

**Comparative Dynamics and Drug Targeting against
Multidrug Resistant *Bordetella Pertussis***



By

RAHEELA MAJEED

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-I-Azam University,

Islamabad, Pakistan

2023

**Comparative Dynamics and Drug Targeting against
Multidrug Resistant *Bordetella Pertussis***



By

Raheela Majeed

A thesis submitted in the partial fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY

In

BIOINFORMATICS

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-I-Azam University,

Islamabad, Pakistan

2023

CERTIFICATE

This study, entitled “Comparative Dynamics and Drug Targeting against Multidrug Resistant “*Bordetella Pertussis*” submitted by **Miss. Raheela Majeed** to the National Center for Bioinformatics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan, is accepted in its present form as satisfying the thesis requirements for the Degree of Master of Philosophy in Bioinformatics.

Internal Examiner: _____

Dr. Syed Sikander Azam

Professor & Supervisor

Quaid-i-Azam University Islamabad.

External Examiner: _____

Chairman: _____

Dr. Syed Sikander Azam

Professor

Quaid-i-Azam University Islamabad.

Date:

DECLARATION

The work reported in this study, entitled “Comparative Dynamics and Drug Targeting against Multidrug Resistant *Bordetella Pertussis*” was carried out by **Raheela Majeed** under the supervision of **Dr.Syed Sikander Azam**. I, hereby, declare that the title of thesis and all the contents presented in the following study are product of my own effort and no part has been copied from any published source (except the references, standard mathematical or genetic models /equations /formulas /protocols, etc.). None of the work has been submitted for award of any other degree /diploma. The University may take action if the information provided is found inaccurate at any stage.

RaheelaMajeed

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in this dissertation entitled, Comparative Dynamics and Drug Targeting against Multidrug Resistant *Bordetella Pertussis* is solely my research work with no significant contributions from any other person. Small contribution/help wherever taken has been duly acknowledged and that this M.Phil. This dissertation has been written by me. I understand the zero-tolerance policy of the HEC and Quaid-i-Azam University towards plagiarism. Therefore, I as an author of the above-titled dissertation declare that no portion of my thesis has been plagiarized and any material used as a reference is properly referred/cited. I undertake that if I found guilty of any formal plagiarism in the above-titled thesis ever after award of M.Phil degree, the University reserves the right to withdraw/revoke my M.Phil. degree and that HEC and the University has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarizedthesis.

Raheela Majeed

Signature: _____

DEDICATION

I dedicate this dissertation to my parents. I hope that this achievement will complete the dream that you had for me all those many years ago when you choose to give me the best education you could.

ACKNOWLEDGMENTS

With the kind support and assistance of many people, this thesis becomes a reality. I would like to express my heartfelt gratitude to each and every one of them.

At the start of my dissertation, I would like to thank **Allah Almighty** for the intelligence, strength, mental serenity, and good health that he has bestowed upon me in order to complete this research.

I would like to thank my supervisor, **Dr. Syed Sikander Azam**, Chairman of National Centre for Bioinformatics at Quaid-i-Azam University in Islamabad, for allowing me to conduct research and providing me with valuable guidance throughout the process.

I am also grateful to **Dr. Amir Ali Abbasi** and **Dr. Sajid Rashid** for his continued support and encouragement during my studies.

I would like to thank **Afifa Navid, Ghulam Abbas, Faisal Ahmed, Iqra Aziz, Fouzia Gul, Kainat Gul and Rimsha Yousaf** for sharing their knowledge and experience with me during this research. I will never forget her unwavering support and inspirational encouragement.

And all my lab fellows who have provided invaluable guidance and unwavering support. I must express my gratitude to my classmates and friends, for their motivation, excitement, and assistance during this research.

Last but not least, no words can adequately describe my thanks for my loving family's encouragement and motivation. I am only here due of their love and affection, without them, I would not have been able to achieve my goals.

I am grateful to Pakistan-United States Science and Technology Cooperation Program, International Foundation for Science (IFS) under grant number 55461 and Higher Education commission (HEC), Pakistan for granting the financial assistance to our Lab.

CONTENTS

LIST OF FIGURES	xi
LIST OF TABLES	xii
ABBREVIATIONS	xiii
ABSTRACT	xiv
CHAPTER: 1 INTRODUCTION	1
1. Introduction	1
1.1 Molecular Docking.....	5
1.2 Molecular Dynamic Simulations	5
1.2.1 Classical Mechanics.....	6
1.2.2 Molecular Mechanics.....	7
1.2.3 Statistical Mechanics	7
1.3 Aims and Objectives	8
CHAPTER: 2 METHODOLOGY.....	9
2.1 SUBTRACTIVE PROTEOMICS	10
2.1.1 Proteome Retrieval and Drug Candidate Prioritization.....	10
2.1.2 Virulent Factors and Sub-Cellular Localization Analysis.....	11
2.1.4 Drug Target Selection.....	12
2.2 Cellular Interactome Analysis	12
2.3 Structure Modeling and Validation	12
2.4 Minimization of Protein Structure.....	13
2.5 Binding Site Prediction	14
2.5.1 Active Site Prediction.....	14
2.6 Antibacterial Library Preparation	14
2.7 Molecular Docking	14
2.8 Pharmacokinetic Profile Evaluation	15
2.8.1 ADMET and Toxicity Analysis	15
2.9 2D Interaction Analysis	16
2.10 Molecular Dynamics Simulation	16
2.10.1 Root Mean Square Deviation.....	17

2.10.2 Root Mean Square Fluctuation	17
2.10.3 Beta-Factor.....	17
2.10.4 Radius of Gyration	17
2.10.5 Hydrogen Bond Analysis.....	18
2.10.6 Radial Distribution Function.....	18
2.10.7 Binding Free Energy Calculation.....	18
CHAPTER: 3 RESULTS.....	19
3.1 SUBTRACTIVE PROTEOMICS	19
3.1.1 Retrieval of Proteome and Drug Candidate Prioritization.....	19
3.1.2 Sub-cellular Localization Analysis	21
3.1.3 Physicochemical Properties of Target Proteins	21
3.1.4 Drug Target Selection.....	23
3.2 CPS Drug Target Proteins	23
3.2.1. Role of Capsular Polysaccharide Proteins	24
3.2.2. Mode of action of CPS	24
3.2.3 Unique Metabolic Pathway.....	25
3.3 Cellular Interactome Analysis	25
3.4 Structure Modeling and Validation	26
3.5 Minimization of Modeled Structure	28
3.6 Functional Domains of the CPS Protein	29
3.7 Prediction of Binding Site.....	29
3.7.1 Active Site Prediction.....	29
3.8 Anti-bacterial Library Preparation and Molecular Docking	29
3.9 Pharmacokinetic Profile Evaluation	32
3.9.1 ADMET analysis.....	32
3.9.2 Toxicity Analysis	33
3.10 2D Interaction Analysis	34
3.11 Molecular Dynamics Simulations.....	36
3.11.1 Root Mean Square Deviation.....	36
3.11.2 Root Mean Square Fluctuation	36
3.11.3 Beta-factor.....	37

3.11.4 Radius of Gyration	37
3.12 Hydrogen Bond Analysis	38
3.13 Radial Distribution Function	40
3.14 Binding Free Energy Calculation	43
CHAPTER: 4 DISCUSSION.....	45
4. Discussion.....	45
CHAPTER: 5 CONCLUSION.....	48
CONCLUSION	49
REFERENCES.....	50

LIST OF FIGURES

Figure 1.1: Variants of proteins found in vaccines of pertussis.....	3
Figure 2.1: Flowchart of the methodology used in current study	9
Figure 3. 1: Subtractive proteomics analysis of the proteome of	20
Figure 3. 2: Protein-protein interactions of target protein	26
Figure 3. 3: A) Predicted 3D structure of CPS B) Ramachandran plot of CPS showing residues in favoured, allowed and disallowed region.....	27
Figure 3. 4: Errat Plot with overall quality factor	28
Figure 3. 5: A) shows the Z-score B) shows the energy calculated for the predicted model	28
Figure 3. 6: Active site residues of WbpP-CPS protein in light blue.....	29
Figure 3. 7: 2D interactions of Top complexes using Discovery Studio.....	35
Figure 3. 8: Superimposed A) Root Mean Square Deviation B) Root Mean Square Fluctuation C) Beta-Factor D) Radius of Gyration of top two complexes	38
Figure 3. 9: Hydrogen Bond Analysis of A) CPS/CHEMBRIDGE-10000290 and B) CPS/CHEMBRIDGE-10008002	40
Figure 3. 10: A) Radial Distribution Function of CPS/CHEMBRIDGE-10000290 and B) CPS/CHEMBRIDGE-10008002	43

LIST OF TABLES

Table 3. 1: Physicochemical properties of selected proteins	22
Table 3. 2: Physicochemical characterization of best selected protein	23
Table 3. 3: Structural evaluation of predicted 3D structures	27
Table 3. 4: Compounds with the highest GOLD scores	30
Table 3. 5: Predicted Lipinski descriptors of all selected top compounds	33
Table 3. 6: Toxicity properties of top five docked compounds	34
Table 3. 7: Binding Free Energy Calculations of selected complexes via MM/GBSA	44
Table 3. 8: Binding Free Energy Calculations of selected complexes via MM/PBSA	44

ABBREVIATIONS

WHO	(World Health Organization)
ADMET	(Absorption, Distribution, Metabolism, Excretion and Toxicity)
AMBER	(The Assisted Model Building with Energy Refinement)
MDR	(Multidrug Resistance)
AMR	(Antimicrobial Resistance)
BLAST	(Protein Basic Local Alignment Search Tool)
CD-HIT	(Cluster Database at High Identity with Tolerance)
CYP2D6	(Cytochrome P450 2D6 enzyme inhibition)
DEG	(Database of Essential Genes)
DS	(Discovery Studio)
GAFF	(General AMBER Force Field)
GRAVY	(Grand Average Hydrophobicity)
HBD	(Hydrogen Bond Donors)
HIA	(Human Intestinal Absorption)
KAAS	(KEGG Automatic Annotation Server)
MDR	(Multi-Drug Resistant)
MMGBSA	(Molecular Mechanics Generalized Born Surface Area Continuum Solvation)
MMPBSA	(Molecular Mechanics Poisson Boltzmann Surface Area Continuum Solvation)
RDF	(Radial Distribution Function)

ABSTRACT

The *Bordetella Pertussis* is a gram-negative bacterium and it belongs to the family of Alcaligenaceae. It is a causative agent of several types of infections mainly pertussis which infects human respiratory tracts. Pertussis is linked to the high degree of resistance to various class of antibiotics. The focus of the current study is to identify the potential drug target and to categorize the best drug candidate by employing subtractive proteomics approach and structure based virtual screening. Subtractive proteomics pipeline revealed that Capsular Polysaccharide Biosynthesis protein (CPS) that is a promising drug target. An AI based approach; Alpha fold was used for acquiring 3D structure of CPS. The Molecular Docking was performed against Chembridge library via GOLD software. Pharmacokinetic profiling was performed and several *in-silico* analysis such as ADMET and Toxicity analysis, Molecular Dynamics simulations, Radial Distribution Function and Binding Free Energy calculations were performed. The top inhibitors 10000290 and 10008002 with GOLD dock score 81 and 73 respectively, were used as promising drug candidates against CPS.

CHAPTER: 1 INTRODUCTION

1. Introduction

Bordetella pertussis occurs as a stringent human pathogen. Bordet and Gengou first reported its isolation in 1906 (Mattoo & Cherry, 2005). *B. pertussis* is very monomorphic. Guillaume de Baillou provided the first account of an epidemic, which happened in Paris in 1578. The ISE IS481, which makes up 6.2% of the genome of the strain Tohama I of *B. pertussis* and is found in 238 copies (Mooi, 2010). One of the eight species in the *Bordetella* genus, *B. pertussis* is the main cause of pertussis. *Bordetella* is an essential genus of bacteria required for multidrug-resistance with a genomic size of 4MB (Lin et al., 2021). *B. pertussis* is oxidase and catalase-positive, with a low metabolic activity. It is only isolated from humans only (Lutwick et al., 2014). The cause of pertussis, sometimes known as whooping cough, has gram stain negative, multidrug resistant, requires oxygen, virulent, disease causing, fastidious and coccobacillus shape containing capsule around it. Numerous pathogenic species commonly linked to upper respiratory tract diseases in warm-blooded animals can be found in this genus resulting in about three hundred thousand fatalities across the world (Locht, 1999). Convulsions, encephalopathy, encephalitis, lifelong brain damage, and death are all complications of infection. Following a natural infection, acquired immunity to when *B. pertussis* appears, it provides a comparably robust defence against reinfection (Mills & Th, 2001). Currently, one of the top 10 infectious disease killers in the world is whooping cough. According to reports, severe pertussis death rates in affluent nations can range from 19.7 to 31%. A study of babies found that the mortality rate was 70% overall and 84% for infants under 6 weeks (Shi et al., 2021). Although it can affect anyone at any age, pertussis has a significant death rate in newborns under the age of one. 95% of the 16000000 documented victims were of this disease in worldwide developing countries, and the illness caused about 195,000 child deaths (Jamal et al., 2022). Despite a high vaccination rate, this disease has made a comeback in a number of nations and only humans are susceptible to the illness caused by *B.pertussis*, which is spread by respiratory droplets (Kerr & Matthews, 2000). Aerosols are used to spread *B. pertussis*, which infects the ciliated airway epithelium. *B. pertussis* is a mobile organism with a

characteristic that resembles a flagellum. The toxins filamentous haemagglutinin, adenylate cyclase toxin, pertactin, fimbria, and tracheal cytotoxin are some of its virulence components (Be et al., 2019). Adhesins and toxins fall into two broad categories that can be used to classify these virulence factors. The majority of virulence factors are coordinately regulated by a signal transduction system of two-components consisting of the regulator protein BvgA and the sensor protein BvgS. Together, the adhesins and toxins work to create an infection (Locht, 1999).

The bacteria produce the following toxins:

1. ADP ribosyltransferase pertussis toxin (Ptx), a complex hexameric protein.
2. Adenylate cyclase toxin (ACT), which can promote apoptosis and cytotoxicity due to post-translational modifications.
3. BteA, a type III secretion system effector protein, which causes rapid, non-apoptotic cell death.
4. The disaccharide tetra peptide tracheal cytotoxic (TCT), which induces ciliostasis and harms respiratory epithelial cells, is generated from the organism's cell wall (Lutwick et al., 2014).

Early in the 1960s, whole-cell pertussis vaccination was made available in China (L. Zhang et al., 2010). In 1997, acellular vaccinations for *B.pertussis* were made available in Australia. These vaccinations took the place of whole-cell vaccines by the year 2000. There was a significant pertussis outbreak between 2008 and 2012. During this time, 30 percent of its segregates lacked pertactin (Prn) (Lam et al., 2014).

It has been demonstrated that high PT antibody titres can predict recent *B.pertussis* infection. Between 1994 and 1998, the prevalence rate of standardised anti-pertoxin antibodies was assessed in few Western European nations then linked to monitoring and vaccination programmes having relevant data. Recent infection was significantly more common in adults and adolescents (10–19 years old) in countries with high coverage (Finland, The Netherlands, France, East Germany), while it was even more common in children (3–9 years old) than teenagers in nations with low coverage (Italy, West Germany, and the United Kingdom; 90%) (Si & Ki, 2005). Even though it is predicted

that after three doses, 86% of people worldwide have received the DTP vaccine, the World Health Organization (WHO) reported 139,786 cases of pertussis in 2014. Pertussis was the most newly emerging disease in the United States with too many yearly reported cases (Nieves & Heininger, 2016). When the DPT vaccination was released in US during late 1940s, number of cases gradually decreased over the decades later on (Lutwick et al., 2014).

Vaccine	Fim2	Fim3	Prn	PtxA
Pw	Fim2-1, Fim2-2 ^b	Fim3-1	Prn1, Prn7, Prn10 ^c	PtxA1, PtxA2, PtxA4
Pa	Fim2-1	Fim3-1	Prn1, Prn7	PtxA1, PtxA2

Figure 1.1: Variants of proteins found in vaccines of pertussis

Three stages make up the clinical course of the 6–12-week course or procedure:

1. The insidious emergence of minor symptoms of upper respiratory tracts, such as fever, sneezing, and a moderate, infrequent coughing, that are comparable to rhinovirus infections characterize the catarrhal phase. The cough steadily worsens during the course of the first few of weeks.
2. The paroxysmal phase is characterized by spasmodic coughing bouts or paroxysms of up to 10 or more uninspiring coughs. These spasms can occasionally be followed by a protracted whooping sound or posttussive vomiting. At night, paroxysmal attacks are more common, and eating may trigger one. Cyanosis can develop during paroxysms, and a spasmodic cough can lead to a variety of other health issues, such as cerebral hypoxia, severe alkalosis, convulsions, subcutaneous emphysema, subconjunctival hemorrhage, umbilical or inguinal hernias, rib fractures, and umbilical or inguinal hernias. Following an incident, young children and newborns may appear particularly agitated and worn out. Surprisingly, the child can seem fine in between assaults. This phase can extend up to 10 weeks, however it often lasts 2 to 6 weeks. In the absence of subsequent bacterial infection, fever and pharyngitis are uncommon in pertussis patients.
3. The third phase, known as the convalescent phase, is when healing happens gradually. Paroxysms first become less frequent and then less severe, and the cough may go away

in two to three weeks. Other respiratory illnesses may cause paroxysmal episodes to recur (Lutwick et al., 2014).

Unfortunately, there is no proven method of easing pertussis symptoms. The authors of the most recent Cochrane Center comprehensive review of pertussis treatment trials discovered no appreciable positive benefit of treatment with salbutamol (a bronchodilator), dexamethasone (an anti-inflammatory steroid), or diphenhydramine (an antihistamine). Patients with pertussis receive macrolide antibiotics, although often only to stop the spread of the disease as taking antibiotics seldom alters the infected person's clinical course (Carbonetti, 2016). Convulsions, encephalopathy, encephalitis, lifelong brain damage, and death are all complications of infection.

For more than three decades, the generation of therapeutically significant small molecules had greatly benefited from computer-aided drug designing techniques. Drug designing aimed in the identification of potential drug candidates for the pathogenic bacteria (Leelananda & Lindert, 2016). Computational resources and methods had been widely employed to explore the biological mechanisms for discovering the new antibiotics and to combat the extremely drug resistant pathogens (Mahtab et al., 2021). The proteome analysis technique was employed to filter proteins which shows a high specificity to pathogens (Garg & Gupta, 2008). Advance tactics must be used to create the therapeutic drugs against the prospective therapeutic targets, which could have numerous consequences for immunologists in order to achieve a protection against *B. pertussis*. The goal of this work is to find a potential therapeutic candidate while taking into account the limitations of computational drug creation (Rosini et al., 2020). For the identification of proteins as therapeutic targets, the proteome of *B. pertussis* was used in screening process. It was then filtered based on non-human homologous nature, essentiality, virulence, and sub cellular localization. After drug target selection, the potential drug candidate was identified using *in silico* techniques (Jamal et al., 2022). The comparative analysis was performed on protein via the anti-bacterial library Cambridge. The pharmacokinetic profile was checked along with admet properties to check the reliability for effective drug against *B. pertussis*, and if necessary changes are incorporated in the drug for maintainance (Othman et al., 2022). Therefore, the results

suggested at the end of this study might be helpful for designing an effective therapeutic drug against *B.pertussis*.

1.1 Molecular Docking

Molecular docking techniques examine the behaviour of small molecules at a target protein's binding site. As more experimentally determined protein structures are discovered using X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, molecular docking is being used increasingly frequently as a tool in drug development (Pagadala et al., 2017). For quantification of the intensity during binding and accurately forecast the ligand inside binding site of receptor, docking is used (Yuriev & Ramsland, 2013). The number of tools available for structure-based drug design is expanding quickly, driven by improvements in the determination of molecular structure. An appealing substitute for high-throughput random screening is lead discovery using molecular docking techniques to scan ligand databases (Knegtel et al., 1997). The preferred orientation of the bound molecules can be used to predict the strength and stability of complexes, as well as their energy profile (such as their binding free energy). Utilizing the molecular docking scoring function, this is possible. These days, molecular docking is frequently used to predict how tiny compounds (drug candidates) would bind to their targets including biomolecules such as proteins, carbohydrates, and nucleic acids, in order to establish their preliminary binding characteristics (Agarwal, 2016).

For molecular docking, the chembridge-5900 library of compounds as an anti-bacterial target was used along with the drug target protein via the gold score algorithm of GOLD and 2D interactions using DS (Discovery Studio). After docking, the top compounds were selected and subjected to MD Simulations for exploring behavior in functions and structure of the protein-ligand inclusion complexes.

1.2 Molecular Dynamic Simulations

Biomolecular structures are alive through MD simulation, which also provides insights into the natural dynamics of biomolecules in solution on various timescales. Second, MD simulation provides molecular property thermal averages (Hansson et al., 2002). Molecular dynamic simulations help to comprehend the dynamics of protein and drugs in

bacteria and human. MD simulations can be used to observe the protein folding pattern, conformational changes, and protein stability (Hospital et al., 2015).

The three approaches upon which MD simulation is based are as follows:

1. Classical Mechanics
2. Molecular Mechanics
3. Statistical Mechanics

1.2.1 Classical Mechanics

In MD simulations, laws of classical mechanics are employed to handle tiny systems. (Phys, 2021). Classical mechanics uses Newton's 2nd law of motion. It states when a force field is given to the body of mass 'm', acceleration 'a' produces. Classical mechanics effectively explain the behavior for 'N' number of particles. A specific particle 'i' facing a cumulative force 'F' due to the interaction of all other particles with it. The force depends upon the position of interacting particles moving with velocities 'v' which produces acceleration 'a' through the system over a certain timescale.

The following equation expresses Newton's 2nd Law of Motion for 'i' th particles:

$$F_i = m_i a_i \dots\dots\dots (1.1)$$

Here 'F_i' represents the force exerted on particle 'i', while mass and acceleration are represented by 'm_i' and 'a_i' respectively.

Second derivative for acceleration represented by distance 'r' and time 't' can be substituted by these quantities and the above equation is represented as follows:

$$F_i = m_i \frac{d^2 r_i}{dt^2} \dots\dots\dots (1.2)$$

A force F_i when applied on particle of mass m_i in terms of potential energy 'v' it can be described to give the following equation:

$$F_i = -dv / dr_i \dots\dots\dots (1.3)$$

As both equations represent the expression of force, equation 1.3 and 1.2 can be equated to give the following equation:

$$-dv/dri = mid^{2ri}/dt^2 \dots\dots\dots (1.4)$$

So, equation for the next set of coordinates for particles in the system from the previous set of coordinates is given as follows:

$$A = dv/dt \dots\dots\dots (1.5)$$

1.2.2 Molecular Mechanics

Energy calculation of the system and dynamic behavior of protein both aspects are dealt by molecular mechanics. The biological system's potential energy can be computed based on non-bonded interaction, dihedral and torsion angles, and bond stretching. The sum of individual energy can be used to describe the force field of molecular mechanics as follows:

$$E_{total} = E_{covalent} + E_{non-covalent} \dots\dots\dots (1.6)$$

$$E_{covalent} = E_{bond} + E_{angle} + E_{dihedral} \dots\dots\dots (1.7)$$

$$E_{non-covalent} = E_{electrostatic} + E_{vanderwaals} \dots\dots\dots (1.8)$$

1.2.3 Statistical Mechanics

The thermodynamic behavior of the biological system like volume, temperature, energy, heating is described by statistical mechanics. In statistical mechanics, there are four ensembles that are used to highlight the thermodynamic properties, which are listed below:

1. Canonical Ensemble (NVT) – fixed volume, fixed atoms, and constant temperature.

2. Micro Canonical Ensemble (NVE) – fixed volume, fixed atoms, fixed energy amount E.
3. Isothermal isobaric Ensemble (NPT) – fixed pressure and atoms, and constant value of temperature.
4. Grand Canonical Ensemble (uVT) – fixed volume, constant chemical potential with constant temperature.

1.3 Aims and Objectives

This study includes *in silico* techniques applied on the extensively drug resistant pathogen, *B. Pertussis* with four key objectives:

1. Identifying the potential drug target protein by employing the subtractive proteomics approach.
2. Structure prediction of protein via computational approaches.
3. Molecular Docking and Simulation of protein
4. Designing a therapeutic drug against the target protein Capsular Polysaccharide Biosynthesis (CPS) or UDP-N-acetylglucosamine C4 epimerase (UDP-GlcNAc).
5. Comparative dynamics and drug targeting against *B.pertussis*.

CHAPTER: 2 METHODOLOGY

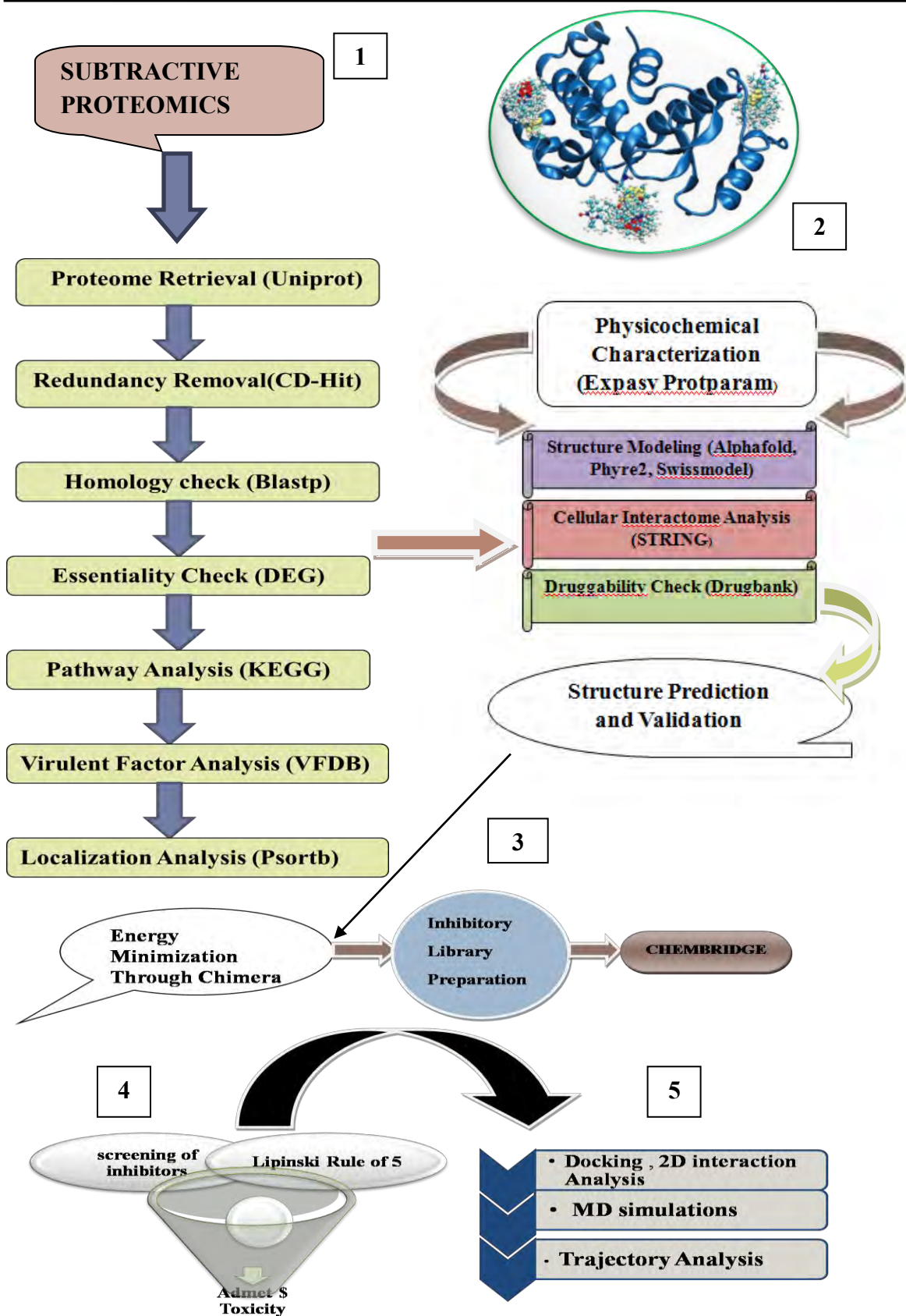


Figure 2.1: Flowchart of the methodology used in current study

2.1 SUBTRACTIVE PROTEOMICS

2.1.1 Proteome Retrieval and Drug Candidate Prioritization

The study consists of retrieval of pan proteome of *B. Pertussis* reference strain Tohama I from uniprot database (Consortium, 2015). The proteome was then analyzed by various standard criterias for an effective therapeutic drug against multidrug resistant *B. Pertussis*. The proteome was then clustered through CD-Hit web server which removed redundant protein. This programme is frequently used in grouping and comparing biological sequences in order to remove sequences showing identity greater than threshold value and enhance the efficiency of other sequence studies as it is fast and reliable (Fu et al., 2012). The CD-HIT was applied using Perl script. The percentage identity threshold was set at 0.8% so the sequences with identity greater than 80% will be clustered reducing the redundancy (Å & Godzik, 2006). The chosen protein sequences were BLAST with respect to the reference human proteome, with 9606 as taxonomic ID, according to BLASTp of NCBI, which has eliminated the proteins with a sequence identity of greater than 30 (Telkar et al., 2013). Genes essential to maintain life of cells are known as essential genes. These genes make up the basic gene set needed by a live cell. All the proteins showing no homology or the sequence identity < 30% were BLASTp against DEG (R. Zhang et al., 2004). The proteins considered essential for the survival of pathogens were shortlisted by Geptop server crucial for cell survival. Finding essential proteins, it is believed they play a role in cell viability. The value used as cutoff for essentiality, considered as 0.24. Geptop offers a platform for determining which proteins are necessary for bacteria through comparison of the proteins' orthology and phylogeny to DEG that have been created experimentally (Shahid et al., 2018). That information is used by Geptop (DEG). Geptop 2.0, predicts gene essentiality having a mean value of 0.84 and is most recent and stable version available (Wen et al., 2019). The threshold for essentiality was set at a sequence identity > 30% for the proteins and highest bit score in druggability greater than 100 was considered to be pathogen essential proteins. The next essential step was to check if any of the essential protein shares the same human metabolic pathway. The identified pathogen proteins were all mapped to the metabolic pathway in the next phase (Prioleau & MacAlpine, 2016). KAAS was employed to map pathways of protein metabolism (Moriya et al., 2007). Successively, the

protein was shortlisted by applying the virulence analysis. All the virulent proteins were screened as they help the pathogen to adhere, colonize, and disrupt the host immune system. For this purpose, the VFDB was used (Chen et al., 2012).

2.1.2 Virulent Factors and Sub-Cellular Localization Analysis

After that, the filtered proteins were examined for virulence by being compared to the database of virulence factors (VFDB). This check is crucial because pathogenic proteins are thought to be potential vaccine targets due to their capacity to spread illness. Additionally, they support the bacterium's ability to survive in a hostile environment. The presence of these proteins on the surface can initiate linkage, preserve the pathogenic cell as carbohydrates, and support the pathogenicity of the bacterium as hydrolytic enzymes and toxins. Even in small amounts, these proteins may enable the bacterium to reproduce and continue growing in the cells. Proteins with identity levels above 50% and maximum bit scores above 100 were chosen (Chen et al., 2005). Proteins were then evaluated according to their sub cellular localizations. The cytoplasmic proteins were prioritized as they are the attractive target of drug agents. Localization of proteins was then checked using PSORTb which is the localization prediction tool. It shortlists the proteins confined in the cytoplasm. The results were then crosschecked and verified with CELLO and CELLO2GO (C. Yu et al., 2014). Gneg-mPLOC is also used predict the localization of proteins in bacteria (Shen & Chou, 2010).

2.1.3 Physicochemical Properties of Filtered Proteins

The next step was evaluation of the physicochemical properties of the proteins validated on the basis of atomic composition, weight, instability, and GRAVY (Sahay et al., 2020). The protein sequences were then used in ExPASy ProtParam which computes several physical as well as chemical properties of a protein (Suhaibun et al., 2020). The instability index, which was set at 40 and used as the key measurement in this representation, was the key factor. Sequences having instability > 40 were not stable. Effective and efficient pharmacological targets are thought to have molecular weights below 110 kDa, preferably between 50 and 60 kDa (Sahay et al., 2020).

2.1.4 Drug Target Selection

Druggability of the shortlisted proteins can be checked through drugbank (Wishart et al., 2018). It is a special bioinformatics database used to combine drug target (protein) information to detailed data of drugs. Among greater than 4100 drug listings in the database are >800 small molecule and biotech pharmaceuticals that have received FDA approval, as well as >3200 investigational drugs. Additionally, these medication entries are connected to about 14,000 protein or drug target sequences (Wishart et al., 2006). By running BLASTp against the drugbank database, all hypothetical, non-homologous, and non-essential protein sequences are screened (Wadood et al., 2018). A number of proteins targets with IDs of medicines approved by the FDA are included in the drug bank database. The drugs having a maximum bit score greater than 100 and the highest sequence identity and coverage in the druggability study can be referred to as drug targets. The literature was cross-checked and used to confirm a protein's druggability (Rahman et al., 2020).

2.2 Cellular Interactome Analysis

Protein protein interaction (PPIs) was employed to check the importance, validity, essentiality and necessity of the chosen target protein through above mentioned parameters. The STRING database was used for determining the interactions of target protein (Mering et al., 2003). A complex web of functional relationships between biomolecules is essential for cellular life. Due to their variety, selectivity, and adaptability, protein-protein interactions among these associations are highly significant (Szklarczyk et al., 2021). It offers access to information about expected and experimental interactions with a remarkably thorough coverage. STRING interactions are having a confidence score as well as auxiliary data like protein domains and 3D structures (Szklarczyk et al., 2011).

2.3 Structure Modeling and Validation

A protein's 3D structure is required for Molecular Dynamic Simulations (Bliznyuk & Gready, 2006). As PDB does not provide the 3D structure of Capsular Polysaccharide

CPS or UDP-N-acetyl glucosamine C4 Epimerase, structure prediction method were used for determining the structure of proteins (Angamuthu & Piramanayagam, 2017). Two main strategies prior to the development of the Alpha Fold algorithm were homology modeling and ab-initio. The most successful and popular approach is homology modeling (or template-based approach). AlphaFold structures are the most accurate. Alpha Fold structures has at 95% residue coverage and 95% confidence interval (Ronneberger et al., 2021). It is an Artificial Intelligence based approach, Deep mind used to determine the 3D structure of protein. For this purpose, AlphaFold Deepmind Database is used in this study. Residues with $90 < \text{pLDDT}$ are categorised confident, while those with $\text{pLDDT} > 90$ are model with very high confidence. Low confidence are those having $\text{pLDDT} < 70$. pLDDT scores that have very low confidence have recently been shown to be correlated with increased propensities for intrinsic psychopathology. The Alpha Fold technique also produces the Predicted Aligned Error (PAE). If the predicted along with actual structures are aligned on residue y, it shows the anticipated positioning error at residue x (Varadi et al., 2022). The 3-Dimensional structure was also predicted by different web based automated modeling tools for comparative analysis of all the models. Parallel to Alpha Fold, SWISS-MODEL and Phyre2 tools were also used for the prediction. Thermodynamic stability of all the predicted models was evaluated using online tools such as Verify3D, ERRAT, PROCHECK and ProsA-web for the least Z-score (Hema et al., 2021). The structure in the Ramachandran plot that has the most residues in core area, highest quality factor, most residues in favourable area, and the fewest amino acid residues in the unfavourable one is thought to be the most appropriate and correct (Read et al., 2011).

2.4 Minimization of Protein Structure

The energy minimization was performed using UCSF Chimera of the selected model to improve the quality of structure by eradicating steric clashes. Minimization was a two-way process, in first get rid of any unfavourable clashes and in next we remove the remaining ones (Sarkar & Id, 2020). After that, the 3D model was rearranged and structurally optimized using 750 steepest descent minimization steps (step size of 0.02 Å), 750 conjugate gradient steps (step size of 0.02 Å), with the AMBER ff14SB force

field (Pourseif et al., 2019). Tripos Force Field was employed to perform the minimization with a step size at 0.02 Å (Keser & Stupp, 2000).

2.5 Binding Site Prediction

2.5.1 Active Site Prediction

The active sites of enzymes are complex environments created by a variety of functional groups that enable the catalysis of activities with unmatched selectivity and efficiency. A group of residues present in the active site that has evolved having different purposes, to stabilise the protein or to assist production, may have led to the catalytic function (Sun et al., 2001). If 3D structure of the protein is present in the PDB, then active site residues can easily be identified. While in the absence, there are various computational tools. However, the literature also explores active site residues of a relevant protein during drug designing (Abduljalil, 2022). The active site of the CPS was also predicted using computational tools such as CASTp, DoGSiteScorer, Prankweb, Motif Finder and Prosite (Hulo et al., 2006; Jendele et al., 2019; W. Tian et al., 2018; Volkamer et al., 2012).

2.6 Antibacterial Library Preparation

For this study, two datasets were used which are: FDA approved drug library and chembridge-5600 antibacterial library. The FDA approved drugs were taken from the drugbank with the possibility that they could be considered as potential drug candidates against CPS. The FDA library consists of 2000 inhibitors (Kocisko et al., 2003). Chembridge library was comprised of 5600 compounds. A specific force field was employed for the minimization of these compounds (Brokl et al., 2013).

2.7 Molecular Docking

Chembridge library was docked against minimized GOLD (Zou et al., 2019). For docking in GOLD software, grid size was 10 Å. The docking was performed using active site residues A209, N195, S306, S143, S142, K170, Y166, S103, G102, Q201 as cavity residues, for binding of the inhibitors. Total ten iterations were used for each inhibitor. Finally, the top docked complexes showing highest GOLD scores were selected for

further evaluation (Onodera et al., 2007). The 20 protein complexes of top scores were imported. In order to visualize the docking results, UCSF-Chimera, Discovery Studio (DS) Visualizer and Ligplus were used (Pettersen et al., 2004), (Modeling, n.d.), (Siraj et al., 2015).

2.8 Pharmacokinetic Profile Evaluation

2.8.1 ADMET and Toxicity Analysis

After docking, the top compounds from library were further analyzed for drug likeliness with respect to Lipinski's rule of 5 using PH, LogP value, M.W, H.B.A, H.B.D's and refractivity as well as other parameters (Lagorce et al., 2017). After implementing Lipinski rule of five, the compounds were analyzed by using ADMET properties for the gastrointestinal absorption, solubility, plasma protein binding ability, hepatotoxicity, Inhibition of cytochrome P450 and BBB (Nuez & Rodríguez, 2008). Swissadme was used for finding various properties essential for meeting criteria of drug designing like Canonical SMILES, Formula, MW, Rotatable bonds, H.B.A, H.B.D, TPSA, all LOG values, gastrointestinal absorption, permeable BBB, CYP2D6 inhibitor, log Kp (skin permeability), Lipinski #violations, Score for Bioavailability and Synthetic Accessibility (Aqeel & Majid, n.d.). A better drug candidate should have appropriate ADMET properties. As a result, numerous *in silico* models are created to predict the features of chemical ADMET. This study includes the ADMET-score, a scoring system to assess a compound's drug-likeness as these properties together have a huge impact on the pharmacology and pharmacokinetics of a drug (Kaplita et al., 2005).

Toxicity is another step to perform for selection of reliable therapeutic drug. The toxicity analysis was done using ProTox-II: a webserver for the toxicity in chemicals (Banerjee et al., 2018). This web server takes Canonical Smiles as input and evaluates several tests as shown in table in results section. Toxim is a small molecule toxicity prediction tool developed using chemoinformatics and machine learning techniques. Any molecule with a categorization score of 0.8 or higher is regarded as toxic, according to Matthews' correlation coefficient on 10-fold cross-validation (0.84), and according to the logarithm

of water solubility (LogS) , LogPapp - Logarithm of caco-2 permeability (range from -3.65 to -7.49) (Sharma et al., 2017).

2.9 2D Interaction Analysis

Discovery studio was used to visualize the hydrogen bonds, Vander Waals forces, electrostatic interactions, and hydrophobic interactions involved in top docked protein-ligand complexes. However, ligplus was also used to explore the interactivity (Sadeghi et al., 2021).

2.10 Molecular Dynamics Simulation

The dynamic nature of the complex and the dynamic behavior of ligand present in the active region of the protein were investigated using MD simulation. AMBER 16, a software tool (Case et al., 2017) was used to simulate an antibacterial drug target complex of 100 ns for each top docked CPS complexes. For system preparation, the antechamber leap software was used to analyze the geometry of a complex using an ff14SB as a force field. The addition of Na⁺ and Cl⁻ ions neutralize the system during preparation. After being neutralized, the system was placed in a TIP3P water box to study protein dynamics in a real-time environment. The system was then preprocessed, which included a number of stages. The first step in preprocessing is to minimize the docked complex to remove steric clashes caused by atom movement using a steepest descent algorithm and conjugate gradient algorithm of 750-steps. The heating of 10 picoseconds using dynamics of Langevin, with the (NVT) ensemble maintaining a constant temperature and volume, then pre-equilibrated for 100 picoseconds with a 2 fs time step. NVT ensemble, a run of 100 ns production for docked complexes with an 8.0 cut-off value for non-bonded interactions. The CPPTRAJ package of AMBER (Roe & Cheatham, 2013) was used to examine the trajectory after the production run analysis, and the stability of the ligand was depicted using VMD (Humphrey, W., Dalke, A., & Schulten, 1996). Different parameters including RMSD, RMSF, Rg, and B-Factor were measured, and graphical representations were done in Xmgrace and Qtgrace to validate the conformational stability of ligand in protein complexes.

2.10.1 Root Mean Square Deviation

RMSD is used to calculate the similarity between atomic coordinates of two superimposed structures, which can be implemented by using the formula as follows:

$$\text{RMSD} = \frac{\sum_{i=1}^N d_i}{N} \dots\dots\dots (2.1)$$

Here in the formula N represents the total number of atoms, while root mean square distance between two structures is represented by d_i .

2.10.2 Root Mean Square Fluctuation

RMSF is used to calculate C atoms in individual residues on the receptor–ligand docked complex. The variation of C atoms in each residue was calculated from the mean position. Following formula is used.

$$\text{RMSF} = \sqrt{\frac{\sum_t^T j=1 (x_i(tj) - x)}{T}} \dots\dots\dots (2.2)$$

While in formula, x_i represents the position of C-alpha atoms, x represents the atom's average location, and T represents the atom's time interval.

2.10.3 Beta-Factor

The beta factor, also known as atomic pressure, is a measurement of the disordered changes in a system that occurs as a result of an atom's temperature-dependent oscillation. Calculated in the same way as RMSF, but with a tiny difference in the squared fluctuation is multiplied by $(8/3)\pi^2$. Beta-factor can be calculated the following formula:

$$\text{Beta-factor} = \sqrt{\frac{\sum_t^T j=1 (x_i(tj) - x) \left(\frac{8}{3} \pi^2 \right)}{T}} \dots\dots\dots (2.3)$$

2.10.4 Radius of Gyration

It represents the compactness involved in protein structure and dictates shape, and is calculated using the formula:

$$\text{Rg} = \frac{\sum r^2 m}{\sum m} \dots\dots\dots (2.4)$$

2.10.5 Hydrogen Bond Analysis

The hydrogen bond graphs of the docked complexes were calculated using the AMBER 16 CPPTRAJ module. The hydrogen bond requires fractional analysis of both the donor and acceptor atoms with a cut-off value of ≥ 0.005 . X-axis represents the timescale, and Y-axis represents donor or acceptor atoms.

2.10.6 Radial Distribution Function

RDF is the probability to find a group involving N inhibitor atoms of r (distance) from any other specific atom in protein in r (radius) (Hemmer *et al.*, 1999). Conformation of amino acids close to a protein and its ligand's active site is calculated using the CPPTRAJ. RDF is represented as:

$$g(r) = \frac{p_{ij}(r)}{\langle p_j \rangle} = \frac{n_{ij}(r)}{\langle p_j \rangle 4\pi r^2 \sigma} \dots \dots \dots (2.5)$$

2.10.7 Binding Free Energy Calculation

The molecule's binding free energy is the sum of G_{solv}, G_{gas}, E_{vdw}, and electrostatic forces of interactions in MM (PBSA) as well as MM (GBSA) methods (Homeyer & Gohlke, 2012). In computational drug design, predicting the binding free energy of a ligand to a protein has become overly important since it enables the discovery of new compounds that can bind to a target and serve as therapeutic medications (Woo & Roux, 2005).

MMGBSA and MMPBSA compute binding energies of individual frames of top complexes of proteins using MMPBSA.py module/package in AMBER 16. Van der Waals interactions, electrostatics forces of interactions and changes in solvation energy in relation to ligand and protein binding affinity are all included in the MMPBSA approach (Wang *et al.*, 2016). The Ante-MMPBSA.py script was used to construct the ligand, receptor, and complex topology files (prmtop).

CHAPTER: 3 RESULTS

3.1 SUBTRACTIVE PROTEOMICS

3.1.1 Retrieval of Proteome and Drug Candidate Prioritization

Proteome of a bacterium was examined using subtractive proteomics methods. To start, the entire pan proteome of the *B. pertussis* strain Tohama I, which contained 4021 proteins, was downloaded from uniprot as a reference proteome (Consortium, 2021). To remove sequence redundancy and enhance the effectiveness, biological sequences are often grouped using the CD-HIT tool (Fu et al., 2012). This tool generated a file with non-redundant 3825 proteins and 90% sequence identity. The output of the redundancy check now only contains sequences that are found in the whole proteome after the vast numbers of similar protein sequences were deleted. The creation of drugs and vaccines is thought to benefit most from the use of non-redundant protein sequences as precursors. After that, BLASTp was run for these non-redundant proteins. It was used to make sure that none of the non-redundant protein sequences that we filtered out in the previous phase shared any similarities with the human proteome. It was crucial to do a homology study in order to choose the non-human homologues. When doing BLASTp on these protein sequences, Homo Sapiens (taxid: 9606) was used for the organism check and reference non-redundant proteins in the database check. The number of proteins was further decreased. This stage must be carefully completed since if the human homologues are chosen, the autoimmunity will be triggered, negatively damaging the host's tissues and cells (Brennan et al., 2010).

Executing the essentiality check of non-homologue proteins was a crucial next step in the pipeline. The proteins that are known to be essential are taken from the entire set of previously chosen human non homologous proteins since they form the basis of therapeutic targets. The remaining 3825 proteins' essential genes are found using DEG, 423 were proven essential by the geptop server when the essential genes were compared to the geptop database (Hua et al., 2016). The percentage was set as the result having percentage identity $\geq 30\%$ are considered. These proteins are vital to life and are frequently expressed by cells, making them well-known as potent therapeutic targets. These proteins are recognised as being crucial due to their localization in pathogens and ability to overcome effective immune response against infection. A total of 33 proteins

participating in metabolic pathways were mapped using the KAAS KEGG (Okuda et al., 2008).

Using the VFDB (Virulent Factor Database), which offers easy access to virulent factors and proteins in bacterial pathogens, a filter for virulence factors was applied on clustered proteome sequence. The VFDB identified 172 pathogenic proteins, while the remaining ones were disregarded because no hits were found (Chen et al., 2012). Virulent proteins are crucial therapeutic and vaccine targets since they are fascinating in the development of infections and initiating pathological conditions. The pathogen can survive inhospitable environmental circumstances because of the virulent proteins. The bacteria could reproduce, attack, and survive in the host cells with just a tiny amount of these proteins. These proteins are positioned on the cell surface to attract adhesion, protect bacterial cells with carbohydrates, and bestow pathogen pathogenicity with hydrolytic enzymes and toxins. 63 non-homologous proteins were detected as a result of BLASTp involved in an unique metabolic pathway (Peer-reviewed, 2020).

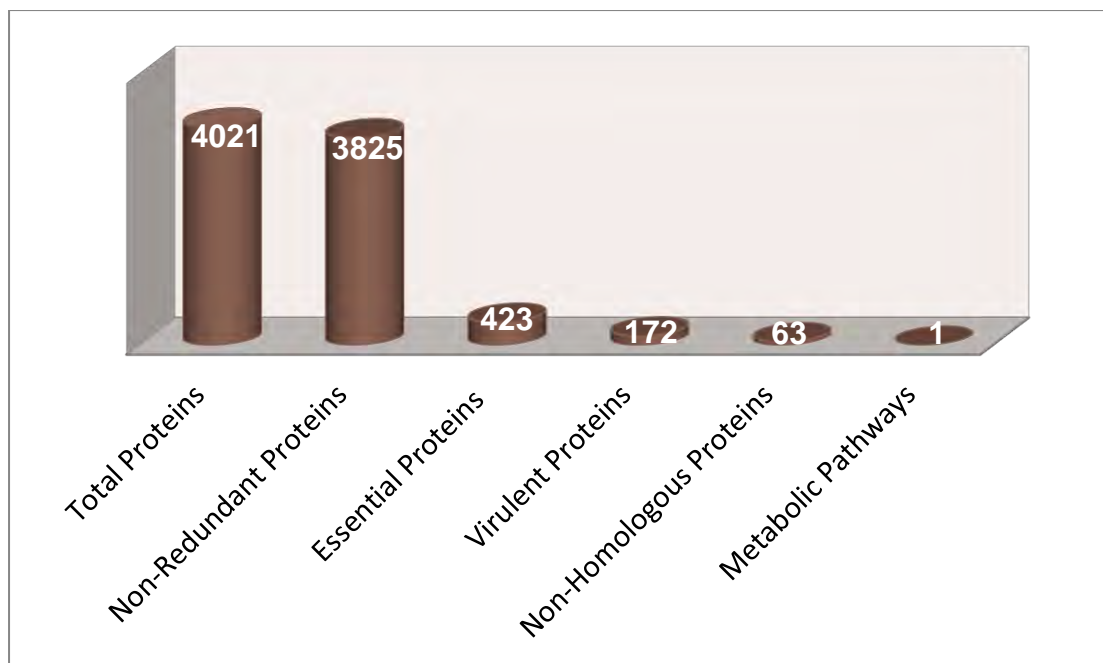


Figure 3. 1: Subtractive proteomics analysis of the proteome of bacterium executed by servers and databases

3.1.2 Sub-cellular Localization Analysis

The virulent and essential proteins were then subjected to shortlist the proteins which were confined in the cytoplasmic space via PSORTb, sub-cellular localization prediction tool. 35 shortlisted protein sequences were uploaded on PSORTb under the check of negative gram strain bacteria. Subcellular localization of proteins revealed that total 14 proteins which were localized in the cytoplasmic space (N. Y. Yu et al., 2010). CELLO and CELLO2GO predictors were used to cross-check the localizations predicted by PSORTb and the results are the same (C. Yu et al., 2014). Cytoplasmic proteins are typically enzymatic in nature hence all the cellular processes are catalyzed by these proteins. Also they have hydrophobic pockets for ligands/substrates as well as drugs and they are able to modulate biological functions.

3.1.3 Physicochemical Properties of Target Proteins

To extract the appropriate targets that can be used for the development of drug, various physicochemical properties are measured. The various physical and chemical properties of these remaining virulent proteins which was performed using ExPASy ProtParam tool (Sahay et al., 2020). Tools for analysis include Calculate ProtParam to determine various physicochemical characteristics and pI/Mw to predict the protein isoelectric point and molecular weight of molecules (Gasteiger et al., n.d.). It allows calculation of various parameters like atomic composition, instability index, amino acid, estimated half-life, aliphatic index, extinction coefficient, and grand average of hydropathicity. The protein stability is another vital factor to be measured using instability index calculator by envisaging the existence of specific dipeptides, absent in vivo unstable protein but present in protein having stability. The cut-off value for the protein instability index is set at 40. The protein showing prediction value > 40 are categorized as unstable. Proteins having low molecular weight are easy to purify and isolate to identify their structure and function. Higher stability proteins are preferred. The isoelectric points (PI) are estimated for the proteins. GRAVY values are calculated for virulent cytoplasmic proteins. Some protein show positive GRAVY values and some negative. For example, Acyl-CoA synthetases has 0.031 and capsular polysaccharide biosynthesis protein has -0.150

GRAVY value. Positive GRAVY value specifies hydrophobic nature whereas negative value specifies hydrophilic value.

Table 3. 1: Physicochemical properties of selected proteins

S. No	Protein	Subcellular Localization	Bit-score /Percentage Identity	Physiochemical characterization
1	Acyl-CoA synthetases	Cytoplasmic	212.616 /39.44%	Stable
2	Capsular polysaccharide biosynthesis protein	Cytoplasmic	485.723/70.71%	Stable
3	Acyl-CoA dehydrogenase fadE12	Cytoplasmic	105.531/29.32%	Unstable
4	Short chain dehydrogenase	Cytoplasmic	no matches/ 34.87%	Stable
5	High-affinity amino acid transport ATP-binding protein	Cytoplasm	75.8702/ 51.91%	Stable
6	ATP-binding component of ABC transporter protein	Cytoplasmic	87.8113/ 45.8%	Unstable
7	Chemotaxis methyltransferase	Cytoplasmic	147.517/ 64.22%	Stable
8	Protein-glutamate Omethyltransferase	Cytoplasmic	146.747/ 63.71%	Stable
9	Short chain dehydrogenase	Cytoplasmic	no matches/ 34.87%	Stable
10	LysR family regulatory protein	Cytoplasmic	No matches/ 28.9%	Stable
11	family regulatory protein	Cytoplasmic	No matches /36.3%	Unstable
12	Putative LysR family transcriptional regulator	Cytoplasmic	no matches/ 28.26%	Stable
13	Probable two-component response regulator	Cytoplasmic	53.5286/ 35.84%	Unstable
14	Cyclopropane mycolic acid synthase 1	Cytoplasmic	161.77 / 35.5%	Unstable

3.1.4 Drug Target Selection

Druggability of all the filtered proteins can be checked easily by exploring all the protein in drugbank database (Murugan et al., 2020). Drug Bank is special bioinformatics database which combines thorough target (protein) information with detailed drug (i.e. chemical) data. Among greater than >4100 drug listings in the database are >800 small molecule and biotech pharmaceuticals that have received FDA approval, as well as >3200 investigational drugs. Additionally, these medication entries are connected to about 14 000 protein sequences. Several protein targets for drugs that have received FDA approval are included in the Drug bank database (Wishart et al., 2006). Druggability of all the proteins was determined in order to identify the best therapeutic target with highest druggability bit-score. Bit score > 100 and identity > 30 can be suitable drug targets. However, a protein having score upto 485.723 with query coverage 99% and identity 70.71% as in case of capsular polysaccharide biosynthesis protein can be considered as a best target. CPS shows involvement in essential pathway, resides in cytoplasm, having a reliable GRAVY index, theoretical PI, molecular weight, and instability index. Several *in silico* approaches such as structure modeling, docking, and molecular dynamic simulations provide an excellent platform to design an antibiotic against the target protein, CPS. However, the druggability of the protein was also cross checked in literature and it was confirmed that CPS is a good anti-bacterial target.

Table 3. 2: Physicochemical characterization of best selected protein

Protein	Subcellular Localization	Drugbank Bit-score	Physiochemical characterization
Capsular Polysaccharide biosynthesis protein	Cytoplasmic	485.723	Stable

3.2 CPS Drug Target Proteins

CPS or UDP-N-acetyl glucosamine C4 epimerase for which structural and biochemical data is available is WbpP (Demendi et al., 2005) . WbpP has super family involving

short-chain dehydrogenases and reductases, including variety of oxidoreductases. WbpP has been identified as an actual UDP-GlcNAc 4-epimerase involved in interconversion of N-acetylated UDP-linked galactose and glucose (Udp-hexose et al., 2004). WbpP is made up of two domains in three dimensions. The domain at N-terminal has (residues ranging from 1-192 and 238-264). It has a modified Rossmann fold made up of nine helices on either side parallel-sheets having seven-strands. The UDP-linked hexose substrate is held in C-terminal domain (residues ranging from 193-237 and 265-343) which also has a /motif made up of four -helices and four-strands. WbpP has a structural resemblance to UDP-Gal 4-epimerases.

3.2.1. Role of Capsular Polysaccharide Proteins

PS capsules, which are the outermost covering of few bacteria, are crucial for defending them against hostile or unfavorable circumstances. The ability of bacterial capsules to mediate interactions of host-pathogen and immunity, resistance in antimicrobes, inhibits neutrophils recruitment, phagocytosis resistant and complement killing resistant, has been recognised as an important virulence determinant in addition to serving as a physical barrier for protection. Additionally, capsules also have stages of complex biofilm structures that are highly resistant to antibiotics. In BvgS-mediated signal transduction, transport machinery of capsule spanning across the envelope of cell presumably plays a role (Hoo et al., 2014). Humans lack the biosynthetic machinery necessary to produce the CP (capsular polysaccharide), making it a prime target for those seeking to combat bacteria like *S. aureus*. *S. aureus* can live in the blood of an infected host by building up a thick layer of sugars on the cell surface known as CP. The bacteria acquire anti-phagocytic properties from this layer (Miyafusa et al., 2013).

3.2.2. Mode of action of CPS

When PS haptens are covalently coupled to protein carriers, they acquire the ability to elicit humoral immune reactions that exhibit T cell-dependent antigen-like properties: responses in children's memory, affinity maturation, and most importantly, immunogenicity (Finn, 2004). Apart from the fact that CPS is a suitable target for vaccines, unencapsulating bacteria might be a good drug design technique by interfering with the process for CPS production and surface assembly. A Gram-negative bacterium

produces CPS through ATP-dependent and Wzy-polymerase dependent mechanisms. CPS provides protection from phagocytosis, and component of the innate immunological response of the host. It prevents the beginning of the phagocytic process by decreasing opsonin binding and obscuring ligand for phagocytic cell attachment (Sachdeva et al., 2017).

3.2.3 Unique Metabolic Pathway

CPS is involved in unique metabolic pathways of Galactose metabolism, Amino sugar and nucleotide sugar metabolism derived from KEGG analysis.

3.3 Cellular Interactome Analysis

The target protein CPS can be seen involved in a string interaction network via STRING. The total 10 interactions were observed and the proteins involved in this interaction with interaction scores as BP3144(0.832), BP3145(0.811), BP3151(0.611), BP3146 (0.804), Glycosyltransferase (0.877),BP3147(0.906),WbpO(0.957),WbpT(0.453), BP3148(0.854), BP3149, WbpO(0.748). Query protein is shown in the interactome is displayed by red bubble and the remaining color nodes represent interacting proteins. The wideness of blue interacting lines depict strength of association whereas the confidence score greater than 0.8 indicates functional significance of target protein (Szklarczyk et al., 2021).

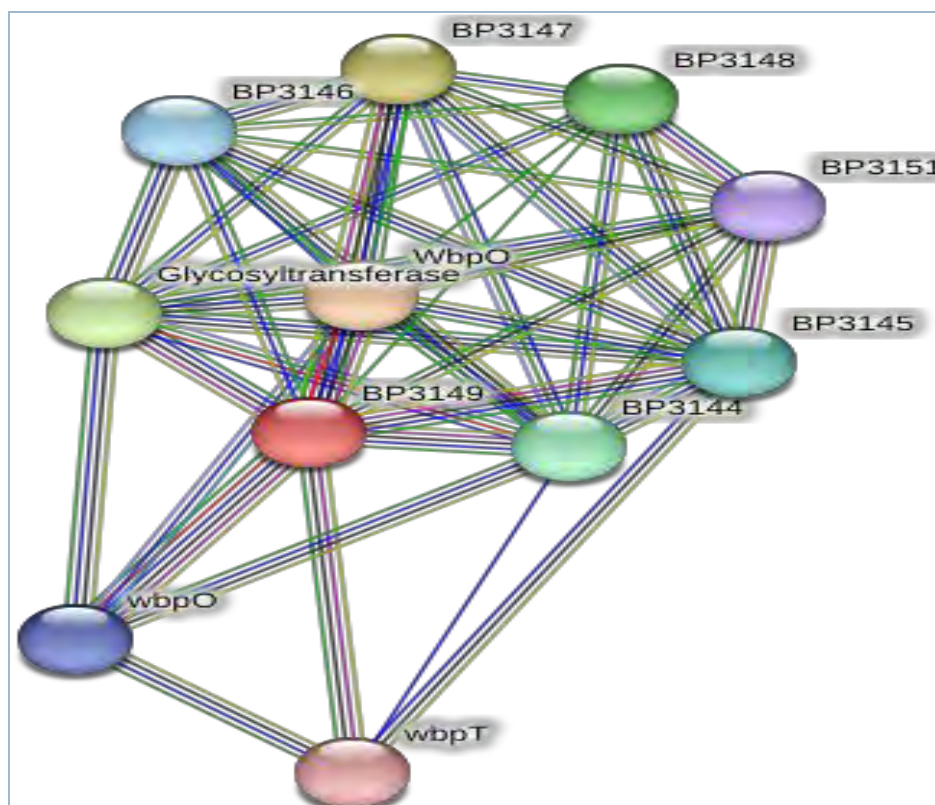


Figure 3. 2: Protein-protein interactions of target protein

3.4 Structure Modeling and Validation

The protein three-dimensional structure is required for performing molecular docking, simulations, trajectory analysis, and binding free energy. The 3D structure of a protein CPS is predicted using structure prediction methods. Swiss Model, Phyre2 and Alpha fold is used to predict the 3D structure of protein (Ruff & Pappu, 2021), (Kelley et al., 2015), (Kiefer et al., 2009). Stereo chemical properties of all the modeled 3D structures is compared for selecting the reliable model of the protein. The model which shows maximum favorable region in Ramachandran plot is selected. Along the Ramachandran plot, the model which shows maximum ERRAT and Verify3D score from Saves is preferred to be selected as the 3D structure of CPS as its overall quality factor is 95.7958 and 96.48% respectively. Based on the quality evaluation criteria, the Alpha Fold predicted 3D structure is selected as 93.3% core residues in the favorable region. The score is 6.4% in allowed, 0.3% in generally allowed and 0.0% disallowed. ProSA has -8.96 Z-score. The 3D structure of Alpha fold comes out to be the most reliable and

optimal model for the screening of antibiotics. It is given by Alpha fold protein structure database on the basis of PI used to determine an intra domain confidence which was very high ($pLDDT > 90$) indicating that the structure is modeled to high accuracy with high confidence level and ($pLDDT < 50$) shows low confidence and accuracy with ribbon like appearance with a high inclinations for intrinsic disorder whereas PAE is used for determining confidence level between the domains and chains of a protein.

Table 3. 3: Structural evaluation of predicted 3D structures

Structure Resources	Ramachandran Plot Analysis	Verify-3D	Errat Quality Factor	ProSA-Z score
Alpha-Fold	93.3% core	96.48%	95.7958	-8.96
Swiss-model	92.1% core	90.74%	94.8758	-8.85
Phyre2	90.9% core	91.50%	91.2913	-9.02

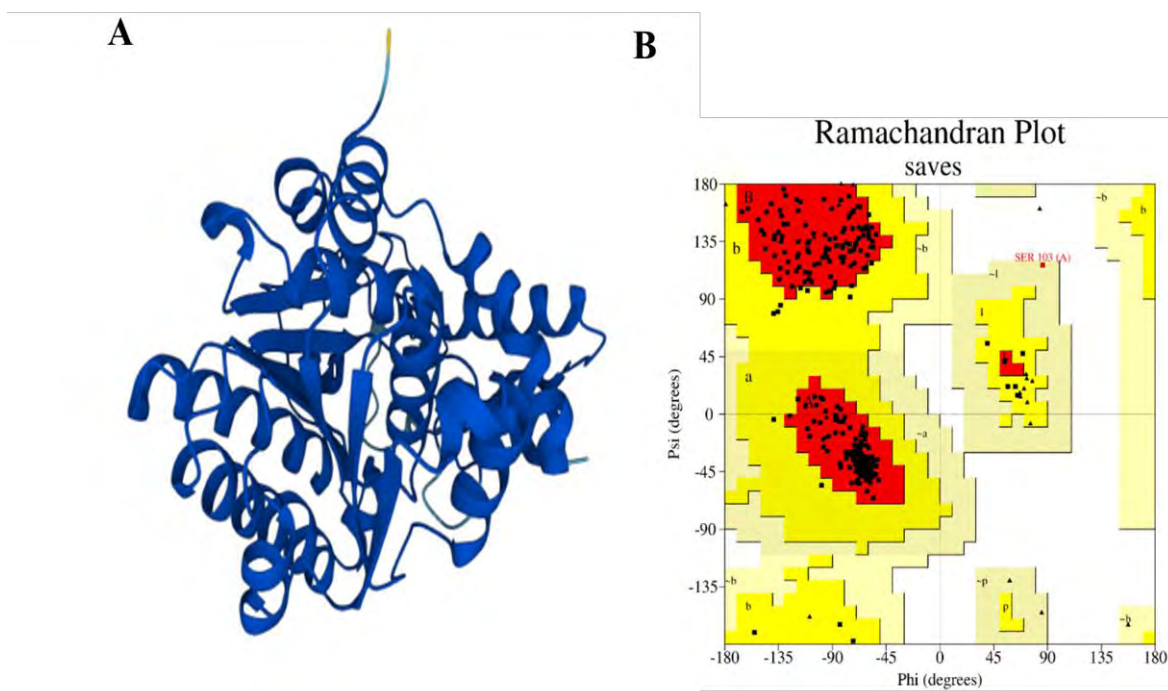


Figure 3. 3: A) Predicted 3D structure of CPS B) Ramachandran plot of CPS showing residues in favoured, allowed and disallowed region

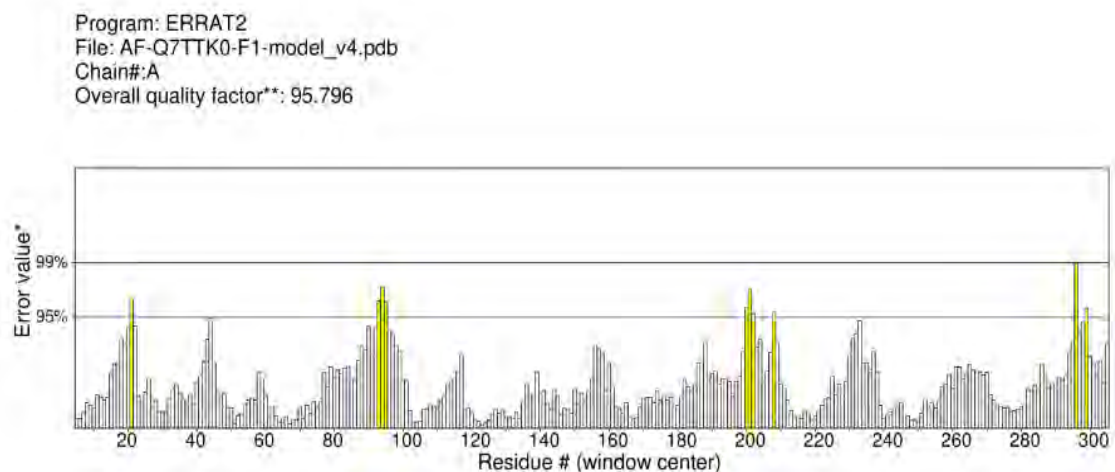


Figure 3. 4: Errat Plot with overall quality factor

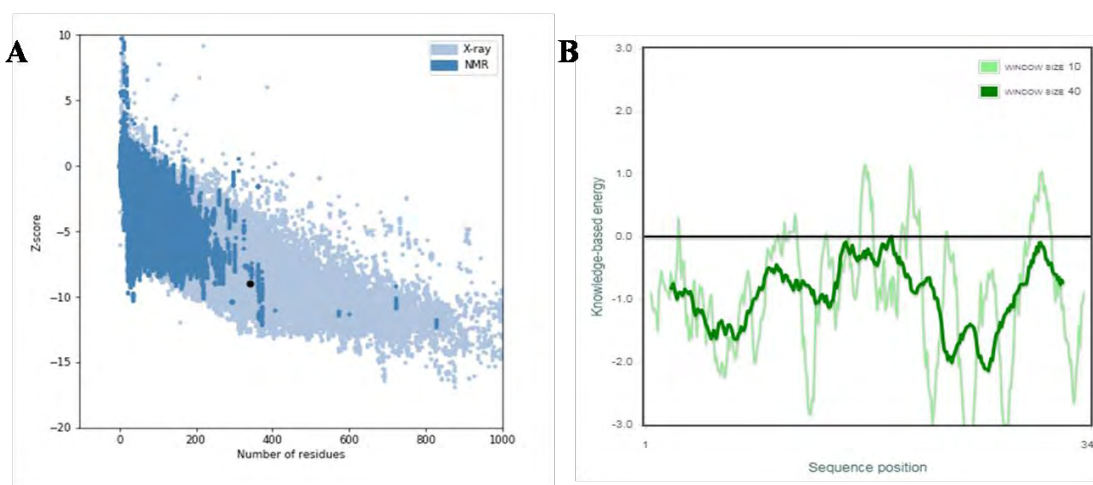


Figure3. 5: A) shows the Z-score B) shows the energy calculated for the predicted model

3.5 Minimization of Modeled Structure

The selected model is then minimized to improve the structure quality by eradicating steric clashes. For the minimization of protein model, UCSF Chimera is used. Minimization is executed in a two step-way process; in first one, the protein is undergone 750 steps of steepest descent to get rid of any unfavorable clashes, succeeded by the 750 steps of conjugate gradient method to remove the clashes left during the first phase of minimization. Tripos Force Field is employed to perform the minimization with a step size at 0.02A (Pettersen et al., 2004).

3.6 Functional Domains of the CPS Protein

Epimerase_deHydtase (19-264) belongs to Rossmann-fold domains (14-339). WbpP is made up of two domains in three dimensions, N-Terminal domain and C-terminal domain.

3.7 Prediction of Binding Site

3.7.1 Active Site Prediction

Active site of the protein involves sequence of amino acids involved in catalytic process and substrate binding. Therefore, known active site is a vital step to understand the protein structure and ultimately its function. The active site of CPS is predicted using computational tools such as CASTp, DoGSiteScorer, and Prankweb (W. Tian et al., 2018), (Volkamer et al., 2012), (Jendele et al., 2019). The active site residues is verified by literature A209, N195, S306, S143, S142, K170, Y166, S103, G102, Q201.

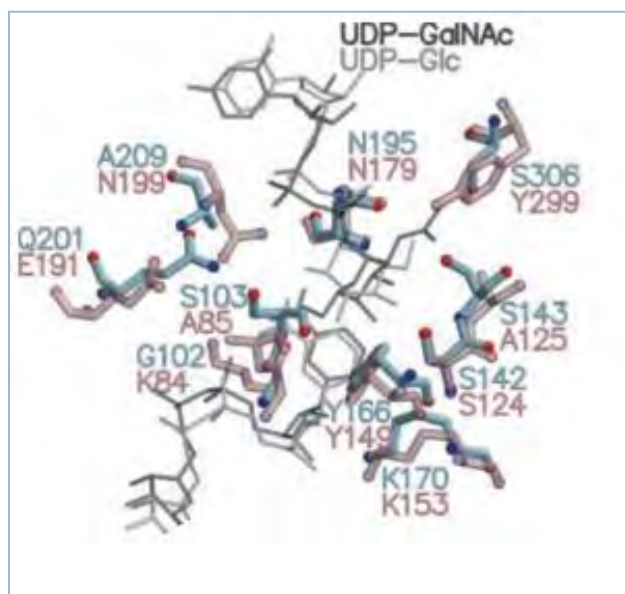


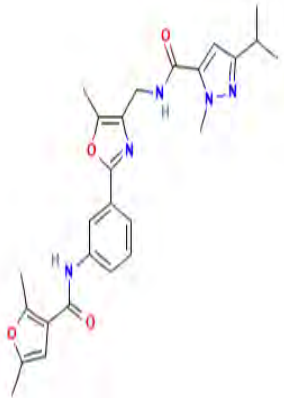
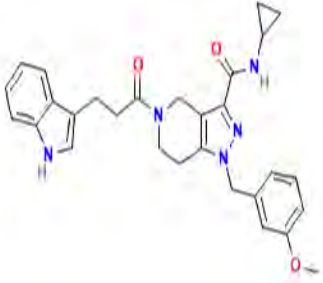
Figure 3. 6: Active site residues of WbpP-CPS protein in light blue

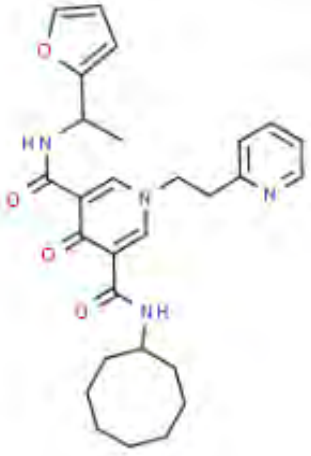
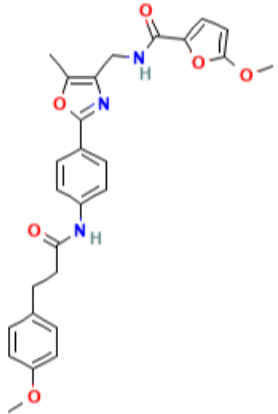
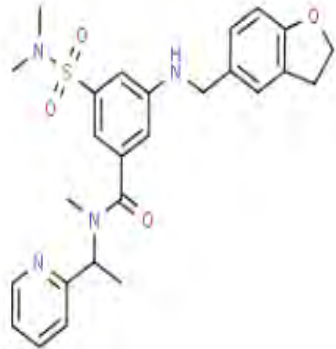
3.8 Anti-bacterial Library Preparation and Molecular Docking

A library chembridge-5900 is docked against CPS in GOLD. Active site residues arouse for molecular docking as mentioned above. Total ten iterations are produced for each

inhibitor. The Top 20 complexes having highest GOLD scores are imported. However, they are shortlisted on the basis of 2D interaction analysis and Swissadme tool analysis. For visualization of the docking results in order to find the interactions among the ligand and protein in complexes, UCSF-Chimera, VMD, and Discovery Studio (DS) visualize was used (Modeling, n.d.).

Table 3. 4: Compounds with the highest GOLD scores

Compounds Structure	Compound Name/formula	GOLD score
	<p>N-[[2-[3-[(2,5-dimethylfuran-3-carbonyl)amino]phenyl]-5-methyl-1,3-oxazol-4-yl]methyl]-2-methyl-5-propan-2-ylpyrazole-3-carboxamide/$C_{26}H_{29}N_5O_4$</p>	81
	<p>N-cyclopropyl-5-[3-(1<i>H</i>-indol-3-yl)propanoyl]-1-[(3-methoxyphenyl)methyl]-6,7-dihydro-4<i>H</i>-pyrazolo[4,3-<i>c</i>]pyridine-3-carboxamide/ $C_{29}H_{31}N_5O_3$</p>	73

	<p>N-Cyclooctyl-N'-[1-(2-furyl)ethyl]-4-oxo-1-[2-(2-pyridinyl)ethyl]-1,4-dihydro-3,5-pyridinedicarboxamide/$C_{28}H_{34}N_4O_4$</p>	77
	<p>ZINC11838605 5-methoxy-N- {[2-(4-{[3-(4-methoxyphenyl)propanoyl]amino} phenyl)-5-methyl-1,3-oxazol-4-yl]methyl}-2-furamide / $C_{27}H_{27}N_3O_6$</p>	75
	<p>3-[(2,3-Dihydro-1-benzofuran-5-ylmethyl)amino]-5-(dimethylsulfamoyl)-N-methyl-N-[1-(2-pyridinyl)ethyl]benzamide / $C_{26}H_{30}N_4O_4S$</p>	74

3.9 Pharmacokinetic Profile Evaluation

3.9.1 ADMET analysis

Following docking, the top compounds from the library are further tested for ADMET characteristics. Top-docked compounds in the library include ADMET descriptors and the Lipinski Rule of Five. The medicine will be deemed perfect if it complies with LR05 for physical and chemical parameters. It reveals if a chemical compound is drug-like or not, i.e., whether it has biological action intended for oral administration. A drug-like chemical compound should have hydrogen bond donors (HBD) > 5, molecular weight (MW) < 500g/mol, hydrogen bond acceptors (HBA) sites >10, and log p-value < 5 representing hydrophobicity of a compound, according to rule of thumb (RO5).. According to rule of thumb (RO5), a drug like chemical compound should have hydrogen bond donors > 5, molecular weight < 500g/mol, hydrogen bond acceptors (HBA) sites >10, and log p-value <5 represents hydrophobicity of a compound. The other parameters are bonds of rotation < 12 and a polar surface area < 140Å² which are simultaneous with the drug flexibility and permeability, respectively. The ADMET properties calculates absorption, distribution, metabolism, elimination, and toxicity (Lagorce et al., 2017).

The ADMET absorption descriptor predicts the human intestinal absorption (HIA). The solubility of each chemical molecule that resembles a medication in water at 25 °C is predicted using the ADMET aqueous solubility. The ADMET blood-brain barrier is used to assess the blood brain barrier (BBB) penetration of a chemical substance following oral digestion. The ADMET plasma protein binding descriptor is used to determine if a medicinal molecule would demonstrate strong bonding to the blood carrier protein. A 2D chemical structure of a molecule is used to determine a cytochrome P450 2D6 enzyme inhibition using the ADMET CYP2D6 binding model. The ADMET hepatotoxicity test quantifies the likelihood that a wide range of structurally different substances will cause hepatotoxicity in humans (Kaplita et al., 2005). The logP value must be between -0.4 and +5.6. The typical range of -8.0 to -1.0 applies to the log Kp for skin permeability. The top drug candidates from the library show good ADMET properties. Also, the results of Lipinski Rule of 5 of all the inhibitors are mentioned in the table.

Table 3. 5: Predicted Lipinski descriptors of all selected top compounds

Name	M.W	H.B.D	H.B.A	Rotatable Bonds	GI absorption
Chem-1	475.54	2	6	9	High
Chem-2	497.59	2	4	10	High
Chem-3	491.6	3	4	10	High
Chem-4	489.52	2	7	12	High
Chem-5	495.61	2	5	9	High
Name	BBB per - Meant	TPSA	logKp (cm/s)	ILOGP	CYP2D6 Inhibitors
Chem-1	No	120.07	-8.64	-3.12	No
Chem-2	No	90.43	-7.98	0	No
Chem-3	No	111.41	-6.72	0	No
Chem-4	No	123.16	-7.91	4.4	No
Chem-5	No	101.47	3.62	-7.28	Yes

3.9.2 Toxicity Analysis

For Rodent Oral Toxicity, ProTox-II and toxim is employed to check the toxicity of all the inhibitors from the library. Toxicity analysis of all the compounds is done to check the following properties like Molecular Refractivity, LogP, Hepatotoxicity, Carcinogenicity, and Immunotoxicity. Toxicity results depicted that the top compounds filtered from the docking results on the basis of GOLD score is found to be a good target for drug designing. The toxicity results are mentioned in the table. However, the further analysis during MD simulations will confirm the potency of these candidates to be used as a drug target against CPS.

Table 3. 6: Toxicity properties of top five docked compounds

Name	Molecular Refractivity	LogP	Hepatotoxicity	Carcinogenicity	Immunotoxicity
Chem-1	132.36	4.57	True	True	False
Chem-2	152.8	2.6	False	True	True
Chem-3	139.04	4.61	False	False	True
Chem-4	133.25	5.23	False	False	False
Chem-5	136.26	4.29	True	False	False

3.10 2D Interaction Analysis

The hydrogen bonds and hydrophobic interactions in the CPS and its corresponding ligand atoms are analyzed in Discovery Studio and LigPlus. After docking, it is essential to find out the interactions which hold the ligand in binding pocket of protein. Therefore, in order to perform 2D interaction analysis, the 2D diagram is generated. Significant interactions including hydrogen and hydrophobic bonds are observed between proteins and ligands. For complex1, the hydrogen bond residues are Asn195 (A), Phe27 (A) while hydrophobic residues are Ala209, Val104, Ser143, Arg234, Val303, Ser144, Thr51, Tyr193, Val196, Gly29, Gln98, Ile28, Gly23, Asp302, and Gly26. Second top docked score complex has Arg299, Tyr166, Gln98, Asn195 residues forming hydrogen bonds with hydrophobic residues as Ala209, Arg234, Ser306, Pro105, Ser103, Ala208, Tyr193, Ala100, Val196, Tyr207, Ser143, Thr51, Ile28, Phe27. Next complex has hydrogen bond residues Ser142, Ser143, Asn195 with hydrophobic residues Gln98, Ile28, Ala100, Thr51, Tyr166, Tyr193, Val196, Ala209, Ala140, Arg234, Phe194, Asp235. Complex 1 has Tyr166, Asn195 residues as hydrogen bond while hydrophobic one as Arg299, Pro105, Gly301, Val104, Ala300, Arg234, Ser143, Asp302, Phe194, Ser142, Tyr193,

3.11 Molecular Dynamics Simulations

The conformational dynamics of a protein molecule are equally vital to its function. A protein's structure contains all of the information necessary for the protein to perform its function. The detailed knowledge and better understanding of a structure of a protein and its conformational dynamics is important for its functional evaluation and variability. The top two complexes were undergone MD simulations which were CPS/ CHEM-10000290, CPS/CHEM-10008002. MD simulation of 100 ns were employed to understand the dynamics of the protein-ligand complexes, protein-ligand interactions, dynamical shifts, conformational fluctuations and hydrogen bond residues, hydrophobic residues and other amino acid residues of a protein playing crucial role in protein-ligand binding.

Several analysis were performed such as trajectory analysis, root mean square deviation (RMSD), root mean square fluctuations (RMSF), radius of Gyration (Rg) and beta factor were calculated.

3.11.1 Root Mean Square Deviation

RMSD is an essential parameter for estimating the deviation of the backbone C α atoms of proteins is observed for the complete simulation run. The mean RMSD calculated for the inclusion complex1 (CPS/ CHEM-10000290) and complex2 (CPS/CHEM-10008002) are 2.31 Å whereas maximum RMSD of 3.1536. Its highest peaks were 2.91 Å at 14ns, 2.97Å at 22ns, 2.98 Å at 27, 3.04 Å at 31ns, 3.1 at 45ns and 59ns. The other complex has 1.49 Å at 24 ns and max 2.1 Å at 44ns. Highest peaks were 2.05 Å at 21ns, 2.1 Å at 31ns and 2.1 Å at 53 respectively. Both complexes shows highest peaks at 2.98 Å at 22ns (coinciding peak), 3.06 Å at 31ns, 3.1 Å at 45ns, 2.7 Å at 81ns and 2.8 Å at 98ns. No extensive structural rearrangements were observed that explicates the complex stability and second complex is even more stable.

3.11.2 Root Mean Square Fluctuation

RMSF is an important factor in MD simulations to measure the fluctuations and motion of each residue in protein which explains protein flexibility during simulation. The

maximum RMSF for CPS/complex1 was 5.6696 Å and minimum was 0.3492 Å. The mean RMSF reported for complex1 was 0.837 Å. the mean RMSF for second complex was 0.78 Å with minimum at 0.3312 Å and maximum at 4.3493 Å. The regions depicting higher oscillations are highly flexible regions showing systematic conversion of helix into loop and loop into helix. The higher peak of the graph in all complexes specifies higher oscillations and fluctuations. As both complexes are stable throughout the simulation period because the active site residues are not showing prominent fluctuations as the binding pocket of the protein is stable. However, the second complex is more stable. Almost all the fluctuations were present in chain's loop and endpoints.

3.11.3 Beta-factor

The backbone and side chains of the proteins are in constant motion due to possession of thermal and kinetic energy of the atoms. With reference to RMSF, the thermal disorderness and the structural stability of a protein at an atomic level can be calculated by Beta-factor which provides essential information about the protein dynamics, as it measures the protein atomic fluctuations about their regular position. The graph pattern of beta-factor is coherent with the pattern of RMSF graph by identifying the same flexible regions which indicates the flexibility of a protein. The mean value of Beta-factor for first and second complex is 25.1838 Å and 20.3548 Å respectively. The minimum Beta-factor calculated for complex1 (CPS/ CHEM-10000290) is 3.2088 Å whereas the maximum is 846.007 Å. The minimum Beta-factor calculated for complex2 (CPS/CHEM-10008002) is 2.8875 Å whereas the maximum was 497.851 Å.

3.11.4 Radius of Gyration

Stability and compactness of protein structure is further examined by the calculation of another essential parameter of trajectory analysis which is radius of gyration (Rg). Rg represents the mass-weighted scalar length of each atom from the center of the mass of molecule. The higher Rg value indicates less compact structure whereas the lowest value illustrates the tight packing of a protein structure. It helps to understand the protein equilibrium conformation. The minimum Rg is 19.5184 Å for complex1 (CPS/ CHEM-10000290) whereas the maximum is 20.3657 Å. The mean value is 20.0487 Å. The

minimum Rg for complex2 (CPS/10008002) is 19.5798 Å whereas the maximum is 20.0727 Å. The mean value is 19.8174 Å.

Almost both the complexes shows similar Rg. The pattern of Rg is stable during the simulation run despite few variations in the graph (coherence with the results of RMSD) that could be due to some flexible loop regions of the protein.

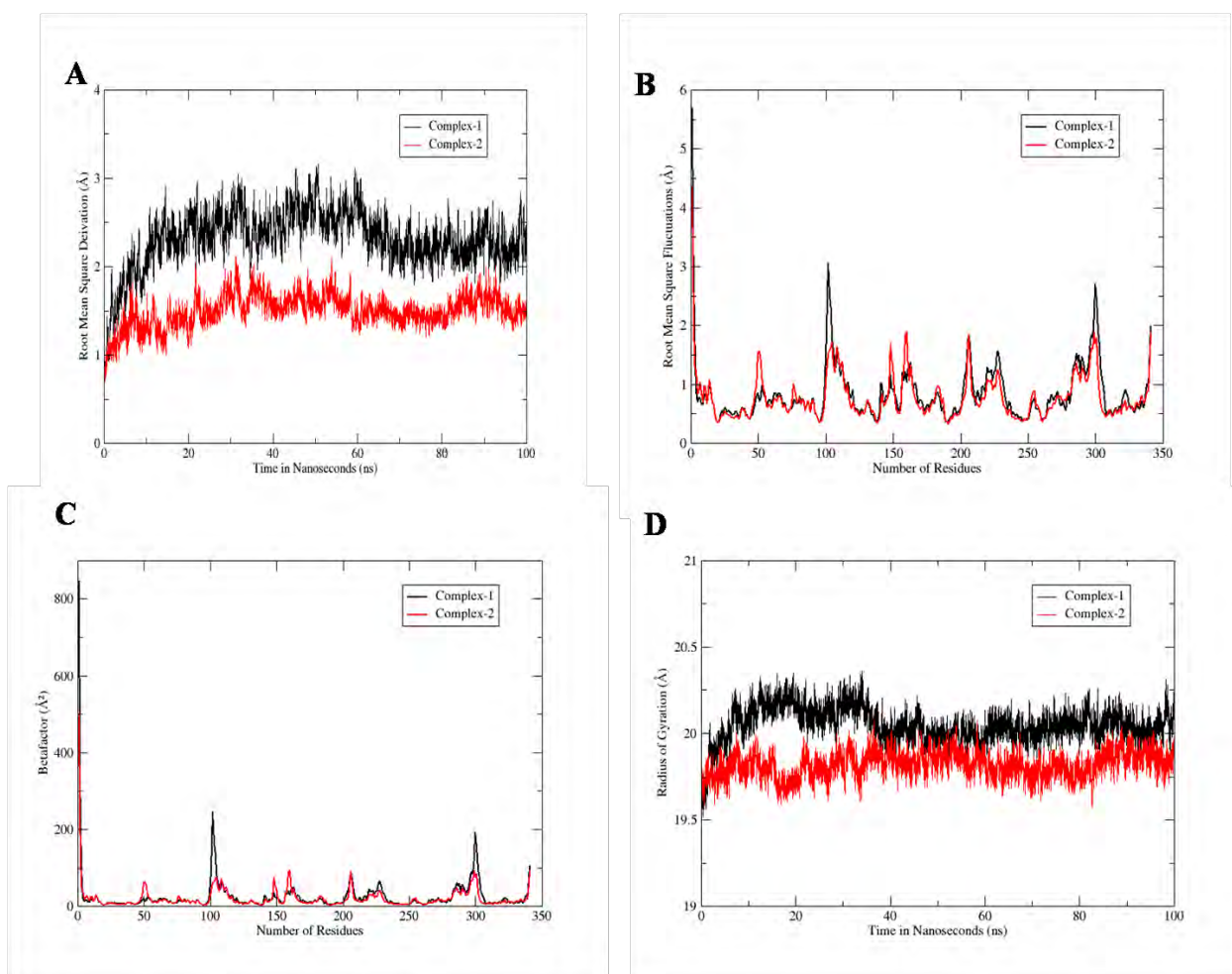


Figure 3. 8: Superimposed **A)** Root Mean Square Deviation **B)** Root Mean Square Fluctuation **C)** Beta-Factor **D)** Radius of Gyration of top two complexes

3.12 Hydrogen Bond Analysis

Hydrogen bond plays an essential role in the formation of stable and actively favorable receptor-ligand inclusion complex. To evaluate and analyze the influence and impact of a ligand on inhibitory mechanism of an enzyme, the complex conformation is required. To

B)

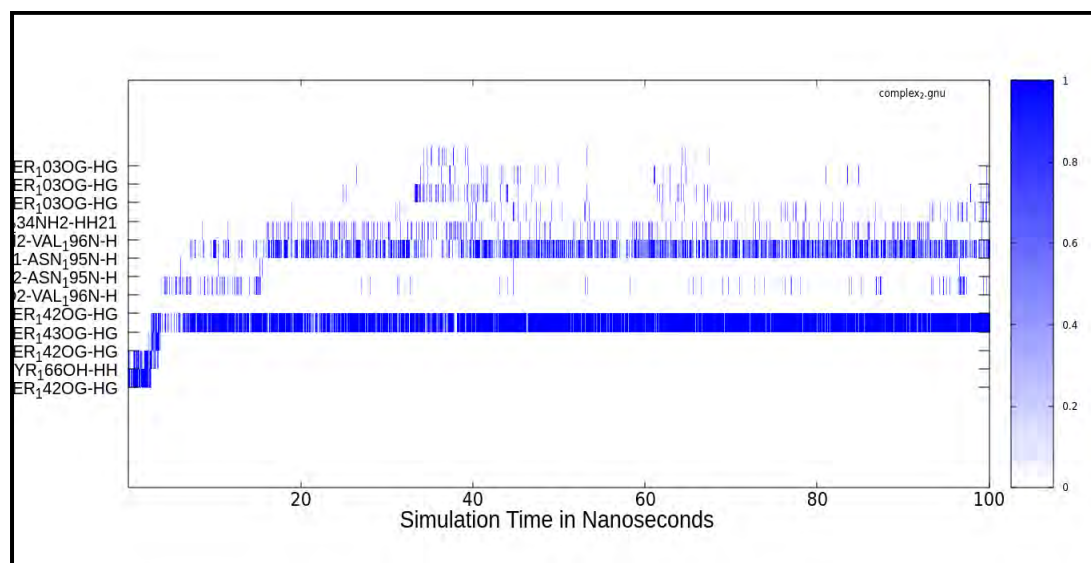


Figure 3. 9: Hydrogen Bond Analysis of A) CPS/CHEMBRIDGE-10000290 and B) CPS/CHEMBRIDGE-10008002

3.13 Radial Distribution Function

Both the complexes showing residues with constant and strongest hydrogen bonding are further subjected to radial distribution function (RDF) analysis. To reveal the interactions that mainly governs the protein-ligand affinity and complex stability, RDF is used by analyzing the interactions at the start, middle, and end of MD simulations. Therefore, depending on strong, continuous and stable hydrogen bonds, CPS/CHEMBRIDGE-10000290 and CPS/CHEMBRIDGE-10008002 are further subjected to radial distribution frequency analysis.

In case of CPS/ CHEMBRIDGE-10000290 in dataset operations, the N3 of CHEMBRIDGE -10000290 bonded with NE of Arg234 at 100ns is showing maximum peak at 0.21 at 4.95 Å, 0 at 0.025 Å while mean value of 0.032 at 2.5 Å. At 10ns, the minimum value is 0 at 0.025 Å, while maximum of 0.088376 at 4.975 Å. The mean value is 0.00734745 at 2.5 Å. At 100ns, H18 of CHEMBRIDGE -10000290 bonded with N of Gly29 is showing minimum peak of 0 at 0 Å, 0 at 0.475 Å while mean value of 0 at 0.025 Å. At 10ns, the RDF is 0 at 0.025 Å at its minimum peak, and 0.150494 at 4.975 Å, the

mean value is 0.0327431 at 2.5 Å. At 50ns, the N3 of CHEMBRIDGE -10000290 bonded with NE of Arg234 is showing its minimum peak at 0.025 at 0 Å, and maximum peak of RDF of 0.130188 at 4.975 Å. The mean value is 0.0222468 at 2.5 Å. At 50ns, H18 of CHEMBRIDGE -10000290 bonded with N of Gly29, the minimum value is 0 at 0.025 Å and the maximum value of RDF is 0.038872 at 4.975 Å. The mean value is 0.00512696 at 2.5 Å.

The N3 of CHEMBRIDGE-10000290 bonded with NE of Arg234 at 100ns, shows highest peak at 0.21 at 3.22 Å and lowest at 0.021 at 4.86 Å. At 10ns, 0.088 at 3.016 Å is observed at maximum and minimum of 0.0036 at 3.612 Å. The maximum peak observed is 0.129 at 3.776 Å at 50ns while the minimum is 0.004 at 2.82 Å. The H18 of CHEMBRIDGE -10000290-bonded with N of Gly29, shows its highest peak at 0.062 at 4.12 Å, and lowest at 0.042 at 4.41 Å at 100ns. At 10ns, 0.15 at 2.96 Å for highest one and 0.035 at 4.81 Å for lowest one. At 50ns, RDF is 0.037 at 4.421 Å at its highest one, while lowest at 0.007 at 3.866 Å.

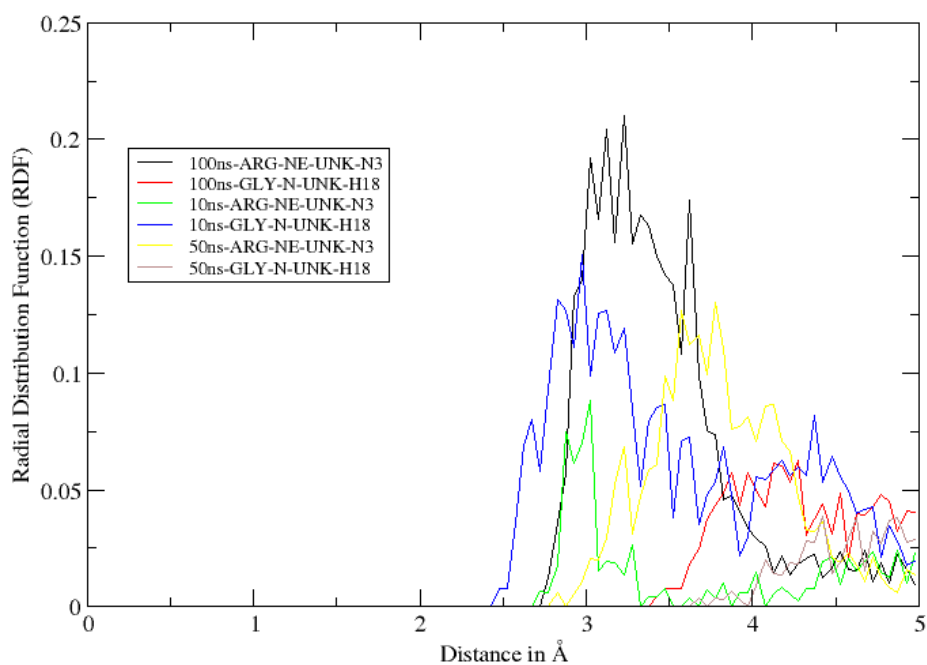
In case of CPS/CHEMBRIDGE-10008002, the H2 of CHEMBRIDGE -10000290-bonded with N of Asn195, the highest peak is observed at 0.229 RDF at 2.66 Å at 10ns, 0.028 RDF at 4.797 Å was shown as the lowest peak. At 10ns, the O1 bonded with OG of Ser142 shows its highest peak 0.59 at 2.67 Å while lowest at 0.014 at 4.135 Å. At 50ns, the H2 bonded with N of Asn195, shows its highest peak at 0.659 at 2.763 Å and lowest at 0.01 at 3.56 Å. At 50ns, the O1 bonded with OG of Ser142 shows its highest peak, 1.24 at 2.61 Å (the highest one in entire RDF) and lowest at 0.066 at 3.024 Å. At 100ns, the H2 bonded with N of Asn195, the highest peak observed was 0.561 at 2.624 Å and 0.561 at 2.79 Å. At 100ns, the O1 bonded with OG of Ser142 shows its highest peak, 1.154 at 2.60 Å while lowest at 0.014 at 3.114 Å.

The data set operations shows 0 at 0.025 Å at minimum of the H2 of CHEMBRIDGE -10000290 bonded with N of Asn195 at 10ns while maximum at 0.232677 at 4.975 Å. The mean observed was 0.041 at 2.5 Å. At 50ns, it showed 0 at 0.025 Å for minimum and 0.654 at 4.975 Å for maximum, the mean value was 0.0611 at 2.5 Å. At 100ns, it also

shows 0 at 0.025 Å at minimum while maximum of 0.566 at 4.975 Å, the same mean value 0.060 at 2.5 Å.

At 10ns, the O1 bonded with OG of Ser142, shows its minimum RDF of 0 at 0.025 Å, and maximum of 0.598 at 4.975 Å. The mean value is 0.0547 at 2.5 Å. At 50ns, RDF is 0 at 0.025 Å, as a minimum value while maximum is 1.249 at 4.975 Å. The mean value is 0.065 at 2.5 Å. the 100ns, the minimum value is 0 at 0.025 Å, maximum of 1.1529 at 4.975 Å. The mean value is 0.0653.

(A)



B)

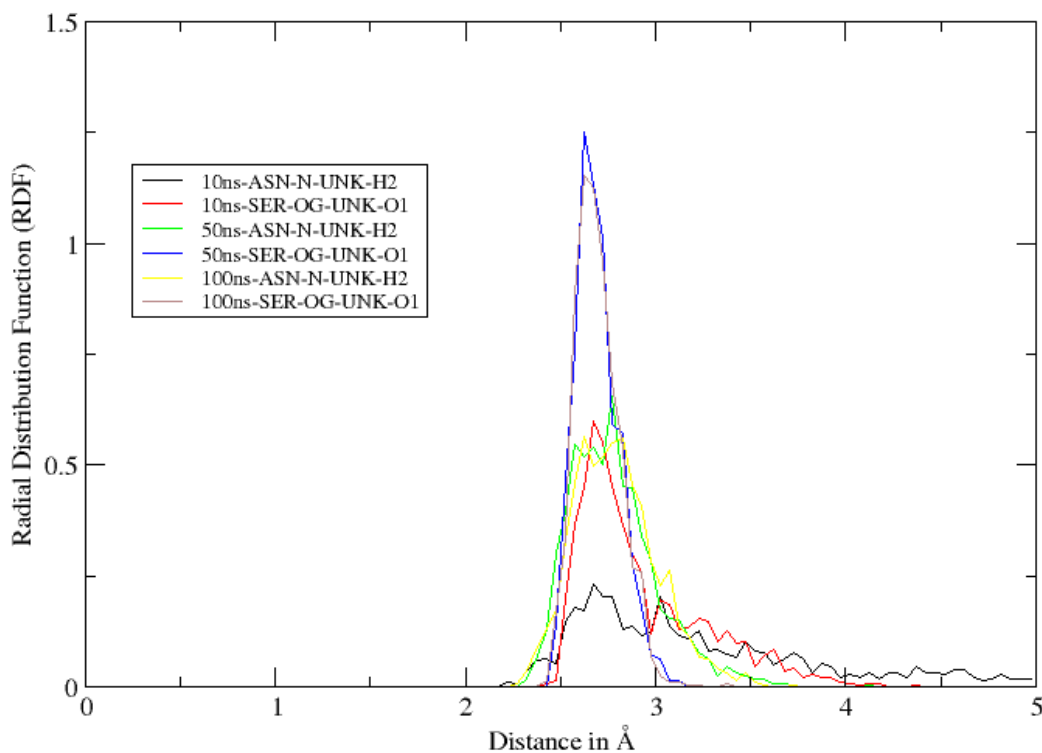


Figure 3. 10: A) Radial Distribution Function of CPS/CHEMBRIDGE-10000290 and B) CPS/CHEMBRIDGE-10008002

3.14 Binding Free Energy Calculation

The binding affinity between the ligand and the protein molecule in all the inclusion complexes is investigated by MMGB/PBSA based binding free energy (ΔG_{bind}). MMGB/PBSA is a package of AMBER16 which is applied on MD simulation trajectories of both complexes. The electrostatic interaction (ΔE_{EL}), van der Waals interaction (ΔV_{DW}), and the ΔG_{bind} values of MMGB/PBSA are calculated. Following values shows the binding free energy calculations for both complexes.

Table 3. 7: Binding Free Energy Calculations of selected complexes via MM/GBSA

Contribution (GB)	Energy values (kcal/mol) of CHEMBRIDGE-1000290-40.4_81	Energy values (kcal/mol) of CHEMBRIDGE-10008002-720_73
ΔE_{vdw}	-10.5500	-6.8932
ΔE_{ele}	-111.1270	-102.3875
ΔG_{gas}	-45.6847	-35.0992
ΔG_{solve}	42.9649	34.4252
ΔG_{total}	-2.7198	-0.6740

Table 3. 8: Binding Free Energy Calculations of selected complexes via MM/PBSA

Contribution(PB)	Energy values (kcal/mol) of CHEMBRIDGE-1000290-40.4_81	Energy values (kcal/mol) of CHEMBRIDGE-10008002-720_73
ΔE_{vdw}	-10.5500	-6.8932
ΔE_{ele}	-111.1270	-102.3875
ΔG_{gas}	-45.6847	-35.0992
ΔG_{solve}	39.1813	30.8535
ΔG_{total}	-6.5034	-4.2458

CHAPTER: 4 DISCUSSION

4. Discussion

In this study, an *in silico* subtractive proteomic approach was employed to categorize and identify potential drug targets. The cause of pertussis, sometimes known as whooping cough, is a gram-negative, aerobic, pathogenic, encapsulated coccobacillus, opportunistic pathogen, extremely drug resistant, nonfermentative and catalase positive of the genus *Bordetella* that causes serious human infections. It belongs to the family of Alcaligenaceae. It is responsible for upper respiratory tracts, such as fever, sneezing, and a moderate and infrequent coughing. cerebral hypoxia, severe alkalosis, convulsions, subcutaneous emphysema, subconjunctival hemorrhage, umbilical or inguinal hernias, rib fractures, and umbilical or inguinal hernias, a light, infrequent cough (babies do not do this) newborns and young children who experience apnea, which are potentially fatal breathing pauses, and cyanosis, which is a blue or purple coloration. The bacteria *Bordetella pertussis* is the cause of pertussis, an acute respiratory illness that is extremely contagious. There are three stages to this condition: catarrhal, paroxysmal, and convalescent. The proteome wide identification begins with the retrieval of the proteome of *B.Pertussis* from uniprot which comprised of 4021 proteins. The complete proteome was used to identify the essential, virulent, and therapeutic protein by applying several *in silico* methods. After the removal of non-essential and homologous proteins, 423 essential proteins, involved in different pathways were discovered. The virulence factors, localization, druggability, and physicochemical properties of proteins were checked to identify a protein involved in the cytoplasmic space of pathogen and found to be the most druggable target. By applying the various subtractive proteomics filters, CPS was discovered and selected. The CPS has a virulent nature therefore; it can be used as a target of interest for the development of novel therapeutic drug against *B.Pertussis*. It has various unique functions, such as protection and primarily by interfering with host opsonophagocytic clearance processes, plays a crucial part in pathogenicity which ultimately disrupts the host immune system. The 3D structure of the protein was Alpha fold predicted using AI based approach which is recent cutting edge technology for the best modeling with refined results. SWISS-MODEL and Phyre2 were utilized for

structure validation of the CPS protein. By comparing these three models developed by different tools, the model created for Alpha Fold was selected because of its high credibility. The hypothesized structure was minimized and optimized with 750 steps of steepest descent and 750 steps of conjugate gradient method to remove any unfavorable clashes by using Tripos Force Field. CASTp, DoGSiteScorer, and Prankweb were employed to predict the catalytic site of CPS. The active site residues A209, N195, S306, S143, S142, K170, Y166, S103, G102, and Q201 were present in CPS with epimerase domain. Chembridge-5900 Library of inhibitor compounds was created. The binding poses and interactions between inhibitors and protein were determined using GOLD. The top inhibitors on the basis of GOLD scores were then evaluated for ADMET, Lipinski Rule of Five, and toxicity analysis. The top five inhibitors were shortlisted. The aim of this study, was to perform the comparative analysis on novel inhibitors from Chembridge library. The top compounds of chembridge (IUPAC name: N-[[2-[3-[(2,5-dimethylfuran-3-carbonyl)amino]phenyl]-5-methyl-1,3-oxazol-4-yl]methyl]-2-methyl-5-propan-2-ylpyrazole-3-carboxamide/C₂₆H₂₉N₅O₄,N-cyclopropyl-5-[3-(1*H*-indol-3-yl)propanoyl]-1-[(3-methoxyphenyl)methyl]-6,7-dihydro-4*H*-pyrazolo[4,3-*c*]pyridine-3-carboxamide/ C₂₉H₃₁N₅O₃) were undergone MD simulations of 100ns to examine the dynamics. The structural deviations were analyzed using trajectory analysis which includes RMSD, RMSF, Radius of Gyration, and Beta-Factor. The RMSD graph of both complexes exhibited variations at different nanoseconds which explain the structural changes in the backbone of CPS. Second complex showed more stability as mentioned in the results section. By interpreting all the graphs, it was concluded from trajectory analysis, that both the complexes were showing the constant pattern of fluctuations because of alterations in loop, helix, and Beta sheets but the ligand remained at the targeted site. Further, the hydrogen bond analysis was performed. Both the complexes had hydrogen bonds but second one was showing more stable hydrogen bond at different nanosecond scale as explained in results section. The both complexes has shown strong interactions and binding with water and ligand. Although second complex hydrogen bond was more stable, both were further subjected to RDF analysis. The results of RDF were showing compatibility the results of docking and MD simulations. The binding free energy of both the complexes was calculated using MMGB/PBSA approaches. The

strong binding affinity of both the complexes for ligand and receptor was observed. The more negative values indicate more binding. Consequently, it is concluded that they can be used as a potential drug targets to inhibit the growth of multidrug resistant *B.Pertussis* by inhibiting the function of CPS. The simulations results indicate that giving stability to the complex can be used as a therapeutic drug target. The complexes showed the best binding at the active site of CPS predicted in this study. Therefore, by incorporating the changes in their structure, its absorption, solubility, and pharmacokinetic profile can be maintained so that it can be used as a therapeutic drug against CPS in future. Thus, this study helps us to understand how to choose a potential drug target protein by using comparative dynamics, proteome wide identification technique and to develop a therapeutic drug to inhibit the pathogen growth by inhibiting the target protein function.

CHAPTER: 5 CONCLUSION

CONCLUSION

Alcaligenes is a phylogenetic subcategory of proteobacteria comprised of diverse species. *B. Pertussis* belongs to the family of Alcaligenaceae. It is a gram-negative bacterium and extremely drug resistant pathogen that causes serious human infections. It causes infections of respiratory tracts, vomiting, whooping cough, urinary infections and blood stream infections. In the current study, proteome wide identification approach, comparative dynamics and subtractive proteomics was used that highlighted CPS as a promising drug target. The 3D structure of the protein Alpha Fold predicted and comparison with other structures predicted via Swiss-Model and Phyre2 further validated and emphasized the credibility of the model generated by Alpha Fold. To perform the comparative analysis and identify the potential drug candidates, the top inhibitors from Chembridge library were undergone *in silico* analysis which includes Docking, ADMET and Toxicity Analysis, MD simulations, Trajectories, Hydrogen Bond Analysis and Radial Distribution Function. The results showed more stability of second complex as indicated by hydrogen bonding and binding free energy calculations. The *in silico* study revealed that both complexes can be used as potential drug targets to inhibit the role of CPS. The binding of the inhibitor at the domain of Capsular Polysaccharide protein will inhibit its activity thus halting the virulence and protection mechanism ultimately killing the extensively drug resistant *B. Pertussis*. As the ligand is fixed at the binding site and did not move away from the binding, it exhibited prominent interactions with the solvent, therefore it can be subjected to further investigation. These results revealed that both inhibitor complexes can be used as a potential therapeutic drug against CPS after improving and analyzing their pharmacokinetic profile. Although these inhibitors showed good binding efficacy for the CPS, the experimental in-vivo and in-vitro validation is still required to authenticate the therapeutic and prophylactic effects of these drugs.

REFERENCES

- Ã, W. L., & Godzik, A. (2006). *Cd-hit : a fast program for clustering and comparing large sets of protein or nucleotide sequences*. 22(13), 1658–1659.
<https://doi.org/10.1093/bioinformatics/bt1158>
- Abduljalil, J. M. (2022). *Serine / threonine kinase of human Monkeypox virus : computational modeling and structural analysis*. 1–17. <https://doi.org/10.21203/rs.3.rs-2315316/v1>
- Angamuthu, K., & Piramanayagam, S. (2017). *Evaluation of In silico Protein Secondary Structure Prediction Methods by Employing Statistical Techniques*. 29–36.
<https://doi.org/10.4103/bbrj.bbrj>
- Azam, S. S., Abro, A., & Raza, S. (2015). Binding pattern analysis and structural insight into the inhibition mechanism of Sterol 24-C methyltransferase by docking and molecular dynamics approach. *Journal of Biomolecular Structure and Dynamics*, 1102, 1–15.
<https://doi.org/10.1080/07391102.2014.1002423>
- Banerjee, P., Eckert, A. O., Schrey, A. K., & Preissner, R. (2018). *ProTox-II : a webserver for the prediction of toxicity of chemicals*. 46(April), 257–263.
<https://doi.org/10.1093/nar/gky318>
- Brennan, F. R., Morton, L. D., Spindeldreher, S., Allenspach, R., Hey, A., Müller, P., Frings, W., Brennan, F. R., Morton, L. D., Spindeldreher, S., Allenspach, R., Hey, A., Müller, P., Frings, W., Safety, J. S., Brennan, F. R., Morton, L. D., Spindeldreher, S., Kiessling, A., ... Sims, J. (2010). *Safety and immunotoxicity assessment of immunomodulatory monoclonal antibodies Safety and immunotoxicity assessment of immunomodulatory monoclonal antibodies*. 0862(May). <https://doi.org/10.4161/mabs.2.3.11782>
- Brokl, M., Bishop, L., Wright, C. G., Liu, C., & Mcadam, K. (2013). *Analysis of mainstream tobacco smoke particulate phase using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry*. 1037–1044.
<https://doi.org/10.1002/jssc.201200812>
-

-
- Carbonetti, N. H. (2016). *treatment*. 29(3), 287–294.
<https://doi.org/10.1097/QCO.0000000000000264>. *Bordetella*
- Chen, L., Xiong, Z., Sun, L., Yang, J., & Jin, Q. (2012). *VFDB 2012 update : toward the genetic diversity and molecular evolution of bacterial virulence factors*. 40(November 2011), 641–645. <https://doi.org/10.1093/nar/gkr989>
- Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., & Jin, Q. (2005). *VFDB : a reference database for bacterial virulence factors*. 33, 325–328. <https://doi.org/10.1093/nar/gki008>
- Consortium, T. U. (2015). *UniProt : a hub for protein information*. 43(October 2014), 204–212. <https://doi.org/10.1093/nar/gku989>
- Demendi, M., Ishiyama, N., Lam, J. S., Berghuis, A. M., & Creuzenet, C. (2005). *Towards a better understanding of the substrate specificity of the UDP-N - acetylglucosamine C4 epimerase WbpP*. 180, 173–180.
<https://doi.org/10.1042/BJ20050263>
- Desjardins, M., Mak, W. S., Brien, T. E. O., Carlin, D. A., Tantillo, D. J., & Siegel, J. B. (2017). *Systematic Functional Analysis of Active-Site Residues in*.
<https://doi.org/10.1021/acsomega.7b00519>
- Finn, A. (2004). *Bacterial polysaccharide – protein*. 1–14. <https://doi.org/10.1093/bmb/ldh021>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). *BIOINFORMATICS APPLICATIONS NOTE Sequence analysis CD-HIT : accelerated for clustering the next-generation sequencing data*. 28(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Garg, A., & Gupta, D. (2008). *VirulentPred : a SVM based prediction method for virulent proteins*. 12, 1–12. <https://doi.org/10.1186/1471-2105-9-62>
- Hansson, T., Oostenbrink, C., & Gunsteren, W. F. Van. (2002). *Molecular dynamics simulations*. 190–196. [https://doi.org/10.1016/S0959-440X\(02\)00308-1](https://doi.org/10.1016/S0959-440X(02)00308-1)
- Hema, K., Ahamad, S., Joon, H. K., Pandey, R., & Gupta, D. (2021). *Atomic Resolution Homology Models and Molecular Dynamics Simulations of Plasmodium falciparum*
-

- Tubulins*. <https://doi.org/10.1021/acsomega.1c01988>
- Hemmer, M. C., Steinhauer, V., & Gasteiger, J. (1999). *Deriving the 3D structure of organic molecules from their infrared spectra*. 151–164. [https://doi.org/10.1016/S0924-2031\(99\)00014-4](https://doi.org/10.1016/S0924-2031(99)00014-4)
- Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area method. *Molecular Informatics*, 31(2), 114–122. <https://doi.org/10.1002/minf.201100135>
- Hoo, R., Lam, J. H., Huot, L., Pant, A., & Li, R. (2014). *Evidence for a Role of the Polysaccharide Capsule Transport Proteins in Pertussis Pathogenesis*. 1–31. <https://doi.org/10.1371/journal.pone.0115243>
- Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). *Molecular dynamics simulations : advances and applications* *Molecular dynamics simulations : advances and applications*. <https://doi.org/10.2147/AABC.S70333>
- Hua, H., Zhang, F., Labena, A. A., Dong, C., Jin, Y., & Guo, F. (2016). *An Approach for Predicting Essential Genes Using Multiple Homology Mapping and Machine Learning Algorithms. 2016*. <https://doi.org/10.1155/2016/7639397>
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Castro, E. De, Langendijk-genevaux, P. S., Pagni, M., & Sigrist, C. J. A. (2006). *The PROSITE database*. 34, 227–230. <https://doi.org/10.1093/nar/gkj063>
- Jamal, A., Jahan, S., Choudhry, H., & Rather, I. A. (2022). *A Subtraction Genomics-Based Approach to Identify and Characterize New Drug Targets in Bordetella pertussis : Whooping Cough*. 1–10. <https://doi.org/10.3390/vaccines10111915>
- Jendele, L., Krivak, R., Skoda, P., Novotny, M., & Hoksza, D. (2019). *PrankWeb : a web server for ligand binding site prediction and visualization*. 47(May), 345–349. <https://doi.org/10.1093/nar/gkz424>
- Kaplita, P. V, Hu, H., Liu, L., Farrell, T. M., Grbic, H., & Spero, D. M. (2005). *Automatable*

-
- Formats of Higher Throughput ADMET Profiling for Drug Discovery Support*. June.
<https://doi.org/10.1016/j.jala.2005.03.004>
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling , prediction and analysis. *Nature Protocols*, 10(6), 845–858. <https://doi.org/10.1038/nprot.2015-053>.
- Keser, M., & Stupp, S. I. (2000). *Genetic algorithms in computational materials science and engineering : simulation and design of self-assembling materials*. 186.
[https://doi.org/10.1016/S0045-7825\(99\)00392-8](https://doi.org/10.1016/S0045-7825(99)00392-8)
- Khan, A. S., Miah, I., & Rahman, S. R. (2022). Informatics in Medicine Unlocked Subtractive proteomic analysis for identification of potential drug targets and vaccine candidates against Burkholderia pseudomallei K96243. *Informatics in Medicine Unlocked*, 35(October), 101127. <https://doi.org/10.1016/j.imu.2022.101127>
- Kiefer, F., Arnold, K., & Ku, M. (2009). *The SWISS-MODEL Repository and associated resources*. 37(October 2008), 387–392. <https://doi.org/10.1093/nar/gkn750>
- Knegtel, R. M. A., Kuntz, I. D., & Oshiro, C. M. (1997). *Molecular Docking to Ensembles of Protein Structures*. 424–440. <https://doi.org/10.1006/jmbi.1996.0776>
- Kocisko, D. A., Baron, G. S., Rubenstein, R., Chen, J., Kuizon, S., & Caughey, B. (2003). *New Inhibitors of Scrapie-Associated Prion Protein Formation in a Library of 2 , 000 Drugs and Natural Products*. 77(19), 10288–10294.
<https://doi.org/10.1128/JVI.77.19.10288>
- Lagorce, D., Douguet, D., Miteva, M. A., & Villoutreix, B. O. (2017). Computational analysis of calculated physicochemical and ADMET properties of protein- protein interaction inhibitors. *Nature Publishing Group, December 2016*, 1–15.
<https://doi.org/10.1038/srep46277>
- Lam, C., Octavia, S., Ricafort, L., Sintchenko, V., Gilbert, G. L., Wood, N., McIntyre, P., Marshall, H., Guiso, N., Keil, A. D., Lawrence, A., Robson, J., Hogg, G., & Lan, R. (2014). *Rapid Increase in Pertactin-deficient Bordetella pertussis Isolates , Australia*.
-

- 20(4). <https://doi: 10.3201/eid2004.131478>
- Leelananda, S. P., & Lindert, S. (2016). *Computational methods in drug discovery*. 2694–2718. <https://doi.org/10.3762/bjoc.12.267>
- Lin, D., Wang, X., Li, Y., Wang, W., Li, Y., Yu, X., Lin, B., Lei, C., Zhang, X., Zhang, X., Huang, J., Lin, B., Yang, W., & Zhou, J. (2021). *Sputum microbiota as a potential diagnostic marker for multidrug-resistant tuberculosis*. 18. <https://doi.org/10.7150/ijms.53492>
- Lutwick, L., Preis, J., & Presentation, C. (2014). Pertussis. In *Emerging Infectious Diseases*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-416975-3.00027-3>
- Mahtab, T., Jyoti, A., Khusro, A., Redwan, B. M., Zidan, M., Mitra, S., Bin, T., Dhama, K., Hossain, K., Gajdács, M., Umar, M., Sahibzada, K., Hossain, J., & Koirala, N. (2021). *Journal of Infection and Public Health Antibiotic resistance in microbes : History , mechanisms , therapeutic strategies and future prospects*. 14, 1750–1766. <https://doi.org/10.1016/j.jiph.2021.10.020>
- Mattoo, S., & Cherry, J. D. (2005). Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. In *Clinical Microbiology Reviews* (Vol. 18, Issue 2, pp. 326–382). <https://doi.org/10.1128/CMR.18.2.326-382.2005>
- Mering, C. Von, Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., & Snel, B. (2003). *STRING : a database of predicted functional associations between proteins*. 31(1), 258–261. <https://doi.org/10.1093/nar/gkg034>
- Mills, K. H. G., & Th, T. (2001). *Immunity to Bordetella pertussis*. [https://doi.org/10.1016/S1286-4579\(01\)01421-6](https://doi.org/10.1016/S1286-4579(01)01421-6)
- Miyafusa, T., Caaveiro, J. M. M., Tanaka, Y., Tanner, M. E., & Tsumoto, K. (2013). *Crystal structure of the capsular polysaccharide synthesizing protein CapE of Staphylococcus aureus*. <https://doi.org/10.1042/BSR20130017>
- Modeling, L. S. (n.d.). *Discovery Studio Life Science Modeling and Simulations*.

<https://doi.org/10.2174/1570180819666220330122542>

- Mooi, F. R. (2010). *Infection , Genetics and Evolution Bordetella pertussis and vaccination : The persistence of a genetically monomorphic pathogen*. 10, 36–49.
<https://doi.org/10.1016/j.meegid.2009.10.007>
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). *KAAS : an automatic genome annotation and pathway reconstruction server*. 35, 182–185.
<https://doi.org/10.1093/nar/gkm321>
- Murugan, N. A., Kumar, S., Jeyakanthan, J., & Srivastava, V. (2020). Searching for target - specific and multi - targeting organics for Covid - 19 in the Drugbank database with a double scoring approach. *Scientific Reports*, 1–16. <https://doi.org/10.1038/s41598-020-75762-7>
- Nieves, D. J., & Heininger, U. (2016). *Bordetella pertussis*.
<https://doi.org/10.1128/9781555819453.ch17>
- Okuda, S., Yamada, T., Hamajima, M., & Itoh, M. (2008). *KEGG Atlas mapping for global analysis of metabolic pathways*. 36(May), 423–426. <https://doi.org/10.1093/nar/gkn282>
- Onodera, K., Satou, K., & Hirota, H. (2007). *Evaluations of Molecular Docking Programs for Virtual Screening*. 1609–1618. <https://doi.org/10.1021/ci7000378>
- Othman, D. I. A., Hamdi, A., Abdel-aziz, M. M., & Elfeky, S. M. (2022). Bioorganic Chemistry Novel 2-arylthiazolidin-4-one-thiazole hybrids with potent activity against Mycobacterium tuberculosis. *Bioorganic Chemistry*, 124(December 2021), 105809.
<https://doi.org/10.1016/j.bioorg.2022.105809>
- Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking : a review. *Biophysical Reviews*, 91–102. <https://doi.org/10.1007/s12551-016-0247-1>
- Peer-reviewed, N. O. T. (2020). © 2020 by the author(s). Distributed under a Creative Commons CC BY license. February. <https://doi.org/10.20944/preprints202002.0071.v1>

- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). *UCSF Chimera — A Visualization System for Exploratory Research and Analysis*. <https://doi.org/10.1002/jcc.20084>
- Phys, J. C. (2021). *Connecting the quantum and classical mechanics simulation world : Applications of reactive step molecular dynamics simulations* *Connecting the quantum and classical mechanics simulation world : Applications of reactive step molecular dynamics simulations*. 194105. <https://doi.org/10.1063/5.0048618>
- Pourseif, M. M., Yousefpour, M., Aminianfar, M., & Moghaddam, G. (2019). A multi-method and structure-based *in silico* vaccine designing against *Echinococcus granulosus* through investigating enolase protein. *Tabriz University of Medical Sciences*, 9(3), 131–144. <https://doi.org/10.15171/bi.2019.18>
- Rahman, N., Ajmal, A., Ali, F., & Rastrelli, L. (2020). Genomics Core proteome mediated therapeutic target mining and multi-epitope vaccine design for *Helicobacter pylori*. *Genomics*, 112(5), 3473–3483. <https://doi.org/10.1016/j.ygeno.2020.06.026>
- Read, R. J., Adams, P. D., Arendall, W. B., Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lu, T., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., Vriend, G., & Zwart, P. H. (2011). *Ways & Means A New Generation of Crystallographic Validation Tools for the Protein Data Bank*. 1395–1412. <https://doi.org/10.1016/j.str.2011.08.006>
- Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Ballard, A. J., Cowie, A., Romera-paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., ... Kavukcuoglu, K. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(May). <https://doi.org/10.1038/s41586-021-03819-2>
- Rosini, R., Nicchi, S., Pizza, M., & Rappuoli, R. (2020). *Vaccines Against Antimicrobial Resistance*. 11(June), 1–14. <https://doi.org/10.3389/fimmu.2020.01048>
- Ruff, K. M., & Pappu, R. V. (2021). AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of Molecular Biology*, 433(20), 167208.
-

<https://doi.org/10.1016/j.jmb.2021.167208>

- Sachdeva, S., Palur, R. V., & Sudhakar, K. U. (2017). *E . coli Group 1 Capsular Polysaccharide Exportation Nanomachinery as a Plausible Antivirulence Target in the Perspective of Emerging Antimicrobial Resistance*. 8(January), 1–19.
<https://doi.org/10.3389/fmicb.2017.00070>
- Sadeghi, M., Moradi, M., Madanchi, H., & Johari, B. (2021). *In silico study of garlic (Allium sativum L .)- derived compounds molecular interactions with α - glucosidase*. *In Silico Pharmacology*, 7–14. <https://doi.org/10.1007/s40203-020-00072-9>
- Sarkar, M., & Id, S. S. (2020). *Structural insight into the role of novel SARS- CoV-2 E protein : A potential target for vaccine development and other therapeutic strategies*. 1–25. <https://doi.org/10.1371/journal.pone.0237300>
- Shahid, F., Ashfaq, U. A., Saeed, S., Munir, S., & Almatroudi, A. (n.d.). *In Silico Subtractive Proteomics Approach for Identification of Potential Drug Targets in Staphylococcus saprophyticus*. 1–10. <https://doi.org/10.3390/ijerph17103644>
- Sharma, A. K., Srivastava, G. N., Roy, A., & Sharma, V. K. (2017). *ToxiM : A Toxicity Prediction Tool for Small Molecules Developed Using Machine Learning and Chemoinformatics Approaches*. 8(November), 1–18.
<https://doi.org/10.3389/fphar.2017.00880>
- Shen, H., & Chou, K. (2010). Gneg-mPLoc : A top-down strategy to enhance the quality of predicting subcellular localization of Gram-negative bacterial proteins. *Journal of Theoretical Biology*, 264(2), 326–333. <https://doi.org/10.1016/j.jtbi.2010.01.018>
- Shi, T., Wang, L., Du, S., Fan, H., Yu, M., Ding, T., & Xu, X. (2021). Mortality risk factors among hospitalized children with severe pertussis. *BMC Infectious Diseases*, 1–10.
<https://doi.org/10.1186/s12879-021-06732-1>
- Singh, J. K., Adams, F. G., & Brown, M. H. (2019). *Diversity and Function of Capsular Polysaccharide in Acinetobacter baumannii*. 9(January), 1–8.
<https://doi.org/10.3389/fmicb.2018.03301>
-

- Siraj, F., Natarajan, S., Huq, A., & Kim, Y. J. (2015). Structural investigation of ginsenoside Rf with PPAR α major transcriptional factor of adipogenesis and its impact on adipocyte. *Journal of Ginseng Research*, 39(2), 141–147. <https://doi.org/10.1016/j.jgr.2014.10.002>
- Sun, D., Jones, L. H., Mathews, F. S., & Davidson, V. L. (2001). *Active-site residues are critical for the folding and stability of methylamine dehydrogenase*. 14(9), 675–681. <https://doi.org/10.1093/protein/14.9.675>
- Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguetz, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L. J., & Mering, C. Von. (2011). *The STRING database in 2011 : functional interaction networks of proteins , globally integrated and scored*. 39(November 2010), 561–568. <https://doi.org/10.1093/nar/gkq973>
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & Mering, C. Von. (2021). *The STRING database in 2021 : customizable protein – protein networks , and functional characterization of user-uploaded gene / measurement sets*. 49(November 2020), 605–612. <https://doi.org/10.1093/nar/gkaa1074>
- Telkar, S., Kumar, S., & Deeplanaik, N. (2013). *Strategic genome-scale prioritization of unique drug targets : A case study of Streptococcus gordonii*. 9(19). <https://doi: 10.6026/97320630009983>
- Tian, H., Jiang, X., & Tao, P. (2021). *PASSer : prediction of allosteric sites server OPEN ACCESS*. <https:// doi 10.1088/2632-2153/abe6d6>
- Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). *CASTp 3 . 0 : computed atlas of surface topography of proteins*. 46(June), 363–367. <https://doi.org/10.1093/nar/gky473>
- Udp-hexose, S. S. I. N., Ishiyama, N., Creuzenet, C., Lam, J. S., & Berghuis, A. M. (2004). Crystal Structure of WbpP , a Genuine UDP- N -acetylglucosamine 4-Epimerase from Pseudomonas aeruginosa. *Journal of Biological Chemistry*, 279(21), 22635–22642. <https://doi.org/10.1074/jbc.M401642200>
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D.,

-
- Stroe, O., Wood, G., Green, T., Tunyasuvunakool, K., Petersen, S., Laydon, A., Augustin, Z., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., ... Velankar, S. (2022). *NAR Breakthrough Article AlphaFold Protein Structure Database : massively expanding the structural coverage of protein-sequence space with high-accuracy models*. 50(November 2021), 439–444. <https://doi.org/10.1093/nar/gkab1061>
- Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). *DoGSiteScorer : a web server for automatic binding site prediction , analysis and druggability assessment*. 28(15), 2074–2075. <https://doi.org/10.1093/bioinformatics/bts310>
- Wadood, A., Jamal, A., Riaz, M., Khan, A., Uddin, R., & Jelani, M. (2018). Microbial Pathogenesis Subtractive genome analysis for *in silico* identification and characterization of novel drug targets in Streptococcus pneumonia strain JJA. *Microbial Pathogenesis*, 115(December 2017), 194–198. <https://doi.org/10.1016/j.micpath.2017.12.063>
- Wen, Q., Liu, S., Dong, C., Guo, H., Gao, Y., & Guo, F. (2019). *Geptop 2 . 0 : An Updated , More Precise , and Faster Geptop Server for Identification of Prokaryotic Essential Genes*. 10(June), 1–6. <https://doi.org/10.3389/fmicb.2019.01236>
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., ... Wilson, M. (2018). *DrugBank 5 . 0 : a major update to the DrugBank database for 2018*. 46(November 2017), 1074–1082. <https://doi.org/10.1093/nar/gkx1037>
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., Woolsey, J., & Tg, C. (2006). *DrugBank : a comprehensive resource for in silico drug discovery and exploration*. 34, 668–672. <https://doi.org/10.1093/nar/gkj067>
- Woo, H., & Roux, B. (2005). Calculation of absolute protein – ligand binding free. *Proceedings of the National Academy of Sciences*, 102(19), 6825–6830. <https://doi.org/10.1073/pnas.0409005102>
- Woodward, L., & Naismith, J. H. (n.d.). ScienceDirect Bacterial polysaccharide synthesis and
-

- export. *Current Opinion in Structural Biology*, 40, 81–88.
<https://doi.org/10.1016/j.sbi.2016.07.016>
- Yu, C., Cheng, C., Su, W., Chang, K., & Huang, S. (2014). *CELLO2GO : A Web Server for Protein subCELLular LOCALization Prediction with Functional Gene Ontology Annotation*. 9(6). <https://doi.org/10.1371/journal.pone.0099368>
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). *PSORTb 3.0 : improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes*. 26(13), 1608–1615.
<https://doi.org/10.1093/bioinformatics/btq249>
- Yuriev, E., & Ramsland, P. A. (2013). *Latest developments in molecular docking : 2010 – 2011 in review*. October 2012, 215–239. <https://doi.org/10.1002/jmr.2266>
- Zhang, L., Xu, Y., Zhao, J., Kallonen, T., Cui, S., Xu, Y., Hou, Q., Li, F., Wang, J., He, Q., & Zhang, S. (2010). *Effect of Vaccination on Bordetella*. 16(11), 13–16.
<https://doi.org/10.3201/eid1611.100401>
- Zhang, R., Ou, H., & Zhang, C. (2004). *DEG : a database of essential genes*. 32, 271–272.
<https://doi.org/10.1093/nar/gkh024>
- Zou, Y., Hu, Y., Ge, S., Zheng, Y., Li, Y., Liu, W., Guo, W., Zhang, Y., Xu, Q., & Lai, Y. (2019). European Journal of Medicinal Chemistry Effective Virtual Screening Strategy toward heme-containing proteins : Identification of novel IDO1 inhibitors. *European Journal of Medicinal Chemistry*, 184, 111750.
<https://doi.org/10.1016/j.ejmech.2019.111750>

ORIGINALITY REPORT

13%

SIMILARITY INDEX

9%

INTERNET SOURCES

7%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

- 1** Submitted to Higher Education Commission Pakistan 2%
Student Paper
- 2** prr.hec.gov.pk 2%
Internet Source
- 3** Afifa Navid, Sajjad Ahmad, Rida Sajjad, Saad Raza, Syed Sikander Azam. " Structure Based in Silico Screening Revealed a Potent Ftsz Inhibitor From Asinex Antibacterial Library ", IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2022 1%
Publication
- 4** link.springer.com <1%
Internet Source
- 5** Rimsha Yousaf, Afifa Navid, Syed Sikander Azam. "Discovery of novel Glutaminase allosteric inhibitors through drug repurposing and comparative MMGB/PBSA and molecular dynamics simulation", Computers in Biology and Medicine, 2022 <1%
Publication
- 6** Saba Ismail, Noorah Alsowayeh, Hyder Wajid Abbasi, Aqel Albutti et al. "Pan-Genome-Assisted Computational Design of a Multi-Epitopes-Based Vaccine Candidate against Helicobacter cinaedi", International Journal of Environmental Research and Public Health, 2022 <1 85
Publication