.0

Table 6.3 Free energy of interaction of the top four drug-likeness (D₁-D₄) molecules and Pyrazinamide (R₁) with target protein.

The results are shown in Table 5.3, which reveal that D₁, D₂, and D₃ have lower complex free energy values than the reference molecule pyrazinamide (R₁). These data imply that D₁, D₂, and D₃ may have higher binding affinity and preferential interactions with the protein than the reference molecules.

6.3 Conclusions

LORD is thoroughly investigated in this work on the ribosomal protein S1 of mycobacterium tuberculosis. We discovered five putative binding sites, one of which was experimentally active. LORD uses five possible binding sites to construct compounds, yielding 25 potential therapeutic candidates. The experimentally active site in CBE values is more prominent than the other four binding sites. LORD 25 molecules exhibit superior BE (kcal/mol) and physicochemical qualities and better ADME features than the currently marketed medication pyrazinamide. In addition, unlike less stable pyrazinamide medicines, MD simulations of the top four potential drug-like compounds are investigated and shown to be protein

stable during lengthy simulation runs. The study focuses on using LORD on the ribosomal protein S1 of mycobacterium tuberculosis, which is linked to TB. The researchers found five possible binding sites, one experimentally active, and utilized LORD to generate 25 prospective drug options. The prospective therapeutic candidates' binding energy (BE) values were determined to be good, and they demonstrated better physicochemical and ADME qualities when compared to the market-available medication pyrazinamide.

Furthermore, MD simulations on the top four potential drug-like compounds revealed that they were protein stable during lengthy simulation runs. Overall, the study indicates that LORD might be a valuable method for discovering new potential drug-likeness molecules and improving their characteristics to facilitate drug development. The study also emphasizes the potential of the discovered drug-likeness molecules for treating TB.

REFERENCES

1. World Health Organization. Global tuberculosis report **2019**.
2. Yamurai Bishi, L.; Chaitanya Vedithi, S.; L. Blundell, T.; Chitima Mugumbate, G. Computational Deorphaning of Mycobacterium Tuberculosis Targets. In *Drug Discovery and Development - New Advances*; IntechOpen, 2020.
3. Yang, L.; Hu, X.; Chai, X.; Ye, Q.; Pang, J.; Li, D.; Hou, T. Opportunities for Overcoming Tuberculosis: Emerging Targets and Their Inhibitors. *Drug Discov. Today* **2022**, 27 (1), 326–336.
4. Rehman, A. U.; Khan, M. T.; Liu, H.; Wadood, A.; Malik, S. I.; Chen, H.-F. Exploring the Pyrazinamide Drug Resistance Mechanism of Clinical Mutants T370P and W403G in Ribosomal Protein S1 of Mycobacterium Tuberculosis. *J. Chem. Inf. Model.* **2019**, 59 (4), 1584–1597.
5. Junaid, M.; Khan, M. T.; Malik, S. I.; Wei, D.-Q. Insights into the Mechanisms of the Pyrazinamide Resistance of Three Pyrazinamidase Mutants N11K, P69T, and D126N. *J. Chem. Inf. Model.* **2019**, 59 (1), 498–508.
6. Thapa, S.; Nargund, S. L.; Biradar, M. S.; Banerjee, J.; Karati, D. In-Silico Investigation and Drug Likelihood Studies of Benzimidazole Congeners: The New Face of Innovation. *Inform. Med. Unlocked* **2023**, 38 (101213), 101213.
7. Akinpelu, O. I.; Lawal, M. M.; Kumalo, H. M.; Mhlono, N. N. Computational Studies of the Properties and Activities of Selected Trisubstituted Benzimidazoles as Potential Antitubercular Drugs Inhibiting MTB-FtsZ Polymerization. *J. Biomol. Struct. Dyn.* **2022**, 40 (4), 1558–1570.
8. Yamurai Bishi, L.; Chaitanya Vedithi, S.; L. Blundell, T.; Chitima Mugumbate, G. Computational Deorphaning of Mycobacterium Tuberculosis Targets. In *Drug Discovery and Development - New Advances*; IntechOpen, **2020**.
9. Yang, J.; Liu, Y.; Bi, J.; Cai, Q.; Liao, X.; Li, W.; Guo, C.; Zhang, Q.; Lin, T.; Zhao, Y.; Wang, H.; Liu, J.; Zhang, X.; Lin, D. Structural Basis for Targeting the Ribosomal Protein S1 of Mycobacterium Tuberculosis by Pyrazinamide. *Mol. Microbiol.* **2015**, 95 (5), 791–803.
10. Ejalonibu, M. A.; Ogundare, S. A.; Elrashedy, A. A.; Ejalonibu, M. A.; Lawal, M. M.; Mhlono, N. N.; Kumalo, H. M. Drug Discovery for Mycobacterium Tuberculosis Using Structure-Based Computer-Aided Drug Design Approach. *Int. J. Mol. Sci.* **2021**, 22 (24), 13259.
11. Dai, Y.; Xu, Y.; Guo, C.; Xue, X.; Lin, D.; Lin, K. Discovery and Evaluation of New Compounds Targeting Ribosomal Protein S1 in Antibiotic-Resistant Mycobacterium Tuberculosis. *Eur. J. Med. Chem.* **2020**, 196 (112317), 112317.
12. Wadood, A.; Shareef, A.; Ur Rehman, A.; Muhammad, S.; Khurshid, B.; Khan, R. S.; Shams, S.; Afridi, S. G. In Silico Drug Designing for Ala438 Deleted Ribosomal Protein S1 (RpsA) on the Basis of the Active Compound Zrl15. *ACS Omega* **2022**, 7 (1), 397–408.
13. Rehman, A. U.; Ali, S.; Rafiq, H.; Rasheed, S.; Nouroz, F.; Wadood, A. Computational Insight into the Binding Mechanism of Pyrazinoic Acid to RpsA Protein. *Curr. Chin. Sci.* **2021**, 1 (2), 207–215.
14. Thapa, S.; Nargund, S. L.; Biradar, M. S.; Banerjee, J.; Karati, D. In-Silico Investigation and Drug Likelihood Studies of Benzimidazole Congeners: The New Face of Innovation. *Inform. Med. Unlocked* **2023**, 38, 101213.

15. Koseki, Y.; Aoki, S. Computational Medicinal Chemistry for Rational Drug Design: Identification of Novel Chemical Structures with Potential Anti-Tuberculosis Activity. *Curr. Top. Med. Chem.* **2013**, *14* (1), 176–188.
16. Tam, B.; Sherf, D.; Cohen, S.; Eisdorfer, S. A.; Perez, M.; Soffer, A.; Vilenchik, D.; Akabayov, S. R.; Wagner, G.; Akabayov, B. Discovery of Small-Molecule Inhibitors Targeting the Ribosomal Peptidyl Transferase Center (PTC) of M. Tuberculosis. *Chem. Sci.* **2019**, *10* (38), 8764–8767.
17. Zhi, Y.; Dai, Y.; Yang, J.; Tan, S.; Lin, D.; Lin, K. Lead Compounds and Key Residues of Ribosomal Protein S1 in Drug-Resistant Mycobacterium Tuberculosis. *Bioorg. Chem.* **2019**, *82*, 58–67.
18. Yasir, M.; Singh, P.; Chohan, S.; Shrivastava, R. Structure Based Computational Exploration of Beilschmiedia Compounds with Selected Targets against Multidrug-Resistant Mycobacterium Tuberculosis. *Ind. J. Pharm. Educ.* **2019**, *53* (2s), s143–s150.
19. Khan, M. T.; Khan, A.; Rehman, A. U.; Wang, Y.; Akhtar, K.; Malik, S. I.; Wei, D.-Q. Structural and Free Energy Landscape of Novel Mutations in Ribosomal Protein S1 (RpsA) Associated with Pyrazinamide Resistance. *Sci. Rep.* **2019**, *9* (1), 7482.
20. Rajasekhar, S.; Karuppasamy, R.; Chanda, K. Exploration of Potential Inhibitors for Tuberculosis via Structure-Based Drug Design, Molecular Docking, and Molecular Dynamics Simulation Studies. *J. Comput. Chem.* **2021**, *42* (24), 1736–1749.
21. Lata, S.; Akif, M. Structure-Based Identification of Natural Compound Inhibitor against M. Tuberculosis Thioredoxin Reductase: Insight from Molecular Docking and Dynamics Simulation. *J. Biomol. Struct. Dyn.* **2021**, *39* (12), 4480–4489.
22. Yang, L.; Hu, X.; Chai, X.; Ye, Q.; Pang, J.; Li, D.; Hou, T. Opportunities for Overcoming Tuberculosis: Emerging Targets and Their Inhibitors. *Drug Discov. Today* **2022**, *27* (1), 326–336.

CHAPTER 7

Conclusions and Scope for Future Work

7.1 Conclusions

We have developed an automatic drug designer called “LSTM-Enabled On-Site Recurrent Molecular Designer” (LORD). The LORD designer scans the entire target protein and finds the potential binding energy sites using MESP cavity analysis. The target binding energy sites facilitate LORD to design the potential drug-like molecule at that specific protein target binding site in the following step. While developing the drug molecule, LORD encapsulates the Physicochemical properties such as Lipinski’s rule of five. Drug-likeness molecules are filtered through ADME properties and molecular dynamics (MD) studies. The LORD designer applied AD, COVID-19, PD, and TB, essential in the potential drug target to combat the various diseases. We generated potential drug-like molecules and studied their physicochemical properties, ADME properties, and MD studies on top potential drug-likeness molecules. In these studies show that the 125 drug-like molecules are promising to combat those diseases compared with available market drugs. Generally, the designer can be used for any target protein to design drug-like molecules and define binding sites. While identifying the potential binding sites, it uses the MESP calculation, which is performed at a lower computational cost using the substrate structure approach. Design strategy mainly depends on the target binding site environment; hence, LORD can accurately design the drug.

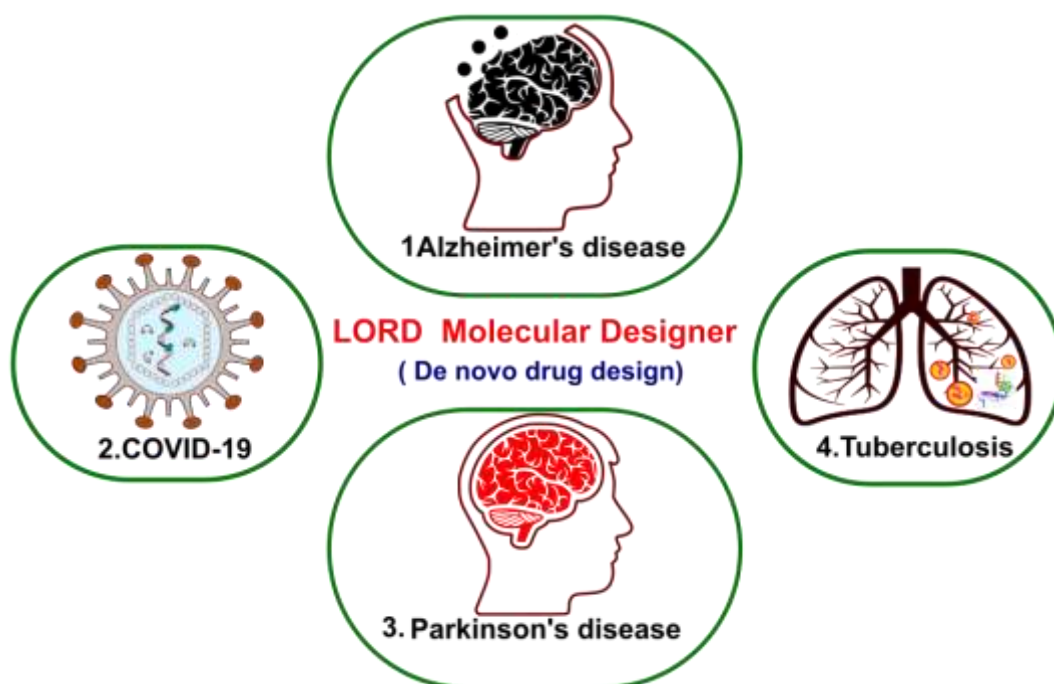


Figure 7.1 Current working chapters.

LORD is a computational strategy for designing novel compounds with particular characteristics, like high affinity and selectivity for a target protein. Computer methods and simulations forecast how a target protein will interact with various small-molecule drugs.

The subsequent step involves utilizing a computational algorithm to produce and assess numerous potential drug-likeness molecules. The biological potential of these molecules is assessed, including potency, selectivity, and bioavailability. Next, the most promising compounds are produced and tested in vitro and in vivo to determine their effectiveness and safety.

7.2 Scope for Future Work

LORD molecular design has several advantages over traditional drug discovery methods, including the possibility of developing incredibly selective medicines for a specific target protein. Also, it makes it possible to quickly test many compounds, which might help save both time and money. De novo drug design will move towards combining computational and experimental methods, leading to greater precision and effectiveness. Here are some possible directions for the future of de novo drug design.

It is predicted that de novo drug design will adopt this strategy in the future, including a range of computational and experimental techniques to enhance the accuracy and efficacy of the drug development process. The following directions for LORD drug design might develop in the future:

1. Artificial intelligence and machine learning may be used to analyze vast amounts of data from experimental assays and improve the accuracy of predictions of molecular interactions. This may speed up the search for potential new drug candidates.
2. Using big data: The expanding availability of detailed biological data, like genomic and proteomic data, may offer crucial knowledge about disease pathophysiology and potential treatment targets.
3. Improving the efficiency of the drug discovery process: LORD molecular design can benefit from advancements in high-throughput screening, which allow for the rapid testing of large compounds. In addition, advances in synthetic chemistry and automation can improve the efficiency of compound synthesis and testing.

- 4.Targeting complex disorders: Since many diseases have complex or poorly known underlying systems, it may be challenging to create effective therapies. LORD design offers a systematic strategy for exploring novel therapeutic targets and designing molecules with the most desirable features, which can assist in overcoming these difficulties.
- 5.The future of LORD de novo molecular designer will help develop potential drug-likeness compounds to treat various target proteins such as HIV and breast cancer.

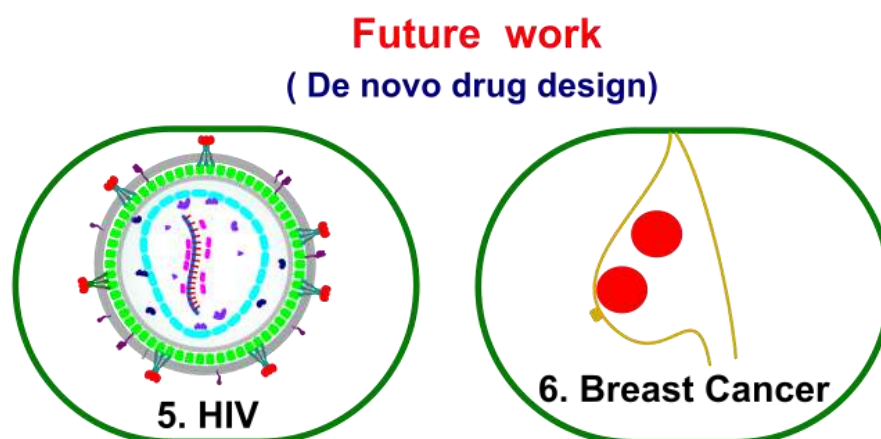


Figure 7.2 Future direction of the thesis work.

Development and Application of a Long Short-Term Memory- Based Drug-likeness Molecular Designer Algorithm

by Srinivasarao Mande

Submission date: 27-Jun-2023 04:44PM (UTC+0530)

Submission ID: 2123428220

File name: Srinivas_thesis_Plagiarism_check.pdf (987.65K)

Word count: 25592

Character count: 139581



Librarian
Indira Gandhi Memorial Library
UNIVERSITY OF HYDERABAD
Central University P.O.
HYDERABAD-500 046.

Development and Application of a Long Short-Term Memory-Based Drug-likeness Molecular Designer Algorithm

ORIGINALITY REPORT

6%

SIMILARITY INDEX

4%

INTERNET SOURCES

4%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1

www.ncbi.nlm.nih.gov

Internet Source

<1 %

2

www.mdpi.com

Internet Source

<1 %

3

Sakshi Singh, Qanita Bani Baker, Dev Bukhsh Singh. "Molecular docking and molecular dynamics simulation", Elsevier BV, 2022

Publication

<1 %

4

"Forcefields for Atomistic-Scale Simulations: Materials and Applications", Springer Science and Business Media LLC, 2022

Publication

<1 %

5

www.researchgate.net

Internet Source

<1 %

6

Submitted to University of Sydney

Student Paper

<1 %

7

pubmed.ncbi.nlm.nih.gov

Internet Source

<1 %

8	scholarworks.gsu.edu Internet Source	<1 %
9	manual.gromacs.org Internet Source	<1 %
10	www.tandfonline.com Internet Source	<1 %
11	Submitted to University of Edinburgh Student Paper	<1 %
12	Shalini Mathpal, Tushar Joshi, Priyanka Sharma, Veena Pande, Subhash Chandra. "Assessment of activity of chalcone compounds as inhibitors of 3-chymotrypsin like protease (3CLPro) of SARS-CoV-2: in silico study", Structural Chemistry, 2022 Publication	<1 %
13	mts.intechopen.com Internet Source	<1 %
14	Markus A. Lill. "Efficient Incorporation of Protein Flexibility and Dynamics into Molecular Docking Simulations", Biochemistry, 2011 Publication	<1 %
15	Thomas Simonson. "Chapter 12 Free Energy Calculations: Approximate Methods for Biological Macromolecules", Springer Science and Business Media LLC, 2007	<1 %

16	iopscience.iop.org Internet Source	<1 %
17	www.nature.com Internet Source	<1 %
18	archipel.uqam.ca Internet Source	<1 %
19	link.springer.com Internet Source	<1 %
20	research.snu.edu.in Internet Source	<1 %
21	"Target Discovery and Validation", Wiley, 2019 Publication	<1 %
22	Submitted to University of Hyderabad, Hyderabad Student Paper	<1 %
23	L. Yan. "LOOPER: a molecular mechanics- based algorithm for protein loop prediction", Protein Engineering Design and Selection, 01/24/2008 Publication	<1 %
24	www.kma.org.kw Internet Source	<1 %
25	repository.nwu.ac.za Internet Source	<1 %

26	warm.dovepress.com Internet Source	<1 %
27	Ryo Hatada, Koji Okuwaki, Yuji Mochizuki, Yuma Handa, Kaori Fukuzawa, Yuto Komeiji, Yoshio Okiyama, Shigenori Tanaka. "Fragment Molecular Orbital Based Interaction Analyses on COVID-19 Main Protease – Inhibitor N3 Complex (PDB ID: 6LU7)", Journal of Chemical Information and Modeling, 2020 Publication	<1 %
28	Jung-Hsin Lin, Alexander L. Perryman, Julie R. Schames, J. Andrew McCammon. "The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme", Biopolymers, 2003 Publication	<1 %
29	Submitted to Sparsholt College Hampshire Student Paper	<1 %
30	par.nsf.gov Internet Source	<1 %
31	academicrepository.khas.edu.tr Internet Source	<1 %
32	Submitted to Medizinischen Universität Wien Student Paper	<1 %
33	pure.coventry.ac.uk Internet Source	<1 %

34	d-nb.info Internet Source	<1 %
35	Ritu ., Prakash Chandra, Asmita Das. "Cholinergic Side Effect-Free Anticancer Drugs: Paving the Way for Safer and More Effective Cancer Treatment", Research Square Platform LLC, 2023 Publication	<1 %
36	Submitted to University of Keele Student Paper	<1 %
37	docserv.uni-duesseldorf.de Internet Source	<1 %
38	freidok.uni-freiburg.de Internet Source	<1 %
39	repository.up.ac.za Internet Source	<1 %
40	virologyj.biomedcentral.com Internet Source	<1 %
41	www.hindawi.com Internet Source	<1 %
42	www.iflscience.com Internet Source	<1 %
43	Gabriel Corrêa Veríssimo, Simone Queiroz Panteleão, Jadson Castro Gertrudes, Thales Kronenberger et al. "MASSA Algorithm:	<1 %

automated rational sampling of training and test subsets for QSAR modelling", American Chemical Society (ACS), 2023

Publication

44

Seketoulie Keretsu, Swapnil P. Bhujbal, Seung Joo Cho. "Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation", Scientific Reports, 2020

Publication

<1 %

45

Submitted to Swinburne University of Technology

Student Paper

<1 %

46

Submitted to University College London

Student Paper

<1 %

47

advances.sciencemag.org

Internet Source

<1 %

48

de.slideshare.net

Internet Source

<1 %

49

Yasuaki Anami, Nobutaka Shimizu, Toru Ekimoto, Daichi Egawa, Toshimasa Itoh, Mitsunori Ikeguchi, Keiko Yamamoto. "Apo- and Antagonist-Binding Structures of Vitamin D Receptor Ligand-Binding Domain Revealed by Hybrid Approach Combining Small-Angle X-

<1 %

ray Scattering and Molecular Dynamics", Journal of Medicinal Chemistry, 2016

Publication

50

era.library.ualberta.ca

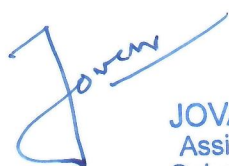
Internet Source

<1%

Exclude quotes ☒ On

Exclude matches ☒ < 14 words

Exclude bibliography ☒ On



JOVAN JOSE K. V.
Assistant Professor
School of Chemistry
University of Hyderabad
Hyderabad - 500 046.