SRI VENKATESWARA INTERNSHIP PROGRAM FOR RESEARCH IN ACADEMICS (SRI-VIPRA)

SRI-VIPRA

Project Report SVP-2429

"Identification of novel potential inhibitors against the PPE/PE proteins of *Mycobacterium tuberculosis* using *in silico* approach"

IQAC

Sri Venkateswara College

University of Delhi

Benito Juarez Road, Dhaula Kuan, New Delhi

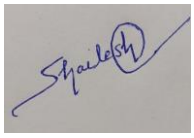
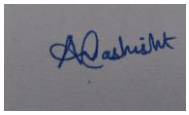
New Delhi -110021

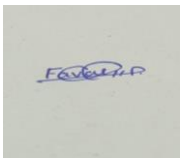
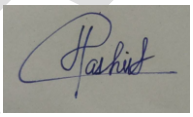
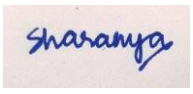
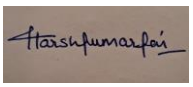
SRIVIPRA PROJECT 2024

Title: ***Identification of novel potential inhibitors against the PPE/PE proteins of Mycobacterium tuberculosis using in silico approach***

Name of Mentor: Dr Nimisha Sinha **Name of Department: Department of Biochemistry** **Designation: Assistant Professor**	
Name of Mentor: Dr Vandana Malhotra **Name of Department: Department of Biochemistry** **Designation: Assistant Professor**	

List of students under the SRIVIPRA Project

S.No	Photo	Name of the student	Roll number	Course	Signature
1		ANKITA PATEL	1222017	B.Sc (H) Biochemistry Semester V	
2		SHAILESH VERMA	1222046	Bsc. (H) Biochemistry Semester V	
3		VIVEK BHARDWAJ	1222038	B.Sc (H) Biochemistry Semester V	
4		AARAV VASHISHT	1223017	B.Sc (H) Biochemistry Semester III	
5		AASTHA GUPTA	1223036	B.Sc (H) Biochemistry Semester III	

6		MUHAMMED FAVAS AK	1223040	B.Sc (H) Biochemistry Semester III	
7		HARSHIT CHAUDHARY	1223064	B.Sc.(H) Biochemistry Semester III	
8		SHARANYA PATNAIK	1223059	B.Sc (H) Biochemistry Semester III	
9		HARSH RAI	1322026	B.Sc (H) Biological sciences Semester- V	
10		SUSMITA VERMA	1322034	B.Sc (H) Biological sciences Semester- V	

Certificate of Originality

This is to certify that the aforementioned students from Sri Venkateswara College have participated in the summer project SVP2429 titled ***"Identification of novel potential inhibitors against the PPE/PE proteins of Mycobacterium tuberculosis using in silico approach"***. The participants have carried out the research project work under my guidance and supervision from 1st July, 2024 to 30th September 2024. The work carried out is original and carried out in an online/offline/hybrid mode.

Signature of Mentor **Signature of Mentor**

Acknowledgement

We express our heartfelt gratitude to Almighty God for blessing us with good health throughout this journey. Our deepest thanks go to our families, whose unwavering support and unconditional love have been our constant source of strength.

We extend our sincere appreciation to our teachers and mentors, Dr. Nimisha Sinha and Dr. Vandana Malhotra, for their invaluable guidance, insightful suggestions, and innovative ideas throughout the course of our work. Their encouragement and open communication greatly fueled our curiosity and deepened our engagement with the project. We are profoundly thankful for the time and assistance she generously provided. It has been a privilege to work under their mentorship in accomplishing this endeavor.

We are also deeply thankful to everyone who contributed to the project, offering their assistance and correcting our mistakes along the way. Their support was crucial in completing the project within the designated time frame.

Finally, we would like to thank our college for offering us the necessary resources, knowledge, and the unforgettable experience of working on this project.

TABLE OF CONTENTS

Abstract

Tuberculosis (TB), caused by *Mycobacterium tuberculosis,* is a global health threat, with drug-resistant strains complicating treatment efforts. Recent studies have highlighted the potential of PE/PPE family proteins, such as PPE18 and PPE25, as novel drug targets due to their roles in virulence and immune modulation. These PPE proteins have been reported to contribute to M.tb pathogenicity by modulating he host immune responses and facilitating persistence in the host. This study aims to identify natural ligands derived from marine organisms and plants sources as potential inhibitors of PPE18 and PPE25 using bioinformatics-based molecular docking.

In this study, 3D structures of PPE18 and PPE25 were obtained from the Protein Data Bank (PDB) and prepared for molecular docking using AutoDock 4.2. A comprehensive library of natural compounds was created, through extensive review of literature, containing marine and plant-derived bioactive compounds known for their antimicrobial properties. The top-ranked ligands were selected based on binding energy scores and then key interactions were assessed, including hydrogen bonding and hydrophobic interactions with critical residues of receptor proteins.

The docking results revealed a few promising ligands with strong binding affinities for PPE18 and PPE25. Some of the ligands showed stable interactions within the critical sites/pockets of the receptor proteins, suggesting potential inhibitory effects on PPE protein function. This study demonstrates the potential of some of the natural compounds as novel inhibitors targeting PPE18 and PPE25, providing a new strategy for combating drug-resistant TB. Future experimental validation is required to confirm these findings and elucidate the mechanisms of action.

Keywords: *Mycobacterium tuberculosis*, PPE18, PPE25, natural ligands, molecular docking, drug resistance.

1. Introduction/Background

In 1882, Robert Koch discovered *Mycobacterium tuberculosis* (*M. tb*), as the causative agent of tuberculosis (TB). It is one of the world's leading causes of infectious disease-related death despite 90 years of vaccination and 60 years of chemotherapy (1). According to the 2022 World Health Organisation (WHO) report, around one quarter of the world's population (2 billion) are latently infected with *M. tb*. In the individuals carrying latent TB infections (LTBI), the estimated lifetime risk for TB reactivation is 5–10%. as shown in the figure below, in 2022, eight countries accounted for more than two thirds of global TB cases: India (27%), Indonesia (10%), China (7.1%), the Philippines (7.0%), Pakistan (5.7%), Nigeria (4.5%), Bangladesh (3.6%) and the Democratic Republic of the Congo (3.0%). Unfortunately, no effective vaccine is currently available to prevent TB disease in adults, either before or after exposure to *M. tb*. Nonetheless, the only licenced TB vaccine, Bacille Calmette-Guérin (BCG) can confer moderate protection in infants and children (2).

Also, the people suffering from HIV have an 18 times higher risk of developing TB than unaffected person. In the context of the coronavirus disease 2019 (COVID-19) pandemic, the impact of TB on global health has become even more severe. Individuals who have recovered from COVID-19 have been found to have a higher risk of developing TB, likely due to the negative impact of COVID-19 on the immune system. Therefore, new TB treatment drugs remain an urgent research priority while the ability of Mtb to survive in the microenvironment of the human host remains as one of the greatest challenges (3).

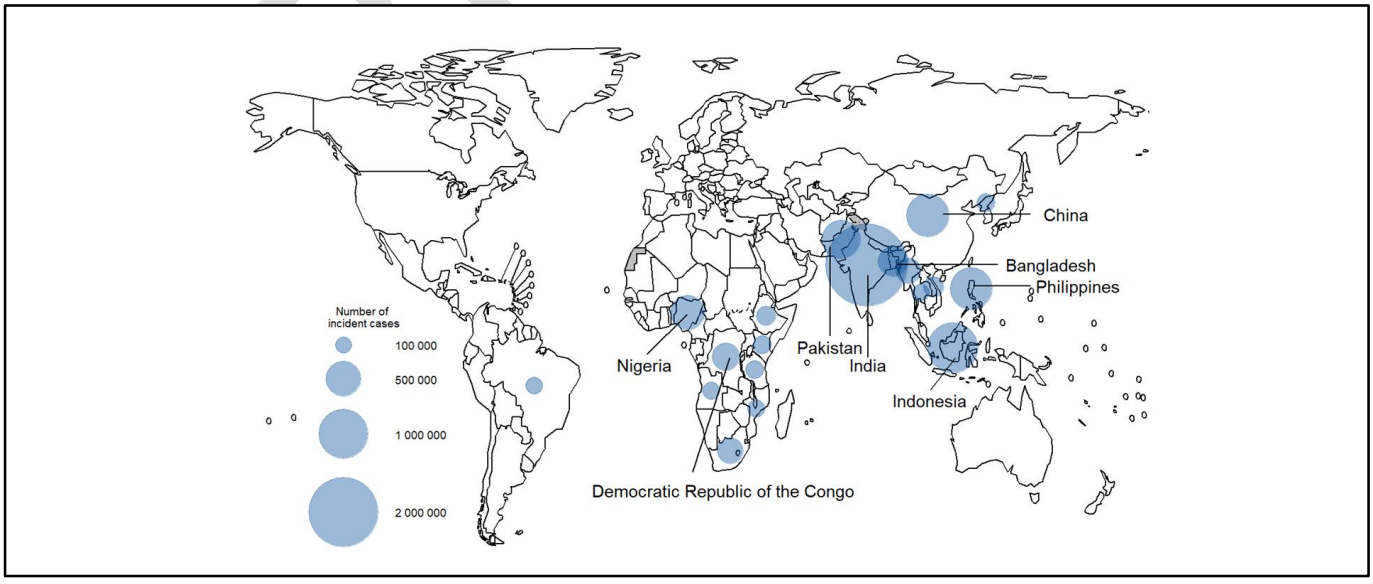


Figure-1: The prevalence of tuberculosis worldwide: Source: WHO report 2023

1.1 Life Cycle and Pathogenesis

The pathogenesis of tuberculosis is connected to the development and progress of the causative bacteria - *Mycobacterium tuberculosis*. The bacterium initiates its life cycle upon engagement with the respiratory tract and lungs, specifically interacting with alveolar macrophages.

Active tuberculosis patients transmit the infection via coughing/sneezing, leading to release of M.tb containing aerosol droplets, which travel through the trachea to bronchi, bronchioles and ultimately alveoli present in the lungs. Occurrence of the bacterium in the alveoli alerts the alveolar macrophages, which recognize it as a foreign particle and initiate phagocytosis by engulfing. Pathogen associated molecular patterns (PAMPS) on TB release Danger associated molecular patterns (DAMPS) that are recognized by Toll Like Receptors (TLRs) on alveolar macrophages and phagocytosis occurs. At this point, intracellular lysosomal enzymes are released in the vesicle containing bacteria, which may digest and destroy the bacilli, leading to the eradication of the disease (4).

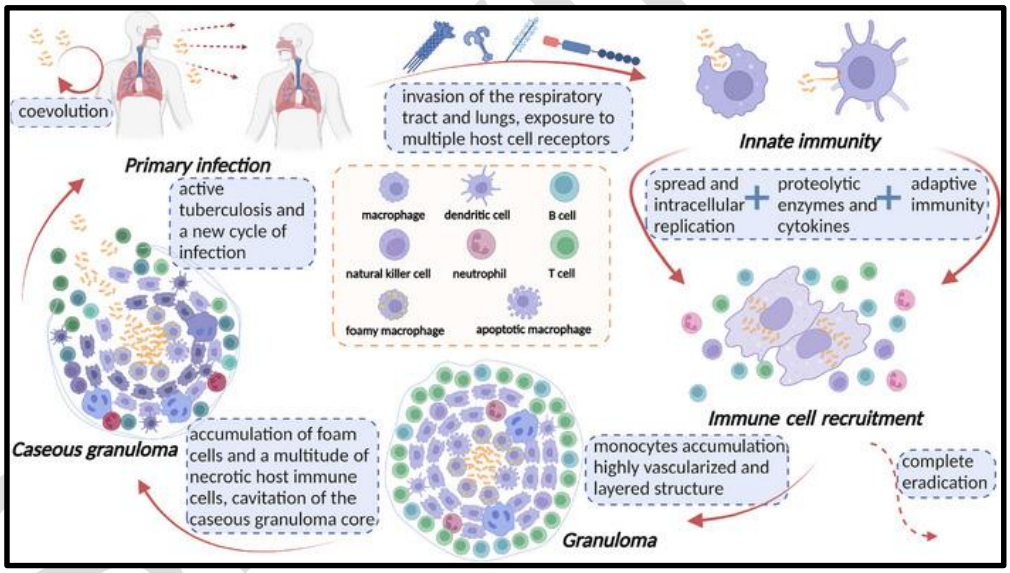


Figure-2: Schematic representation of life cycle of Mycobacterium in the host cell

However, in most cases the bacteria are able to evade this fate by action of complex mechanisms. Prolonged interaction with MTB causes release of inflammatory cytokines that activate and recruit neutrophils, T lymphocytes and monocytes to the site of infection. If the infected macrophages travel to the thoracic lymph nodes, they can cause adverse effects on the immune system and initiate macrophage cell death, which

would finally lead to the advancement of the bacilli through the lymphatic and circulatory system. This would result in the disease spreading to different regions of the body, and is labelled extra pulmonary tuberculosis.

The second fate of mycobacterium tuberculosis infected macrophages is the development of a structure called **granuloma** (an important feature of TB) (5). Macrophages, along with addition of T cells and B cells form a clustered covering in the vicinity of the bacilli, with the intent of containing the bacteria within itself, prohibiting its spread. Unfortunately, this also ensures a refuge for bacterial populations. As a result, the immune system doesn't recognize this danger and brings about the latent state of TB infection, which aids the survival of granulomas and patients are thus asymptomatic (6).

As the granuloma matures, macrophages differentiate into foamy cells. The center of granuloma may undergo necrosis of the host immune cells forming caseum (caseous granuloma) known as Ghon focus. Ghon focus with Hilar Lymph node forms Ghon Complex. This Ghon complex undergoes fibrosis and calcification to form Ranke complex and in some cases the *Mycobacterium tuberculosis* is killed but in many immunocompromised individual reactivation of TB takes place (7).

The immunocompromised individual is often seen suffering from HIV which gives rise to the granuloma cavitation and resulting in the conversion to active form of the disease. A delayed response by T cells triggers necrosis in the center of the granuloma, allowing it to become caseous in nature. This is characterized by the conversion of macrophages into foam cells, leading to their accumulation, along with the presence of necrotic host immune cells. Latent state TB involves the maturation of this state, inducing cavitation, and thereby permitting the release of mycobacterium tuberculosis back into the lung airways. This marks the transition from latent to active tuberculosis, that is symptomatic and infectious.

1.2 Signs and Symptoms

Mycobacterium tuberculosis (MTB) is a rod shaped bacterium, measuring approximately 3-4 micrometers in length and 0.3-0.6 micrometers in width. MTB is a slow growing bacterium, often taking several weeks for a single colony to grow on a culture plate.

Latent MTB infection remains asymptomatic, though it may produce positive results in diagnostic tests. Signs and symptoms of active MTB infection vary depending on the severity of the disease. Common symptoms include a persistent cough lasting for weeks, coughing up blood or phlegm (mucus), chest pain, fatigue, weight loss, fever, night sweats, swellings that persist for weeks, chills, loss of appetite, and shortness of breath.

Depending on where MTB is growing in the body, other symptoms may include swollen, firm, red, or purple lymph nodes under the skin (Lymph nodes), blood in the urine (Kidneys), headache or confusion (Brain), back pain (Spine), and hoarseness (Larynx) (8).

1.3 Current Treatment Regimen

This section reviews various currently essential drugs prescribed to the Drug Sensitive tuberculosis (DS-TB) patients classified according to the basis of their mode of action and their abilities to eradicate specific stages of *Mycobacterium tuberculosis (Mtb)* (7).

1.3.1 Cell Wall assembly in Mycobacterium

Mtb possesses an atypical cell wall structure composed majorly of simple lipids and carbohydrates. This envelope includes Peptidoglycan (PG), Arabinogalactan (AG), Lipoarabinomannan (LAM) and, Mycolic acid (MA) which forms a mycolyl-arabinogalactan peptidoglycan complex (mAGP) with some units of phosphatidyl myo-inositol based lipoglycans. Mycolic acid is one of the most indispensable parts of the cell wall as it envelopes the entirety of the cell and provides the bacterium permeability and ability to anchor it to a foreign object. Although hampering any component of the cell wall of the bacterium renders it vulnerable, interfering with MA and AG synthesis effectively makes it dysfunctional for pathogenesis.

1.3.2 First Line Drugs for Tuberculosis treatment

The first line drugs include those antibiotic drugs which are specifically designed to affect Mtb and its pathogenesis. The drugs include Isoniazid (INH), Rifampicin (RIF), Pyrazinamide (PZA) and Ethambutol (EMB) whose combination is given for the initial DS-TB patients.

- INH exerts an inhibitory effect on an enzyme InhA which is involved in the biosynthesis of MA. InhA is an enoyl acyl carrier protein (ACP) reductase which reduces long chain 2-enoyl acyl by forming covalent adducts with NAD cofactors and enoyl-CoA substrates. INH forms an adduct with NAD cofactors which indirectly inhibits InhA activity ultimately hampering MA biosynthesis. It is most effective against actively growing Mtb.

- RIF is a potent inhibitor of bacterial DNA-dependent RNA polymerase. RIF binds to the β subunit of the RNA pol II preventing the initiation factor (σ) from binding thereby arresting the process of transcription but RIF does not arrest DNA replication. RIF has a rapid bactericidal effect on the bacterial cells making it essential for TB treatment regimen for DS-TB patients.

- PZA is a prodrug which gets activated in the presence of an enzyme pyrazinamidase. This enzyme is found in the latent granuloma formed by Mtb infection and converts PZA into pyrazinoic acid which in turn interferes with fatty acid biosynthesis thereby affecting Mtb cell wall. PZA has a bactericidal effect against both actively growing as well as dormant Mtb cells.

- EMB inhibits the synthesis of arabinogalactan (AG) by targeting the arabinosyl transferases such as EmbA, EmbB and EmbC. EMB inhibits the growth of the actively proliferating Mtb cells but does not directly kill them, rendering a bacteriostatic effect (5, 9).

1.3.3 Current DS-TB Treatment and its consequences

For DS-TB (Drug Sensitive tuberculosis) patients, initially, the treatment lasts for 6 months with use of the first-line drugs (INH, RIF, PZA and EMB) only. All the first-line drugs for the initial two months and then continuation of INH and RIF for the next four months are prescribed for complete eradication of the dormant TB.

The treatment for DS-TB (as well as other forms of TB discussed later) has severe side effects associated with it which includes liver dysfunction, peripheral neuropathy, erythromelalgia, ocular toxicity, central nervous system (CNS) toxicity, gastrointestinal (GI) intolerance and skin rash.

1.4 Challenges faced against TB and the need for a novel approach

Poor patient compliance owing to the above mentioned unwanted side-effects, high pill count and protracted duration of therapy in addition to the overuse/misuse of antibiotics contributed to the emergence of drug resistant (DR) M.tb strains.

1.4.1 MDR-TB and emergence of second-line drugs in TB treatment

Multi Drug Resistant tuberculosis (MDR-TB) is referred to the drug resistant M.tb infection which is immune to the effect of either INH or RIF or both. MDR-TB cure rates are significantly lower as compared to DS-TB as it is resistant to two of the most powerful TB frontline drugs.

MDR-TB requires another auxiliary series of drugs that are referred to as second-line drug targets for its treatment. Second-line drugs are not specific for TB treatment but are antibiotics which support the first-line drugs in their treatment. Second-line drugs are classified into three groups namely GroupA (Linezolid, Bedaquiline, Moxifloxacin and Levofloxacin) GroupB (Clofazimine, Terizidone and Cycloserine) and Group C (Delamanid, Streptomycin, Amikacin, Imipenem, Meropenem, ANSA, Ethionamide and Prothionamide) drugs.

Two of the important second-line drugs include Bedaquiline (BDQ) and Fluoroquinolones (Moxifloxacin and Levofloxacin). BDQ is one of the drugs which functions on a novel mechanism discovered in the last half a century (other being Delamanid (DLM)). BDQ is a bactericidal antibiotic which inhibits the activity of ATP synthase thereby arresting the energy production from metabolism.

Fluoroquinolones is a class of antibiotics which contain a bicyclic structure. These inhibit DNA Gyrase, an enzyme responsible for separating DNA strands during replication (comparable to DNA topoisomerases in eukaryotes) arresting DNA replication in any prokaryotic cell (bacterial cell).

1.4.2 MDR-TB Treatment

The treatment for MDR-TB (after the initial period of six months for DS-TB) lasts for nearly 12 months. In general, the treatment starts with an initial phase of four-six months of administering INH (high doses), ETH/PTH, BDQ, one fluoroquinolone and clofazimine keeping in mind BDQ must be continued for six months regardless of the duration for other drugs. It is then continued with administering one fluoroquinolone, clofazimine, PZA and EMB for 5 months.

The treatment however, is not so simple as it is dependent on various factors such as (1) fluoroquinolone susceptibility (2) no previous history of second-line drug being administered for one month before (3) no other antibiotic resistance other than INH or RIF (4) no pregnancy (5) age 6 or above.

1.4.3 Co-Infection with other diseases and XDR

TB, in particular, is a disease known to remain in the pulmonary alveoli as a latent granuloma even after the treatment and the Mtb cells always try to find a way to become active whenever the immune response of the person is diverted or in a dysfunctional state. Hence, a co-infection with another severe disease is a perfect opportunity for Mtb cells to become active and perform its pathogenic processes which has become a highly potent challenge.

TB and HIV co-infection is near to impossible to treat as both the pathogens completely disarm the immune response. There is no proper treatment plan that can be devised for both the pathogens increasing the pill burden, overlapping side effects and drug-drug interactions.

The pandemic of the COVID-19 in 2020 also caused an increase in the number of deaths due to TB as their co-infection had catastrophic effects on any infected person. Since both the pathogens targeted the respiratory tract there was an increase in deaths from TB with 1.5 million in 2020 to 1.6 million in 2021. The pandemic affected the progress in controlling TB disease as all the efforts for the provision of preventive therapy and DR-TB treatment significantly declined.

Extensive Drug-Resistant Tuberculosis (XDR-TB) is a subset of MDR-TB where the Mtb infection has drug resistance towards at least one of the fluoroquinolone or any injected second line drug in addition to INH or RIF drug resistance. This has been a very severe challenge as the pill burden increases for a very limited option of drugs which may also cause resistance to these antibiotics too. There is neither a general treatment nor a plan to tackle XDR-TB. The only procedure that can be done is reducing or delaying the effects of the infection because such an infection cannot be eradicated (10, 11).

1.4.4 Need for a Novel Approach in drug discovery

The above mentioned challenges including MDR-TB, XDR-TB, various co-infections, etc has lead the research towards some novel findings in the *Mycobacterium tuberculosis* pathogenesis which includes binding to the alveolar sacs, evasion of immune system and other modes of absorption of nutrients and metal ions. Our research thus came across a family of proteins known as PE/PPE proteins which are potentially responsible for such effects and are present in different strains of Mycobacterium.

1.5. Introduction to PE/PPE family of proteins in Mycobacterium

Analysis of the *M. tuberculosis H37Rv* genome sequence revealed the presence of two novel gene families that comprise almost 10% of the coding capacity of the genome [12]. These were designated the PE and PPE genes, after highly conserved Proline-Glutamate and Proline-Proline-Glutamate residues near the start of their encoded proteins. The proteins can be categorized into subgroups, encompassing members with highly variable length and sequence features (Figure 2) [12].

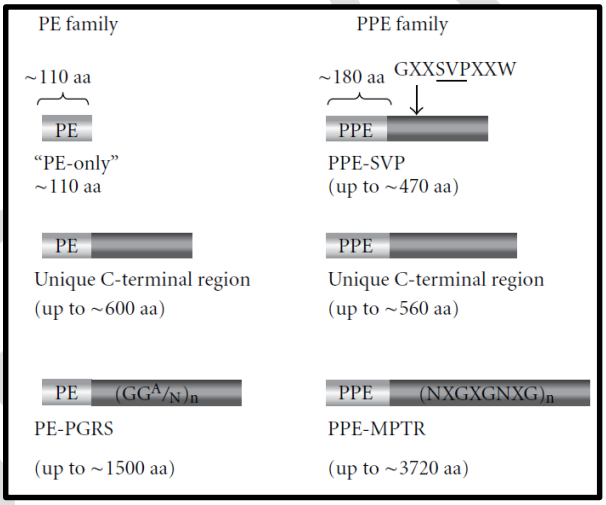


Figure-3: Schematic representation of PE and PPE family Sub-groups.

The relatively conserved N-terminal is approximately 110 amino acids (aa) and 180 aa in the PE and PPE families, respectively. Recent studies have indicated that the ESX system contributes to PE/PPE protein export, and, likewise, ESX system protein secretion is related to that of PE/PPE proteins.

1.5.1 Evolution of PE/PPE proteins

Obligate symbiotic bacteria and obligate intracellular pathogens, such as *Mycobacterium*, often undergo reductive evolution, eliminating genes that are non-essential for survival while duplicating those critical for persistence, especially within a host. However, the PE/PPE gene families within *Mycobacterium* have exhibited constructive evolution, expanding significantly in number. This gene expansion is particularly notable in *Mycobacterium tuberculosis*, which possesses the highest number of PE/PPE proteins, along with species such as *M. marinum*, *M. leprae*, and *M. avium*. These PE/PPE genes comprise around 7-10% of the total coding capacity in these species, indicating their importance in pathogenicity, immune evasion, and survival within host cells (13,14, 15).

The PE-PPE gene families are intricately associated with the ESAT-6 (ESX) gene clusters, which encode the type VII secretion system (T7SS) in *Mycobacterium* species (16). These ESX regions are believed to have originated from an ancestral plasmid found in fast-growing Mycobacteria, carrying virulence factors that facilitated interactions with host macrophages. This co-evolutionary relationship between PE-PPE genes and the ESX system suggests that both evolved in concert to enhance *Mycobacterium's* ability to survive within host cells, contributing to its pathogenicity. The ESX secretion system comprises five distinct ESX clusters: ESX-1, ESX-2, ESX-3, ESX-4, and ESX-5. Among these, ESX-1, ESX-3, and especially ESX-5 are critically involved in the secretion of PE/PPE proteins. ESX-5 is particularly notable as it is primarily responsible for exporting PE/PPE proteins into the host environment. This secretion process plays a vital role in immune modulation and the pathogen's ability to evade the host's immune system, thereby promoting survival within the host. ESX-1, while essential for virulence, and ESX-3, involved in iron acquisition, are less directly associated with PE/PPE secretion compared to ESX-5, which is uniquely linked to the expansion and functional diversification of the PE/PPE gene families.

The evolutionary expansion of PE/PPE genes is reflected in five distinct sublineages: PE_PGRS, PPE-PPW, PPE-SVP, and PPE-MPTR. Among these, the PE_PGRS and PPE-MPTR subfamilies are believed to have evolved from ESX-5, indicating a specialized evolutionary path tied to the secretion functions of this system. The ancestral ESX cluster, ESX-4, does not contain any PE/PPE genes, suggesting that these genes were integrated later, likely into ESX-1, and subsequently expanded via gene duplication.

The PE-PPE genes associated with ESX-1, specifically PE35 (Rv3872) and PPE68 (Rv3873), are considered the ancestral members of the PE-PPE family. These genes are thought to be the progenitors

from which the entire PE/PPE family evolved through multiple rounds of duplication and diversification. This expansion, driven by the selective pressures of host interaction and survival, highlights the complex co-evolution of the PE/PPE genes and the ESX secretion system, with ESX-5 being a major driver of the PE/PPE gene family's evolution and functional diversification within pathogenic Mycobacteria (15).

1.5.3 Significance of PE/PPE proteins family

The PE (proline-glutamate) and PPE (proline-proline-glutamate) protein family represents a cluster of proteins present in the cell wall of Mycobacteria, including human pathogen *Mycobacterium tuberculosis.* The genes for PE and PPE proteins, which accounts for about 10% of the coding sequence, are closely related to bacterial virulence [17]. The N-terminal sequence of this family is relatively conserved and C-terminal sequence is highly polymorphic and the variation in the C-terminal sequence might be the molecular basis of mutations of PE/PPE proteins. Most of the PE/PPE family proteins are localized in the cell wall and can inhibit macrophage apoptosis [18].

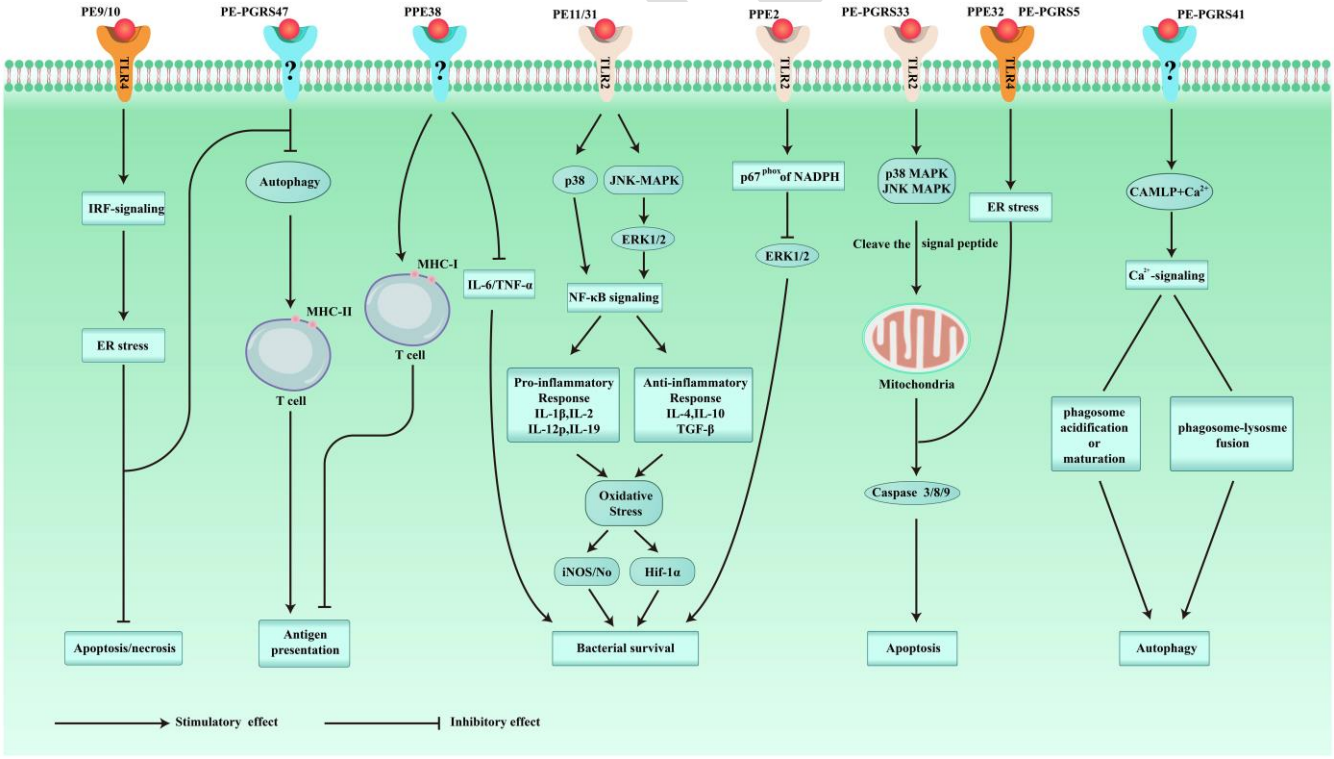


Figure-4: Schematic representation of some immunomodulatory roles played by PE/PPE proteins

PE/PPE family proteins regulate the immune function of host cells. Secretion of PE/PPE proteins dependent on the early secreted antigenic target 6 kDa (ESAT-6) secretion system (ESX). Mutations in PPE38 can

block the secretion of two major substrates of ESX-5, thereby increasing the virulence [19]. During the Mtb infection PE/PPE proteins can regulate various cell death pathways, such as apoptosis and pyroptosis. PE/PPE family proteins, integral to the pathogenicity of Mtb, present novel therapeutics against Mtb [20]. These proteins due to their significant representation in the Mtb genome and multiple roles in pathogenesis, particularly in mechanisms of immune invasion, offer a unique target for drug development. The conceptualization of small molecule inhibitors targeting distinct PPE/PE proteins holds significant promise in disrupting pathophysiological processes of Tuberculosis.

1.6 Selection of Ligands from natural sources: marine organisms and plant derived secondary metabolites

The significance of natural products in antibacterial drug treatment has been indisputable. Historically, natural products have been important in therapy against TB. Natural compounds form the basis for many commonly used medications. Bioactive molecules derived from widespread naturally occurring plant substances, including alkaloids, organosulfur compounds, phenolic acids, flavonoids, carotenoids, coumarins, terpenes, tannins, and some primary metabolites (amino acids, peptides, organic acids) have been reported to exhibit antimicrobial properties (21, 22). Flavonoids are secondary polyphenolic metabolites occurring commonly in many plants and fungi. Their effectiveness to treat tuberculosis was documented six decades back with the intravenous use of rutin to relieve pulmonary tuberculosis. According to the reported data, flavonoids and other phenolic compounds can disrupt specific mycobacterial mechanisms that are essential for the pathogen's survival (23). For instance, some of them impede mycolic acid synthesis, which aids in the formation of a highly impenetrable bacterial cell wall, limiting antibiotic effectiveness. Moreover, other flavonoids are reported to inhibit nucleic acid synthesis, energy metabolism, and reverse antibiotic resistance, which can improve the efficacy of currently available drugs (24).

Traditionally terrestrial microorganisms were explored as a source of biologically active natural products, however, natural products sourced from the marine environment are becoming increasingly important as a source of structurally novel and biologically active compounds. The oceans, with their unique aquatic environment and rich biodiversity, have proven to be a plentiful source of diverse natural products with significant antimicrobial, antiviral, antimalarial, antitumor, anti-inflammatory, and anti-oxidant activities [25, 26].

2. Objectives

2.1 Identification and characterization of PE/PPE proteins as potential drug targets in *Mycobacterium tuberculosis.*

2.2 Characterization of novel ligands derived from natural sources, plant based and marine organisms derived to be used as potential drug molecules.

2.3 Docking study of the target proteins with the selected ligands to identify potential drug molecules.

3. Materials/Methods

3.1. Selection of PE/PPE Proteins:

Through extensive literature search, selection of PPE proteins that play a vital role in virulence in *Mycobacterium tuberculosis H37RV* was done. It was found that PPE18, PPE25, PE-PGRS47 and PE35-PPE68 complex proteins are highly responsible for the virulence in the host organism. These proteins play crucial roles in the pathogenicity, immune evasion, and antigenic variation in bacteria (27, 28, 29).

3.2. Sequence of selected proteins:

To obtain the sequences of selected proteins i.e, PPE18, PPE25, PE-PGRS47 and PE35-PPE68, we used the **NCBI** (National Center for Biotechnology Information) database and derived the FASTA sequence of each protein.

3.3. Prediction of Protein Structure:

To obtain the structure of each selected protein, Protein Data Bank (PDB) database was used but respective X-ray crystallography structures were not available on PDB. So to predict the 3D structure we proceeded with online structure prediction tools such as Swiss-Modeller. **Swiss-Modeller** uses the principle of homology modeling where the query sequence is matched with a template sequence already present in the database and a model is built around that. We gave the FASTA format sequence of our PE and PPE proteins and a 3D structure was built by the software whose validation was done by studying the Ramachandran Plot of the given protein. A **Ramachandran plot** is a two-dimensional graphical representation that displays the allowed and disallowed regions of dihedral angles (ϕ and ψ) in a polypeptide backbone. It is a

valuable tool in structural biology for understanding protein folding, secondary structure, and overall conformation.

3.4. Comparison of virulent and non-virulent strains

We selected ten virulent and ten non-virulent Mycobacterium strains as shown in Table 2, to study and compare how the sequences of specific proteins (PPE18, PPE25, PE-PGRS47 and PE35-PPE68) are distributed across different strains. To analyze the sequence of these proteins in different virulent and non-virulent strains, we used Uniprot and Clustal Omega for sequence comparison and alignment. Through comparison and alignment analysis, it was found out that PPE18 and PPE25 were showing greater degree of similarity with each of the virulent strains while their sequences were absent in non- virulent strains leading to the conclusion that these were the major proteins responsible for virulence in Mycobacterium. Other PE-PGRS47 and PE35-PPE68 proteins were excluded from the study as they were showing lesser degree of similarity in virulent and non-virulent strains of Mycobacterium.

Table 1: List of Virulent and Non-Virulent Species for analysis

Virulent Species	**Non-Virulent Species**
M.tuberculosis	*M.smegmatis*
M. africanum	*M.indicus pranii*
M. canettii	*M. parafortuitum*
M. microti	*M.sphagni*
M.marinum	*M.elephantis*
M.leprae	*M.flavescens*
M.ulcereans	*M.gilvum*
M.bovis	*M.vanbaalenii*
M.orygis	*M.pulveris*
M.haemophilum	*M.phlei*
M.avium	*M.chobuense*
M.paratuberculosis	*M.obuense*
M.abscessus	*M.riyadhense*

3.5. Active Site Prediction

For screening of the selected proteins with several drug molecules, information about the active sites is an important parameter. The active sites of the models can be predicted by software like COACH and PROSITE. COACH generates complementary ligand binding site predictions using two comparative methods, TM-SITE and S-SITE, which recognize ligand-binding templates from the BioLiP protein function database by binding-specific substructure and sequence profile comparisons. Though we used blind docking for screening drug molecules, we still used COACH for active site prediction of our target proteins. We provided the FASTA sequence of the target protein and obtained the results from COACH in 24-48 hrs.

3.6. Selection of ligands

In order to select a competent ligand different secondary metabolites like flavonoids, alkaloids, plant derived molecules and marine compounds were studied. Through an extensive literature research a compounds library was created of the compounds which showed antagonistic effects on various strains of Mycobacterium. The structure of these compounds was found on Pubchem. For a ligand to be a suitable drug Lipinski's rule of 5 is the criteria which indicates 5 properties to be in a molecule to be a drug. For a compound to be a drug, it should have some basic properties. It should not have more than 5 H bond donors, it should not have more than 10 H bond acceptors, its molecular mass should be less than 500 Dalton, partition coefficient should not be greater than 5. This is called the Lipinski Rule of 5. This set of rules was studied using a software called Swiss ADME which is a free software and uses the SMILES formula to study a compound. Ligands not following Lipinski's rule were removed from the list. The structures of selected ligands were downloaded from PubChem and saved in .sdf format (Spatial Data File format). For drugs to work on Autodock .pdb format is required. The .sdf format was converted to .pdb format using software Open Babel. It is desirable to download the 3D conformers of the ligands from PubChem.

3.7. Screening of Ligand Molecules

For screening of ligand molecules Autodock4.2 was used as the docking tool. The receptor molecules (selected proteins) were read by autodock and were converted to **.pdbqt** file format by first removing water molecules. As our models were created by software so there was no water of crystallization but we did it as an exercise, then by adding polar hydrogens, then adding Kohlman Charges and finally assigning molecules AD-4 type. Same steps were taken to save ligands in .**pdbqt** format. While selecting the ligand,

the detect root function was done and then it was again saved in **.pdbqt** file format. The autogrid function was first run where the grid box was made around the whole protein for blind docking or a grid box was created around the specific amino acids in the active site making sure that it was at the center of the grid box. The grid box was first saved as a text file and then saved as .**gpf** file and then the autogrid.exe was run which generated the **.glg** file. After successful completion of the autogrid function then the autodock function was executed. The macromolecule and the ligand were selected for the autodock. The Genetic Algorithm was selected for performing the autodock which uses the Darwinian method of natural selection and a total of ten runs were performed. The file was saved as a **.dpf** file and the autodock command was executed and the .dlg file was generated. This **.dlg** file was analyzed and the binding energies were calculated to find the best ligand for a receptor.

4. Results

4.1 Sequence of proteins (in FASTA Format)

4.1.1 PPE18 (*Mycobacterium tuberculosis* H37Rv)

>CCP43952.1 PPE family protein PPE18 [Mycobacterium tuberculosis H37Rv]
MVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIGSSAGLMVA
AASPYVAWMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMILIATNLLGQNTPAIAVNE
AEYGEMWAQDAAAMGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPQ
ALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISNMVSMANNHMSMTNSGVSMTNTLSSMLKGFAPAAAA
QAVQTAAQNGVRAMSSLGSSLGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSL
TSAAERGPGQMLGGLPVGQMGARAGGGLSGVLRVPPRPYVMPHSPAAG

4.1.2 PPE25 (*Mycobacterium tuberculosis* H37Rv)

>CCP44553.1 PPE family protein PPE25 [Mycobacterium tuberculosis H37Rv]
MDFGALPPEINSGRMYCGPGSGPMLAAAAAWDGVAVELGLAATGYASVIAELTGAPWVGAASLSMVAA
ATPYVAWLSQAAARAEQAGMQAAAAAAAYEAAFVMTVPPPVITANRVLVMTLIATNFFGQNSAAIAVA
EAQYAEMWAQDAVAMYGYAAASASASRLIPFAAPPKTTNSAGVVAQVAAVAAMPGLLQRLSSAASVS
WSNPNDWWLVRLLGSITPTERTTIVRLLGQSYFATGMAQFFASIAQQLTFGPGGTTAGSGGAWYPTPQFA
GLGASRAVSASLARANKIGALSVPPSWVKTTALTESPVAHAVSANPTVGSSHGPHGLLRGLPLGSRITRRS
GAFAHRYGFRHSVVARPPSAG

4.1.3 PPE68 (*Mycobacterium tuberculosis* H37Rv)

>YP_178022.1 PPE family protein PPE68 [Mycobacterium tuberculosis H37Rv]

MLWHAMPPELNTARLMAGAGPAPMLAAAAGWQTLSAALDAQAVELTARLNSLGEAWTGGGSDKALA
AATPMVVWLQTASTQAKTRAMQATAQAAAYTQAMATTPSLPEIAANHITQAVLTATNFFGINTIPIALTE
MDYFIRMWNQAALAMEVYQAETAVNTLFEKLEPMASILDPGASQSTTNPIFGMPSPGSSTPVGQLPPAAT
QTLGQLGEMSGPMQQLTQPLQQVTSLFSQVGGTGGGNPADEEAAQMGLLGTSPLSNHPLAGGSGPSAGA
GLLRAESLPGAGGSLTRTPLMSQLIEKPVAPSVMPAAAAGSSATGGAAPVGAGAMGQGAQSGGSTRPGL
VAPAPLAQEREEDDEDDWDEEDDW

4.1.4 PE-PGRS47 (*Mycobacterium tuberculosis* H37Rv)

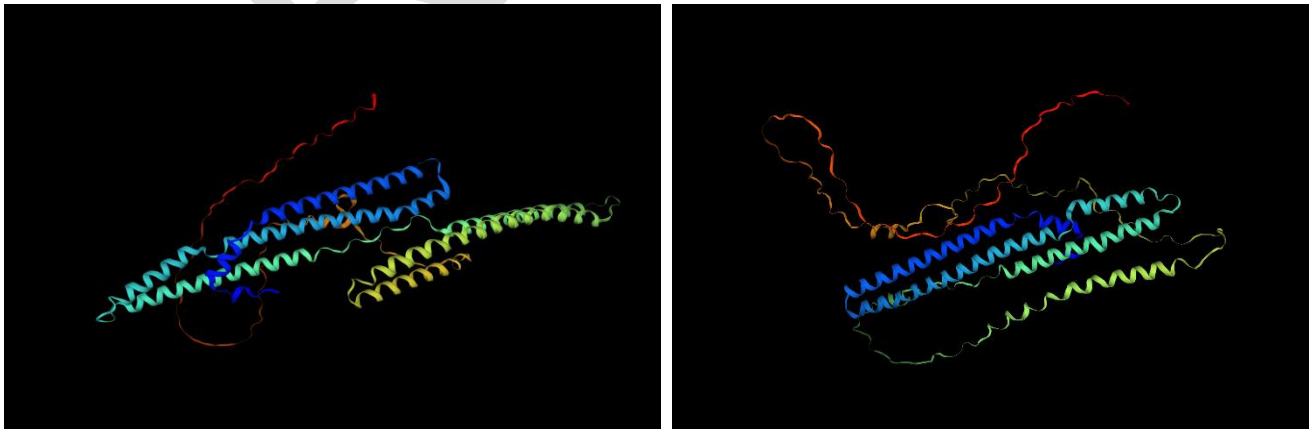
>CCP45540.1 PE-PGRS family protein PE_PGRS47 [Mycobacterium tuberculosis H37Rv]

MSFVIAAPEFLTAAAMDLASIGSTVSAASAAASAPTVAILAAGADEVSIAVAALFGMHGQAYQALSVQAS
AFHQQFVQALTAGAYSYASAEAAAVTPLQQLVDVINAPFRSALGRPLIGNGANGKPGTGQDGGAGGLLY
GSGGNGGSGLAGSGQKGGNGGAAGLFGNGGAGGAGASNQAGNGGAGGNGGAGGLIWGTAGTGGNGG
FTTFLDAAGGAGGAGGAGGLFGAGGAGGVGGAALGGGAQAAGGNGGAGGVGGLFGAGGAGGAGGFS
DTGGTGGAGGAGGLFGPGGGSGGVGGFGDTGGTGGDGGSGGLFGVGGAGGHGGFGSAAGGDGGAGGA
GGTVFGSGGAGGAGGVATVAGHGGHGGNAGLLYGTGGAGGAGGFGGFGGDGGDGGIGGLVGSGGAG
GSGGTGTLSGGRGGAGGNAGTFYGSGGAGGAGGESDNGDGGNGGVGGKAGLVGEGGNGGDGGATIAG
KGGSGGNGGNAWLTGQGGNGGNAAFGKAGTGSVGVGGAGGLLEGQNGENGLLPS

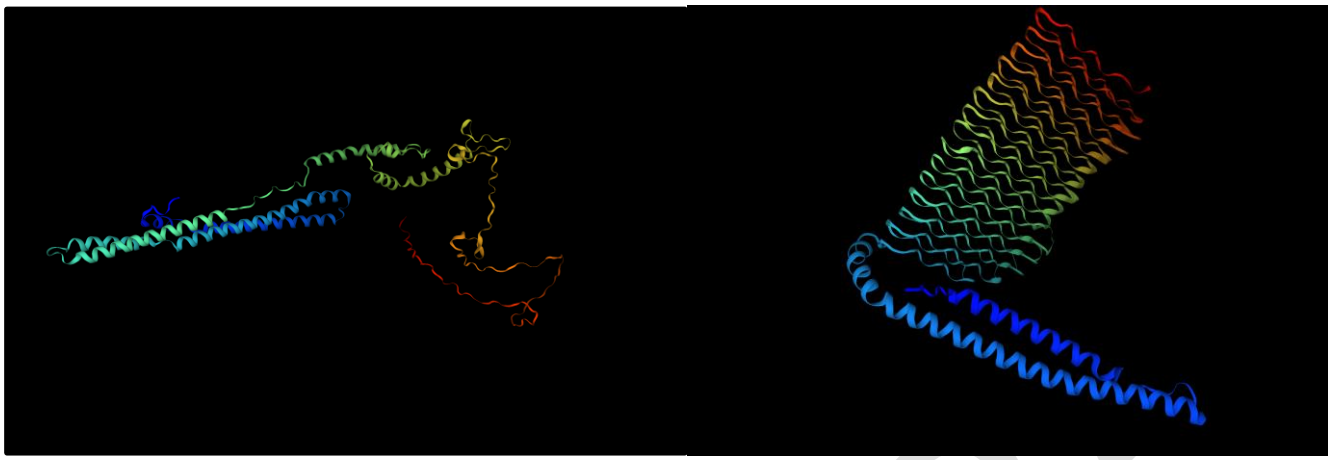
4.2 Protein Model

4.2.1 Building the Model

The models for the proteins were built by using Swiss-Model as shown in the figure 5.



A) PPE18 **B) PPE68**

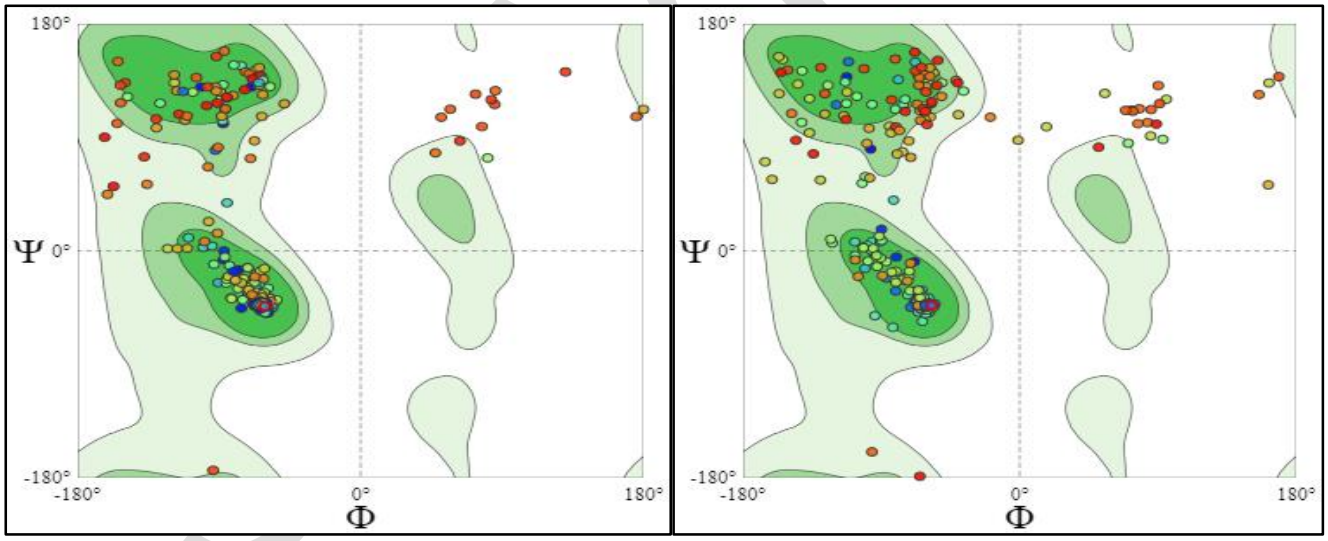


C) PPE25 **D) PE-PGRS47**

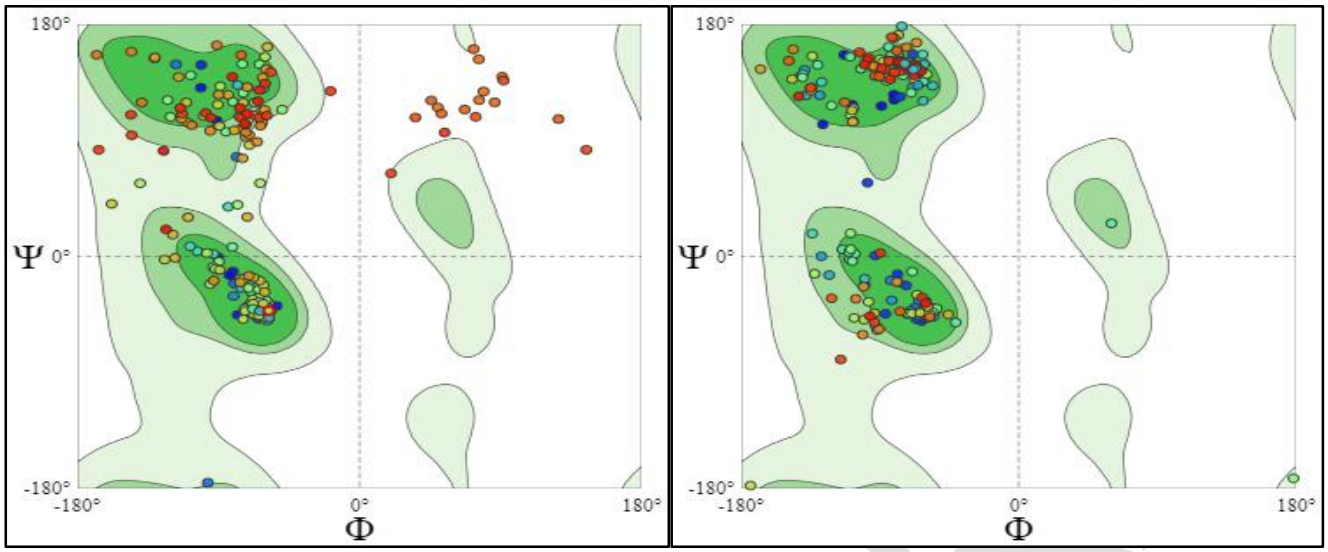
Figure-5: Proteins models obtained from Swiss-Modeller

4.2.2 Validation of models

The validation of PPE18, PPE25, PE-PGRS47 and PPE68 protein model described above was done using Ramachandran plot with the allowed and not allowed amino acids to have been plotted.



A) PPE18 **B) PPE68**



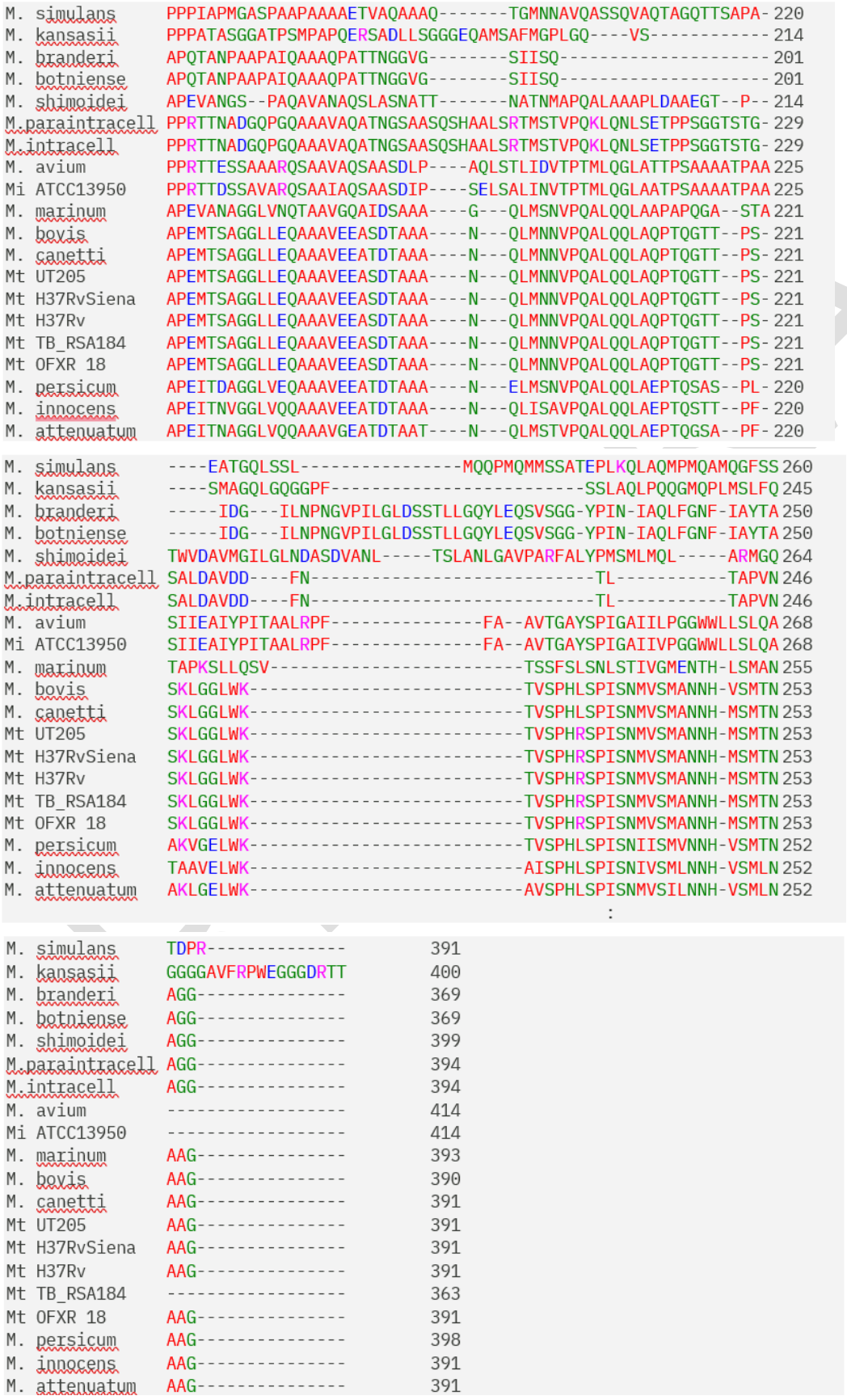
C) PPE25 **D) PE-PGRS47**

Figure-6: Ramachandran Plots for the modeled protein structures. The Ramachandran plot is the 2d plot of the φ-ψ torsion angles of the protein backbone. It provides a simple view of the conformation of a protein. The φ-ψ angles cluster into distinct regions in the Ramachandran plot where each region corresponds to a particular secondary structure.

The Ramachandran plot of PPE18 protein showed that almost 90% of the amino acid residues are situated in the favored areas and in PPE25 protein almost 85% amino acid residues are present in the allowed regions. In PPE68 many amino acids were falling in the not allowed region and in PE-PGRS47 which is a complex protein containing two different chains was difficult to dock. After the protein modeling and studying the Ramachandran plot two proteins PPE18 and PPE25 were selected for further docking studies. So, the docking exercise was done on two selected protein models with a list of ligands shown in section 4.4.

4.2.3 CLUSTAL OMEGA analysis of PPE proteins in virulent and non-virulent proteins

4.2.3.1 PPE 18


M. simulans PPPIAPMGASPAAPAAAAETVAQAAAQ-------TGMNNAVQASSQVAQTAGQTTSAPA- 220
M. kansasii PPPATASGGATPSMPAPQERSADLLSGGGEQAMSAFMGPLGQ----VS---------- 214
M. branderi APQTANPAAPAIQAAAQPATTNGGVG--------SIISQ---------- 201
M. botniense APQTANPAAPAIQAAAQPATTNGGVG--------SIISQ---------- 201
M. shimoidei APEVANGS--PAQAVANAQSLASNATT-------NATNMAPQALAAAPLDAAEGT--P-- 214
M.paraintracell PPRTTNADGQPGQAAAVAQATNGSAASQSHAALSRTMSTVPQKLQNLSETPPSGGTSTG- 229
M.intracell PPRTTNADGQPGQAAAVAQATNGSAASQSHAALSRTMSTVPQKLQNLSETPPSGGTSTG- 229
M. avium PPRTTESSAAARQSAAVAQSAASDLP----AQLSTLIDVTPTMLQGLATTPSAAAATPAA 225
Mi ATCC13950 PPRTTDSSAVARQSAAIAQSAASDIP----SELSALINVTPTMLQGLAATPSAAAATPAA 225
M. marinum APEVANAGGLVNQTAAVGQAIDSAAA----G---QLMSNVPQALQQLAAPAPQGA--STA 221
M. bovis APEMTSAGGLLEQAAAVEEASDTAAA----N---QLMNNVPQALQQLAQPTQGTT--PS- 221
M. canetti APEMTSAGGLLEQAAAVEEATDTAAA----N---QLMNNVPQALQQLAQPTQGTT--PS- 221
Mt UT205 APEMTSAGGLLEQAAAVEEASDTAAA----N---QLMNNVPQALQQLAQPTQGTT--PS- 221
Mt H37RvSiena APEMTSAGGLLEQAAAVEEASDTAAA----N---QLMNNVPQALQQLAQPTQGTT--PS- 221
Mt H37Rv APEMTSAGGLLEQAAAVEEASDTAAA----N---QLMNNVPQALQQLAQPTQGTT--PS- 221
Mt TB_RSA184 APEMTSAGGLLEQAAAVEEASDTAAA----N---QLMNNVPQALQQLAQPTQGTT--PS- 221
Mt OFXR 18 APEMTSAGGLLEQAAAVEEASDTAAA----N---QLMNNVPQALQQLAQPTQGTT--PS- 221
M. persicum APEITDAGGLVEQAAAVEEATDTAAA----N---ELMSNVPQALQQLAEPTQSAS--PL- 220
M. innocens APEITNVGGLVQQAAAVEEATDTAAA----N---QLISAVPQALQQLAEPTQSTT--PF- 220
M. attenuatum APEITNAGGLVQQAAAVGEATDTAAT----N---QLMSTVPQALQQLAEPTQGSA--PF- 220
M. simulans ----EATGQLSSL----------MQQPMQMMSSATEPLKQLAQMPMQAMQGFSS 260
M. kansasii ----SMAGQLGQGGPF----------SSLAQLPQQGMQPLMSLFQ 245
M. branderi -----IDG---ILNPNGVPILGLDSSTLLGQYLEQSVSGG-YPIN-IAQLFGNF-IAYTA 250
M. botniense -----IDG---ILNPNGVPILGLDSSTLLGQYLEQSVSGG-YPIN-IAQLFGNF-IAYTA 250
M. shimoidei TWVDAVMGILGLNDASDVANL-----TSLANLGAVPARFALYPMSMLMQL-----ARMGQ 264
M.paraintracell SALDAVDD----FN----------TL----------TAPVN 246
M.intracell SALDAVDD----FN----------TL----------TAPVN 246
M. avium SIIEAIYPITAALRPF----------FA--AVTGAYSPIGAIILPGGWWLLSLQA 268
Mi ATCC13950 SIIEAIYPITAALRPF----------FA--AVTGAYSPIGAIIVPGGWWLLSLQA 268
M. marinum TAPKSLLQSV----------TSSFSLSNLSTIVGMENTH-LSMAN 255
M. bovis SKLGGLWK----------TVSPHLSPISNMVSMANNH-VSMTN 253
M. canetti SKLGGLWK----------TVSPHLSPISNMVSMANNH-MSMTN 253
Mt UT205 SKLGGLWK----------TVSPHRSPISNMVSMANNH-MSMTN 253
Mt H37RvSiena SKLGGLWK----------TVSPHRSPISNMVSMANNH-MSMTN 253
Mt H37Rv SKLGGLWK----------TVSPHRSPISNMVSMANNH-MSMTN 253
Mt TB_RSA184 SKLGGLWK----------TVSPHRSPISNMVSMANNH-MSMTN 253
Mt OFXR 18 SKLGGLWK----------TVSPHRSPISNMVSMANNH-MSMTN 253
M. persicum AKVGELWK----------TVSPHLSPISNIISMVNNH-VSMTN 252
M. innocens TAAVELWK----------AISPHLSPISNIVSMLNNH-VSMLN 252
M. attenuatum AKLGELWK----------AVSPHLSPISNMVSILNNH-VSMLN 252
M. simulans TDPR---------- 391
M. kansasii GGGGAVFRPWEGGGDRTT 400
M. branderi AGG---------- 369
M. botniense AGG---------- 369
M. shimoidei AGG---------- 399
M.paraintracell AGG---------- 394
M.intracell AGG---------- 394
M. avium ---------- 414
Mi ATCC13950 ---------- 414
M. marinum AAG---------- 393
M. bovis AAG---------- 390
M. canetti AAG---------- 391
Mt UT205 AAG---------- 391
Mt H37RvSiena AAG---------- 391
Mt H37Rv AAG---------- 391
Mt TB_RSA184 ---------- 363
Mt OFXR 18 AAG---------- 391
M. persicum AAG---------- 398
M. innocens AAG---------- 391
M. attenuatum AAG---------- 391

4.2.2 PE-PGRS47

1.Mycobacterium Tuberculosis variant Africanum: KBF91755.1
2.Mycobacterium Canetti : CCC45099.1
3.Mycobacterium Tuberculosis H37Rv : AOE37152.1
4.Mycobacterium Tuberculosis variant Microti : AMC60453.1
5. Mycobacterium tuberculosis variant Bovis : QCU70083.1

```
KBF91755.1      ---------------MDLASIGSTVSAASAAASAPTVAILAAGADEVSIAVAALFGMYGQ    45
CCC45099.1      MSFVIAAPEFLTAAAMDLASIGSTVSAASAAASAPTVAILAAGADEVSIAVAALFGLHGQ    60
AOE37152.1      MSFVIAAPEFLTAAAMDLASIGSTVSAASAAASAPTVAILAAGADEVSIAVAALFGMHGQ    60
AMC60453.1      ---------------MDLASIGSTVSAASAAASAPTVAILAAGADEVSIAVAALFGMHGQ    45
QCU70083.1      ---------------MDLASIGSTVSAASAAASAPTVAILAAGADEVSIAVAALFGMHGQ    45
                              ****************************************::**

KBF91755.1      AYQALSVQASAFHQQFVQALTAGAYSYASAEAAAVTPLQQLVDVINAPFRSALGRPLIGN    105
CCC45099.1      AYQALSVQASAFHQQFVQALTAGAYSYASAEAAAVTPLQQLVDVINAPFRSALGRPLIGN    120
AOE37152.1      AYQALSVQASAFHQQFVQALTAGAYSYASAEAAAVTPLQQLVDVINAPFRSALGRPLIGN    120
AMC60453.1      AYQALSVQASAFHQQFVQALTAGAYSYASAEAAAVTPLQQLVDVINAPFRSALGRPLIGN    105
QCU70083.1      AYQALSVQASAFHQQFVQALTAGAYSYASAEAAAVTPLQQLVDVINAPFRSALGRPLIGN    105
                ************************************************************

KBF91755.1      GANGKPGTGQDGGAGGLLYGSGGNGGSGLAGSGQKGGNGGAAGLFGNGGAGGAGASNQAG    165
CCC45099.1      GANGKPGTGQDGGAGGLLYGSGGNGGSGLAGSGQKGGNGGAAGLFGNGGAGGAGASNQAG    180
AOE37152.1      GANGKPGTGQDGGAGGLLYGSGGNGGSGLAGSGQKGGNGGAAGLFGNGGAGGAGASNQAG    180
AMC60453.1      GANGKPGTGQDGGAGGLLYGSGGNGDQGWPAPARRAVTEELPDCLATAGPAVPARPTKPA    165
QCU70083.1      GANGKPGTGQDGGAGGLLYGSGGNGGSGLAGSGQKGGNGGAAGLFGNGGAGGAGASNQAG    165
                ************************..*  . .::. .    . :...* .  .  .: .

KBF91755.1      NGG-AGGNGGAGGLIWGTAGTGGNGGFTTFLDAAGGAGGAGGAGGLFGAGGAGGVGGAAL    224
CCC45099.1      NGG-AGGNGGAGGLIWGTAGTGGNGGFTTFLDAAGGAGGAGGAGGLFGAGGAGGVGGAAL    239
AOE37152.1      NGG-AGGNGGAGGLIWGTAGTGGNGGFTTFLDAAGGAGGAGGAGGLFGAGGAGGVGGAAL    239
AMC60453.1      TAAPAETAAPVG-------------------------------------------------    177
QCU70083.1      NGG-AGGNGGAGGLIWGTAGTGGNGGFTTFLDAAGGAGGAGGAGGLFGAGGAGGVGGAAL    224
```

```
                ... *   . .*

KBF91755.1      GGGAQAAGGNGGAGGVGGLFGAGGAGGAGGFGDTGGTGGAGGAGGLFGPGGGSGGVGGFG    284
CCC45099.1      GGGAQAAGGNGGAGGVGGLFGAGGAGGAGGFGDTGGTGGAGGAGGLFGPGGGSGGVGGFG    299
AOE37152.1      GGGAQAAGGNGGAGGVGGLFGAGGAGGAGGFGDTGGTGGAGGAGGLFGPGGGSGGVGGFG    299
AMC60453.1      ------------------------------------------------------------    177
QCU70083.1      GGGAQAAGGNGGAGGVGGLFGAGGAGGAGGFG--------------------------    256

KBF91755.1      DTGGTGGDGGSGGLFGVGGAGGHGGFGSAAGGDGGAGGAGGTVFGSGGAGGAGGVATVAG    344
CCC45099.1      DTGGTGGDGGSGGLFGVGGAGGHGGFGSAAGGDGGAGGAGGTVFGSGGAGGAGGVATVAG    359
AOE37152.1      DTGGTGGDGGSGGLFGVGGAGGHGGFGSAAGGDGGAGGAGGTVFGSGGAGGAGGVATVAG    359
AMC60453.1      ------------------------------------------------------------    177
QCU70083.1      DTGGTGGDGGSGGLFGVGGAGGHGGFGSAAGGDGGAGGAGGTVFGSGGAGGAGGVATVAG    316

KBF91755.1      HGGHGGNAGLLYGTGGAGGAGGFGGFGGDGGDGGIGGLVGSGGAGGSGGTGTLSGGRGGA    404
CCC45099.1      HGGHGGNAGLLYGTGGAGGAGGFGGFGGDGGDGGIGGMVGSGGAGGSGGTGTQSGGRGGA    419
AOE37152.1      HGGHGGNAGLLYGTGGAGGAGGFGGFGGDGGDGGIGGLVGSGGAGGSGGTGTLSGGRGGA    419
AMC60453.1      ------------------------------------------------------------    177
QCU70083.1      HGGHGGNAGLLYGTGGAGGAGGFGGFGGDGGDDGIGGLV--------------------    355

KBF91755.1      GGNAGTFYGSGGAGGAGGESDNGDGGNGGVGGKAGLVGEGGNGGDGGATIAGKGGSGGNG    464
CCC45099.1      GGNAGTFYGSGGAGGAGGESDNGDGGNGGVGGKAGLVGEGGNGGDGGATIAGKGGSGGNG    479
AOE37152.1      GGNAGTFYGSGGAGGAGGESDNGDGGNGGVGGKAGLVGEGGNGGDGGATIAGKGGSGGNG    479
AMC60453.1      ------------------------------------------------------------    177
QCU70083.1      ----------------------------------GEGGNGGDGGATIAGKGGSGGNG    378

KBF91755.1      GNAWLTGQGGNGGNAAFGKAGTGSVGVGGAGGLLEGQNGENGLLPS  510
CCC45099.1      GNAWLTGQGGNGGNAAFGKAGTGSVGVGGAGGLLEGQNGENGLLPS  525
AOE37152.1      GNAWLTGQGGNGGNAAFGKAGTGSVGVGGAGGLLEGQNGENGLLPS  525
AMC60453.1      ------------------------------------------------  177
QCU70083.1      GNAWLTGQGGNGGNAAFGKAGTGSVGVGGAGGLLEGQNGENGLLPS  424
```

4.2.3 PPE25

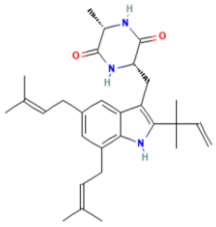
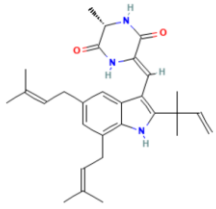
```
CLUSTAL O(1.2.4) multiple sequence alignment


SHOTTSII            MVVDTSPVDAGGYVSVPVLLSDFGSLSPEINSGRIYAGPGSGPVLAAAGAWDALATELGA      60
AVIUM               -------------------------------------------------------MAAAAAWDEVAAELGI     16
MARINUM             ------------------------------ MFDFGALPPEINSGRMYTGPGSGPMMAAASAWDALAAELSS     41
LACUS               ------------------------------MDFGALPPEVNSGRMYSGPGMGPMLAAAAAWDGLAAELGS     40
CANETTII            ------------------------------MDFGALPPEINSGRMYSGPGSGPMLAAAAAWDGVAVELGL     40
BOVIS               ------------------------------MDFGALPPEINSGRMYCGPGSGPMLAAAAAWDGVAVELGL     40
BOVIS BCG           ------------------------------MDFGALPPEINSGRMYCGPGSGPMLAAAAAWDGVAVELGL     40
PASTEUR
TUBERCULOSIS        ------------------------------MDFGALPPEINSGRMYCGPGSGPMLAAAAAWDGVAVELGL     40
H37RV
TUBERCULOSIS        ------------------------------MDFGALPPEINSGRMYCGPGSGPMLAAAAAWDGVAVELGL     40
CDC1551
KANSASII            ------------------------------MDFGALPPEINSGRMYAGPGSGSMMAAAAAWDDLAAELAS     40
SHINJUKUENSE        MPQ-AGS-LTRSRSAEEGRVLDFGALPPEVNSGRMYSGPGSGPMLAAAAAWDQLAAELAW             58
                                                                                :***.*** :*.**.

SHOTTSII            TACGYSCVITQLTSLPWSGPAAAAMLTAVTPYVSWLTTMATRAEQTAMQARAAAAAYEAA      120
AVIUM               AASGYHSVIAELTSGPWVGPASLSMVSAITPYVGWLSAVAAQAEETASQGRAAAAAFEAA       76
MARINUM             AASGYSSVIAELTSSPWLGPASRSMMSAVAPYVTWLSAAAAQTETAASQARAAAAAYETA      101
LACUS               AASGYASVISELSSSPWLGPASMAMVASATPYFTWLDSAAELAEQAGMQARAAAAAYEAA      100
CANETTII            AATGYASVIAELTGSPWVGAASLSMVAAATPYVAWLSQAAARAEQAGMQAAAAAAAYEAA      100
BOVIS               AATGYASVIAELTGAPWVGAASLSMVAAATPYVAWLSQAAARAEQAGMQAAAAAAAYEAA      100
BOVIS BCG           AATGYASVIAELTGAPWVGAASLSMVAAATPYVAWLSQAAARAEQAGMQAAAAAAAYEAA      100
PASTEUR
TUBERCULOSIS        AATGYASVIAELTGAPWVGAASLSMVAAATPYVAWLSQAAARAEQAGMQAAAAAAAYEAA      100
H37RV
TUBERCULOSIS        AATGYASVIAELTGAPWVGAASLSMVAAATPYVAWLSQAAARAEQAGMQAAAAAAAYEAA      100
CDC1551
KANSASII            AASGYESVIAELTGSPWMGPASLSMVAAARPYIAWLSGASALAEQAGMQGRAAAAAYETA      100
SHINJUKUENSE        AASAYESVIAELTSSWWVGPTAMSMVASATPYIAWLRTAATRAEQAGVQARAAAAAYETA      118
                    :* .* .**::*:.  * * :: :*:::  **. **   :  :* :. *. *****:*:*

SHOTTSII             FAMTVPPPAIAVNRVRLLGLIATNFFGQNTPSIAATEAEYAEFWAQDATAMYCYASTSTS      180
AVIUM                FAMTVPPPVIAANRVLLATLVATNFFGQNTPAIAATEAQYMEMWAQDAAAMYGYAAASQT      136
MARINUM              FVMTVPPPVIAANRVLLMTLVATNFFGQNTPAIAATEAQYMEMWAQDAAAMYSYAGSSAL      161
LACUS                FAMTVPPPVIAANRVLLMTLIATNFFGQNTPAIAATEAQYAEMWAQDAAAMYGYAGASAT      160
CANETTII             FVMTVPPPVITANRVLVMTLIATNFFGQNSAAIAVAEAQYAEMWAQDAVAMYGYAAASAS      160
BOVIS                FVMTVPPPVITANRVLVMTLIATNFFGQNSAAIAVAEAQYAEMWAQDAVAMYGYAAASAS      160
BOVIS BCG            FVMTVPPPVITANRVLVMTLIATNFFGQNSAAIAVAEAQYAEMWAQDAVAMYGYAAASAS      160
PASTEUR
TUBERCULOSIS         FVMTVPPPVITANRVLVMTLIATNFFGQNSAAIAVAEAQYAEMWAQDAVAMYGYAAASAS      160
H37RV
TUBERCULOSIS         FVMTVPPPVITANRVLVMTLIATNFFGQNSAAIAVAEAQYAEMWAQDAVAMYGYAAASAS      160
CDC1551
KANSASII             FVMTVPPPAIAANRLLLMTLIATNFFGQNTPAIAATEAHYAEMWAQDAAAMYGYAGASAT      160
SHINJUKUENSE         FAMTVPPPVIAANRVLLQTLCATNFFGQNTAAIAAAEAQYAEMWAQDAIAMYGYAGSSAG      178
                     *.******.*:.**: :   * ********: :**.:**.* *:***** *** **.:*

SHOTTSII            FLALPAFPGPPRTTDPAGLAGQAAATSQTAASATSSSA----------------------------IQALTQV      225
AVIUM                ASTLSPFAAPPNTTAPEGESDQAAAVAQAAAEPAGNSAQTAAQASSQLATSQAVTTGSQA      196
MARINUM              ATELPRFAPPPDTTSVNAAGGQSVVVAQAAATPAGDSAQTITATIPQLLSSAGVPGLQQL      221
LACUS                ASELTPFTEPAQATDATGLAAQAAAVSKAAATPGAASAQTVAAAAPELTIGSAVPGLLQQ      220
CANETTII            ASRLIPFAAPPKTTNSAGVVAQVAAV---------------------------------------- AAMPGLLQR      195
BOVIS               ASRLIPFAAPPKTTNSAGVVAQVAAV---------------------------------------- AAMPGLLQR      195
BOVIS BCG           ASRLIPFAAPPKTTNSAGVVAQVAAV---------------------------------------- AAMPGLLQR      195
PASTEUR
TUBERCULOSIS        ASRLIPFAAPPKTTNSAGVVAQVAAV---------------------------------------- AAMPGLLQR      195
H37RV
TUBERCULOSIS        ASRLIPFAAPPKTTNSAGVVAQVAAV---------------------------------------- AAMPGLLQR      195
CDC1551
KANSASII             ASELTPFADPGTTTDPAGLAAQSAAVAKAAATPAGTSAPALSSADPGLVSTAAVPQLLQQ      220
SHINJUKUENSE        AAQLTPFPKPAQTSNPAGVAAQAATVAKAAATAAGTSAPTVSSMDAELFATAAVPQLLRH      238
                       *  *  *  ::   .   * ...                                :    :

SHOTTSII            -SSS--------LAAAGHWL-----------------------------------PILSTAEWNTLVNTWGLA      254
AVIUM                VAQA--STTTASQPGPFTWL--TSALQNLQQSGLPTPTNNWLGLNPSFYTVMLKQNTGLA      252
MARINUM              SLSSVSSMSPSAASSPTVWQIIQTQVQNFLTYGLPTPDNNYAGLYPGMYNAL--RQTLQA      279
LACUS               LSTA----ASAAMLDLNNSQIAQLLG-------TITPAM----------------RTSLV--RTMGLS      259
```

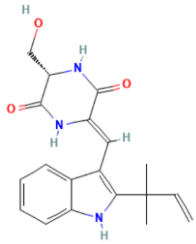
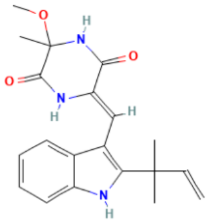
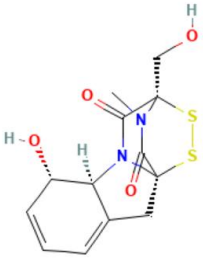
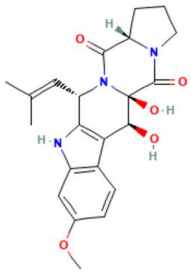
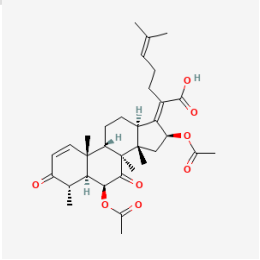
```
SHOTTSII          -SSS--------LAAAGHWL-------------------------------------PILSTAEWNTLVNTWGLA     254
AVIUM              VAQA--STTTASQPGPFTWL--TSALQNLQQSGLPTPTNNWLGLNPSFYTVMLKQNTGLA     252
MARINUM           SLSSVSSMSPSAASSPTVWQIIQTQVQNFLTYGLPTPDNNYAGLYPGMYNAL--RQTLQA     279
LACUS            LSTA----ASAAMLDLNNSQIAQLLG-------TITPAM---------------RTSLV--RTMGLS     259
CANETTII         LSSA----ASVSWPNPNDWWLVRLLG-------SITPTE---------------RTTIV--RLLGES     234
BOVIS            LSSA----ASVSWSNPNDWWLVRLLG-------SITPTE---------------RTTIV--RLLGQS     234
BOVIS BCG        LSSA----ASVSWSNPNDWWLVRLLG-------SITPTE---------------RTTIV--RLLGQS     234
PASTEUR
TUBERCULOSIS     LSSA----ASVSWSNPNDWWLVRLLG-------SITPTE---------------RTTIV--RLLGQS     234
H37RV
TUBERCULOSIS     LSSA----ASVSWSNPNDWWLVRLLG-------SITPTE---------------RTTIV--RLLGQS     234
CDC1551
KANSASII         LSTS----TLAGSLNPNDWWIVQTLG-------SINSAQ---------------RVSVV--RTLGLS     259
SHINJUKUENSE     LSTS----ALSAWPNPNDWWIVQLLG-------SLTPAN ---------- RTTIV--RFLGLS     277
                    :                                                          .    :

SHOTTSII         YFGAGIFQLGTLFAQQLLPDAAAAPSAAAASTAVAAQ---LAAP---------- VAATPVWAVFA     306

MARINUM          YFGVGIGNSGWSIAQQLTFGPGGTTAGAGGAWYPTPEFAALGADTW ---------- HWHPMASFA     334
LACUS            YFAVGIPQFFASIGQQLTFGPGGTTAGSGGAWYPTPQFAGLGLGGG---------- GPASASLA     313
CANETTII         YFATGMAQFFASIAQQLTFGPGGTTAGSGGAWYPTPQFAGLGAS---------- RAVSASLA     286
BOVIS            YFATGMAQFFASIAQQLTFGPGGTTAGSGGAWYPTPQFAGLGAS---------- RAVSASLA     286
BOVIS BCG        YFATGMAQFFASIAQQLTFGPGGTTAGSGGAWYPTPQFAGLGAS---------- RAVSASLA     286
PASTEUR
TUBERCULOSIS     YFATGMAQFFASIAQQLTFGPGGTTAGSGGAWYPTPQFAGLGAS---------- RAVSASLA     286
H37RV
TUBERCULOSIS     YFATGMAQFFASIAQQLTFGPGGTTAGSGGAWYPTPQFAGLGAS---------- RAVSVSLA     286
CDC1551
KANSASII         YFAMGIPQFIASIGQQLTFGSG-TTAGSGGAWYATPQFAGLGTGAG---------- RAAASASLA     313
SHINJUKUENSE     YFGMGIAQFFASIGQQLTFGPGGTTAGSGGAWYPTPQFVGLGLGLG ---------- GGKMSASLA     332
                 **. *: .    :.***  . . : :.:..:   : :   *.                .  *

SHOTTSII         RADTIGPMSVPPSWTTAAPAPTAGGAVISGIDVAA-----GRPQAPNALLPNLRDNS---     358
AVIUM            GAGRVGALSVPQQWATLTSAVSPAVSEEGSAVQAA----AG---GAPG---VRPTGCCAA     362
MARINUM          SSSKVGGLSVPASWGTAPGSVEQASTKLVSTNFTTSPGEANPANGLNAALRGFPVGSRGA     394
LACUS            SAVKVGPLSVPPTWIATETALEQRVARLVGANVAA---------AGSPNAANGVLNGLPMRGAGR     369
CANETTII         RANKIGALSVPPSWVKTTALTESPVAHAVSANPTV---------- GSSYGPHGLLRGLPLGSRIT     341
BOVIS            RANKIGALSVPPSWVKTTALTE-PGAHAVSANPTV--------- GSSHGPHGLLRGLPLGSRIT     340
BOVIS BCG        RANKIGALSVPPSWVKTTALTE-PGAHAVSANPTV--------- GSSHGPHGLLRGLPLGSRIT     340
PASTEUR
TUBERCULOSIS     RANKIGALSVPPSWVKTTALTESPVAHAVSANPTV---------- GSSHGPHGLLRGLPLGSRIT     341
H37RV
TUBERCULOSIS     RANKIGALSVPPSWVKTTALTESPVAHAVSANPTV---------- GSSHGPHGLLRGLPLGSRIT     341
CDC1551
KANSASII         SANKIGALSVPPTWVGSSAEVERAAAGLVSADVGT---------- NSQGGVQGLLRGIPLGSGAA     368
SHINJUKUENSE     SANTIGRLSVPPSWATSPVAAEHATPAVVGADVVT---------- TGPSGAAGLLRGMPLGSGLG     387
                  :  :* :***  *                 .    .          .  .

SHOTTSII          RATTFARRRYGIRLTVMLRPPEAG---------        382
AVIUM             C----------RWG-RWAGAARPRAMSTNTAFATAS        387
MARINUM           QRTGNLGVRYGFRYPVLTRPPSAG---------        418
```

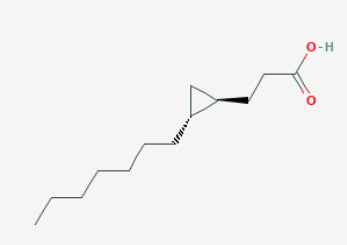
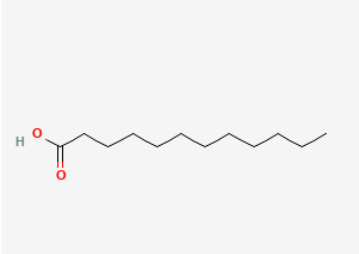
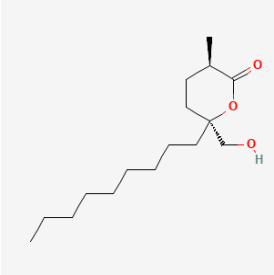
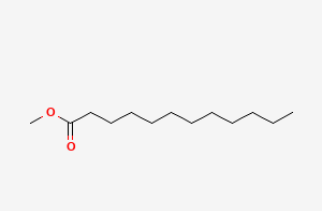
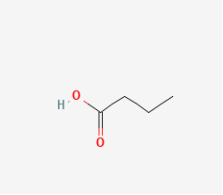
4.3 Ligand Library

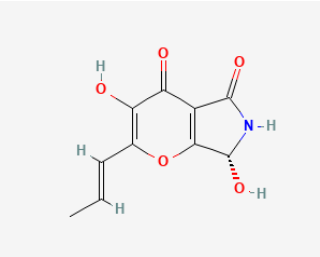
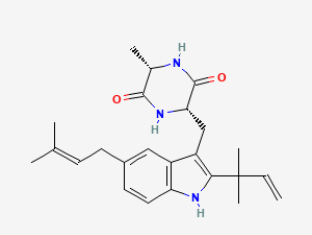
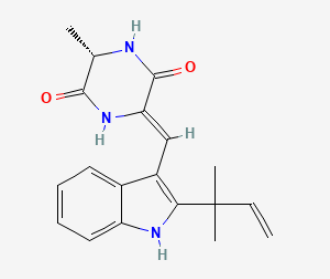
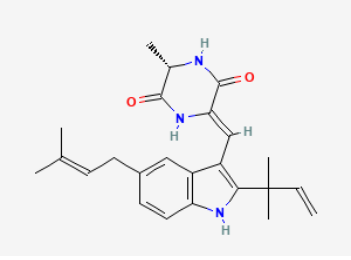
The table 2 below shows the ligands that were selected through literature review. These ligands are obtained from natural sources primarily derived from marine organisms and plant sources. Natural products play a pivotal role in contemporary drug development, particularly as antibacterial. Plant derived alkaloids and flavonoids have emerged as crucial sources of novel pharmacologically active compounds, directly or indirectly contributing to the development of numerous drugs. Similarly, the marine compounds, obtained from the marine environment that has vast species diversity, exhibit structural diversity at the level of secondary metabolites. It has been reported that marine natural products demonstrate a higher incidence of significant bioactivity and structural novelty when compared to their terrestrial counterparts. This emphasis on marine organisms as a source of novel compounds underscores the potential for discovering unique and biologically active molecules in the quest for new therapeutic agents.

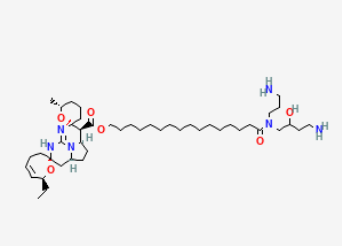
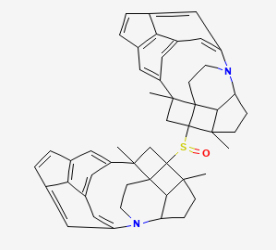
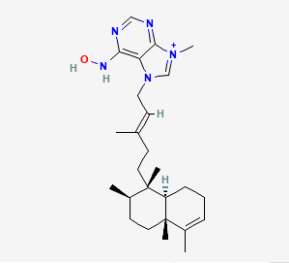
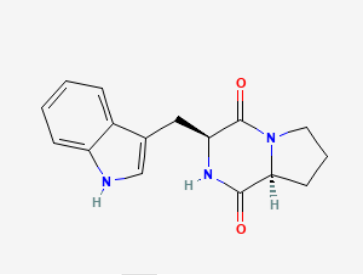
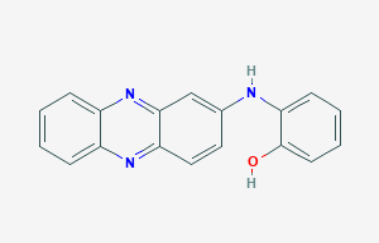
Table 2 List of ligands with molecular weight, structure and adherence to Lipinski rule

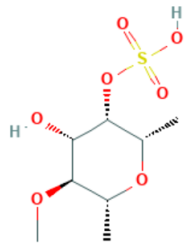
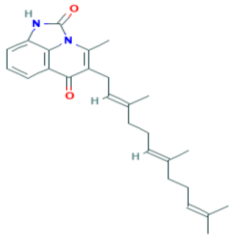
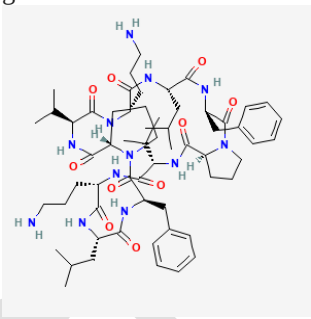
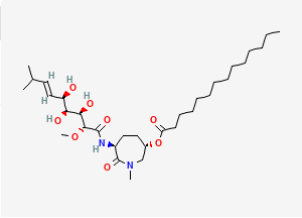
S.No.	Name of the ligand	Molecular weight (g/mol)	Lipinski Rule
1	Echinulin 	461.6	HBD= 3 HBA= 2 LogP= 4.6
2	Dehydroechinulin 	459.6	HBA= 2 HBD= 3 LogP= 4.61
3	Cristatumin A	339.4	HBA= 3 HBD= 4 LogP= 2.16

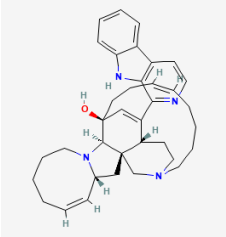
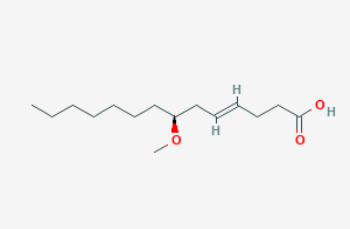
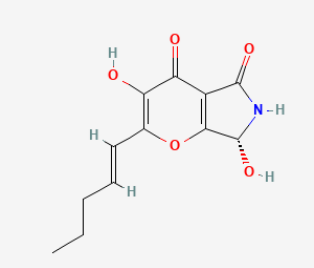
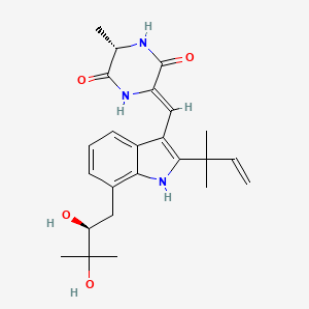
			
4	Variecolorin H 	353.4	HBA= 3 HBD= 3 LogP= 2.86
5	Gliotoxin 	326.39	HBA= 4 HBD= 2 LogP= 1.92
6	12,13-dihydroxyfumitremorgin C 	411.45	HBA= 5 HBD= 3 LogP= 2.61
7	Helvolic acid 	568.7	HBA= 8 HBD= 1 LogP=3.33
9	Lyngbyoic acid	212.33	HBA= 2 HBD= 1

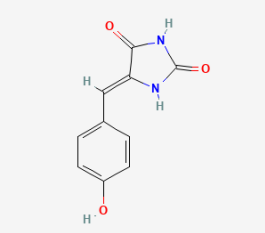
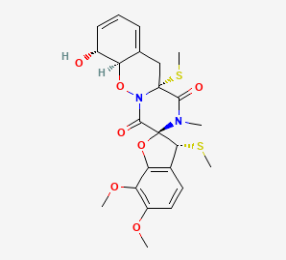
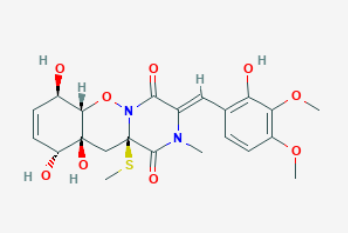
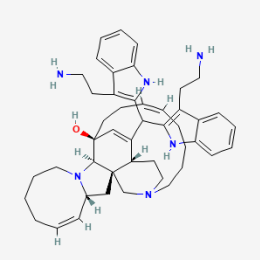
			LogP= 2.75
10	Dodecanoic acid 	200.32	HBA= 2 HBD= 1 LogP= 2.7
11	Malyngolide 	270.41	HBA= 3 HBD= 1 LogP= 3.59
12	Methyl ester of dodecanoic acid 	214.34	HBA= 2 HBD= 0 LogP= 3.48
13	Butyric acid 	88.11	HBA= 2 HBD= 1 LogP= 1.1

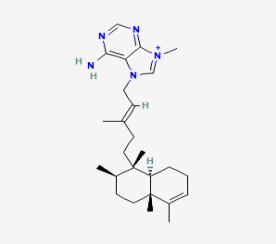
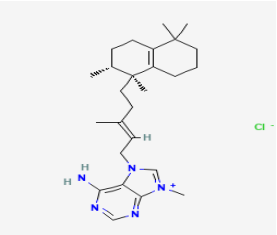
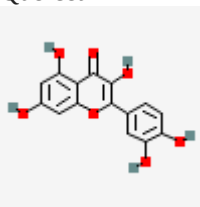
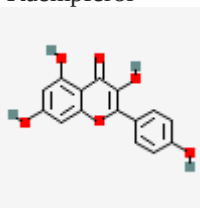
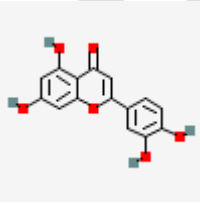
14	Geodisterol-3-O-sulfite	506.69	HBA= 6 HBD= 3 LogP= 3.11
15	Pyranonigrin A 	223.18	HBA= 5 HBD= 3 LogP= 1.07
16	Tardioxopiperazine A 	393.5	HBA= 2 HBD= 3 LogP= 3.25
17	Neoechinulin A 	323.4	HBA= 2 HBD= 3 LogP= 2.57
18	Isoechinulin A 	391.5	HBA= 2 HBD= 3 LogP= 3.66

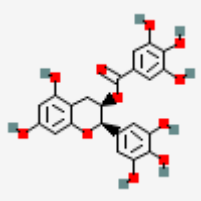
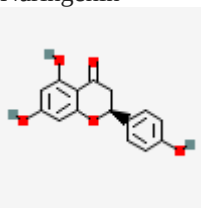
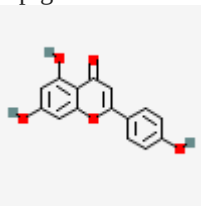
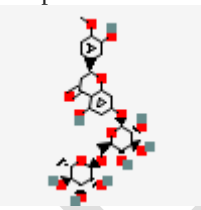
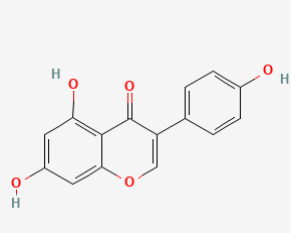
19	Crambescidin 800 	801.2	HBA= 9 HBD= 4 LogP= 7.62
20	Xinghaiamine A 	724.99	HBA= 1 HBD= 0 LogP= 4.93
21	Ageloxime B 	438.6	HBA= 3 HBD= 2 LogP= 0.8
22	Brevianamide F 	422.63	HBA= 2 HBD= 1 LogP= 0.69
23	*N*-(2-hydroxyphenyl)-2-phenazinamine 	283.33	HBA= 2 HBD= 2 LogP= 1.77

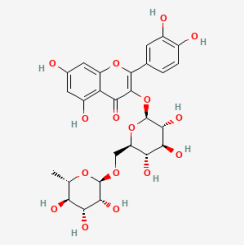
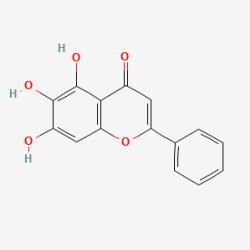
24	Axinellamine A and B	287.32	HBA= 3 HBD= 2 LogP= 2.64
25	Fucoidan 	256.27	HBA= 7 HBD= 2 LogP= 1.02
26	Aurachin 	404.54	HBA= 2 HBD= 1 LogP= 1.02
27	gramicidin S 	404.54	HBA= 2 HBD= 1 LogP= 1.02
28	Bengamide B 	598.8	HBA= 8 HBD= 4 LogP= 5.51

29	Manzamine A 	548.8	HBA= 4 HBD= 2 LogP= 4.62
30	Lyngbic acid 	256.38	HBA= 3 HBD= 1 LogP= 3.44
31	Pyranonigrin F 	251.23	HBA= 5 HBD= 3 LogP= 1.54
32	Rubrumazine B 	425.5	HBA= 4 HBD= 5 LogP= 2.92

33	(Z)-5-(4-hydroxybenzylidene)imidazolidine-2,4-dione 	204.18	HBA= 3 HBD= 3 LogP= 0.97
34	Peniciadametizine A 	506.6	HBA= 7 HBD= 1 LogP= 3.44
35	Peniciadametizine B 	494.5	HBA= 9 HBD= 4 LogP= 2.18
36	Zamamidine D 	713.01	HBA= 5 HBD= 5 LogP= 4.98

37	Agelasine B 	422.6	HBA= 2 HBD= 1 LogP= 0.83
38	Agelasine J 	458.1	HBA= 2 HBD= 1 LogP= -2.26
39	Quercetin 	302.1	HBA= 7 HBD= 5 LogP= 1.63
40	Kaempferol 	286.24	HBA= 6 HBD= 4 LogP= 1.7
41	Luteolin 	286.24	HBA= 6 HBD= 4 LogP= 1.86

42	Epigallocatechin Gallate (EGCG) 	458.37	HBA= 11 HBD= 8 LogP= 1.53
43	Naringenin 	272.25	HBA= 5 HBD= 3 LogP= 1.75
44	Apigenin 	270.24	HBA= 5 HBD= 3 LogP= 1.89
45	Hesperidin 	610.6	HBA= 15 HBD= 8 LogP= 0.89
46	Genistein 	270.24	HBA= 5 HBD= 3 LogP= 1.91
47	Rutin	610.52	HBA= 16 HBD= 10 LogP= 0.46

			
48	Baicalein 	270.24	HBA= 5 HBD= 3 LogP= 2.43

4.4 Docking Results

In this study, blind docking of several ligands was carried out and after successful docking. We obtained 10 best docked conformations in each docking and then two best conformers based on the minimum binding energy were selected for analysis of the kind of interactions shown between the residues of receptor protein and ligands (Table 3). The results showed significant binding of ligands with target proteins. These interactions were visualized using protein ligand interaction profiler. Protein ligand interactions in general is stabilized by different types of weak interactions but usually we consider hydrogen interactions as the most important interaction. After obtaining results from profiler, it was found that the ligands showed energetically favorable hydrogen and hydrophobic interactions with the amino acid residues present within the protein and not on the surface (Table 4). The best conformers were analyzed using Discovery Studio Visualizer.

Table 3 showing the two best conformers in each docking with the binding energy

S.No.	Receptor	Ligand	Run Number	Binding Energy (Kcal/mol)
1	PPE 25	Echinulin	9 7	-5.52 -5.47

2	PPE25	Dehydroechinulin	2 10	-5.45 -5.34
3	PPE25	Cristatumin A	5 3	-5.71 -5.58
4	PPE 25	Variecolorin H	3 5	-8.50 -8.43
5	PPE18	Atropine	9 5	-5.61 -4.92
6	PPE18	Aurachin	9 2	-3.52 -3.04
7	PPE25	Epigallocatechin Gallate	2 3	-8.21 -6.86
8	PPE18	Palmatine	10 9	-7.22 -7.20
9	PPE18	Sanguinarine	1 2	-7.03 -7.03
10	PPE25	Agelasine J	3 8	-5.70 -5.56
11	PPE25	Ageloxime B	4 10	-6.37 -5.91
12	PPE25	Apegenin	8 7	-5.82 -5.66
13	PPE25	Artemisinin	7 10	-5.40 -5.40
14	PPE25	Aurachin	1 7	-8.16 -5.83
15	PPE25	Bengamide	1 3	0.72 1.01
16	PPE25	Brevianamide F	10 9	-6.31 -6.31
17	PPE25	Butyric acid	8 5	-3.56 -3.46
18	PPE25	Curcumin	3	-7.31

			6	-7.09
19	PPE25	Fucoidan	5 9	-8.41 -7.85
20	PPE25	Gliotoxin	5 1	-2.71 -2.64
21	PPE25	Harmine	6 5	-3.74 -3.73
22	PPE25	Helvolic acid	4 9	-5.84 -4.77
23	PPE25	Lyngbioic acid	5 7	-5.05 -4.83
24	PPE25	Manzamine	5 1	-5.05 -5.05
25	PPE25	Naringenin	6 3	-5.36 -5.14
26	PPE25	N-(2-hydroxyphenyl)-2-phenazamine	1 5	-7.32 -7.31
27	PPE25	Peniciadametizine A	4 8	-5.11 -5.06
28	PPE25	Rutin	4 1	-7.59 -7.51
29	PPE18	Baicalin	3 6	-6.24 -6.23
30	PPE18	Benzamide F	3 6	-5.76 -5.46
31	PPE18	Butyric acid	2 4	-2.91 -2.76
32	PPE18	Cristatumin A	7 5	-4.66 -4.30
33	PPE18	Echinulin	1 7	-3.58 -3.35

34	PPE18	Fumetri	5 1	-4.03 -3.81
35	PPE18	Gliotoxin	1 8	-5.07 -4.58
36	PPE18	Helvolic acid	7 8	-4.08 -3.59
37	PPE18	Kaempferol	5 1	-4.95 -4.83
38	PPE18	Luteolin	5 6	-2.73 -2.57
39	PPE18	Lutin	7 5	-6.53 -5.63
40	PPE18	Lyngbioic acid	6 7	-2.97 -2.52
41	PPE18	Malyngolide	6 9	-2.61 -2.39
42	PPE18	N-(2-hydroxyphenyl)-2-phenazamine	8 4	-6.13 -6.10
43	PPE18	Pyranonigrin	1 8	-4.04 -3.41
44	PPE18	Quercetin	1 7	-4.19 -4.18
45	PPE25	Quercetin	2 3	-7.23 -7.03
46	PPE18	Quinine	9 8	-9.62 -8.51
47	PPE18	Rubruzamine	2 7	-4.96 -4.57
48	PPE18	Varieclorin H	9 10	-5.30 -5.02

Table 4 shows the interacting residues of different ligands with the receptor proteins PPE18and PPE25

S.no.	Name of ligand	Interacting Amino acid of protein	Hydrogen bond forming residues	Hydrophobic interaction showing residues
1	Agalasine J	PRO19 SER21 PRO23 TRP31 TYR97 PHE101 ILE110 SER162 AGR224		
2	Ageloxime B	PRO19 SER21 PRO23 MET24 TYR97 PHE101 PRO106 PRO107 ILE116 ARG224		
3	Artimisinin	MET15 GLY18 PRO19 TYR97 PHE101 PRO106 PRO107		
4	Apegenin	GLY18 PRO19 SER21 PRO23 MET24 TYR97 PHE101 PRO107		
5	Aurachin	MET15 CYS17 GLY18	GLU174 THR176 ARG235	

		PRO19 TYR97 PHE101 PRO107 ILE110 ARG224	TRP228	
5	Gliotoxin	GLU142 GLN146 ASP147 VAL149 ALA150 GLY153 TYR154 LEU193 LEU216		
6	Epigallocatechin Gallate	GLY18 TYR97 PHE101 PRO107 PRO108 ILE110 THR111 ARG214 ARG224	MET15 MET24 SER21	THR111 ARG224 ILE110 PRO106 PHE101 CYS17 ARG14 TYR16 GLY18 GLY20 ALA94
7	Butyric acid	HIS328 LEU331 ARG332 PRO335		
8	Hesperidin	ALA42 TYR45 SER64 ALA68 TYR72 TRP75 ALA168 ALA169 ALA288 ASN289		
9	Rutin	ARG14 CYS17		

		PRO19 GLY20 SER21 TYR97 ALA100 PHE101 PRO106 PRO107 ILE110 TYR154 ALA155		
10	Curcumin	PRO19 GLY20 SER21 PRO23 ALA27 TRP31 ALA90 ALA93 TYR97 PHE101 SER162 ARG224 THR225 ILE227 VAL228		
11	Echinulin	PRO19 SER21 TYR97 PHE101 PRO106 PRO107 THR111 ARG114 ARG224		
12	Dihydroechinulin	PRO19 SER21 TYR97 PHE101 PRO106 PRO107 ILE110		

13	Naringenin	PRO19 SER21 PRO23 MET24 TYR97 PHE101 PRO107 ILE110		
14	Brevianamide F	MET15 CYS17 PRO19 GLY20 SER21 PRO23 TYP97 PHE101 PRO107		
15	Cristatumin A	MET15 CYS17 PRO19 SER21 PHE101 PRO107 ILE110		
16	Fucoidan	TYR45 ALA68 TYR72 TRP75 ALA168 ALA169 PRO170 ASN289		
19	Harmine	PRO19 SER21 MET24 ALA27 TRP31 ALA90 ALA93 TYR97 SER162		
20	Helvoic acid	MET15		

		GLY18 PRO19 SER21 PRO23 TYR97 PRO107 ARG224		
21	Lyngbioic acid	ARG14 GLY18 PRO19 GLY20 SER21 TYR97 PHE101 PRO107		
22	Manzamine	CYS17 PRO19 SER21 PRO23 PYR97 PHE101 PRO107		
23	Nhydroxyphenylphenazinamine	MET15 CYS17 GLY18 PRO19 SER21 PRO23 MET24 TYR97 PHE101 PRO107		
24	Quercetin	VAL48 LEU52 ARG287 ILE291 ALA293 TYR351 GLY352 ARG354 HIS355		
25	Variecolorin H	ILE110	GLY18	ILE10

		PRO107 MET15 ARG14 CYS17 GLY18 SER21 PRO19 PRO23 TYR97 PHE101 ALA100 PRO106 ALA155		PRO107 MET15 ARG14 CYS17 SER21
26	Epigallocatechin Gallate	THR111 ILE110 ARG224 PRO106 PRO107 PRO19 PHE101 GLY18 MET15 CYS17 ARG14 TYR16 PRO23 TYR97 SER21 ALA94 TYR97 PRO23	MET15 MET24 SER21	ALA94 GLY20 ARG14 CYS17 GLY18 PHE101 TYR16 PRO106 ILE110 ARG224 THR111
26	Atropine	PRO173 GLU174 GLU170 THR176 SER177 LEU181 MET175 TRP228 ALA178 VAL231 SER232 ARG 235	THR176 TRP228 SER232 PRO173 ARG235	PRO173 GLU170 GLU174 SER177 LEU181 MET175 ALA178 VAL231 ARG235

27	Aurachin	GLU174 THR 176 TRP 228 PRO173 ALA178 ARG235	ARG235 GLU174 THR176 TRP228	ALA178 PRO173
28	N-(2- Hydroxyphenyl)-2-Phenazamine	MET144 ALA112 PRO108 PRO109 GLU113 ALA116 GLU183 GLU143 GLY180 GLN147 THR176 LEU104 SER177 GLY103 GLY179	GLU183 PRO109 SER177	GLU113 GLU143 GLU180 ALA116 GLN147 GLY179 THR176 LEU104 GLY103
29	Palmatine	MET152 MET16 TYR155 ARG15 GLY5 ALA18 GLY19 GLY21 TYR102 GLU99 ALA101 TYR98	ARG15 TYR98 GLU99	GLY5 GLY19 GLY21 ALA101 ALA18
30	Sanguinarine	ALA101 TYR98 MET16 TYR155 ASN12 MET152 ARG15 PHE4 LEU7 ALA6	TYR98	ASN12 PHE4

		GLY5		
31	Variecolorin H	PRO220 LEU224 PHE169 SER221 LEU167 GLU170 LEU166 THR165		ILE110 PRO107 MET15 ARG14 CYS17 SER21
32	Quinine	ILE111 PRO109 PRO108 GLY179 GLN147 GLY180 ALA112 GLY103 SER177 LEU104 THR176 GLU183 GLU143 MET144 ALA146 LEU182	PRO108 THR176	GLY103 GLY179 GLY180 GLN147 MET144 LEU104 LEU182 ALA146 GLU143 GLU183 SER177
28	Quercetin	ALA178 SER177 THR176 GLU174 PHE169 ARG92 GLU170		
29	Kaempferol	PHE169 PRO173 GLU174 THR176 LEU166 LEU167 ALA178 GLU170		
30	Luteolin	TRP228 GLU170		

		PHE169 GLU174 PRO173 ALA96 ARG92		
31	Baicalein	LEU7 ASN12 MET16 MET152 ALA6 ALA18 TYR98 TYR102 TYR155 GLY19 ARG15 ILE11 GLY5 PHE4	ASN12	TYR102 GLY5 GLY19 ALA6 ALA18 PRO20 TYR155 PHE4 MET152 ILE11 LEU7
32	Butyric acid	PRO233 SER236 ARG235 TRP228 SER232		
33	Helvolic acid	ARG235 ALA178 SER177 THR176 PRO173 PHE169 ARG92 GLU170 TRP228		
34	Gliotoxin	GLU99 GLY21 GLY19 PRO20 LEU25 TYR98 ALA18		

		TYR102 SER22		
35	Echinulin	GLN147 GLU143 GLY179 GLU183 SER177 ARG113 PRO173 THR100 GLY103 ARG115		
36	Cristumin	GLN276 ALA275 ALA334 GLN329 ALA330 ALA272 PHE269 ALA273		

5. Discussion

PPE18 and PPE25 are proteins from the PPE (Proline-Proline-Glutamate) family in Mycobacterium tuberculosis H37Rv, which are considered important due to their roles in host-pathogen interactions, immune modulation, and potential involvement in bacterial virulence. Their unique properties make them intriguing candidates for drug targeting. Both PPE18 and PPE25 have been identified as virulence factors. They can modulate host immune responses, potentially helping the pathogen persist within the host. PPE18 has been linked to inhibiting host immune mechanisms, making it a key player in immune evasion. PPE18 and PPE25 **are both surface-exposed proteins** and are secreted, making them accessible targets for drug molecules. Targeting such proteins could weaken the pathogen's ability to persist and evade the immune response. Interestingly, it was observed that the expression patterns and functional roles of PPE 18 and PPE 25 showed variation between virulent and non-virulent strains of Mycobacterium species. In pathogenic strains of *Mycobacterium tuberculosis* such as H37Rv, the expression levels of PPE18 and PPE25 have been reported to be higher compared to non-virulent or attenuated strains like H37Ra. On comparing the gene sequence of these two proteins from different virulent and non-virulent strains it was observed that there are variations in the sequence of both these proteins. These variations include mutations, deletions, or

duplications in the PPE gene loci, which can alter their expression and function. This observation has a functional implication and explains the role of PPE18 and PPE25 in modulating host immune responses, which is critical for virulence. In virulent strains, they contribute to immune evasion by interacting with host immune receptors, inhibiting antigen presentation, or altering cytokine responses. But in non-virulent strains these proteins are either absent completely or are attenuated leading to reduced capacity of these strains to inhibit host immune responses and exhibit less effective intracellular survival in host cells. Due to their differential expression in virulent versus non-virulent strains, PPE18 and PPE25 can also serve as potential biomarkers for distinguishing pathogenic Mycobacterium strains. This study highlights the therapeutic potential of natural compounds, especially those derived from marine sources, as novel inhibitors of PE/PPE family proteins in M.tb. By targeting PPE18 and PPE25, these natural ligands may disrupt M.tb's immune evasion mechanisms and reduce its pathogenicity. However, to confirm these findings and determine the mechanisms by which these compounds exert their inhibitory effects need to be validated experimentally through in vitro and in vivo studies.

6. References

[1] Kaufmann SH, Schaible UE. 100th anniversary of Robert Koch's Nobel Prize for the discovery of the tubercle bacillus. Trends Microbiol. 2005 Oct;13(10):469-75. doi: 10.1016/j.tim.2005.08.003. PMID: 16112578.

[2] Holmes KK, Bertozzi S, Bloom BR, Jha P, Gelband H, DeMaria LM, Horton S. Major Infectious Diseases: Key Messages from *Disease Control Priorities*, Third Edition. In: Holmes KK, Bertozzi S, Bloom BR, Jha P, editors. Major Infectious Diseases. 3rd ed. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2017 Nov 3. Chapter 1. PMID: 30212102.

[3] Bagcchi S. WHO's Global Tuberculosis Report 2022. Lancet Microbe. 2023 Jan;4(1):e20. doi: 10.1016/S2666-5247(22)00359-7. Epub 2022 Dec 12. PMID: 36521512.

[4]. Chandra P, Grigsby SJ, Philips JA. Immune evasion and provocation by Mycobacterium tuberculosis. Nat Rev Microbiol. 2022 Dec;20(12):750-766. doi: 10.1038/s41579-022-00763-4. Epub 2022 Jul 25. PMID: 35879556; PMCID: PMC9310001.

[5]. Chandra P, Grigsby SJ, Philips JA. Immune evasion and provocation by Mycobacterium tuberculosis. Nat Rev Microbiol. 2022 Dec;20(12):750-766. doi: 10.1038/s41579-022-00763-4. Epub 2022 Jul 25. PMID: 35879556; PMCID: PMC9310001.

[6]. Cadena AM, Fortune SM, Flynn JL. Heterogeneity in tuberculosis. Nat Rev Immunol. 2017 Nov;17(11):691-702. doi: 10.1038/nri.2017.69. Epub 2017 Jul 24. PMID: 28736436; PMCID: PMC6247113.

[7]. Yang J, Zhang L, Qiao W, Luo Y. *Mycobacterium tuberculosis*: Pathogenesis and therapeutic targets. MedComm (2020). 2023 Sep 4;4(5):e353. doi: 10.1002/mco2.353. PMID: 37674971; PMCID: PMC10477518.

[8]. Toth A, Fackelmann J, Pigott W, Tolomeo O. Tuberculosis prevention and treatment. Can Nurse. 2004 Nov;100(9):27-30. PMID: 15623010.

[9]. Maiolini M, Gause S, Taylor J, Steakin T, Shipp G, Lamichhane P, Deshmukh B, Shinde V, Bishayee A, Deshmukh RR. The War against Tuberculosis: A Review of Natural Compounds and Their Derivatives. Molecules. 2020 Jun 30;25(13):3011. doi: 10.3390/molecules25133011. PMID: 32630150; PMCID: PMC7412169.

[10]. Hamada Y, Getahun H, Tadesse BT, Ford N. HIV-associated tuberculosis. Int J STD AIDS. 2021 Aug;32(9):780-790. doi: 10.1177/0956462421992257. Epub 2021 Feb 20. PMID: 33612015; PMCID: PMC8236666.

[11].TB/COVID-19 Global Study Group. Tuberculosis and COVID-19 co-infection: description of the global cohort. Eur Respir J. 2022 Mar 24;59(3):2102538. doi:

[12]. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin

K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998 Jun 11;393(6685):537-44. doi: 10.1038/31159. Erratum in: Nature 1998 Nov 12;396(6707):190. PMID: 9634230.

[13]. Moran NA. Microbial minimalism: genome reduction in bacterial pathogens. Cell. 2002 Mar 8;108(5):583-6. doi: 10.1016/s0092-8674(02)00665-7. PMID: 11893328.

[14]. Murray GGR, Charlesworth J, Miller EL, Casey MJ, Lloyd CT, Gottschalk M, Tucker AWD, Welch JJ, Weinert LA. Genome Reduction Is Associated with Bacterial Pathogenicity across Different Scales of Temporal and Ecological Divergence. Mol Biol Evol. 2021 Apr 13;38(4):1570-1579. doi: 10.1093/molbev/msaa323. PMID: 33313861; PMCID: PMC8042751.

[15]. Fishbein S, van Wyk N, Warren RM, Sampson SL. Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity. Mol Microbiol. 2015 Jun;96(5):901-16. doi: 10.1111/mmi.12981. Epub 2015 Mar 30. PMID: 25727695.

[16]. Gey van Pittius NC, Sampson SL, Lee H, Kim Y, van Helden PD, Warren RM. Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol Biol. 2006 Nov 15;6:95. doi: 10.1186/1471-2148-6-95. PMID: 17105670; PMCID: PMC1660551.

[17]Medha, Sharma S, Sharma M. Proline-Glutamate/Proline-Proline-Glutamate (PE/PPE) proteins of Mycobacterium tuberculosis: The multifaceted immune-modulators. Acta Trop. 2021 Oct;222:106035. doi: 10.1016/j.actatropica.2021.106035. Epub 2021 Jul 3. PMID: 34224720.

[18]. Guo F, Wei J, Song Y, Li B, Qian Z, Wang X,Wang H and Xu T (2023) Immunological effects of the PE/PPE family proteins ofMycobacterium tuberculosis and related vaccines.Front. Immunol. 14:1255920.doi: 10.3389/fimmu.2023.1255920

[19]. Abdallah AM, Verboom T, Weerdenburg EM, Gey van Pittius NC, Mahasha PW, Jiménez C, Parra M, Cadieux N, Brennan MJ, Appelmelk BJ, Bitter W. PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5. Mol Microbiol. 2009 Aug;73(3):329-40. doi: 10.1111/j.1365-2958.2009.06783.x. Epub 2009 Jul 7. PMID: 19602152.

[20]. Liu CH, Liu H, Ge B. Innate immunity in tuberculosis: host defense vs pathogen evasion. Cell Mol Immunol. 2017 Dec;14(12):963-975. doi: 10.1038/cmi.2017.88. Epub 2017 Sep 11. PMID: 28890547; PMCID: PMC5719146.

[21] Özçelik, B.; Kartal, M.; Orhan, I. Cytotoxicity, antiviral and antimicrobial activities of alkaloids, flavonoids, and phenolic acids. Pharm. Biol. 2011,49, 396–402.

[22) Liu S , Han W, Sun Cet al. . Subtractive screening with the Mycobacterium tuberculosis surface protein phage display library. Tuberculosis 2011; 91:579–86.

[23]. Rabaan, A.A.; Alhumaid, S.; Albayat, H.; Alsaeed, M.; Alofi, F.S.; Al-Howaidi, M.H.; Turkistani, S.A.; Alhajri, S.M.; Alahmed, H.E.; Alzahrani, A.B.; et al. Promising Antimycobacterial Activities of Flavonoids against Mycobacterium Sp. Drug Targets: A Comprehensive Review. Molecules 2022, 27, 5335. https://doi.org/10.3390/ molecules 27165335

[24]. Gygli, S.M.; Borrell, S.; Trauner, A.; Gagneux, S. Antimicrobial resistance in Mycobacterium tuberculosis: Mechanistic and evolutionary perspectives. FEMS Microbiol. Rev. 2017, 41, 354–373
10. X.M. Wei et al. Neopetrosiamine A, biologically active bis-piperidine alkaloid from the Caribbean Sea sponge Neopetrosia proxima, Bioorg. Med. Chem. Lett (2010)

[25]. E. Fahy et al. Haliclonadiamine, an antimicrobial alkaloid from the sponge Haliclona sp Tetrahedron Lett. (1988)

[26]. R.R. Yadav et al. Antimalarial and antitubercular activities of meridianin derivatives Eur. J. Med. Chem. (2015)

[27]. C. D'Souza, U. Kishore, and A. G. Tsolaki, "The PE-PPE Family of Mycobacterium tuberculosis: Proteins in Disguise," 2023. [Online]. Available: https://doi.org/10.1016/j.imbio.2022.152321

[28]. Riley R, Pellegrini M, Eisenberg D. Identifying cognate binding pairs among a large set of paralogs: the case of PE/PPE proteins of Mycobacterium tuberculosis. PLoS Comput Biol. 2008 Sep 12;4(9):e1000174. doi: 10.1371/journal.pcbi.1000174. PMID: 18787688; PMCID: PMC2519833.

[29]. He Z, Wang C, Guo X, Sun H, Bi Y, Pitt ME, Li C, Song J, Coin LJM, Li F. MERITS: a web-based integrated *Mycobacterial* PE/PPE protein database. Bioinform Adv. 2024 Mar 2;4(1) :vbae 035. doi: 10.1093/bioadv/vbae035. PMID: 38549946; PMCID: PMC10973932.