Metallophiles as sources of antimycobacterial agents

by
Kudakwashe Nyambo

Dissertation presented for the degree of
Doctoral of Molecular Biology in the Faculty of
Medicine and Health Sciences at Stellenbosch University

Supervisor: Dr Vuyo Mavumengwana
Co-supervisor: Dr Mhkuseli Ngxande
Co-supervisor: Dr Liezel Smith
Co-supervisor: Prof Idah Sithole-Niang

December 2023

December 2023

Declaration

By submitting this thesis/dissertation, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2023

General Abstract

Mycobacterial pathogens present a significant complication to disease control globally due to their resistance to numerous antibiotics. The rise in resistant strains to current chemotherapeutic treatments has prompted the search, development and implementation of new strategies to address this challenge. Harnessing the bioactivity of natural products found in the vast chemical space by using multi-disciplinary approaches has emerged as a promising way to discover new Tuberculosis drugs. This study aimed to evaluate the potential antimycobacterial activity of secondary metabolites from bacteria, fungi, and plants *in-vitro* and *in-silico*. In addition to mining for *Mycobacterium tuberculosis* targets, this study went further to explore other druggable targets associated with cancer in order to fully explain exhaustive *in-silico* bioactivity profiles.

The following experiments were conducted to satisfy the aims: (i) bacteria from gold mine tailings were isolated and identified using 16S rRNA sequencing. The crude extracts from the bacteria were screened for potential activity against *Mycobacterium tuberculosis (M. tb) H*37Rv, *Mycobacterium smegmatis* MC^2155, and *Mycobacterium aurum* A+ *in-vitro*. The active extracts were tentatively identified using HPLC-qTOF, GNPS, and Ms Dial. The identified compounds were virtually screened against *Mycobacterium* Pks13 and PknG. The natural compound that displayed high affinity was subjected to modification through multiple synthetic routes using reaction-driven enumeration. (ii) A total of 15 fungi compounds from fungi isolated from gold mine tailings were evaluated for their potential activity against *M. tb* PknA, PknB, PknD, and PknE proteins using extra precision molecular docking, molecular dynamics simulations, and molecular mechanics generalized born surface area (MM-GBSA) binding free energy calculation. (iii) Genomic DNA of one bacterial colony that showed activity against M. *tb*, was isolated and

sequenced by Illumina's NextSeq platform. The genes responsible for producing metabolites that may have antimycobacterial activity were determined using antiSMASH and PARTIC. (iv) Predictive machine learning-based quantitative structure-activity relationship models were developed with a pIC_{50} as the dependable variable, while features extracted from compounds found to be active against InhA were the independent variable. Another approach in developing a multitargeted SMILES-based Long Short-term Memory (LSTM) based on $pIC_{50,}$ and small, skewed datasets was attempted. (v) Medicinal plant species indigenous to South Africa namely *Schotia brachypetala, Rauvolfia caffra, Schinus molle, Ziziphus mucronate,* and *Senna petersiana* were evaluated for their potential antimycobacterial activity against *Mycobacterium smegmatis* MC^2155, *Mycobacterium aurum* A+, and *M. tb* H37Rv. Although the study was specific to mycobacteria, further exploration into cytotoxic activity against MDA-MB 231 triple-negative breast cancer cells was also attempted to see if druggable targets could also be identified in eukaryotic cells as a test of the utility and robustness of the method. The constituents of the extracts possessing antimycobacterial activity were virtually screened using a rigorous Virtual Screening Workflow. The compounds exhibiting good binding, and ADME properties were returned and subjected to molecular dynamics simulations. MM-GBSA calculations were performed to evaluate the affinity of the selected compound/s to pantothenate kinase (PanK).

Crude extracts from three bacterial isolates, namely *Bacillus subtilis* and *Bacillus licheniformis,* exhibited activity against *M. tb* H37Rv, *Mycobacterium smegmatis* MC^2155, and *Mycobacterium aurum* A+. The classes of secondary metabolites identified in this study are known to possess antibacterial activity. Virtual screening of the secondary metabolites against PknG and Pks13, returned cyclo-(L-Pro-4-OH-L-Leu) and vazabitide A with pre-MD MM-GBSA values of -42.81

kcal/mol and -47.62 kcal/mol, respectively. The modification of vazabitide A yielded a compound with a higher affinity of -85.80 kcal/mol to the Pks13, binding as revealed by the post-MD MM-GBSA. SAMN36381076 was assigned to be *B. licheniformis* whole genome analysis. The genome length of *B. licheniformis SAMN36381076* was estimated to be 4.213156 Mb, with a G+C content of 46.08%, comprising 58 contigs and exhibiting an N50 length of 165,033 bp. The biosynthetic gene clusters identified included fengycin, butirosin A, butirosin B, schizokinen, pulcherriminic acid, bacillibactin, bacillibactin E, bacillibactin F, lichenicidin VK21 A1, Lichenicidin VK21 A2, and thermoactinoamide A. These gene clusters are known for producing secondary metabolites with antimicrobial activity. Furthermore, The *B. licheniformis* SAMN36381076 possesses genes that encode for six diverse antibiotic resistance mechanisms, with efflux pumps as the predominant mechanism of resistance. Metabolic analysis of *B. licheniformis* SAMN36381076 showed that the presence of genes involved carbohydrate degradation and assembly processes, oxyanion biogeochemical cycling, and nitrogen cycling.

In-silico evaluation of fungi compounds against ser/thr kinases showed the lowest ΔG_{Bind} values of aurovertin D against PknA (-50.9 kcal/mol), aurovertin D against PknB (-50.7 kcal/mol), verticillin A against PknD (-36.8 kcal/mol), and roquefortine C against PknE (-53.4 kcal/mol). Molecular dynamics simulation showed that the PknD-verticillin A exhibited the highest stability. Furthermore, the post-MD MMGBSA ΔG_{Bind} showed that verticillin A has a high affinity for PknD -53.67 kcal/mol. The results indicated that verticillin A is a potential hit compound that can be further optimized and modified to develop a potent antimycobacterial inhibitor.

The classical machine learning models developed from logistic regression and multi-layer perceptron were identified to have significant performance metrics on the InhA dataset. The results

(- R^2) from the multitarget Long Short Term Memory (LSTM) model indicated the need for hyperparameter tuning. However, further external validation of the two classification models is needed.

In this study, the bioactive compounds present in *R. caffra* and *S. molle* showed average activity against *M. tb* H37Rv (MIC 0.25-0.125 mg/mL). Norajmaline with a docking score of -7.47 kcal/mol, and pre-MM-GBSA of -37.64 kcal/mol was returned from the rigorous virtual screening. Molecular dynamics simulation and post-MD MM-GBSA revealed the stable binding of norajmaline to PanK (-58. 73 kcal/mol). Results from the Flow cytometry analysis of treated MDA-MB 231 cells revealed that the dichloromethane extracts from *S. petersiana*, *Z. mucronate,* and ethyl acetate extracts from *R. caffra* and *S. molle* induced higher levels of apoptosis than the control cisplatin.

In conclusion, this study serves as a starting point for the *in-silico* discovery of potent antimycobacterial compounds from metallophiles (fungi and bacteria) and plants. Virtual screening accelerates the drug discovery process by identifying compounds that may possess activity, thus they can be modified to increase potency. The incorporation of a large dataset of compounds comprising different biological conditions but with the same endpoint can be used to develop robust models with exceptional generalization capabilities.

Opsomming

Mikobakteriese patogene bied 'n beduidende komplikasie tot siektebeheer wêreldwyd as gevolg van hul weerstand teen talle antibiotika. Die toename in weerstandige stamme teen huidige chemoterapeutiese behandelings het aanleiding gegee tot die soektog, ontwikkeling en implementering van nuwe strategieë om hierdie uitdaging aan te spreek. Die benutting van die bioaktiwiteit van natuurlike produkte in die chemiese ruimte deur die gebruik van multidissiplinêre benadering het na vore gekom as 'n belowende manier om nuwe tuberkulose middels te ontdek. Die doel van hierdie studie was om die potensiële antimikobakteriële aktiwiteit van sekondêre metaboliete van bakterieë, swamme en plante te evalueer. Hierdie studie het ook die antikankeraktiwiteit van die plante geëvalueer.

Die volgende eksperimente is uitgevoer om die doelwitte te vervul: (i) isolasie en identifikasie van bakterieë uit goudmyn uitskot deur gebruik te maak van 16S rRNA volgorde bepaling. Die ru-ekstrakte van die bakterieë is gesif vir potensiële aktiwiteit teen *Mycobacterium tuberculosis (M. tb)* H37Rv, *Mycobacterium smegmatis* MC ^{2} 155 en *Mycobacterium aurum* A+ *in vitro* . Die aktiewe ekstrakte is voorlopig geïdentifiseer deur gebruik te maak van HPLC- qTOF , GNPS en Ms Dial. Die geïdentifiseerde verbindings is virtueel gesif teen *Mycobacterium* Pks13 en PknG. Die natuurlike verbinding wat hoë affiniteit getoon het, is onderworpe aan modifikasie deur die gebruik van verskeie sintetiese roetes deur reaksie-gedrewe konfigurasies. (ii) 'n Totaal van 15 swam verbindings van swamme van goudmyn uitskot is geëvalueer vir hul potensiële aktiwiteit teen *M. tb* PknA-, PknB-, PknD- en PknE-proteïene deur ekstra presiesheid molekulêre koppeling, molekulêre dinamika-simulasies en molekulêre meganika veralgemeende gebore oppervlak area (MM-GBSA) binding vry energie berekeninge. (iii) Genomiese DNA van een bakterie wat

aktiwiteit teen *M. tb getoon het* , is geïsoleer en georden deur Illumina se NextSeq platform. Die gene wat verantwoordelik is vir die vervaardiging van metaboliete wat antimikobakteriële aktiwiteit kan hê, is bepaal deur gebruik te maak van antiSMASH en PARTIC. (iv) Voorspellende masjienleer-gebaseerde kwantitatiewe struktuur-aktiwiteit verwantskap modelle is ontwikkel met 'n pIC_{50} as die betroubare veranderlike, terwyl kenmerke uittreksels uit verbindings wat aktief teen InhA was, is as die onafhanklike veranderlike gebruik. Nog 'n benadering in die ontwikkeling van 'n multi-geteikende SMILES-gebaseerde lang korttermyn geheue (LSTM) gebaseer op $pIC_{50,}$ en klein, skewe datastelle was probeer. (v) Medisinale plantspesies inheems aan Suid-Afrika, naamlik *Schotia brachypetala , Rauvolfia caffra , Schinus molle , Ziziphus mucronate,* en *Senna petersiana* is geëvalueer vir hul potensiële antimikobakteriese aktiwiteit teen *Mycobacterium smegmatis* MC ^{2} 155, *Mycobacterium aurum* A+ en *Mycobacterium tuberculosis* H37Rv, en sitotoksiese aktiwiteit teen MDA-MB 231 trippel-negatiewe bors kankerselle . Die bestanddele van die ekstrakte wat antimikobakteriese aktiwiteit het, is virtueel gesif deur gebruik te maak van 'n streng virtuele siftings werkvloei. Die verbindings wat goeie binding en ADME-eienskappe getoon het, is teruggestuur en aan molekulêre dinamika-simulasies onderwerp. MM-GBSA berekeninge is uitgevoer om die affiniteit van die geselekteerde verbindings vir pantotenaat kinase (PanK) te evalueer.

Ru-ekstrakte van drie bakteriese isolate, naamlik *Bacillus subtilis* en *Bacillus licheniformis,* het aktiwiteit getoon teen *M. tb* H37Rv, *Mycobacterium smegmatis* MC ^{2} 155 en *Mycobacterium aurum* A+. Dit is bekend dat die sekondêre metaboliete klasse wat in hierdie studie geïdentifiseer is, antibakteriese aktiwiteit het. Virtuele sifting van die verbindings teen PknG en Pks13, siklo-(L-Pro-4-OH-L-Leu) en vazabitied A met pre-MD MM-GBSA het teruggekeer met waardes van -

42.81 kcal/mol en -47.62 kcal/mol, onderskeidelik. Die modifikasie van Vazabitide A tot die Pks13 binding het gelei tot 'n verbinding met 'n hoër affiniteit van -85.80 kcal/mol soos geopenbaar deur die post-MD MM-GBSA. *B. licheniformis SAMN36381076 is geïdentifiseer deur h*eelgenoom analise. Die genoom lengte van *B. licheniformis SAMN36381076* is geskat op 4,213156 Mb, met 'n G+C-inhoud van 46,08%, wat 58 kontigs bevat en 'n N50-lengte van 165,033 bp vertoon. Die biosintetiese geen groepe wat geïdentifiseer is, sluit in fengysien, butirosien A , butirosien B, skisokienen, pulcherriminensuur, bacillibactin , bacillibactin E , bacillibactin F, lichenicidin VK21 A1 , Lichenicidin VK21 A2, en thermoactinoclusters. Hieride geen groepe is bekend vir produksie van metaboliete met antimikrobakteriële aktiwiteit. Verder beskik die *B. licheniformis* SAMN36381076 oor gene wat kodeer vir ses diverse antibiotika weerstand meganismes, met uitvloei pompe as die oorheersende meganisme van weerstand. Metaboliese analise van *B. licheniformis* SAMN36381076 het getoon dat die teenwoordigheid van gene koolhidraat afbraak- en samestellings prosesse, oksianion biogeochemiese siklusse en stikstofsiklusse ingesluit het.

In-silico evaluering van swam verbindings teen ser/ threokinases het die laagste ΔG- $_{\text{bindingswaardes}}$ van aurovertien D teen PknA (-50.9 kcal/mol), aurovertien D teen PknB (-50.7 kcal/mol), vertisillien A teen PknD (-36.8 kcal/mol) getoon, en roquefortine C teen PknE (-53,4 kcal/mol). Molekulêre dinamika-simulasie het getoon dat die PknD-vertisillien A die hoogste stabiliteit getoon het. Verder het die post-MD MMGBSA ΔG_{Bind} getoon dat vertisillien A 'n hoë affiniteit vir PknD -53.67 kcal/mol het. Die resultate het aangedui dat vertisillien A 'n potensiële tref verbinding is wat verder geoptimaliseer en aangepas kan word om 'n kragtige antimikobakteriële inhibeerder te ontwikkel.

In hierdie studie het die bioaktiewe verbindings teenwoordig in *R. caffra* en *S. molle* gemiddelde aktiwiteit teen *M. tb* H37Rv (MIC 0.25-0.125 mg/mL) getoon. Norajmaline met 'n dok telling van -7.47 kcal/mol, en pre-MM-GBSA van -37.64 kcal/mol is teruggekeer van die streng virtuele sifting. Molekulêre dinamika-simulasie en post-MD MM-GBSA het die stabiele binding van norajmaline aan PanK (-58. 73 kcal/mol) geopenbaar. Resultate van die vloeisitometrie-analise van behandelde MDA-MB 231-selle het aan die lig gebring dat die dichloormetaan ekstrakte van *S. petersiana* , *Z. mucronate* en etielasetaat ekstrakte van *R. caffra* en *S. molle* het hoër apoptose vlakke geïnduseer as die kontrole-cisplatien.

Laastens, die klassieke masjienleer modelle wat uit logistiese regressie en multi-laag perseptron ontwikkel is, is geïdentifiseer om beduidende prestasiemaatstawwe op die InhA-datastel te hê. Die resultate (- R^2) van die multiteiken-LSTM-model het die behoefte aan hiper parameter verfynning aangedui. Verdere eksterne validering van die twee klassifikasie modelle is egter nodig.

Ter afsluiting, hierdie studie dien as 'n beginpunt vir die in-silico ontdekking van kragtige antimikobakteriële teenmiddels van metallofiele (swamme en bakterieë) en plante. Virtuele sifting versnel die geneesmiddel ontdekkings proses deur verbindings te identifiseer wat aktiwiteit kan hê, en aangepas kan word om hulle kragtigheid te verhoog. Die inkorporering van 'n groot datastel van verbindings wat verskillende biologiese toestande bevat maar met dieselfde eindpunt kan gebruik word om robuuste modelle met buitengewone veralgemenings vermoëns te ontwikkel.

DEDICATION

To my parents: Abraham and Dorothy Nyambo; brothers: Dr. Peter, Dr. Patrick, and Tendai; sisters: Violet and Christine. This work was conducted within the Centre for Tuberculosis Research of the SA Medical Research Council.

Give instruction to a wise man and he will yet be wiser: teach a just man, and he will increase in learning – Proverbs 9 v 9

ACKNOWLEDGEMENTS

It has been a period of intense learning for me, not only in the scientific arena but also on a personal level. Writing this thesis has had a significant impact on me. I want to reflect on the people who have supported and helped me so much in making this dream a reality.

Firstly, I would like to always express gratitude to **Almighty God** for the **ABUNDANT GRACE IN ALL THINGS** that carried me through my studies.

I would particularly like to single out my supervisors, Dr. Vuyo Mavumengwana, Dr. Mkhuseli Ngxande, Dr. Liezel Smith, and Professor Idah Sithole-Niang, and express my heartfelt gratitude for their exceptional guidance, mentorship, soliciting funding and invaluable contributions. They provided me with the tools that I needed to choose the right direction and complete my thesis.

I am deeply thankful to Dr. Krishna Govender, Miss Precious Chiwira, Dr. Lucinda Baatjies, Dr. Kudzanai Tapfuma, Mr. Francis Adu-Amankwa, and personnel in Vuyo Lab, for their excellent assistance in making laboratory and computational facilities available for me to conduct my research.

I also thank the CTR and CBTBR for their financial support.

There are many colleagues who assisted me in various ways, and those that deserve mention are Candice February, Wonder Choga, and Ivan Nkuhairwe.

I would also like to extend my heartfelt gratitude to my parents for their wise counsel, unwavering support, encouragement, boundless love, and sympathetic ear throughout this journey. Additionally, I want to thank my brothers and sisters who frequently traveled to visit and help me

with maintaining my mental health during the study period. I know this will make you proud. This is the product.

Thank you very much to everyone, and may Almighty God richly bless you all.

"The heart of the discerning acquires knowledge, for the ears of the wise seek it out." Proverbs 18:15

Table of Contents

List of Tables

List of Figures

General Introduction

1.1 Challenges in Treating Tuberculosis

Tuberculosis (TB) caused by a highly specialized human intracellular pathogen *Mycobacterium tuberculosis* (*M. tb*), is a severe and life-threatening infectious disease [1]. Standard chemotherapy treatment of drug-susceptible *M. tb* consists of administering a combination of first-line drugs such as isoniazid, ethambutol, rifampicin, and pyrazinamide, for six months with a success rate of 85 % [2, 3]. However, the emergence of drug-resistant strains over the past two decades has become a critical threat to the global campaign to end TB. Multidrug-resistant (MDR-TB) *M. tb* strains, which account for 3.4 % of new TB cases globally, are resistant to at least two first-line drugs, which are isoniazid and rifampicin. Extensively drug resistance (XDR)- *M. tb* strains, on the other hand, are resistant to isoniazid, rifampicin, fluoroquinolone, and at least one of the three injectable second-line drugs [4, 5]. The emergence of drug resistance in *M. tb* strains is caused by genetic mutations to chromosomal genes associated with pro-drug activation, drug permeability, macromolecular target, and drug efflux. Further, metabolic adaptation to the host lung microenvironment and socioeconomic determinants such as poor diagnosis and limited drug availability, also contribute to an increase in drug-resistant strains [6]. This highlights the urgent requirement for developing and applying targeted strategies to address the rapid increase in drug-resistant *M. tb*. New antimycobacterial agents should be compatible with current antiretroviral regimens, reduce treatment duration, and be effective against MDR-TB and XDR-TB. Additionally, antimycobacterial agents should be able to destroy *M. tb* at different stages of its life cycle, including the non-replicating latent stage [4, 7].

Computational strategies such as well-built statistical methods or artificial intelligence aid the analyses of crucial biological properties of specified functional groups and optimization of the chemical structures to enhance or decrease the desired endpoints.

1.2 Rationale and justification of the study

Drug discovery is a complex and time-consuming process involving identifying and developing new therapeutic drugs for numerous diseases. For many years, natural products (NPs) have served as a verified reference point for discovering and developing highly potent antimicrobials and chemotherapeutics [8, 9]. Microorganisms are known as the most prolific bio-factories of diverse and novel NPs that possess valuable pharmaceutical properties [10]. NPs have been structurally optimized through evolution to enhance the survival prospects of microorganisms in their respective niches. The wide variety of bioactive secondary metabolites produced by microorganisms, for example, fungi, actinomycetes, and myxobacteria make them prime sources for mining and discovery of potential novel therapeutic agents [10–12]. Noteworthy, many drugs in current pharmacopeias are derived from NPs. However, the antibiotic industry has experienced a decline in the discovery of novel bioactive drug scaffolds due to the continuous re-discovery of similar compounds from over-investigated ecological niches. Therefore, it is imperative to explore untapped natural reservoirs to discover novel bioactive small molecules. One such potential source is the South African gold mine tailings, which is characterized by low pH, and high metal content. Bacteria and fungi inhabiting the gold mine tailings possess a unique extended biosynthetic skill that enables them to survive under environmental stresses [13]. By harnessing the novel metabolic machinery of these bacteria and fungi through whole genome mining of biosynthetic functional gene clusters, and metabolomics, it may be possible to synthesize small bioactive molecules, thus

providing molecular starting points for the development of anti-TB drugs. A targeted rational approach to TB drug discovery involves a thorough understanding of the pathogen at a macromolecular level of the pathways involved in maintaining its fitness, thus, leading to identifying the druggable macromolecular targets. Identifying the macromolecular targets or pathways that are critically involved in a disease state is essential in solving XDR, and MDR TB [4, 14–16]. Intrinsic understanding of the target on a molecular scale is essential for example, investigating and elucidating the important druggable active sites on essential *M. tb* proteins. The knowledge of the macromolecular target permits us to design safe and potent anti-*M. tb* drugs. Targeting single proteins in TB drug discovery is not sufficient to tackle the MDR-TB challenges [17–19]. Thus, novel drugs to treat MDR-TB should have novel mechanisms of action in either or all of the following ways: (i) a pathogen-directed approach whereby the drugs inhibit essential *M. tb* macromolecular targets, (ii) a host-directed approach focuses on remodeling the host's cell functions to facilitate the clearance of the pathogen, thus, reducing the *M. tb* 's ability to evade immune responses. Advancements in technologies have accelerated the discovery of lead compounds and designing potential drugs. Virtual screening is a computational method used in drug discovery during the hit-lead discovery phase that aims to identify potential hit compounds, i.e., molecular starting points, from large databases of compounds. Molecular dynamics (MD) simulations is one of the techniques used and it enables the comprehension of complex physical dynamics of a biological system at an atomic level, thus, revealing the hidden states of a biological system that cannot be detected experimentally [20, 21]. The molecular starting points are further validated *in-vitro* and *in-vivo* and after each iteration, the chemical properties of the scaffold are progressively optimized. Thus, incorporating artificial intelligence at every stage of drug design, including computer-aided molecular modeling and machine learning-related techniques,

accelerates the discovery and optimization of potent lead scaffolds that may be optimized to assist in fighting the drug resistance of *M. tb* clinical strains. In this regard, integrating metabolomics, *in-vitro* assays, MD simulations, and accurate binding free energy computation provided refined and possible novel mechanisms of action.

1.3 Study Aims and Objectives

The current study aims to identify and evaluate the antimycobacterial activity of crude extracts from microorganisms isolated from the South African gold mine tailings using a multi-omics approach.

To achieve the above aims, the following objectives were formulated:

1. To isolate and purify bacteria, isolate total genomic DNA, amplify and sequence the 16S rRNA gene, and evaluate the evolutional relationship of identified bacteria.
2. To tentatively identify the secondary metabolites produced by bacteria and fungi using Liquid Chromatography-Mass Spectrometry QTOF.
3. To evaluate minimum inhibition concentration against three *Mycobacterium* strains, namely *M. smegmatis* mc^2 155, *M. aurum* A+, and *M. tb H37Rv*.
4. To virtually screen the identified compounds against *M.tb* macromolecular tagerts using molecular docking, molecular dynamics simulations and binding free energy computations.
5. To sequence the whole genome of the bacteria active against *M.tb* and identify the biosynthetic gene clusters responsible for the synthesis of possible anti-TB secondary metabolites.

6. To create a QSAR model using machine learning algorithms and facilitate the precise screening of small bioactive molecules based on antimycobacterial activity.

1.4 Thesis outline

This thesis has been divided into eight chapters. Descriptions of the chapters are as follows:

CHAPTER 1 – General Introduction

This chapter introduces the readers to the problem statement, rationale, justifications, aims, and objectives of the study.

CHAPTER 2 –Integrating Virtual Screening to Aid Tuberculosis Drug Development- Mini-review

This chapter describes the current trends in Tuberculosis drug discovery. In addition, the integration of targeted multi-omics approach to accelerate TB drug discovery, and save operational costs is fully described. The manuscript from this chapter is ready for submission.

CHAPTER 3 – Molecular docking, molecular dynamics simulations, and binding free energy studies of interactions between *Mycobacterium tuberculosis* Pks13, PknG, and bioactive constituents of extremophilic bacteria

This chapter describes the isolation, and identification of bacteria from gold mine tailings. The crude extracts from the bacteria are tentatively identified and evaluated for potential antimycobacterial activity *in-vitro* and *in-silico. In-silico* modification of a compound exhibiting strong binding profile was performed. The profile of the modified compound

exhibited promising binding profiles, and ADME characteristics. A manuscript resulting from this chapter is under review in a peer reviewed journal, Heliyon (manuscript no. HELIYON-D-23-17664).

CHAPTER 4 – *In-silico* screening of fungal secondary metabolites against *Mycobacterium tuberculosis* Ser/Thr kinases

This chapter describes the virtual screening of compounds from fungal against *M.tb* essential proteins. The Serine/Threonine kinases were investigated as potential drug targets for *M.tb*. The study identified a compound that exhibited strong binding to the protein targets. A manuscript resulting from this chapter is under review in a peer reviewed journal, Computational Biology and Chemistry (CBAC-D-23-00802).

CHAPTER 5 – Exploring the Metabolic Potential of *Bacillus licheniformis* for the Production of Antimycobacterial Secondary Metabolites

This chapter describes the metabolic potential of the bacteria to produce metabolites that have antimycobacterial activity. Whole genome sequencing was perfomed on one of the isolates that exhibited antimicrobial activity. The biosynthetic gene clusters were mined using antiSMASH and they showed the ability of the *B.lichenformis* to produce bioactive active compounds that have antimicrobial properties. The manuscript from this chapter is ready for submission.

CHAPTER 6 – Comparison of predictive machine learning-based quantitative structure-activity relationship models for targeting InhA, Pks13 and MptpA

This chapter describes the development of machine-learning-based quantitative structure-activity relationship models for target *Mycobacterium tuberculosis* InhA, Pks13 and MptpA. The supervised classifiers classical machine learning algorithms trained in the study, include Random Forest (RF), decision tree (DT), support vector machines (SVC), KNeighbors (KNN), logistic regression (LR), and multi-layer perceptron (MLP).

CHAPTER 7– General conclusions and future directions

This chapter presents the conclusions, limitations, and recommendations for future studies.

1.5 References

1. López-Agudelo VA, Baena A, Barrera V, Cabarcas F, Alzate JF, Beste DJV, et al. Dual RNA Sequencing of *Mycobacterium tuberculosis*-Infected Human Splenic Macrophages Reveals a Strain-Dependent Host–Pathogen Response to Infection. Int J Mol Sci. 2022;23:1803.

2. Cohen T, van Helden PD, Wilson D, Colijn C, McLaughlin MM, Abubakar I, et al. Mixed-strain *Mycobacterium tuberculosis* infections and the implications for tuberculosis treatment and control. Clin Microbiol Rev. 2012;25:708–19.

3. Miggiano R, Rizzi M, Ferraris DM. *Mycobacterium tuberculosis* pathogenesis, infection prevention and treatment. Pathogens. 2020;9.

4. Singh R, Dwivedi SP, Gaharwar US, Meena R, Rajamani P, Prasad T. Recent updates on drug resistance in *Mycobacterium tuberculosis*. J Appl Microbiol. 2020;128:1547–67.

5. Abbas HS, Baker DHA. Recent challenges in tuberculosis treatments : A REVIEW. 2020;20:3539–47.

6. Singh V, Chibale K. Strategies to Combat Multi-Drug Resistance in Tuberculosis. Acc Chem Res. 2021;54:2361–76.

7. Bie S, Hu X, Zhang H, Wang K, Dou Z. Influential factors and spatial–temporal distribution of tuberculosis in mainland China. Sci Rep. 2021;11.

8. Franco AR, Peri F. Developing New Anti-Tuberculosis Vaccines: Focus on Adjuvants. Cells. 2021;10:78.

9. Gurnani N, Mehta D, Gupta M, Mehta BK. Natural Products: Source of Potential Drugs. African Journal of Basic & Applied Sciences. 2014;6:171–86.

10. Jakubczyk D, Dussart F. Selected fungal natural products with antimicrobial properties. Molecules. 2020;25.

11. Tapfuma KI, Nyambo K, Adu-Amankwaah F, Baatjies L, Smith L, Allie N, et al. Antimycobacterial activity and molecular docking of methanolic extracts and compounds of marine fungi from Saldanha and False Bays, South Africa. Heliyon. 2022;8:e12406.

12. Deering RW, Whalen KE, Alvarez I, Daffinee K, Beganovic M, LaPlante KL, et al. Identification of a bacteria-produced benzisoxazole with antibiotic activity against multi-drug resistant Acinetobacter baumannii. Journal of Antibiotics. 2021. https://doi.org/10.1038/s41429-021-00412-7.

13. Jones SE, Elliot MA. Streptomyces exploration: Competition, volatile communication and new bacterial behaviours. Trends Microbiol. 2017;25:522–31.

14. Ji H, Zhang Y, Bararunyeretse P, Li H. Characterization of microbial communities of soils from gold mine tailings and identification of mercury-resistant strain. Ecotoxicol Environ Saf. 2018;165:182–93.

15. Basarab GS, Ghorpade S, Gibhard L, Mueller R, Njoroge M, Peton N, et al. Spiropyrimidinetriones: a Class of DNA Gyrase Inhibitors with Activity against *Mycobacterium tuberculosis* and without Cross- Resistance to Fluoroquinolones. Antimicrob Agents Chemother. 2022;66.

16. Schön T, Miotto P, Köser CU, Viveiros M, Böttger E, Cambau E. *Mycobacterium tuberculosis* drug-resistance testing: challenges, recent developments and perspectives. Clinical Microbiology and Infection. 2017;23:154–60.

17. Nimmo C, Millard J, van Dorp L, Brien K, Moodley S, Wolf A, et al. Population-level emergence of bedaquiline and clofazimine resistance-associated variants among patients with drug-resistant tuberculosis in southern Africa: a phenotypic and phylogenetic analysis. Lancet Microbe. 2020;1:e165–74.

18. Yang L, Hu X, Chai X, Ye Q, Pang J, Li D, et al. Opportunities for overcoming tuberculosis: Emerging targets and their inhibitors. Drug Discov Today. 2022;27:326–36.

19. Sundar S, Thangamani L, Manivel G, Kumar P, Piramanayagam S. Molecular docking, molecular dynamics and MM/PBSA studies of FDA approved drugs for protein kinase a of *Mycobacterium tuberculosis*; application insights of drug repurposing. Inform Med Unlocked. 2019;16.

20. Székely R, Wáczek F, Szabadkai I, Németh G, Hegymegi-Barakonyi B, Eros D, et al. A novel drug discovery concept for tuberculosis: Inhibition of bacterial and host cell signalling. Immunol Lett. 2008;116:225–31.

21. Lecina D, Gilabert JF, Guallar V. Adaptive simulations, towards interactive protein-ligand modeling. Sci Rep. 2017;7.

22. Wang Q, Wang L, Zhang Y, Zhang XL, Zhang L, Shang W, et al. Probing the Allosteric Inhibition Mechanism of a Spike Protein Using Molecular Dynamics Simulations and Active Compound Identifications. J Med Chem. 2022;65:2827–35.

Chapter 2

Integrating Virtual Screening to aid Tuberculosis Drug Development-Mini-review

Kudakwashe Nyambo[1], Kudzanai Ian Tapfuma[1], Francis Adu-Amankwaah[1], Lucinda Baatjies[1], Idah Sithole Niang[2], Liezel Smith[1], Krishna Govender[4,5], Mkhuseli Ngxande[3], and Vuyo Mavumengwana[1], *

[1] *DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town; K.N: knyambo@sun.ac.za; F.N: fa@sun.ac.za; K.I.T: kudzanait@su.ac.za; L.B: lbaatjies@sun.ac.za; L.S: liezels@sun.ac.za; V.M: vuyom@sun.ac.za*

[2]*Department of Biotechnology and Biochemistry, University of Zimbabwe, B064, Mount Pleasant, Harare, Zimbabwe; sitholeidah2015@gmail.com*

[3]*Computer Science Division, Department of Mathematical Sciences, Faculty of Science University of Stellenbosch, Matieland, South Africa; ngxandem@sun.ac.za*

[4]*Department of Chemical Sciences, University of Johannesburg, P. O. Box 17011, Doornfontein Campus, 2028, Johannesburg, South Africa; krishnag@uj.ac.za*

[5]*National Institute for Theoretical and Computational Sciences (NITheCS), South Africa*

***** Corresponding author.

Email address: *vuyom@sun.ac.za;* Tel: +27 718502949

2.1 Abstract

Tuberculosis (TB) is a chronic disease caused by *Mycobacterium tuberculosis* (*M. tb*) and is the leading cause of mortality in humans. Single nucleotide polymorphisms (SNPs) and insertion-deletions to the genes encoding enzymes that activate drugs and drug macromolecular targets drive the primary mode of drug resistance. Furthermore, nonreplicating *M. tb* contributes to the lengthy time required for the treatment and serves as a reservoir from which drug-resistant bacteria emerge. Drug-resistant strains pose a significant threat to TB control programs because these

strains increase the chances of relapse when the treatment regimen is no longer effective. The primary focus of this review is to highlight the integration of virtual screening, *in-vitro* and *in-vivo,* in developing new strategies to address the rapid increase in drug-resistant *M. tb* strains.

Keywords: *Mycobacterium tuberculosis*, molecular dynamics simulations, molecular docking, small molecules, natural products, quantitative structure-activity relationships, drug targets

2.2 Selection of druggable essential biochemical pathway.

Numerous serious infectious disease states are caused by an infestation of a foreign organism, with *M. tb* being one example. *M. tb* is exclusively transmitted through inhaling droplets infested with bacilli, which are either coughed or sneezed by a person with an active pulmonary infection [1]. Once inhaled, *M. tb* reaches the lung microenvironment and intimately interacts with soluble constituents of lung mucosa after which the pathogen-associated molecular patterns on the *M. tb* cell envelope are recognized by macrophages and other phagocytic cells. The recognition results in the internalization of the pathogen by the alveolar macrophages. Alveolar macrophages are the target cells of the tubercle bacillus infection where it adapts and exponentially proliferates for 2-3 weeks [2].

The delicate interplay between the tubercule and host during the first encounters determines the outcome of TB, that is, either progression to pulmonary infection or clearance by the host. Cell-mediated immunity such as oxidative and inflammatory responses, autophagy, apoptosis, antigen processing, and presentation from M1 macrophages and *M. tb*-specific CD4 T cells play a fundamental role in fighting *M. tb* infection in the host [3]. *M. tb* encounters distinct harsh conditions within the macrophage environment, for instance, acidic pH, nutrient deprivation, and

oxidative and nitrosative stresses [4]. However, *M. tb* is constantly evolving diverse and unique mechanisms of circumventing host defenses, including preferentially targeting and inhibiting phagolysosome formation and manipulating host signaling pathways a hallmark of chronic tuberculosis [5–7].

Studies have revealed that remediation of the disease state is often achieved by killing the tubercle bacillus by inhibiting its molecular machinery used for cell wall biogenesis [43], sensing external signals, and inhibiting the formation of the phagolysosome. Therefore, understanding the upregulated biochemical pathways used by viable *M. tb* cells to survive and to metabolically adapt to harsh host environments may be the key to solving the challenges faced in TB treatments. Essential proteins that play a critical role in driving the bio-catalysis processes in the metabolic pathways of *M. tb* may be targeted to eliminate or eradicate the bacteria. Furthermore, knowledge of the pathophysiology of the mycobacteria after inhibition of the targeted essential proteins is critical. A daring approach to obtaining potent *M. tb* drugs is targeting druggable essential enzymes whose inhibition has not yet been fully established. The probability of discovering and identifying an active compound against a novel target in preclinical development is estimated to be 3% but as for an established drug target, it is approximately 17%. Attempting to uncover a novel mycobacterial drug target can lead to the identification of drugs that possess novel properties that can potentially make a huge difference in treating drug-resistance diseases. Numerous classes of drugs with different modes of action are currently in development (Figure 2.1).

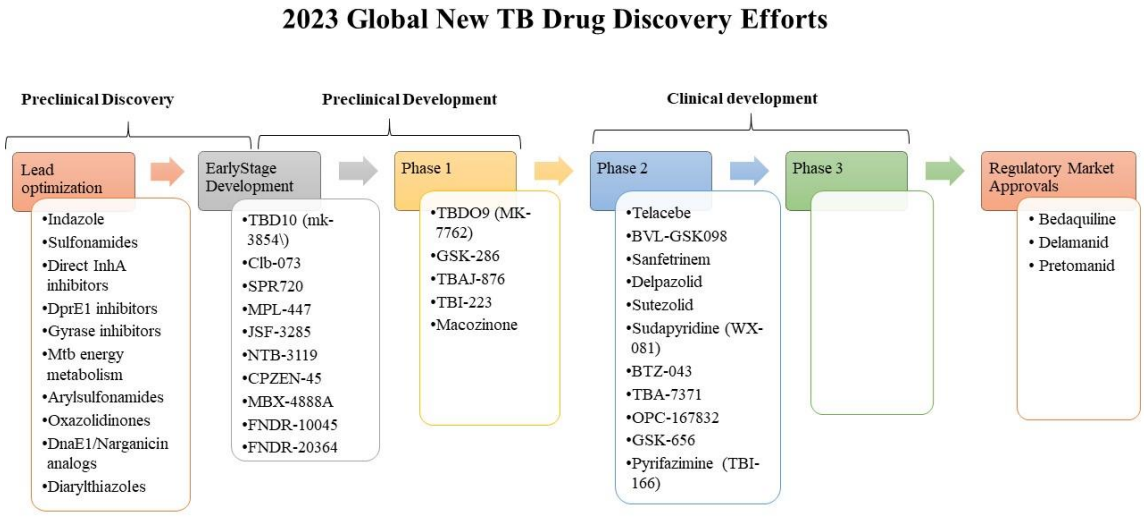


Figure 2.1. Current TB discovery efforts (accessed at https://www.newtbdrugs.org/news 14/6/2023**)**

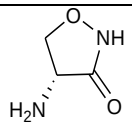
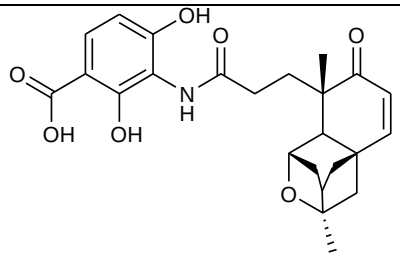
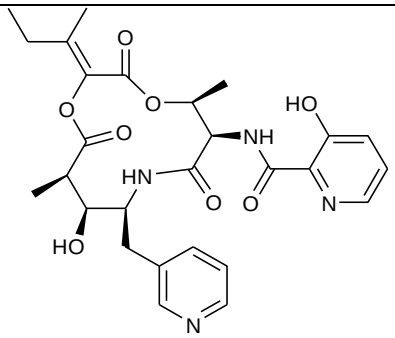
2.3 Accelerating lead scaffolds discovery from natural products with computational tools

Lead compounds in TB drug discovery can be acquired from a chemical space that comprises of astronomical volumes of bioactive compounds (approximately 10^{23}-10^{60}) [8]. However, high throughput screening of the entire chemical space to obtain TB hit compounds is basically impossible. In this regard, studies have zoned in on specific parts, for instance, the natural products portion of the chemical space, as a source of potential lead anti-TB scaffold [9, 10]. Nature is an excellent reservoir of novel drug precursors and approximately half of the drugs approved between 1994 and 2007 were derived from natural products (NPs), with bacteria contributing to more than 30 % of the approved small molecules. Actinomycetes are known to be one of the most efficient producers of secondary metabolites (antibiotics, siderophores, and biosurfactants) that have high medicinal applications [11, 12]. Studies reported the extraction and isolation of bioactive

metabolites from actinomycetes that have anticancer, antibacterial, antifungal, immunomodulating, and herbicidal properties. There are several NPs used as anti-*M. tb* agents in the current pharmacopeias, including, aminoglycosides (amikacin, streptomycin, and kanamycin), ansamycin (rifamycins analogs), and peptides (capreomycin) [13, 14]. Earlier studies have also demonstrated anti-*M. tb* activity from plants and fungi [15–17]. Competition between microorganisms in an environmental niche leads to the evolution of metabolic processes to enhance the fitness of a particular microorganism, particularly, in the production of various secondary metabolites with a broad-spectrum antimicrobial action and different antimicrobial modes of action. NPs from *Streptomyces* sp have been investigated for their potential *M. tb* activity in an earlier study [18]. The authors revealed that compound (**1**) produced by *Streptomyces* sp possesses anti-*M. tb* activity by targeting the cell wall, D-alanine racemase (Alr) and D-alanine:D-alanine ligase (Ddl) with a minimum inhibition concentration (MIC) of 14-900 μM (Table 2.1). However, mutations in various *M. tb* genes, including *rv3423c*, *rv0221*, *rv1403c*, *gabD2* (*rv1731*), *sugI* (*rv3331*), *hisC2* (*rv3772*), etc., have been reported to confer resistance to compound (**1**) [19]. In another study, compound (**2**) isolated from *Streptomyces platensis* was observed to inhibit *M. tb* KasA and KasB with a MIC of 12 μg/mL [20]. Another study reported the anti-*M. tb* activity (MIC = 0.39 μg/ml) of compound (**3**) isolated from *Streptomyces pyridomyceticus* [21]. The authors also revealed that the mode of action was through competitive inhibition of InhA of the NADH- binding site [21].

Table 2.1. Natural products inhibitors of *M. tb*.

N0#	Compound ID	Structure	Activity	Reference

1	D-Cycloserine		MIC (*M. tb*) =14-900 µM	[18]
2	platensimycin		MIC (*M. tb* CDC1551) = 12 µg/mL	[20]
3	pyridomycin		MIC (*M. tb* H37Rv) = 0.39 µg/mL MIC (*M.smeg*) = 0.78 µg/mL	[21]

The traditional "top-down" discovery of novel bioactive compounds from a pool of NPs produced by a microbial consortium involves: the isolation and culturing of microorganisms under favorable laboratory conditions to induce NP production, testing the chemotherapeutic potential of the NPs produced, fractionation of the NPs in the form of crude extracts, chemotherapeutic efficacy assays of the fractions against microbial or human disease models, and purification of active NP/s through multi-fractionation runs [12, 22, 23]. However, the major drawbacks of this gold standard approach include that it is time-consuming, expensive and often results in re-isolation of known compounds produced by microbial species. Furthermore, some bioactive secondary metabolites are produced in trace quantities, thus, detection, purification, and bio-assaying are difficult. In addition, the secondary metabolites produced by bacteria or fungi under environmental micro-stressors are different from those synthesized under laboratory conditions [24]. Previous studies have revealed

that the integration of multiple omics-based strategies such as metabolomics, proteomics, genomics, and meta-omics (machine learning-based tools) has immensely transformed the discovery process from a "top-down" approach to modern multi-omics-based targeted mining. The production of the bioactive secondary metabolites is catalyzed by the multi-modular enzymes assembly that encode (BGCs) [25]. Microbial genome mining enables the deeper exploration of the dark chemical space comprising of cryptic BGCs that encode enzymes that are involved in unprecedented chemical transformations to produce novel bioactive compounds [26, 27]. The integration of genome mining in NP discovery has led to the identification of diverse novel NP biosynthetic pathways, this allows the prioritization of BGCs that encode the production of compounds possessing several bioactivities including anti-TB.

Computational strategies such as well-built statistical methods or artificial intelligence aid the analyses of crucial biological properties of specified functional groups and optimization of the chemical structures to enhance or decrease the desired endpoints [28, 29]. One such strategy is virtual screening which is based on two approaches: ligand-based screening and structure-based screening. Ligand-based screening is based on the hypothesis that structural similarities to a known active ligand should yield similarly active compounds [30, 31]. While structure-based virtual screening is based on a hypothesis of the shape and charge density of the binding pocket that defines what features a complementary ligand should possess [32–34]. Thus, knowledge of one or more known inhibitors for a target or about the three-dimensional structure of the target protein is used in narrowing down large libraries of compounds to small sets comprising only hit compounds, thereby saving screening resources. This form of virtual screening is regarded as a focused screen, which is likely to find the scaffolds with the highest probability of binding to a target molecule.

QSARs models (quantitative structure activity relationship) are mathematical models constructed to reveal active compounds against selected specific biological targets from a chemical library [35, 36]. In other words, it is a regression or a classification technique that quantitatively associates the molecular descriptors derived from chemical scaffolds and their respective endpoint. So far, numerous QSAR models have been created for various endpoints and diverse model constructions techniques, for instance, the stringency of data-processing descriptor types, learning methods, and evaluation metrics have been implemented to achieve this [37–40]. Some of the classical machine learning algorithms have been successfully implemented for developing QSAR models, for instance, support vector machine (SVMs), Naïve Bayesian classifiers, K-Nearest neighbors (K-NN), Random Forest, etc. [28, 41, 42]

2.4 Targeting Cell envelope synthesis machinery using an omics approach

The *M. tb* cell envelope plays a critical role in the delicate immunomodulatory interplay between the bacterium and the host. The mycobacterial cell envelope is a complex permeability barrier that provides natural protection against various external toxins and antibiotics, osmotic protection, mechanical support, and virulence [43]. The envelope of *M. tb* consists of two distinct parts: the plasma membrane and the cell wall. The presence of mycolic acids in the cell wall core is of interest to drug discovery because it forms a hydrophobic permeability barrier which contributes to endogenous resistance of *M. tb* to many drugs [44, 45]. The biochemical pathways of the biosynthesis of mycolic acid constitute more than 20 different multi-enzyme complexes that are well-defined, and the pathway serves as an important reservoir of targets for the development of new TB drugs to combat drug resistance (Figure 2.2) [46].

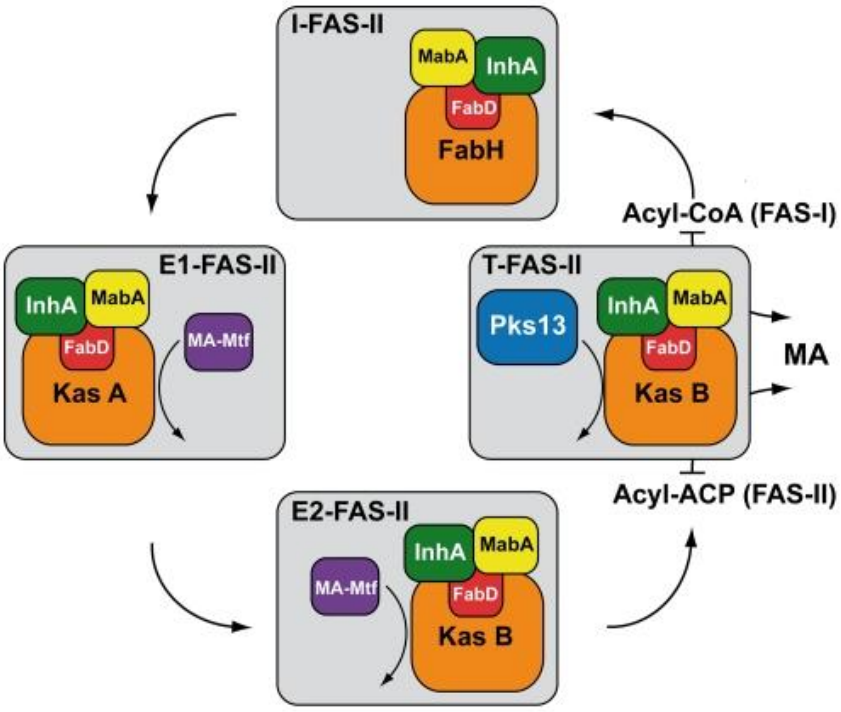
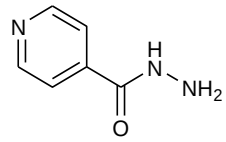
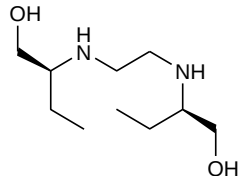
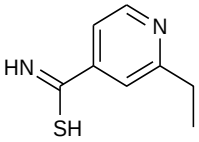


Figure 2.2. Mycolic acid biogenesis pathway comprising of FAS-I and FAS-II systems that are critical for the survival of mycobacteria. FAS-I and FAS-II systems are responsible for the De novo synthesis and elongation. The enzymes involved in the catalysis include InhA (Enoyl ACP reductase), HadABC (β-Hydroxyacyl-ACP dehydratase), KasA and KasB (β-Ketoacyl-ACP synthase), MabA (β-Ketoacyl-ACP reductase), mtFabD (malonyl Co-A-ACP transacylase), and mtFabH (β-Ketoacyl ACP-synthase-III). Adopted from [47].

Several approved drugs available on the market target the mycolic acid biogenesis pathway, including isoniazid [48], ethionamide [49], isoxyl [50], and thiacetazone [50]. The development of these prodrugs was not a targeted approach; hence, they require metabolic activation from specific activating enzymes from *M. tb* such as catalase-peroxidase. However, a rise in the mutation of *katG*, *kasA*, and *inhA* genes encoding pro-drug activating enzymes complicates TB treatment (Table 2.2) [51] . Thus, it is crucial to develop new therapeutic agents that have a targeted mode of action and that do not require prior activation.

Table 2.2. Inhibitors of mycolic acid synthesis and genes conferring resistance to isoniazid, ethambutol, and ethionamide.

Drug Structure	Target/s	Gene/s involved in resistance	Role	References
Isoniazid 	Enoyl-(acyl-carrier-protein) reductase including, catalase peroxidase, NADH-dependent enoyl ACP, 3-Oxoacy ACP, βKetoacyl ACP)	*katG*, *inhA*, *kasA*, *Ndh*, *AhpC*, *niA*, *FadE24*, *FabG1*	Pro-drug conversion, modification overexpression of drug target due to mutations and altered efflux pump activity.	[51]
Ethambutol 	Arabinosyl transferase (inhibition of arabinogalactan synthesis)	*embA*, *embB* ,*embC*, *embR*, *rmLD*, *iniA*, *ubiA*	Change and overexpression of drug target; and altered efflux pump activity	[48] [52]
Ethionamide 	Inhibition of mycolic acid synthesis by binding to the ACP reductase InhA (disrupts cell wall biosynthesis)	*ethA*, *ethR*, *KasA*, *inhA* *inhA pro*	alteration and over-expression drug target due mutation	[49] [53]

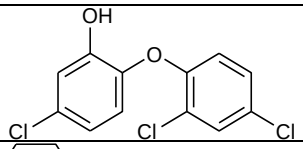
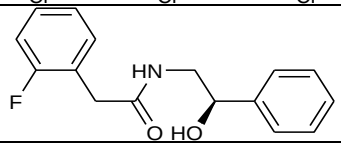
Various proteomics studies have crystallized InhA with small molecule inhibitors that can be accessed on the protein data bank (https://www.rcsb.org). The exploration of the active sites has revealed that some of the small molecule inhibitors occupy the fatty-acyl site and cofactor binding pocket around Tyr158 of InhA [54]. The participation of the conserved residue Tyr158 in hydrogen bonding is regarded to be essential in the development of some InhA inhibitors [55]. Some of these

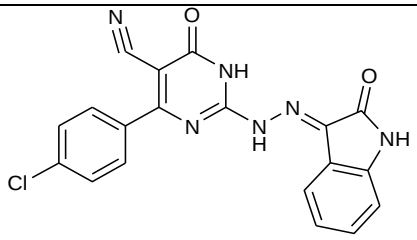
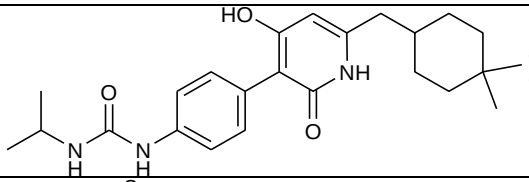
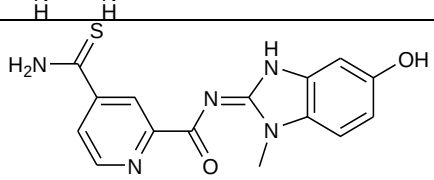
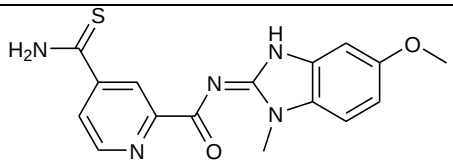
small molecule inhibitors include acetamides, isatin-pyrimidine hybrid derivatives, pyrroles, pyrrolidine carboxamides analogs, hydrazones, triazoles, thiadiazoles, methylthiazoles, and thiourea-based molecules. Other studies have reported small molecules that directly target InhA, including arylamide, diphenyl ethers derivative, and triclosan (**4**) (Table 2.3). In the quest to discover direct InhA inhibitors, a total of 167 550 natural compounds were virtually screened using the Virtual Screening Workflow (VSW) in a Schrödinger suite [56]. The authors reported that compound (**5**) (Table 2.3) exhibited a good extra precision docking score, high affinity to InhA binding site (−109.233 kJ/mol), good pharmacokinetic profile. The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) analysis that describes the reactivity of molecules based on accepting electrons showed that compound **5** exhibited good structural stability [56]. Thus, the compound was regarded as a potential lead that needs further validation of *in-vitro* and *in-vivo* activity against *M. tb.*

In a recent study by Khalifa et al. [54], the authors performed molecular hybridization of isatin-pyrimidine analogs and yielded compound (**6**) (Table 2.3), which was reported to possess antimycobacterial activity against H37Ra (MIC = 0.12 µg/mL), MDR *M. tb* ATCC 35822 (MIC = 0.48 µg/mL) and XDR *M. tb* RCMB 2674 (MIC = 3.9 µg/mL) [54]. Further compound (**6**) inhibited InhA with an IC_{50} of 0.6 ± 0.94 µM. To explore the binding mechanism, molecular docking, and molecular dynamics simulations were performed. The compound (**6**)-InhA complex was revealed to be a relatively stable complex and interacted with Ala198 and Ile202 residues in the hydrophobic pocket of InhA. Phusi et al. [33] performed virtual screening (comprising of Lipinski's five rules, ADME (absorption, distribution, metabolism, and excretion), molecular docking, MD simulations, and *ab initio* fragment molecular orbital calculations) of 128 scaffolds

(4-hydroxy-2-pyridone derivatives) against InhA [33]. The authors observed that compound (**7**) (Table 3) exhibited hydrogen bonds with Tyr158, Thr196, and NADH of InhA and an IFIE of −127.0 kcal/mol, thus, proposed that it may be a potential inhibitor that tightly binds to InhA-NADH [33]. A new class of benzimidazole and imidazoles added to ethionamide analogs that possess antimycobacterial activity was reported previously [57]. Amongst the analogs, (**8**) and (**9**) (Table 2.3), exhibited a MIC of 0.27 ± 0.05 µM and 0.35 ± 0.09 µM against *M. tb* H37Rv, respectively. Further, the efficacy of (**8**) and (**9**) was evaluated against InhA overexpressing *M. tb* strain and they showed MIC values of 0.54 ± 0.03 µM and 0.72 ± 0.19 µM, respectively. The *in-vitro* assays were complemented by molecular docking studies, which provided critical information on the protein-ligand interaction. The docking studies revealed the predicted lowest energy confirmations and the types of interactions between the ligands and InhA, including hydrophobic, hydrogen bond, pi-pi, and pi-alkyl interactions. By incorporating virtual screening techniques during the lead discovery and optimization phase, researchers can efficiently identify, and prioritize scaffolds with interesting properties, thus, speeding up the drug discovery process [57].

Table 2.3. Inhibitors of *M. tb* InhA

N0 #	Compound ID	Structure	Activity	Reference
4	triclosan		MIC = 14-900 µM	[18]
5	Enamine 57340		Binding free energy = −109.233 +/- 84.34 kJ/mol	[20]

6	(E)-4-(4-chlorophenyl)-6-oxo-2-(2-(2-oxoindolin-3-ylidene)hydrazineyl)-1,6-dihydropyrimidine-5-carbonitrile		IC_{50} = 0.6 ± 0.94 µM H37Ra (MIC = 0.12 µg/mL) MDR (MIC = 0.48 µg/mL)	[54]
7	NITD-916–3		IFIE = −127.0 kcal/mol	[33]
8	3i		H37Rv (MIC = 0.27 ± 0.05 µM DS = −9.4 3 kcal/mol	[57]
9	3j		H37Rv (MIC = 0.35 ± 0.09 µM DS = −9.16 kcal/ mol	[57]

DS = Docking score; nd = structure not available

2.5 Targeting Pks13

Polyketide synthase 13 (Pks13) is an essential protein in *M. tb* that catalyzes the final assembly step of mycolic acid biosynthesis to form the direct precursors of mycolic acids by the Clasein-type condensation of a C_{26} α-alkyl branch and $C_{40\text{-}60}$ [58, 59]. Structural proteome analysis of Pks13 revealed that five domains are required for the condensation and release of the polyketide: N-terminal acyl carrier protein domain (N-ACP), β-ketoacyl-synthase domain (KS), acyltransferase domain (AT), C-terminal acyl carrier protein domain (C-ACP), and thioesterase domain (TE) [60]. Thioesterase domain (TE) catalyzes the cleavage of the thioester bond to form an ester linkage between mycolic β-ketoester and the hydroxyl group of Ser1533 of the TE domain. Thus, inhibiting the activity of the TE domain of Pks13 disrupts the mycolic acid pathway, consequently,

the tubercule growth and reproduction are inhibited [61, 62]. Research efforts investigating inhibitors of Pks13-TE have identified and reported different classes, which belong to benzofuran derivatives [61, 62], coumestan derivatives [63, 64] and thiophenes [65]. A study by Lun et al. [63] investigated the antimycobacterial activity of coumestan derivatives using i*n-vitro* assays along with *in-silico* pharmacokinetics analysis. Among the compounds assayed, compound (**10**) (Table 2.4) exhibited a minimum bactericidal concentration (MBC) of 0.0039 to 0.0078 µg/mL. The authors further revealed that *pks13* mutants of *M. tb* with single nucleotide polymorphisms showed resistance to compound **10**, thus, proving the targeting of Pks13 as the mode of action [63]. In this regard, compound **10** is a hit that may further be validated *in-vivo* and *ex-vivo.*

An *in-silico* study screened a diverse natural product library comprising 6208 molecules against Pks13-TE [66]. From this screening, three compounds (**11**, **12**, and **13** Table 2.4) were identified as hits. Molecular dynamics simulations showed the complexes formed with Pks13-TE were stable. Furthermore, the researchers computed the binding free energy (MM-GB\PBSA) ($\triangle G_{bind}$) which revealed strong affinity with a mean energy of approximately −75 kcal/mol observed for all three compounds. The pharmacokinetic properties of the compounds were within the acceptable ranges, thus, the authors concluded that compounds (**11**, **12**, and **13**) can be validated *in-vitro.*

A structure-activity relationship (SAR) research attempted to reduce the cardiotoxicity of a Pks13 inhibitor compound (**14,** hERG IC_{50} = 6.9 µM) [67]. The addition of 2-oxa-6-azaspiro[3.4]octane to compound (**14**) yielded compound (**15)** (Table 2.4). The authors highlighted that compound (**15**) maintained the potency against Pks13 (IC_{50} = 0.3 µM) but reduced the efficacy against H37Rv (MIC = 0.67 µM). Furthermore, compound (**15**) exhibited reduced *in-vitro* cardio-toxicity but

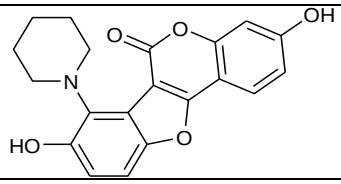
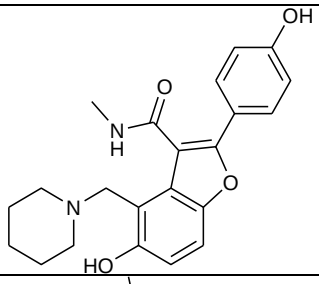
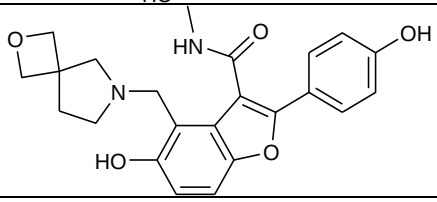
remained cardio-toxic *ex-vivo* [67]. Thus, further structural optimization is required to reduce *ex-vivo* cardiotoxicity.

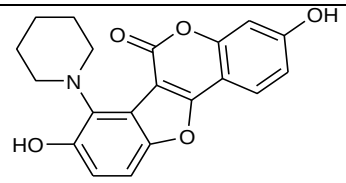
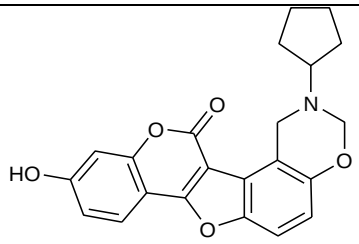
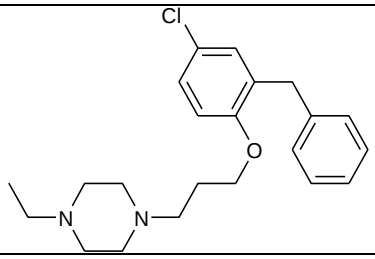
Earlier SAR studies focused on designing a tetracyclic coumestan scaffold (**10**) that possessed antimycobacterial activity. Subsequently, this scaffold was further modified to yield 1,3-oxazine-containing coumestan (**17**) (Table 2.4) [59, 64, 68]. However, it was discovered that compound (**10**) exhibited cardiotoxicity, thus, raising concerns regarding its safety profile. Thus, the researchers were prompted to modify the compound in order to address the cardiotoxicity. Compound (**17**) was evaluated for its antimycobacterial efficacy against various *M. tb* clinical isolates, including H37Rv, DS-TB (V4207), MDR-TB (KZN494 and V2475), and XDR-TB (TF274 and R506). The authors also evaluated its cardiotoxicity by testing its inhibition of the human ether-a-go-go-related gene (hERG) and its potential to bind to Pks13-TE. Compound (**17**) exhibited inhibitory activity against all *M. tb* strains, with MIC values of 0.0039 μg/mL for H37Rv, MDR-TB (KZN494 and V2475), and XDR-TB (TF274 and R506), and 0.0078 μg/mL for DS-TB (V4207). Moreover, it demonstrated an improved cardio cytotoxicity profile (IC_{50} > 30 μM) compared to compound (**10**) (hERG IC_{50} = 0.52 μM). This improved profile is attributed to the presence of the oxazine ring, which disrupts interactions between protonated piperidine and the protein side chain (Y652) [69]. Additionally, compound (**17**) was reported to possess a high affinity for Pks13-TE [69].

In another study conducted by Taira et al. in 2022, a virtual screening of 154,188 compounds was performed using molecular docking against Pks13-TE. The researchers then assessed the stability of the formed complexes using molecular dynamics (MD) simulations and proceeded to evaluate the *in-vitro* efficacy of the identified hit compounds [70]. Among the selected compounds,

compound (**18**) (Table 2.4) exhibited *in-vitro* anti-*Mycobacterium smegmatis* (*M. smeg*) IAM 12065 activity but did not show any activity against *Escherichia coli* JM109 strains. Additionally, the authors reported that Pks13-TE-compound (**18**) complex was relatively stable through MD simulations [70] .

Table 2.4. Inhibitors of Pks13

N0#	Compound ID	Structure	Activity	Reference
10	tetracyclic coumestan scaffold		MBC = 0.00389 to 0.0078 µg/ml	[63]
11	BBB_26582140	nd	DS = − 11.25 kcal/mol $\triangle G_{bind}$ = -75 kcal/mol	[66]
12	BBD_30878599	nd	DS = − 9.87 kcal/mol $\triangle G_{bind}$ = -75 kcal/mol	[66]
13	BBC_29956160	nd	DS = − 9.33 kcal/mol $\triangle G_{bind}$ = -75 kcal/mol	[66]
14	TAM16		H37Rv (MIC = 0.08 µM) Pks13 IC_{50} = 0.3 µM hERG IC_{50} = 6.9 µM	[67]
15	3j		H37Rv (MIC = 0.67 µM)	[67]

10	tetracyclic coumestan scaffold		(MIC = 0.0039 µg/mL) hERG IC_{50} = 0.52 µM Pks13 IC_{50} = 0.3 µM hERG IC_{50} = > 30 µM	[59, 64, 68]
17	oxazine-coumestan analog		H37Rv & MDR KZN494 MDR V2475, (MIC = 0.0039 µg/mL) hERG IC_{50} = 0.67 µM Orally bioavailable in mice	[59, 64, 68]
18	1-[3-(2-Benzyl-4-chlorophenoxy)propyl]-4-ethylpiperazine		DS = -9.7 kcal/mol IC_{50} (IAM 12065) = 8.2 µM	[70]

DS = docking score, nd = the structure was not available

2.6 Targeting signal transduction machinery

M. tb faces a wide range of harsh conditions during the parasitism of macrophages. Its survival in such environments heavily relies on sensing signals from the external environment through extracellular or intracellular sensor domains [71]. Within *M. tb*, Hanks-type serine/threonine kinases play the predominant role in phosphorylation-based transmembrane signal transduction [32]. This reversible phosphorylation activity of serine/threonine kinases is a broadly conserved mechanism of transmembrane signaling in *M. tb* that regulates cell physiology in response to external stimuli. In this regard, blocking the activity of the serine/threonine kinases in *M. tb* has significant consequences. Various metabolic processes enabling the evasion of active immune

surveillance are negatively affected (Table 2.5). Consequently, this inhibition results in impaired bacterial growth, thus, facilitates the efficient clearance of the bacteria by immune cells. Therefore, serine/threonine kinases are legit targets that can be targeted for developing new therapeutic agents aimed at combating TB [72].

Table 2.5. A summary of the properties and function of *M. tb* Serine/threonine kinase.

Name	ORF	MW	Regulatory role	Unique features	Reference
PknA	Rv0015c	45 598	Cell division	-	[72] [73]
PknB	Rv0014c	66 511	Cell division	PonA domain	[74] [73]
PknD	Rv0931c	69 514	Phosphate transport	β-propeller, PQQ domain	[75]
PknE	Rv1743	60 513	Membrane transport	-	[76]
PknF	Rv1746	50 669	Membrane transport	-	[77, 78]
PknG	Rv0410c	81 579	Stationary phase metabolism	Trx motif, TPR motif	[79, 80]
PknH	Rv1266c	66 755	Arabinan metabolism	AfsK like	[81]
Pknl	Rv2914c	61 806	Cell division	Asn in active site	[82]
PknJ	Rv2088	61 564	x	-	[83]
PknK	Rv3080c	119 420	Transcription, secondary metabolites	PDZ and AAA domains	[84]
PknL	Rv21176	42 803	transcription	-	[82]

2.7 Targeting PknA and PknB

PknB is an essential protein involved in the growth and is conserved in clinical strains of *M. tb*, thus, making it an attractive druggable target for developing anti-*M. tb* agents [72]. Previous studies identified hit and lead bio-pharmacophores, including quinazoline derivatives, pyrimidine derivatives, mitoxantrone, staurosporine derivatives, (E)-4-oxo-crotonamide derivatives, and

thiophene amide derivatives as PknB inhibitors [85]. In an earlier study, the *in-vitro* whole cell and targeted efficacy of 12000 compounds obtained from a kinase inhibitor library against *M. tb* and *Corynebacterium glutamicum* were evaluated [86]. The authors initially performed targeted inhibitory assays against the PknA and PknB. This led to the discovery of three compounds: VI-9376 (**19**), VI-7777 (**20**), and VI-18469 (**21**), all of which exhibited inhibitory efficacy against *M. tb* while not inhibiting PknA. Furthermore, the structure-activity relationship (SAR) of VI-9376 (**19**) revealed that it is a potential inhibitor for decaprenylphosphoryl-β-D-ribose 2′-epimerase (DprE1), and the nitro group is responsible for its anti-TB efficacy (Table 2.6). The authors also reported that the SAR optimization of VI-18469 (**21**) showed that the substitution of bromophenyl with 1-ethylpiperidine or 4-ethylmorpholine increases the cytotoxicity activity of the formed derivatives (Table 2.6). Additionally, the researchers suggested that VI-18469 (**21**) exhibits some structural similarity with clofazimine, which is known to possess anti-TB activity [86].

In a previous rational design study, Hologram and three-dimensional QSAR models accurately predicted quinazoline cores as contributors to high affinity and inhibitory activity against PknB [87]. The researchers conducted MD simulations and binding free energy computations using MM-PBSA and MM-GBSA of quinazoline derivatives, reporting that the ligand's affinity to the binding pocket of PknB was mainly influenced by the hydrophobicity nature of the quinazoline cores (Figure 2.3).

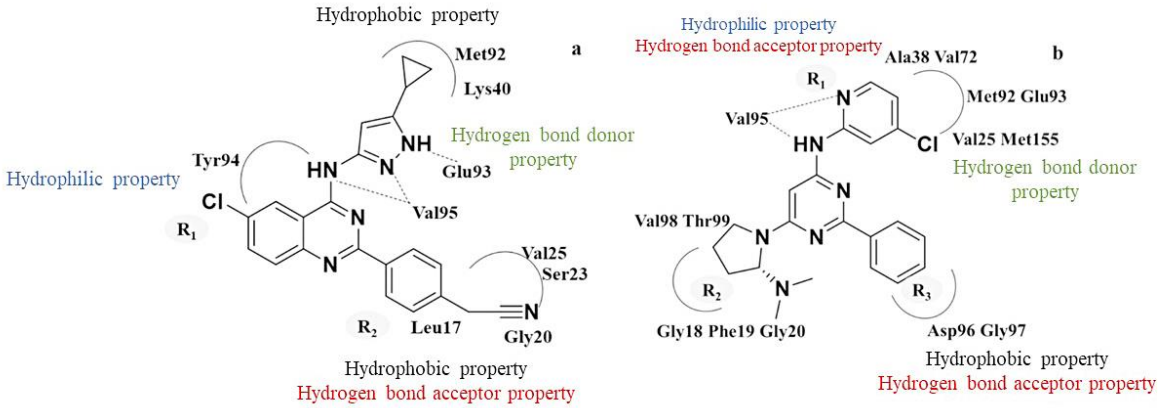


Figure 2.3. The binding properties of quinazoline and pyrimidine derivatives to the binding domain of PknB [87].

Xu and colleagues [74] performed a targeted high-throughput screening of the activity of IMB-YH-8 (**22**) against PknB [74]. They revealed that compound (**22)** exhibited *in-vitro* inhibitory activity of *M. tb* (MIC (H37Rv) = 0.25 μg/ml) and the phosphorylation activity of PknB (Table 2.6). Transcriptional analysis of *M. tb* treated with IMB-YH-8 showed evidence of disruption of the SigH pathway. Furthermore, molecular docking and thermodynamic studies (Gibbs free energy change (ΔG) = −6.64 kcal/mol) of IMB-YH-8 against PknB revealed favorable binding through hydrophobic interactions and hydrogen bonding [74]. In another study by Thongdee et al. [85], virtual screening of 207 369 compounds against PknB identified two hits compounds AE-848/42799159 (**23**), and AP-124/40904362 (**24**) which exhibited good docking scores (-10.4 kcal/mol, and –13.3 kcal/mol) (Table 2.6) [85]. Furthermore, the hit compounds showed potent *in-vitro* inhibitory activity against PknB with a MIC of 6.2 (μg/mL) [85].

The protein kinase (PknA) is another attractive target for the development of new potent drugs to combat TB. Ideally, these anti-TB agents should selectively inhibit *M. tb* kinases only. This review

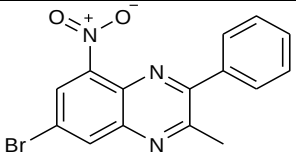
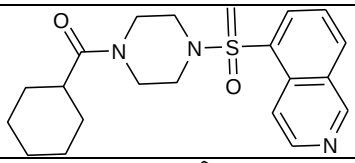
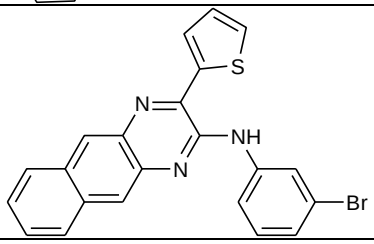
aims to summarize and expand on the findings of recent studies regarding the development and modification of *M. tb* PknA inhibitors. In a previous study by Sundar et al. [88], the affinity of 3176 FDA-approved small molecules against PknA was evaluated using virtual screening tools such as molecular docking, MD simulations, and binding free energy computations [88]. Three compounds namely, ZINC3831425 (**25**), ZINC3871612 (**26**), and ZINC1769096 (**27**), which docked to PknA, were selected for MD simulations and binding free energy calculations based on their docking scores against PknA -12.66 kcal/mol, -12.65 kcal/mol, and -11.78 kcal/mol (Table 2.6). Among the three compounds, ZINC3831425 was predicted to exhibit the highest affinity ($\triangle_{Gbind}$) to PknA (-133.415 ± 35.109 kcal/mol) [88].

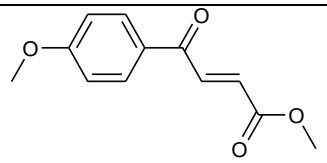
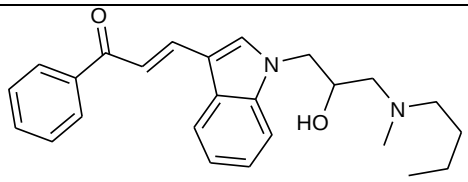
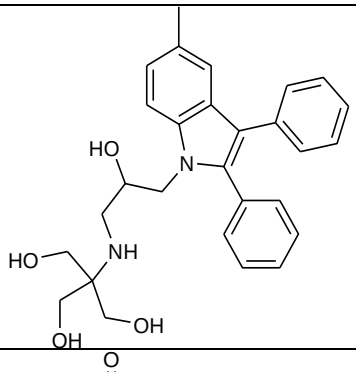
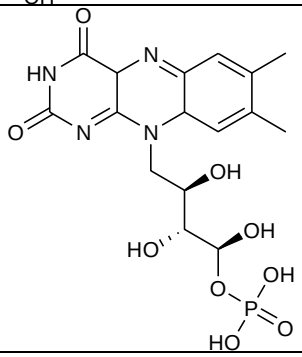
In research conducted by Wang et al. [72], a diverse library of 1078 compounds were screened against PknA and PknB domains. A quinazoline derivative compound (**28**) was identified as a promising molecular starting point for further modification and optimization. Compound (**28**) exhibited inhibition of *M. tb* H37Rv (MIC = 33 μM) and PknB (K_i = 150 nM) (Table 2.6). These findings prompted the researchers to perform further structural modification and optimization [72]. Through a SAR-guided approach, the authors synthesized different analogs of the compound (**28**) for further testing. Notably, one of the derivatives compounds (**29**), with a substituted sulfonamide group, exhibited increased potency against H37Rv (MIC = 4.7 μM) and dual inhibition against both PknA (Ki < 150 nM) and PknB (Ki = 11 nM) (Table 2.6).

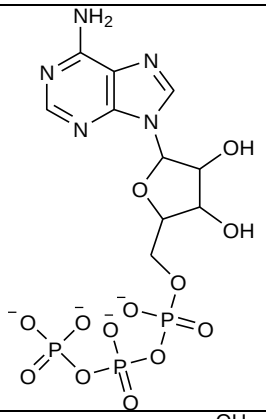
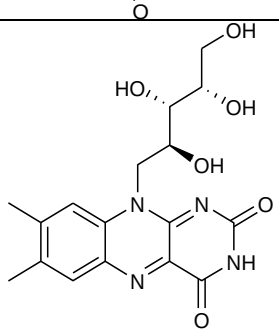
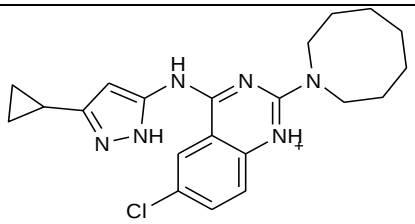
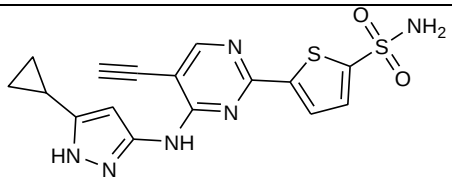
In a separate study by Carette et al. [73], the antimicrobial efficacy of a 5-substituted pyrimidine analog (**30**) was investigated. Compound (**30**) exhibited dual inhibition against PknA (K_i = 0.018 μM) and PknB (K_i = 0.004 μM) (Table 2.6). Additionally, the authors demonstrated the potency of compound (**30**) against *M. tb* H37Ra, with a MIC of 4.5 μM. Further investigations revealed the

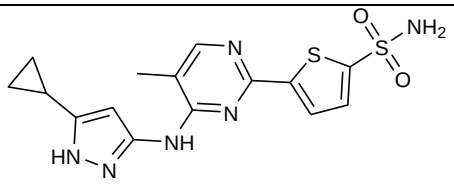
influence of the compound (**30**) on the *M. tb* transcriptome [73]. They observed an early reduction in protein phosphorylation and downregulation of *MtrA*-regulated genes. Conversely, there was increased expression of the SigE regulon, ESXV1 type VII secretion system, and mycobactin biosynthesis. All these findings can be utilized in machine learning-based QSAR modeling for the discovery and synthesis of potent inhibitors of PknA and PknB, thus, accelerating the development of novel therapeutic agents against TB.

Table 2.6. Inhibitors of PknA and PknB

N0 #	Compound ID	Structure	Activity	Reference
19	VI-9376		H37Rv (MIC = 3.1 μM)	[86]
20	VI-7777		H37Rv (MIC = 3.1 μM)	[86]
21	VI-18469		H37Rv (MIC = 6.25 μM)	[86]

22	IMB-YH-8		H37Rv (MIC = 0.25 µg/ml) PknB ΔG = −6.64 kcal/mol	[74]
23	AE-848/42799159		DS = -10.4 kcal/mol PknB (MIC = 6.2 µg/ml)	[85]
24	AP-124/40904362		DS = –13.3 kcal/mol PknB (MIC = 6.2 µg/ml)	[85]
25	ZINC3831425		PknA DS = -12.66 kcal/mol PknA ($\triangle_{Gbind}$ = -133.415 ± 35.109 kcal/mol	[88]

26	ZINC3871612		PknA DS = -12.65 kcal/mol	[88]
27	ZINC1769096		PknA DS = -11.78 kcal/mol	[88]
28	Quinazoline derivative		H37Rv (MIC = 33 µM) PknB (K_i = 150 nM	[72]
29	SCHEMBL116404 3		H37Rv (MIC = 4.7 µM) PknA (K_i < 11 nM) PknB (K_i = 11 nM)	[72]

30		H37Rv (MIC = 4.5 μM) PknA (K_i = 0.018 μM) PknB (K_i = 0.004 μM)

DS = docking score

2.9 Summary and future perspective in drug discovery

Developing and deploying virtual screening approaches in drug discovery is progressively reducing time and costs during drug development. Therefore, it is essential to create multitasking QSAR models that can predict the inhibitory efficacy against multiple *M. tb* essential proteins using different conditions including MIC, K_i, IC_{50}, ($\triangle_{Gbind}$), etc. Some studies have adopted and implemented the Box-Jenkins approach to capture the chemometric information while considering the experimental and theoretical conditions and the different drug targets [89, 90]. Meanwhile, other studies have introduced the construction of sophisticated QSAR models comprising architecture such as combining Transformer and CNN, which enables capturing of intrinsic chemometric data from the SMILES notations of compounds [91]. The advantage of using multi-target virtual screening tools is filtering in the chemical space for versatile *M. tb* inhibitors against diverse essential target macromolecular structures. Biophysical data-driven approaches such as docking, molecular dynamics simulations, and free energy perturbation have been used to extensively validate the QSAR models. Lead compounds often need optimization for their activity against biological targets and various pharmacokinetic parameters, including ADME

characteristics. Predicting these properties in the early stages of drug discovery may increase the chances of the compound not being rejected during clinical trials.

Author Contributions: Conceptualization, V.M., and K.N.; formal analysis, K.N.; investigation, K.N.; resources, V.M., M.N., L.S., K.G., ., and I.SN.; writing original draft preparation, K.N.; writing review and editing, V.M., M.N., K.G., K.I.T., L.B., L.S., and I.S.N.; supervision, V.M., M.G., K.G., L.S, and I.S.N; project administration, L.S.; funding acquisition, V.M., and L.S. All authors have read and agreed to the version of the manuscript.

Funding: This work was supported by the South African Medical Research Association Council (SAMRC), the Centre of Excellence for Biomedical Tuberculosis Research (CBTBR), and the National Research Fund (NRF) [NRF GRANT UID129364]. K.N. is highly indebted to the CTR and CBTBR for financial support.

Ethical Approval: No ethical clearance is needed.

2.8. References

1. Nagpal, priya; Jamal, S.; Singh, H.; Ali, W.; tanweer, S.; Sharma, R.; Grover, A.; Grover, S. Long-Range Replica Exchange Molecular Dynamics Guided Drug Repurposing against Tyrosine Kinase PtkA of *Mycobacterium tuberculosis*. **2020**, doi:10.1038/s41598-020-61132-w.

2. Miggiano, R.; Rizzi, M.; Ferraris, D.M. *Mycobacterium tuberculosis* Pathogenesis, Infection Prevention and Treatment. *Pathogens* 2020, *9*.

3. Rahlwes, K.C.; Dias, B.R.S.; Campos, P.C.; Alvarez-Arguedas, S.; Shiloh, M.U. Pathogenicity and Virulence of *Mycobacterium tuberculosis*. **2023**, *14*, doi:10.1080/21505594.2022.2150449.

4. Watson, K.; Russell, C.D.; Baillie, J.K.; Dhaliwal, K.; Fitzgerald, J.R.; Mitchell, T.J.; Simpson, A.J.; Renshaw, S.A.; Dockrell, D.H. Developing Novel Host-Based Therapies Targeting Microbicidal Responses in Macrophages and Neutrophils to Combat Bacterial Antimicrobial Resistance. *Front Immunol* 2020, *11*.

5. Zeng, J.; Platig, J.; Cheng, T.Y.; Ahmed, S.; Skaf, Y.; Potluri, L.P.; Schwartz, D.; Steen, H.; Branch Moody, D.; Husson, R.N. Protein Kinases PknA and PknB Independently and Coordinately Regulate Essential *Mycobacterium tuberculosis* Physiologies and Antimicrobial Susceptibility. *PLoS Pathog* **2020**, *16*, doi:10.1371/journal.ppat.1008452.

6. Carette, X.; Platig, J.; Young, D.C.; Helmel, M.; Young, A.T.; Wang, Z.; Potluri, L.P.; Moody, C.S.; Zeng, J.; Prisic, S.; et al. Multisystem Analysis of *Mycobacterium tuberculosis* Reveals Kinase-Dependent Remodeling of the Pathogen-Environment Interface. *mBio* **2018**, *9*, doi:10.1128/mBio.02333-17.

7. Ruddraraju, K.V.; Aggarwal, D.; Zhang, Z.Y. Therapeutic Targeting of Protein Tyrosine Phosphatases from *Mycobacterium tuberculosis*. *Microorganisms* 2021, *9*, 1–13.

8. Xu, X.; Dong, B.; Peng, L.; Gao, C.; He, Z.; Wang, C.; Zeng, J. Anti-Tuberculosis Drug Development via Targeting the Cell Envelope of *Mycobacterium tuberculosis*. *Front Microbiol* **2022**, *13*, 1056608, doi:10.3389/FMICB.2022.1056608/BIBTEX.

9. Kirkpatrick, P.; Ellis, C. Chemical Space. *Nature* **2004**, *432*, 823, doi:10.1038/432823A.

10. Martín-González, D.; Bordel, S.; Solis, S.; Gutierrez-Merino, J.; Santos-Beneit, F. Characterization of Bacillus Strains from Natural Honeybee Products with High Keratinolytic

Activity and Antimicrobial Potential. *Microorganisms* **2023**, *11*, 456, doi:10.3390/MICROORGANISMS11020456/S1.

11. Parkinson, E.I.; Tryon, J.H.; Goering, A.W.; Ju, K.S.; McClure, R.A.; Kemball, J.D.; Zhukovsky, S.; Labeda, D.P.; Thomson, R.J.; Kelleher, N.L.; et al. Discovery of the Tyrobetaine Natural Products and Their Biosynthetic Gene Cluster via Metabologenomics. *ACS Chem. Biol.* **2018**, *13*, 1029–1037, doi:10.1021/acschembio.7b01089.

12. Yuan, J.; Wang, L.; Ren, J.; Huang, J.P.; Yu, M.; Tang, J.; Yan, Y.; Yang, J.; Huang, S.X. Antibacterial Pentacyclic Polyketides from a Soil-Derived Streptomyces. *J Nat Prod* **2020**, *83*, 1919–1924.

13. Wu, C.; Zacchetti, B.; Ram, A.F.J.; Van Wezel, G.P.; Claessen, D.; Choi, Y.H. Expanding the Chemical Space for Natural Products by *Aspergillus-Streptomyces* Co-Cultivation and Biotransformation. *Sci Rep* **2015**, *5*, doi:10.1038/srep10868.

14. Sekaggya-Wiltshire, C.; Dooley, K.E. Pharmacokinetic and Pharmacodynamic Considerations of Rifamycin Antibiotics for the Treatment of Tuberculosis. https://doi.org/10.1080/17425255.2019.1648432 **2019**, *15*, 615–618,

15. Kim, T.K.; Hewavitharana, A.K.; Shaw, P.N.; Fuerst, J.A. Discovery of a New Source of Rifamycin Antibiotics in Marine Sponge Actinobacteria by Phylogenetic Prediction. *Appl Environ Microbiol* **2006**, *72*, 2118–2125.

16. Tapfuma, K.I. Anti-Mycobacterial Activity of Superparamagnetic Iron Oxide Nanoparticles Derived from Ascidian Fungal Metabolites and Their Potential to Activate Neutrophils; 2020;

17. Tapfuma, K.I.; Nyambo, K.; Adu-Amankwaah, F.; Baatjies, L.; Smith, L.; Allie, N.; Keyster, M.; Loxton, A.G.; Ngxande, M.; Malgas-Enus, R.; et al. Antimycobacterial Activity and Molecular Docking of Methanolic Extracts and Compounds of Marine Fungi from Saldanha and False Bays, South Africa. *Heliyon* **2022**, *8*, e12406, doi:10.1016/J.HELIYON.2022.E12406.

18. Nyambo, K.; Adu-Amankwaah, F.; Tapfuma, K.I.; Baatjies, L.; Julius, L.; Smith, L.; Ngxande, M.; Govender, K.; Mabasa, L.; Traore, A.; et al. *In-Silico* and *in-Vitro* Assessments of Some Fabaceae, Rhamnaceae, Apocynaceae, and Anacardiaceae Species against *Mycobacterium tuberculosis* H37Rv and Triple-Negative Breast Cancer Cells. *BMC Complementary Med Ther 2023 23:1* **2023**, *23*, 1–21, doi:10.1186/S12906-023-04041-5.

19. Bruning, J.B.; Murillo, A.C.; Chacon, O.; Barletta, R.G.; Sacchettini, J.C. Structure of the *Mycobacterium tuberculosis* D-Alanine:D-Alanine Ligase, a Target of the Antituberculosis Drug D-Cycloserine. *Antimicrob Agents Chemother* **2011**, *55*, 291–301.

20. Chen, J.; Zhang, S.; Cui, P.; Shi, W.; Zhang, W.; Zhang, Y. Identification of Novel Mutations Associated with Cycloserine Resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* **2017**, *72*, 3272–3276, doi:10.1093/JAC/DKX316.

21. Brown, A.K.; Taylor, R.C.; Bhatt, A.; Fütterer, K.; Besra, G.S. Platensimycin Activity against Mycobacterial β-Ketoacyl-ACP Synthases. *PLoS One* **2009**, *4*, e6306, doi:10.1371/JOURNAL.PONE.0006306.

22. Hartkoorn, R.C.; Sala, C.; Neres, J.; Pojer, F.; Magnet, S.; Mukherjee, R.; Uplekar, S.; Boy-Röttger, S.; Altmann, K.H.; Cole, S.T. Towards a New Tuberculosis Drug: Pyridomycin – Nature's Isoniazid. *EMBO Mol Med* **2012**, *4*, 1032–1042, doi:10.1002/EMMM.201201689.

23. Valgas, C.; De Souza, S.M.; Smânia, E.F.A.; Smânia, A. Screening Methods to Determine Antibacterial Activity of Natural Products. *Braz J Microbiol* **2007**, *38*, 369–380, doi:10.1590/S1517-83822007000200034.

24. Latif, Z.; Sarker, S.D. Isolation of Natural Products by Preparative High Performance Liquid Chromatography (Prep-HPLC). *Methods Mol Biol* **2012**, *864*, 255–274, doi:10.1007/978-1-61779-624-1_10.

25. Gurnani, N.; Mehta, D.; Gupta, M.; Mehta, B.K. Natural Products: Source of Potential Drugs. *Afr J Basic Appl Sci* **2014**, *6*, 171–186, doi:10.5829/idosi.ajbas.2014.6.6.21983.

26. Belknap, K.C.; Park, C.J.; Barth, B.M.; Andam, C.P. Genome Mining of Biosynthetic and Chemotherapeutic Gene Clusters in Streptomyces Bacteria. *Sci. Rep.* **2020**, *10*, 2003, doi:10.1038/s41598-020-58904-9.

27. Jackson, S.A.; Crossman, L.; Almeida, E.L.; Margassery, L.M.; Kennedy, J.; Dobson, A.D.W. Diverse and Abundant Secondary Metabolism Biosynthetic Gene Clusters in the Genomes of Marine Sponge Derived Streptomyces Spp. Isolates. *Mar Drugs* **2018**, *16*, doi:10.3390/MD16020067.

28. Su, Z.; Chen, X.; Liu, X.; Guo, Q.; Li, S.; Lu, X.; Zhang, X.; Wang, P.; Dong, L.; Zhao, W.; et al. Genome Mining and UHPLC–QTOF–MS/MS to Identify the Potential Antimicrobial Compounds and Determine the Specificity of Biosynthetic Gene Clusters in Bacillus Subtilis NCD-2. *BMC Genomics* **2020**, *21*, 1–16, doi:10.1186/S12864-020-07160-2/FIGURES/8.

29. Chaube, S.; Goverapet Srinivasan, S.; Rai, B. Applied Machine Learning for Predicting the Lanthanide-Ligand Binding Affinities. *Sci Rep* **2020**, *10*, doi:10.1038/s41598-020-71255-9.

30. Kovács, D.P.; McCorkindale, W.; Lee, A.A. Quantitative Interpretation Explains Machine Learning Models for Chemical Reaction Prediction and Uncovers Bias. *Nat Commun* **2021**, *12*, doi:10.1038/s41467-021-21895-w.

31. Fratev, F.; Sirimulla, S. An Improved Free Energy Perturbation FEP+ Sampling Protocol for Flexible Ligand-Binding Domains. *Sci Rep 2019 9:1* **2019**, *9*, 1–13, doi:10.1038/s41598-019-53133-1.

32. Adeniji, S.E.; Adalumo, O.B. Computational Modeling and Ligand-Based Design of Some Novel Hypothetical Compound as Prominent Inhibitors against *Mycobacterium tuberculosis*. *Futur J Pharm Sci* **2020**, *6*, doi:10.1186/s43094-020-00027-z.

33. Bruch, E.M.; Petrella, S.; Bellinzoni, M. Structure-Based Drug Design for Tuberculosis: Challenges Still Ahead. *Appl Sci 2020, Vol. 10, Page 4248* **2020**, *10*, 4248, doi:10.3390/APP10124248.

34. Phusi, N.; Hashimoto, Y.; Otsubo, N.; Imai, K.; Thongdee, P.; Sukchit, D.; Kamsri, P.; Punkvang, A.; Suttisintong, K.; Pungpo, P.; et al. Structure-Based Drug Design of Novel M. Tuberculosis InhA Inhibitors Based on Fragment Molecular Orbital Calculations. *Comput Biol Med* **2023**, *152*, 106434, doi:10.1016/J.COMPBIOMED.2022.106434.

35. Sun, Q.; Biswas, A.; Vijayan, · R S K; Pierrick Craveur, ·; Forli, S.; Arthur, ·; Olson, J.; Andres, ·; Castaner, E.; Kirby, K.A.; et al. Structure-Based Virtual Screening Workflow to Identify Antivirals Targeting HIV-1 Capsid. *J Comput Aided Mol Des* **2022**, *36*, 193–203, doi:10.1007/s10822-022-00446-5.

36. Halder, A.K.; Dias Soeiro Cordeiro, M.N. QSAR-Co-X: An Open Source Toolkit for Multitarget QSAR Modelling. *J Cheminform* **2021**, *13*, doi:10.1186/s13321-021-00508-0.

37. Neves, B.J.; Braga, R.C.; Melo-Filho, C.C.; Moreira-Filho, J.T.; Muratov, E.N.; Andrade, C.H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front Pharmacol* **2018**, *9*, 1275, doi:10.3389/FPHAR.2018.01275/BIBTEX.

38. Ambure, P.; Halder, A.K.; Gonzaíez Díaz, H.; Nataíia, M.; Cordeiro, D.S. QSAR-Co: An Open Source Software for Developing Robust Multitasking or Multitarget Classification-Based QSAR Models. *J. Chem. Inf. Model* **2019**, *59*, 44, doi:10.1021/acs.jcim.9b00295.

39. Gumede, N.J. Pathfinder-Driven Chemical Space Exploration and Multiparameter Optimization in Tandem with Glide/IFD and QSAR-Based Active Learning Approach to Prioritize Design Ideas for FEP+ Calculations of SARS-CoV-2 PLpro Inhibitors. *Molecules* **2022**, *27*, 8569, doi:10.3390/MOLECULES27238569/S1.

40. Chhatbar, P.; Pambhar, K.; Khedkar, V.; Shah, A.; Khunt, R. In Silico (3D-QSAR) Designed, Study, Synthesis and Anti-Tubercular Evaluation of Pyrazolo-Pyrimidine Derivatives. *Antiinfect Agents* **2019**, *18*, 135–143, doi:10.2174/2211352517666190225143923.

41. Alam, S.; Khan, F. 3D-QSAR, Docking, ADME/Tox Studies on Flavone Analogs Reveal Anticancer Activity through Tankyrase Inhibition. *Sci Rep* **2019**, *9*, doi:10.1038/s41598-019-41984-7.

42. Ortiz, C.L.D.; Completo, G.C.; Nacario, R.C.; Nellas, R.B. Potential Inhibitors of Galactofuranosyltransferase 2 (GlfT2): Molecular Docking, 3D-QSAR, and In Silico ADMETox Studies. *Sci Rep* **2019**, *9*, doi:10.1038/s41598-019-52764-8.

43. Tsou, L.K.; Yeh, S.H.; Ueng, S.H.; Chang, C.P.; Song, J.S.; Wu, M.H.; Chang, H.F.; Chen, S.R.; Shih, C.; Chen, C.T.; et al. Comparative Study between Deep Learning and QSAR Classifications for TNBC Inhibitors and Novel GPCR Agonist Discovery. *Sci Rep 2020 10:1* **2020**, *10*, 1–11, doi:10.1038/s41598-020-73681-1.

44. Kuang, W.; Zhang, H.; Wang, X.; Yang, P. Overcoming *Mycobacterium tuberculosis* through Small Molecule Inhibitors to Break down Cell Wall Synthesis. *Acta Pharm Sin B* **2022**, *12*, 3201–3214, doi:10.1016/J.APSB.2022.04.014.

45. Savintseva, L.A.; Steshin, I.S.; Avdoshin, A.A.; Panteleev, S. V.; Rozhkov, A. V.; Shirokova, E.A.; Livshits, G.D.; Vasyankin, A. V.; Radchenko, E. V.; Ignatov, S.K.; et al. Conformational Dynamics and Stability of Bilayers Formed by Mycolic Acids from the *Mycobacterium tuberculosis* Outer Membrane. *Molecules* **2023**, *28*, 1347, doi:10.3390/MOLECULES28031347/S1.

46. Singh, S.; Singh, D.; Hameed, S.; Fatima, Z. An Overview of Mycolic Acids: Structure–Function–Classification, Biosynthesis, and Beyond. *Biol Mycobact Lipids* **2022**, 1–25, doi:10.1016/B978-0-323-91948-7.00016-6.

47. Cantaloube, S.; Veyron-Churlet, R.; Haddache, N.; Daffé, M.; Zerbib, D. The *Mycobacterium tuberculosis* FAS-II Dehydratases and Methyltransferases Define the Specificity of the Mycolic Acid Elongation Complexes. *PLoS One* **2011**, *6*, doi:10.1371/JOURNAL.PONE.0029564.

48. Ge, F.; Zeng, F.; Liu, S.; Guo, N.; Ye, H.; Song, Y.; Fan, J.; Wu, X.; Wang, X.; Deng, X.; et al. In Vitro Synergistic Interactions of Oleanolic Acid in Combination with Isoniazid, Rifampicin or

Ethambutol against *Mycobacterium tuberculosis*. *J Med Microbiol* **2010**, *59*, doi:10.1099/jmm.0.014837-0.

49. Wan, L.; Hu, P.; Zhang, L.; Wang, Z.X.; Fleming, J.; Ni, B.; Luo, J.; Guan, C.X.; Bai, L.; Tan, Y.; et al. Omics Analysis of *Mycobacterium tuberculosis* Isolates Uncovers Rv3094c, an Ethionamide Metabolism-Associated Gene. *Commu Biol 2023 6:1* **2023**, *6*, 1–14, doi:10.1038/s42003-023-04433-w.

50. De Freitas Paulo, T.; Duhayon, C.; De França Lopes, L.G.; Silva Sousa, E.H.; Chauvin, R.; Bernardes-Génisson, V. Further Insights into the Oxidative Pathway of Thiocarbonyl-Type Antitubercular Prodrugs: Ethionamide, Thioacetazone, and Isoxyl. *Chem Res Toxicol* **2021**, *34*, 1879–1889,

51. Penn-Nicholson, A.; Georghiou, S.B.; Ciobanu, N.; Kazi, M.; Bhalla, M.; David, A.; Conradie, F.; Ruhwald, M.; Crudu, V.; Rodrigues, C.; et al. Detection of Isoniazid, Fluoroquinolone, Ethionamide, Amikacin, Kanamycin, and Capreomycin Resistance by the Xpert MTB/XDR Assay: A Cross-Sectional Multicentre Diagnostic Accuracy Study. *Lancet Infect Dis* **2022**, *22*, 242–249, doi:10.1016/S1473-3099(21)00452-7.

52. Riccardi, G.; Pasca, M.R.; Buroni, S. *Mycobacterium tuberculosis*: Drug Resistance and Future Perspectives. *Future Microbiol* 2009, *4*, 597–614.

53. Ushtanit, A.; Kulagina, E.; Mikhailova, Y.; Makarova, M.; Safonova, S.; Zimenkov, D. Molecular Determinants of Ethionamide Resistance in Clinical Isolates of *Mycobacterium tuberculosis*. *Antibiotics* **2022**, *11*, 133, doi:10.3390/ANTIBIOTICS11020133/S1.

54. Khalifa, A.; Khalil, A.; Abdel-Aziz, M.M.; Albohy, A.; Mohamady, S. Isatin-Pyrimidine Hybrid Derivatives as Enoyl Acyl Carrier Protein Reductase (InhA) Inhibitors against *Mycobacterium tuberculosis*. *Bioorg Chem* **2023**, *138*, 106591, doi:10.1016/J.BIOORG.2023.106591.

55. Davoodi, S.; Daryaee, F.; Iuliano, J.N.; Tolentino Collado, J.; He, Y.; Pollard, A.C.; Gil, A.A.; Aramini, J.M.; Tonge, P.J. Evaluating the Impact of the Tyr158 p K a on the Mechanism and Inhibition of InhA, the Enoyl-ACP Reductase from *Mycobacterium tuberculosis* . *Biochem* **2023**, *62*, 1943–1952.

56. Jayaraman, M.; Loganathan, L.; Muthusamy, K.; Ramadas, K. Virtual Screening Assisted Discovery of Novel Natural Products to Inhibit the Catalytic Mechanism of *Mycobacterium tuberculosis* InhA. *J Mol Liq* **2021**, *335*, 116204, doi:10.1016/J.MOLLIQ.2021.116204.

57. Raghu, M.S.; Pradeep Kumar, C.B.; Yogesh Kumar, K.; Prashanth, M.K.; Alshahrani, M.Y.; Ahmad, I.; Jain, R. Design, Synthesis and Molecular Docking Studies of Imidazole and Benzimidazole Linked Ethionamide Derivatives as Inhibitors of InhA and Antituberculosis Agents. *Bioorg Med Chem Lett* **2022**, *60*, 128604, doi:10.1016/J.BMCL.2022.128604.

58. Gavalda, S.; Léger, M.; van der Rest, B.; Stella, A.; Bardou, F.; Montrozier, H.; Chalut, C.; Burlet-Schiltz, O.; Marrakchi, H.; Daffé, M.; et al. The Pks13/FadD32 Crosstalk for the Biosynthesis of Mycolic Acids in *Mycobacterium tuberculosis*. *J Biol Chem* **2009**, *284*, 19255–19264, doi:10.1074/jbc.M109.006940.

59. Zhang, W.; Lun, S.; Liu, L.L.; Xiao, S.; Duan, G.; Gunosewoyo, H.; Yang, F.; Tang, J.; Bishai, W.R.; Yu, L.F. Identification of Novel Coumestan Derivatives as Polyketide Synthase 13 Inhibitors against *Mycobacterium tuberculosis*. Part II. *J Med Chem* **2019**, *62*, 3575–3589.

60. Bon, C.; Cabantous, S.; Julien, S.; Guillet, V.; Chalut, C.; Rima, J.; Brison, Y.; Malaga, W.; Sanchez-Dafun, A.; Gavalda, S.; et al. Solution Structure of the Type I Polyketide Synthase Pks13 from *Mycobacterium tuberculosis*. *BMC Biology 2022 20:1* **2022**, *20*, 1–19, doi:10.1186/S12915-022-01337-9.

61. Wang, S.; Luan, J.; Chen, L.; Liu, H.; Li, W.; Wang, J. Computational Characteristics of the Structure-Activity Relationship of Inhibitors Targeting Pks13-TE Domain. *Comput Biol Chem* **2023**, *104*, 107864, doi:10.1016/J.COMPBIOLCHEM.2023.107864.

62. Zhao, G.; Tian, X.; Wang, J.; Cheng, M.; Zhang, T.; Wang, Z. The Structure-Based Virtual Screening of Non-Benzofuran Inhibitors against M. Tuberculosis Pks13-TE for Anti-Tuberculosis Phenotypic Discovery. *New J Chem* **2021**, *45*, 1286–1300, doi:10.1039/D0NJ03828H.

63. Wilson, R.; Kumar, P.; Parashar, V.; Vilchèze, C.; Veyron-Churlet, R.; Freundlich, J.S.; Barnes, S.W.; Walker, J.R.; Szymonifka, M.J.; Marchiano, E.; et al. Antituberculosis Thiophenes Define a Requirement for Pks13 in Mycolic Acid Biosynthesis. *Nat Chem Biol 2013 9:8* **2013**, *9*, 499–506, doi:10.1038/nchembio.1277.

64. Lun, S.; Xiao, S.; Zhang, W.; Wang, S.; Gunosewoyo, H.; Yu, L.F.; Bishai, W.R. Therapeutic Potential of Coumestan Pks13 Inhibitors for Tuberculosis. *Antimicrob Agents Chemother* **2021**, *65*.

65. Altharawi, A.; Alossaimi, M.A.; Alanazi, M.M.; Alqahatani, S.M.; Tahir ul Qamar, M. An Integrated Computational Approach towards Novel Drugs Discovery against Polyketide Synthase 13 Thioesterase Domain of *Mycobacterium tuberculosis*. *Sci Re 2023 13:1* **2023**, *13*, 1–12, doi:10.1038/s41598-023-34222-8.

66. Wilson, C.; Ray, P.; Zuccotto, F.; Hernandez, J.; Aggarwal, A.; Mackenzie, C.; Caldwell, N.; Taylor, M.; Huggett, M.; Mathieson, M.; et al. Optimization of TAM16, a Benzofuran That Inhibits the Thioesterase Activity of Pks13; Evaluation toward a Preclinical Candidate for a Novel Antituberculosis Clinical Target. *J Med Chem* **2022**, *65*, 409–423.

67. Zhang, W.; Lun, S.; Wang, S.H.; Jiang, X.W.; Yang, F.; Tang, J.; Manson, A.L.; Earl, A.M.; Gunosewoyo, H.; Bishai, W.R.; et al. Identification of Novel Coumestan Derivatives as Polyketide Synthase 13 Inhibitors against *Mycobacterium tuberculosis*. *J Med Chem* **2018**, *61*, 791–803.

68. Zhang, W.; Lun, S.; Wang, S.S.; Cai, Y.P.; Yang, F.; Tang, J.; Bishai, W.R.; Yu, L.F. Structure-Based Optimization of Coumestan Derivatives as Polyketide Synthase 13-Thioesterase(Pks13-TE) Inhibitors with Improved HERG Profiles for *Mycobacterium tuberculosis* Treatment. *J Med Chem* **2022**, *65*, 13240–13252.

69. Zhang, W.; Lun, S.; Wang, S.S.; Cai, Y.P.; Yang, F.; Tang, J.; Bishai, W.R.; Yu, L.F. Structure-Based Optimization of Coumestan Derivatives as Polyketide Synthase 13-Thioesterase(Pks13-TE) Inhibitors with Improved HERG Profiles for *Mycobacterium tuberculosis* Treatment. *J Med Chem* **2022**, *65*, 13240–13252.

70. Taira, J.; Murakami, K.; Monobe, K.; Kuriki, K.; Fujita, M.; Ochi, Y.; Sakamoto, H.; Aoki, S. Identification of Novel Inhibitors for Mycobacterial Polyketide Synthase 13 via in Silico Drug Screening Assisted by the Parallel Compound Screening with Genetic Algorithm-Based Programs. *J Antibiot 2022 75:10* **2022**, *75*, 552–558, doi:10.1038/s41429-022-00549-z.

71. Khan, M.Z.; Kaur, P.; Nandicoori, V.K. Targeting the Messengers: Serine/Threonine Protein Kinases as Potential Targets for Antimycobacterial Drug Development. *IUBMB Life* **2018**, *70*, 889–904, doi:10.1002/IUB.1871.

72. Wang, T.; Bemis, G.; Hanzelka, B.; Zuccola, H.; Wynn, M.; Moody, C.S.; Green, J.; Locher, C.; Liu, A.; Gao, H.; et al. Mtb PKNA/PKNB Dual Inhibition Provides Selectivity Advantages for Inhibitor Design to Minimize Host Kinase Interactions. *ACS Med Chem Lett* **2017**, *8*, 1224–1229.

73. Carette, X.; Platig, J.; Young, D.C.; Helmel, M.; Young, A.T.; Wang, Z.; Potluri, L.P.; Moody, C.S.; Zeng, J.; Prisic, S.; et al. Multisystem Analysis of *Mycobacterium tuberculosis* Reveals Kinase-Dependent Remodeling of the Pathogen-Environment Interface. *mBio* **2018**, *9*.

74. Xu, J.; Wang, J.X.; Zhou, J.M.; Xu, C.L.; Huang, B.; Xing, Y.; Wang, B.; Luo, R.; Wang, Y.C.; You, X.F.; et al. A Novel Protein Kinase Inhibitor IMB-YH-8 with Anti-Tuberculosis Activity. *Scientific Reports 2017 7:1* **2017**, *7*, 1–10, doi:10.1038/s41598-017-04108-7.

75. Be, N.A.; Bishai, W.R.; Jain, S.K. Role of *Mycobacterium* tuberculosis PknD in the Pathogenesis of Central Nervous System Tuberculosis; 2012; Vol. 12.

76. Abdulhamid, A.; Awad, T.A.; Ahmed, A.E.; Koua, F.H.M.; Ismail, A.M. Acetyleugenol from Acacia Nilotica (L.) Exhibits a Strong Antibacterial Activity and Its Phenyl and Indole Analogues Show a Promising Anti-TB Potential Targeting PknE/B Protein Kinases. *Microbiol Res (Pavia)* **2021**, *12*, 1–15, doi:10.3390/microbiolres12010001.

77. Deol, P.; Vohra, R.; Saini, A.K.; Singh, A.; Chandra, H.; Chopra, P.; Das, T.K.; Tyagi, A.K.; Singh, Y. Role of *Mycobacterium tuberculosis* Ser/Thr Kinase PknF: Implications in Glucose Transport and Cell Division. *J Bacteriol* **2005**, *187*, 3415–3420.

78. Rastogi, S.; Ellinwood, S.; Augenstreich, J.; Mayer-Barber, K.D.; Briken, V. *Mycobacterium tuberculosis* Inhibits the NLRP3 Inflammasome Activation via Its Phosphokinase PknF. *PLoS Pathog* **2021**, *17*, e1009712, doi:10.1371/JOURNAL.PPAT.1009712.

79. Bar-Oz, M.; Meir, M.; Barkan, D. Virulence-Associated Secretion in Mycobacterium Abscessus. *Front Immunol* **2022**, *13*, 938895, doi:10.3389/FIMMU.2022.938895/BIBTEX.

80. Shariq, M.; Quadir, N.; Alam, A.; Zarin, S.; Sheikh, J.A.; Sharma, N.; Samal, J.; Ahmad, U.; Kumari, I.; Hasnain, S.E.; et al. The Exploitation of Host Autophagy and Ubiquitin Machinery by *Mycobacterium tuberculosis* in Shaping Immune Responses and Host Defense during Infection. **2022**, *19*, 3–23,

81. Crook, D.W.; Rodrigues, C.; Ismail, N.A.; Mistry, N.; Iqbal, Z.; Merker, M.; Moore, D.; Walker, A.S.; Thwaites, G.; Niemann, S.; et al. Genome-Wide Association Studies of Global *Mycobacterium tuberculosis* Resistance to 13 Antimicrobials in 10,228 Genomes Identify New Resistance Mechanisms. *PLoS Biol* **2022**, *20*, e3001755, doi:10.1371/JOURNAL.PBIO.3001755.

82. Wang, Z.; Xie, J. Phosphoproteomics of Mycobacterium-Host Interaction and Inspirations for Novel Measures against Tuberculosis. *Cell Signal* **2022**, *91*, 110238, doi:10.1016/J.CELLSIG.2021.110238.

83. Mei, Y.M.; Zhang, W.Y.; Sun, J.Y.; Jiang, H.Q.; Shi, Y.; Xiong, J.S.; Wang, L.; Chen, Y.Q.; Long, S.Y.; Pan, C.; et al. Genomic Characteristics of *Mycobacterium tuberculosis* Isolates of Cutaneous Tuberculosis. *Front Microbiol* **2023**, *14*, 1165916, doi:10.3389/FMICB.2023.1165916/BIBTEX.

84. Malhotra, V.; Okon, B.P.; Satsangi, A.T.; Das, S.; Waturuocha, U.W.; Vashist, A.; Clark-Curtiss, J.E.; Saini, D.K. *Mycobacterium tuberculosis* PknK Substrate Profiling Reveals Essential

Transcription Terminator Protein Rho and Two-Component Response Regulators PrrA and MtrA as Novel Targets for Phosphorylation. *Microbiol Spectr* **2022**, *10*.

85. Thongdee, P.; Hanwarinroj, C.; Pakamwong, B.; Kamsri, P.; Punkvang, A.; Leanpolchareanchai, J.; Ketrat, S.; Saparpakorn, P.; Hannongbua, S.; Ariyachaokun, K.; et al. Virtual Screening Identifies Novel and Potent Inhibitors of *Mycobacterium tuberculosis* PknB with Antibacterial Activity. *J Chem Inf Model* **2022**, *62*, 6508–6518.

86. Magnet, S.; Hartkoorn, R.C.; Székely, R.; Pató, J.; Triccas, J.A.; Schneider, P.; Szántai-Kis, C.; Rfi, L.; Chambon, M.; Banfi, D.; et al. Leads for Antitubercular Compounds from Kinase Inhibitor Library Screens. *Tuberculosis* **2010**, *90*, 354–360, doi:10.1016/J.TUBE.2010.09.001.

87. Hanwarinroj, C.; Thongdee, P.; Sukchit, D.; Taveepanich, S.; Kamsri, P.; Punkvang, A.; Ketrat, S.; Saparpakorn, P.; Hannongbua, S.; Suttisintong, K.; et al. In Silico Design of Novel Quinazoline-Based Compounds as Potential *Mycobacterium tuberculosis* PknB Inhibitors through 2D and 3D-QSAR, Molecular Dynamics Simulations Combined with Pharmacokinetic Predictions. *J Mol Graph Model* **2022**, *115*, 108231, doi:10.1016/J.JMGM.2022.108231.

88. Sundar, S.; Thangamani, L.; Manivel, G.; Kumar, P.; Piramanayagam, S. Molecular Docking, Molecular Dynamics and MM/PBSA Studies of FDA Approved Drugs for Protein Kinase a of *Mycobacterium tuberculosis*; Application Insights of Drug Repurposing. *Inform Med Unlocked* **2019**, *16*, 100210, doi:10.1016/J.IMU.2019.100210.

89. Kleandrova, V. v.; Scotti, L.; Bezerra Mendonça Junior, F.J.; Muratov, E.; Scotti, M.T.; Speck-Planche, A. QSAR Modeling for Multi-Target Drug Discovery: Designing Simultaneous

Inhibitors of Proteins in Diverse Pathogenic Parasites. *Front Chem* **2021**, *9*, doi:10.3389/fchem.2021.634663.

90. Ambure, P.; Halder, A.K.; Gonzaíez Díaz, H.; Nataíia, M.; Cordeiro, D.S. QSAR-Co: An Open Source Software for Developing Robust Multitasking or Multitarget Classification-Based QSAR Models. *J. Chem. Inf. Model* **2019**, *59*, 34, doi:10.1021/acs.jcim.9b00295.

91. Karpov, P.; Godin, G.; Tetko, I. V. Transformer-CNN: Swiss Knife for QSAR Modeling and Interpretation. *J Cheminform* **2020**, *12*, 1–12, doi:10.1186/S13321-020-00423-W/FIGURES/9.

Chapter 3

Molecular docking, molecular dynamics simulations and binding free energy studies of interactions between *Mycobacterium tuberculosis* Pks13, PknG and bioactive constituents of extremophilic bacteria

Kudakwashe Nyambo[1], Kudzanai Ian Tapfuma[1], Lucinda Baatjies[1], Idah Sithole Niang[2], Liezel Smith[1], Krishna Govender[4,5], Mkhuseli Ngxande[3], Daniel J Watson[6], Lubbe Wiesner[6], and Vuyo Mavumengwana[1,] *

[1] *DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town; KN: knyambo@sun.ac.za; K.I.T: kudzanait@su.ac.za; LB: lbaatjies@sun.ac.za; LS: liezels@sun.ac.za; VM: vuyom@sun.ac.za*

[2]*Department of Biotechnology and Biochemistry, University of Zimbabwe, B064, Mount Pleasant, Harare, Zimbabwe; sitholeidah2015@gmail.com*

[3]*Computer Science Division, Department of Mathematical Sciences, Faculty of Science University of Stellenbosch, Matieland, South Africa; ngxandem@sun.ac.za*

[4]*Department of Chemical Sciences, University of Johannesburg, P. O. Box 17011, Doornfontein Campus, 2028, Johannesburg, South Africa; krishnag@uj.ac.za*

[5]*National Institute for Theoretical and Computational Sciences (NITheCS), South Africa*

[6]*Division of Clinical Pharmacology, Department of Medicine, Faculty of Health Sciences, University of Cape Town; D.J.W: daniel.watson@uct.ac.za; L.W: lubbe.wiesner@uct.ac.za*

***** Corresponding author.

Email address: *vuyom@sun.ac.za;* Tel: +27 718502949

3.1 Abstract

This study aimed to evaluate the antimycobacterial efficacy of functional crude extracts from bacteria isolated from gold mine tailings in South Africa, using an untargeted metabolomics approach to identify antimycobacterial metabolites. Bacterial strains were isolated from South African gold mine tailings and identified using 16S rRNA sequencing. The crude extracts obtained from the bacteria were tested against *Mycobacterium tuberculosis* H37Rv, *Mycobacterium*

smegmatis MC^2155, and *Mycobacterium aurum* A+. Untargeted HPLC-qTOF and molecular networking using the GNPS platform were used to identify the functional constituents present in extracts that exhibited inhibitory activity. A virtual screening workflow (VSW) comprising three molecular docking stages with increased precision was used to filter compounds that were strong binders to *Mycobacterium* Pks13 and PknG to investigate the potential mode of interaction between the compounds and the target proteins. The ligands returned from the VSW were optimised using density functional theory (DFT) at M06-2X/6-311++(*d,p*) level of theory and basis set implemented in Gaussian16 Rev.C01. The optimised ligands were re-docked against *Mycobacterium* Pks13 and PknG. Molecular dynamics (MD) simulation and molecular mechanics generalised born surface area (MM-GBSA) were used to evaluate the stability of the protein-ligand complexes formed by the identified hits. The hit that showed promising binding characteristics was virtually modified through multiple synthetic routes using reaction-driven enumeration. Three bacterial isolates, identified as *Bacillus subtilis*, *Bacillus licheniformis*, and *Streptomyces mycarofaciens*, showed significant activity against the three strains of *Mycobacterium*, while only *Bacillus subtilis* and *Bacillus licheniformis* exhibited activity against *M. tuberculosis* H37Rv. The tentatively identified compounds from the bacterial crude extracts belonged to various classes of natural compounds associated with antimicrobial activity. Two compounds, namely, cyclo-(L-Pro-4-OH-L-Leu) and vazabitide A, exhibited strong binding profiles against PknG and Pks13, with pre-MD MM-GBSA values of −42.8 kcal/mol and −47.6 kcal/mol, respectively. The DFT-optimized compounds exhibited the same docking scores as the ligands optimised using the OPSL4 force field. MD simulation and Post-MD MM-GBSA binding free energy showed that vazabitide A is a strong binder. Upon modification of Vazabitide A, the affinity to the Pks13 binding site increased to −85.8 kcal/mol, as shown by the post-MD MM-GBSA. Overall, this study revealed the potential use of bacterial isolates from mine tailings as sources of novel scaffolds that are useful in designing and optimising a new set of anti-*Mycobacterium* agents that can further be tested *in-vitro*.

3.2 Introduction

Infectious diseases such as tuberculosis (TB) have a negative global impact and are a major cause of death and disability. Despite the availability of treatment, 1.5 million people die each year from tuberculosis [1,2]. Current TB chemotherapies in clinical use are administered over a lengthy period, affecting compliance with the treatment. *Mycobacterium tuberculosis* (*M. tb*) gene mutation, alongside an incomplete commitment to prolonged treatment regimens and co-infection with HIV (Human Immunodeficiency Virus), encourages the development of drug-resistant

strains. Therefore, it is imperative to immediately develop and explore new effective drugs to address the rapid increase in drug-resistant strains and reduce the duration of TB treatment [3].

Some taxonomically different microorganisms within environmental niches do not amicably interact because of limited space and resources. Researchers have reported that microorganisms dominate environmental niches via the strategic use of diverse arsenal systems, for instance, strain-specific bacteriocins and broad-spectrum antimicrobials [4–6]. Therefore, understanding the mechanisms and driving forces employed by bacteria in antagonistic competition is essential in drug discovery. Studies have shown that the *Bacillus* genus isolated found in soil is a prolific biofactory of a wide range of bioactive small molecules. These include antimicrobial non-ribosomal cyclic lipopeptides, polyketides, and discoipyrole alkaloids. *Bacillus subtilis* and *B. licheniformis, in particular,* have been reported to produce metabolites that inhibit the growth of *Candida albicans*, *Heliobacter pylori,* and *M. tb* [7, 8]. Meanwhile, members of the genus *Streptomycetes* genus, which are filamentous actinobacteria frequently isolated from soil, represent an immeasurable reservoir of novel metabolites. These compounds account for two-thirds of agriculturally and medically essential secondary metabolites. Notable examples of wide-spectrum antimicrobials produced by *Streptomyces* sp. include streptomycin, ivermectin, nystatin, and tetracycline [9, 10].

The extraction and processing of valuable minerals such as gold in the Gauteng province of South Africa have resulted in the deposition of large mine tailing dumps that have been recognized to be a source of secondary environmental contamination. Gold mine tailings are characterized by scarce organic matter, low pH, and high concentrations of heavy metals [11, 12]. The anthropogenic factors in the gold mine tailings negatively impact the ecosystem, particularly the microbial

communities, by acting as a selection pressure. Interestingly, the bacteria inhabiting the gold mine tailings have managed to withstand these extreme conditions, by modifying their metabolic and genomic machinery to adapt to the environmental stresses. Evidence from the whole genome analysis of *Serratia* sp. and *Stenotrophomonas* sp. isolated from waste-rock piles of abandoned gold mine indicated genomic plasticity which is driven by the acquisition of functional gene clusters via horizontal operon transfer [13]. Fundamentally, the transferred functional gene clusters enhance microbial survival fitness by expressing various unique metabolic pathways to extract energy from a wide array of inorganic electron donors and acceptors, thus, compensating for the fluctuating harsh gold mine tailings environments [14]. Another example is the spectacular phenotypic plasticity of some *Actinobacteria*, to the fluctuating tailings environment, elevates their potential to produce novel secondary metabolites. Thus, gold mine tailings are a unique target for biomining of a wide range of natural products that have a medical impact [15, 16].

Mycobacterium tuberculosis PknG is an essential serine-threonine protein kinase that is required for the survival and virulence of *M. tb.* Protein kinase G is an attractive macromolecular druggable target because it is involved in regulating cell wall biosynthesis and cell division [17, 18]. *M. tb* Pks13 is involved in the biosynthesis of mycolic acids, which are essential building blocks of the bacterial cell wall. Mycolic acids are important virulence factors that shield the bacterium from host immune defences [19]. Understanding the molecular mechanism of the interaction of inhibitors with Pks13 and PknG is therefore of great interest in the development of new TB treatments.

Interestingly, the bioactive arsenal from extremophilic microorganisms has been structurally optimized throughout evolution to enhance microbial endogenous and exogenous defence systems.

Therefore, it is a prerequisite to identify and investigate the antimicrobial properties of the unique bioactive secondary metabolites that govern bacterial communities through killing [20, 21]. To date, there are no reports regarding the mycobacterial activity of bioactive secondary metabolites from bacteria isolated from South African gold mine tailings. Thus, South African gold mine tailings may be considered an underexplored reservoir of bacterial genomes with a promising potential of producing unique scaffolds with antimycobacterial activity. Therefore, the present study was performed to isolate bacteria from gold mine tailings to screen their metabolite crude extracts against *M. smegmatis* MC^2155, *Mycobacterium aurum A+*, and *M. tb* H37Rv. The chemical classes present in bacterial crude extract were tentatively identified using high-pressure liquid chromatography coupled to a quadrupole time-of-flight high-resolution mass spectrometer (HPLC-qTOF) followed by molecular networking using the Global Natural Product Social platform. The tentatively identified compounds were virtually screened against *M. tb* H37Rv essential proteins, namely Pks13, and PknG.

3.3 Materials and methods

Samples of gold mine tailings were obtained from five sites located around the Germiston area, Johannesburg, South Africa (26º13'7.08" S, 28º29'8.64" E). At each site, 0.5 kg of sample material was collected at a depth of 12 cm. Samples were stored in polyethylene bags at 4 ℃ until they were used.

3.3.1 Isolation of bacteria

The culturable bacterial community available in the mine tailings samples was isolated by adding one gram of each soil sample to nine mL of sterile saline water (0.85% NaCl *w/v*) and thoroughly

mixed by vortexing. The upper suspension of each mixture was collected and serially diluted at a 10-fold gradient. The bacteria were cultivated by inoculating 50 µL of the 10-fold serial dilution of each sample onto three types of growth media, namely nutrient agar (N.A.), Luria-Bertani agar (L.B.), and tryptic soy agar (TSA) [22]. The three types of growth media were used to increase the likelihood of isolating a broad spectrum of culturable bacteria. The plates were incubated for 24 hours at 37℃ under aerobic conditions. Colonies were randomly picked and streaked onto fresh media plates to obtain pure bacterial cultures. Molecular techniques were used to identify the pure colonies for further identification and phylogenetic studies.

3.3.2 Isolation of genomic DNA and Amplification of 16S rRNA gene

The total bacterial genomic DNA was extracted from the bacterial cultures using the Quick-DNA™ Fungal/Bacterial Miniprep Kit (Zymo Research) following the manufacturer's instructions. The 16S rRNA target gene was amplified by polymerase chain reaction (PCR) using the following., universal primers and enzymes: Forward primer (16S-27F: AGATTTGATCCTGGCT), reverse primer (16S-*1492R*: CGGTACCTTGTTGTTAC), and OneTaq® Quick-load® 2X Master Mix [23].The PCR amplicons were run on an agarose gel and extracted with a Zymoclean™ Gel DNA Recovery Kit (Zymo Research). The extracted DNA fragments were sequenced based on forward and reverse direction (Nimagen, BrilliantDye™ Terminator Cycle Sequencing Kit V3.1, BRD3-100/1000) and purified (Zymo Research, ZR-96 DNA Sequencing Clean-up Kit™). The purified DNA fragments were analyzed on an ABI 3500xl Genetic Analyzer (Applied Biosystems, ThermoFisher Scientific) for each reaction for every sample. The bacterial sequences were edited and aligned using MEGA X (version 10.2.6) [24]. The nBLAST database (http://www.ncbi.nlm.nih.gov/blast) was used to obtain closely matching

sequences and to create consensus sequences for each bacterial isolate. Consensus sequences were then used to construct a phylogenetic tree using the Maximum-likelihood method in MEGA X. The Bootstrap values were generated from 1000 replicates. All the bacteria isolates' 16S rRNA nucleotide sequences were deposited at GenBank.

3.3.3 Secondary metabolite production

The production of secondary metabolites was induced by growing the pure bacterial isolates in liquid culture (broth) as described by Uche-Okereafor et al. [25]. Briefly, a starter culture was prepared by inoculating a loop full of pure bacteria isolates into 50 mL of tryptic soy broth in Erlenmeyer flasks. The starter cultures were incubated at 37℃ for 7 days with constant shaking at 200 revolutions per minute (rpm). An inoculum size of 5% of the starter culture was inoculated into 1 L fermentation broth (tryptic soy broth) for large-scale fermentation. The cultures were then fermented at 37℃ with constant shaking at 90 rpm for 7 days. To obtain cell debris-free crude extracts, the cellular debris was collected by centrifugation at 4000 rpm for 10 minutes and filtered through a 0.2 nm nylon filter. The cellular debris was collected by centrifugation at 4000 rpm for 10 minutes to obtain cell debris-free crude extracts and filtered through a 0.2 nm nylon filter. The liquid mixtures containing bacterial secondary metabolites were then freeze-dried. The dried bacterial extracts were suspended in a solvent at a ratio of 100 mL of methanol to 10 grams of dried bacterial culture [26]. The methanolic extracts were air-dried at room temperature under a constant stream of air and then stored at 4℃ until further analysis.

3.3.4 Minimum inhibition concentration evaluation

The mycobacteria strains, namely, *M. smegmatis* MC^2155, *M. aurum A+*, and *M. tb* H37Rv, were used in this study. *Mycobacterium smegmatis* mc^2155 and *M. aurum A+* were employed as

surrogate strains in the preliminary screening for *M. tuberculosis* H37Rv. These two surrogate strains share some genome similarities with *M. tuberculosis* but are non-virulent. The microorganisms were stored in 50% glycerol at – 80℃. Working stocks were cultured in vials containing Middlebrook 7H9 broth supplemented with 10% bovine albumin, catalase, dextrose, and sodium chloride (OADC-BBL/Becton-Dickinson, USA) and grown at 37℃ [27]. The bacterial crude extract stocks were prepared by dissolving the methanol crude extracts in 100% dimethyl sulfoxide (DMSO) and then diluted to 10% with water. The minimum inhibitory concentration (MIC) of the bacterial extracts was then determined through the broth microdilution technique as described [26]. Briefly, *M. smegmatis* MC^2155, *M. aurum A+*, and *M. tb* H37Rv were subcultured in Middlebrook 7H9 broth to reach an optical density (OD) of 0.2-0.3 at 600 nm. Aliquots 100 μL inoculum of the test organism (*M. smegmatis* MC^2155, *M. aurum A+* and *M. tb* H37Rv) diluted 1:99 were pipetted into 96 well microtiter plates. The crude extracts were pipetted into the first wells and then serially diluted to achieve final concentrations varied from 2500 μg/mL to 19.53 μg/mL. Isoniazid was used as the positive control. The plates were incubated at 37℃ for 72 hours for *M. smegmatis* MC^2155 and 144 hours for *M. tb* H37Rv after the results were evaluated adding 20 μL of 0.015% resazurin dye as a colorimetric indicator of mycobacterial growth. The MIC was defined as the lowest concentration capable of inhibiting bacterial growth and all the assays were performed in technical and biological triplicates.

3.3.5 LC-QTOF-MS/MS analysis

High-resolution mass spectra were obtained using an AB Sciex® X500R QTOF coupled to an AB Sciex® Exion LC system. Spectral data were obtained using information-dependent acquisition (IDA) at a mass range of 50–1500 Da. All methods, batches, and data were processed using OS

Sciex® v3.1. The declustering potential was 80 V, the curtain gas (N2) was at 25 pounds per square inch (psi), the ion spray voltage was 5500 V, and the source temperature was 450℃ . Ion source gases 1 and 2 were at 45 and 55 psi, respectively. The collision energy was 10 eV for the MS scans and 20–50 eV for MS/MS scans. The IDA intensity threshold was 50 cycles per second. The aqueous mobile phase used was 1 mM ammonium formate in water, and the organic mobile phase was 0.5% formic acid dissolved in methanol. The gradient elution program for the organic mobile phase was set to start at 2% and end at 98% between 0-25 minutes, holding for 5 minutes before returning to 2% over 5 minutes to re-equilibrate for the next injection. The flow rate was 700 µL/min and the run time was 35 minutes. A Kinetex® C18 column (5 µm, 100 Å, 150 mm x 6 mm) with a column protector was used. All solvents were sonicated for 10 minutes before use to remove bubbles.

3.3.6 Data processing and Annotation

To determine the metabolite classes, present in the three crude bacterial extracts, molecular networks were computed using the Global Natural Products Social Molecular Networking (GNPS) platform (https://ccms-ucsd.github.io/GNPSDocumentation/). A molecular network was created with an MS/MS fragment ion tolerance of 0.025 Da [28]. The created molecular network was enriched with information from *in-silico* structure annotations from GNPS Library Search, and variable Dereplicator using the GNPS MolNetEnhancer workflow (https://ccmsucsd.github.io/GNPSDocumentation/molnetenhancer/). The chemical class annotations were performed using the ClassyFire chemical ontology [29]. Furthermore, the raw HPLC-qTOF data was converted to a ".abf" format by ABF converter software (http://www.reify.cs.com/AbfConverter) and then annotated using metabolic workflow on MS-

DIAL software version 4.24 [26, 30]. The parameters used for processing the files were as follows: mass range (MS1) m/z 50-1500; MS1 and MS2 tolerance of 0.01 and 0.025 respectively; [M+H] adducts ions with a peak height of 10,000. The tentative prediction of molecular formula and structure elucidation of the bacterial metabolites were processed using MS-FINDER software version 3.50 using the following parameters: MS1 and MS2 tolerances were set to 0.01Da; formula calculation with isotopic ratio tolerance was set to 20%; *in-silico* MS/MS fragmenter tree depth was set to 2; the databases selected were COCONUT(Natural product) (https://coconut.naturalproducts.net/), UNDP(Natural product) (https://github.com/DIFACQUIM/Natural-products-subsets-generation), ChEBI(Biomolecules) (https://www.ebi.ac.uk/), KNApSAcK(Natural product) (http://www.knapsackfamily.com/KNApSAcK/), PubChem(Biomolecules) (https://pubchem.ncbi.nlm.nih.gov/), and LipidMaps(Lipids) (https://lipidmaps.org/). To reveal the differences in metabolic profiles of the bacterial crude extracts, principal component analysis (PCA) plot was generated using Metaboanalyst 5.0 software (http://www.metaboanalyst.ca/) [30]. To expand the characterization of the metabolomic potential of *B. subtilis*, *S. mycarofaciens*, and *B. licheniformis* the Metaboanalyst 5.0 software was further used to perform pathway enrichment as described by [31]. An over-representation analysis (ORA) was implemented using hypergeometric testing to determine whether certain metabolite sets were overrepresented compared to what would have been observed by chance. Pathway topology analysis was conducted based on betweenness and out-of-degree centrality measures, evaluating the significance of each metabolite in each metabolic network. Potential targets were chosen using p-values from pathway enrichment analysis or impact values from pathway topology analysis, with an impact value

threshold of 0.10 and a negative-log p-value threshold of 10. Altered pathways were identified, and potential functional analysis was carried out.

3.3.7 Virtual Screening of bacterial compounds

The potential mechanism of action of all tentatively identified bacterial compounds was evaluated using the Virtual Screening Workflow in Schrödinger Release 2022-1 [32]. Briefly, the three-dimensional crystal structures of *M. tb* proteins (PDB:7Q52: PknG), and (PDB: 7VJT: Pks13) solved with small ligands were downloaded from the Protein Data Bank (PDB) (https://www.rcsb.org/). The resolution for 7Q52 is 2.35 Å, while for 7VJT is 1.94 Å (https://www.rcsb.org/). The raw crystal structures were prepared using the Protein Preparation Preparation Wizard (Schrödinger Release 2022-1), as described by [32]. Hydrogen atoms were added, the loop region was refined, H-bond assignments were optimized, and energy was minimized by an OPLS-4 force field. The co-crystallized heterogeneous ligands and water were removed, while polar hydrogens were added. The Receptor Grid Generator module generated the docking receptor grid configurations for all proteins using the coordinates of the previously co-crystallized ligands. The tentatively identified bacterial compounds were prepared by the LigPrep module (Schrödinger Release 2022-1) using the following parameters: energy minimised by an OPLS4 (Optimized Potentials for Liquid Simulations 4) force field, generated ionisation states at pH 7.0+2.0, and 32 multiple conformers per ligand. For Pks13, **3,8-bis(oxidanyl)-7-(piperidin-1-ylmethyl)-[1]benzofuro[3,2-c]chromen-6-one (**7IJ**) was used as control inhibitor while 2-azanyl-3-(4-fluorophenyl)carbonyl-indolizine-1-carboxamide (8ZC) was used as a control inhibitor for PknG.** The Root Mean Square Deviation (RMSD) between the co-crystallized ligand and the ligand after docking was calculated to validate the docking protocol**.** A Virtual Screening

Workflow (VSW) was used to screen a library of prepared compounds to obtain a hit list [33]. The QikProp module (Schrödinger Release 2022-1) was used to filter compounds QikProp is a Schrodinger package that predicts ADME properties and filters compounds with poor ADME (absorption, distribution, metabolism, and excretion) properties. Further, the returned compounds were subjected to three docking regimes of increasing precision using the Glide module [32]. Briefly, the bacterial compounds were docked against drug targets using a hierarchical approach that employed high-throughput virtual screening (HTVS), followed by standard precision (SP), and ultimately extra-precision (XP). The output hit from HTVS were filtered, and only 20 % were selected for further SP docking. Similarly, from the SP docking outputs, only 20 % were subjected to XP docking. Finally, 30 % of the XP docking hits were retrieved and subjected to MMGBSA free energy calculations. The Gaussian 16 Rev. C01 software was used for geometry optimisation and frequency calculations. The DFT calculations were performed using the M06-2X level of theory and 6-311++G(*d*,*p*) basis set. The DFT computations were used to isolate the minimum energy conformation of a ligand along a potential energy surface. The compounds returned from the VSW and the modified compound were visualised in Gauss-view. The properties of the compounds were evaluated based on the E_{HOMO} and E_{LUMO}. The equations used for the calculations are as follows:

$$\mu = \frac{E_{LUMO}+E_{HOMO}}{2} \tag{1}$$

$$\eta = \frac{E_{LUMO}-E_{HOMO}}{2} \tag{2}$$

$$(s) = \frac{1}{\eta} \tag{3}$$

3.3.8 Molecular dynamics simulations

Molecular dynamics (MD) simulations in the study were performed by using the Desmond v5.3 module implemented in the Maestro interface (Schrödinger 2022-1 suite). A total of 6 molecular dynamics simulation systems were built by solvating the protein-ligand complexes with Transferable Intermolecular Potential 4 Point (TIP4P) explicit water molecules and placed in the center of an orthorhombic box with boundary dimensions of (10 Å × 10 Å × 10 Å). The systems were neutralized by adding counter ions and a 0.15 M NaCl solution. The MD protocol involved minimization, pre-production, and finally production MD steps. In the minimization step, the entire system was allowed to relax for 2500 steps using the steepest descent approach. Then the temperature of the system was raised from 0 to 300 K with a small force constant on the protein to restrict any drastic changes. MD simulations were performed via NPT (constant number of atoms, constant pressure i.e., 1.01325 bar and c, constant temperature i.e., 300K) ensemble. The Nose-Hoover chain method was used as the default thermostat with a 1.0 ps interval and Martyna-Tobias-Klein as the default barostat with a 2.0 ps interval by applying an isotropic coupling style. Long-range electrostatic forces were calculated based on the Particle-mesh-based Ewald approach with the cut-off radius for columbic forces set to 9.0 Å. Finally, the system was subjected to production MD simulations for 200 ns for the free protein and protein-ligand complexes using the OPSL-4. During the simulation, the trajectories were written out every every 1000 ps.. The systems' dynamic behaviour and structural changes were analysed by calculating the root mean square deviation (RMSD) and root mean square fluctuation (RMSF). Subsequently, the energy-

minimized structure calculated from the equilibrated trajectory system was evaluated to investigate each ligand-protein complex interaction.

3.3.9 MM-GBSA calculations

The free energy change (ΔG_{Bind}) for the interaction between the receptor and the ligand to form a complex was computed using Molecular Mechanics Generalized Born Surface Area (MM/GBSA) as the summation of different interactions according to the equation below:

$$\Delta G_{Bind} = E_{Complex} - [E_{Receptor} + E_{Ligand}] \quad (4)$$

In the formula provided:

ΔG_{Bind} represents the calculated relative free energy, which takes into account both ligand and receptor strain energy.

$E_{Complex}$ represents the MM-GBSA energy of the minimised complex,

E_{Ligand} represents the MM-GBSA energy of the ligand after it has been removed from the complex and allowed to relax.

$E_{Receptor}$ represents the MM-GBSA energy of relaxed protein after separating it from the ligand [34].

3.3.10 Reaction-based *in-silico* modification of the selected strong binder compound

A virtual library generation approach was performed using the Enumeration module in Schrödinger. The identified strong binder for Pks13 datasets was selected as the molecular starting point for modification to expand the structural complexity and explore the chemical space using the Pathfinder (Schrödinger Release 2022-1) automated reaction-driven enumeration [35]. To achieve this, the compound was enumerated based on the 100 multiple synthetic routes and filtered

based on the similarity to the active compound, SMiles ARbitrary Target Specification (SMARTS) removed compounds with reactive functional groups, and Pan assay Interfering Structures (PAINS) properties. Regardless of some novelty properties obtained from the new library comprising 1000 generated molecules, it is critical to extensively evaluate their molecular interaction with Pks13 respectively, as well as structural profiles [35]. In this regard, the library created was subjected to a VSW as described in section 2.6 above. The strong binder returned from VSW was then subjected to MD simulation for 200 ns and binding free energy calculations as described in sections 2.7 and 2.8.

3.4 Results

3.4.1 Identification of Bacteria

In this study, 11 bacteria were isolated and identified from gold mine tailings. The 16 S rRNA sequences of the bacterial isolates were compared to the GenBank sequence database information for species identification using BLASTn. The phylogenetic tree was constructed using the Maximum-likelihood in MEGA X for sequences that had high quality and showed the relationships among the bacteria species (Figure A3). The bacterial sequences were submitted to GenBank for curation and assigned accession numbers (OM182829 - OM182840) as shown in Table 3.1. A total of eleven identified bacteria belonging to the genera *Bacillus* (six), *Micrococcus* (two), *Streptomyces* (one), *Staphylococcus* (one), and *Kocuria* (one) were investigated for their secondary metabolite diversity and antimycobacterial activity. The current study also elaborates on the detailed antimycobacterial efficacy of crude bioactive extracts and *in-silico* screening for better-targeted anti-*Mycobacterium* validation. A MIC below 2.5 mg/mL was considered an indicator of inhibitory activity. The results revealed that crude extracts from three bacterial isolates,

specifically *B. subtilis*, *B. licheniformis*, and *S. mycarofaciens*, exhibited significant activity against three strains of *Mycobacterium*, including *M. smegmatis* MC^2155, *M. aurum* A+, and *M. tb* H37Rv. The crude extracts from *B. subtilis* exhibited the most potent efficacy against all three strains, with MIC values ranging from 0.3125-0.625 mg/mL, while the crude extracts from *B. licheniformis* strongly inhibited all three test strains with MIC values ranging from 1.25-2.6 mg/mL. The crude extracts from *S. mycarofaciens* also exhibited strong inhibition against *M. smegmatis* MC^2155 and *M. aurum* A+, with MIC values ranging from 0.325-2.25 mg/mL.

Table 3.1. Minimum inhibition concentration of bacterial crude extracts against *M. smegmatis* MC^2155, *M. aurum* A^+ and *M. tb* H37Rv expressed in mg/mL.

Sample ID	Predicted identity	Accession number	MIC (mg/mL)		
			M. smegmatis mc^2155	*M. aurum* A^+	*M. tb* H37Rv
KN1	*Micrococcus luteus*	OM182829	>2.5	>2.5	>2.5
KN2	***Streptomyces mycarofaciens***	**OM182830**	**2.5-1.25**	**2.5-0.325**	>2.5
KN3	*Bacillus simplex*	OM182831	>2.5	>2.5	>2.5
KN4	*Bacillus sp.*	OM182832	>2.5	>2.5	>2.5
KN5	*Micrococcus luteus*	OM182833	>2.5	>2.5	>2.5
KN6	*Staphylococcus saprophyticus*	OM182834	>2.5	>2.5	>2.5
KN7	***Bacillus licheniformis***	**OM182835**	**2.5-0.3125**	**2.5-0.625**	**2.5-1.25**
KN8	*Kocuria rhizophila*	OM182836	>2.5	>2.5	>2.5
KN10	*Bacillus paralicheniformis*	OM182838	2.5	>2.5	>2.5
KN11	*Bacillus mobilis*	OM182839	>2.5	>2.5	>2.5
KN12	***Bacillus subtilis***	**OM182840**	**0.1625**	**0. 3125**	**0.625-0.3125**
Control	**Isoniazid**		**0.125-0.0625**	**0.125-0.0625**	**0.125-0.0625**

†MIC stands for minimum inhibitory concentration.
*Isoniazid was used a positive control for antimycobacterial activity.

3.4.2 Metabolite profiling of bacterial crude extract

Information regarding the antimycobacterial efficacy of functional crude extracts from bacteria isolated from South African, gold mine tailings is limited. Thus, a comprehensive untargeted metabolomics analysis of the bacterial crude extracts was performed to differentiate the bacteria's

metabolic potential and, also, the metabolites' influence on antimycobacterial activity. A principal component analysis (PCA) approach was constructed from the HPLC-qTOF data of the three active crude extracts against the *Mycobacteria* species. The secondary metabolites produced by the three bacteria were clearly separated and grouped into three clusters, corresponding to *B. subtilis*, *B. licheniformis*, and *S. mycarofaciens* as depicted in Figure. 3.1.A. The general pattern of the tentatively identified secondary metabolites across the three bacterial isolateswas relatively similar with the divergence being primarily attributed to the presence of organic nitrogen compounds and organic oxygen classes contributing to the distinction, as shown by the clustering in the loadings and molecular network in Figure 3.1.B and Figure 3.2.

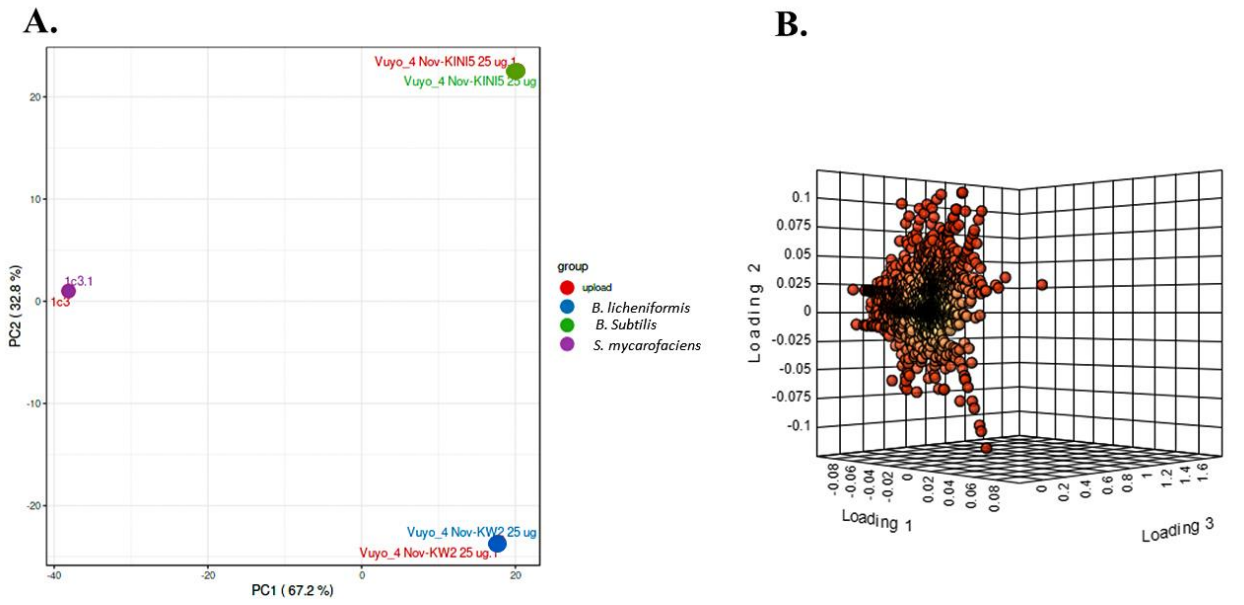


Figure 3.1. Principal component analysis (PCA) of metabolite data acquired by HPLC-qTOF of three bacterial crude extracts in positive ionization mode**. A.** PCA scores plot comparing metabolites present in crude extracts from *B. subtilis, B. licheniformis,* and *S. mycarofaciens.* **B.** Loading plot from PCA analysis.

In the current study, molecular networking was performed to tentatively identify metabolites. GNPS computed the network by grouping metabolites based on spectral similarities and are represented as clusters of different nodes as depicted in Figure 3.2 [36]. The tentatively identified compounds from *S. mycarofaciens*, *B. subtilis,* and *B. licheniformis* belong to various classes of natural compounds, with a high quantitative variation in the pool of compounds strongly correlated with antimicrobial activity in Figure 3.2 and Table A.1. Cyclic peptides which possess antimicrobial potential were putatively identified from *S. mycarofaciens* are shown in cluster F at mass to charge ratio (m/z) 245.128, 211.169, and 235.119 as depicted in Figure 3.2.A. Three classes of the tentatively identified compounds that might have contributed to the antimycobacterial activity were found to have molecular weights similar to cyclic lipopeptides produced by *Bacillus* sp. These included plipastatin and surfactin. In cluster A (depicted in Figure 3.2), the nodes for various surfactin isomers were directly connected. In cluster B, a direct relationship was observed between plipastatin B 1 at m/z 1492.92 and Plipastatin A 2 at m/z 1478.94, with an additional node of another plipastatin analog observed at *m/z* 1506.94. Cluster C featured a node for cyclo(L-Val-L-Pro) at *m/z* 197.0, which was also associated with a possible cyclic-peptide compound at m/z 261.17 (Figure 3.2). In cluster D shown in Figure 3.2.C the node at *m/z* 1057.1 was classified as an unknown compound, but it has an m/z that is associated with iturin analog (1057). In cluster E, the node with *m/z* 1035.63 is corresponding to a surfactin (Figure 3.2.C). The majority of the metabolites present in the bacterial crude extracts have not been fully characterized and remain unknown as depicted in Figure 3.2. The classes of tentatively identified metabolites from the three bacterial crude extracts belong to aminocyclitol glycosides, alpha-amino acids, and derivatives; cyclic depsipeptides, phenylpropanoids, and polyketides; benzenoids, organic acids, and alpha-amino derivatives; organic heterocyclic compounds, and

organic oxygen compounds (Table A.1). GNPS did not assign structure to the respective nodes.Thus, MS-FINDER was used to predict the molecular formulas (Table A.1).

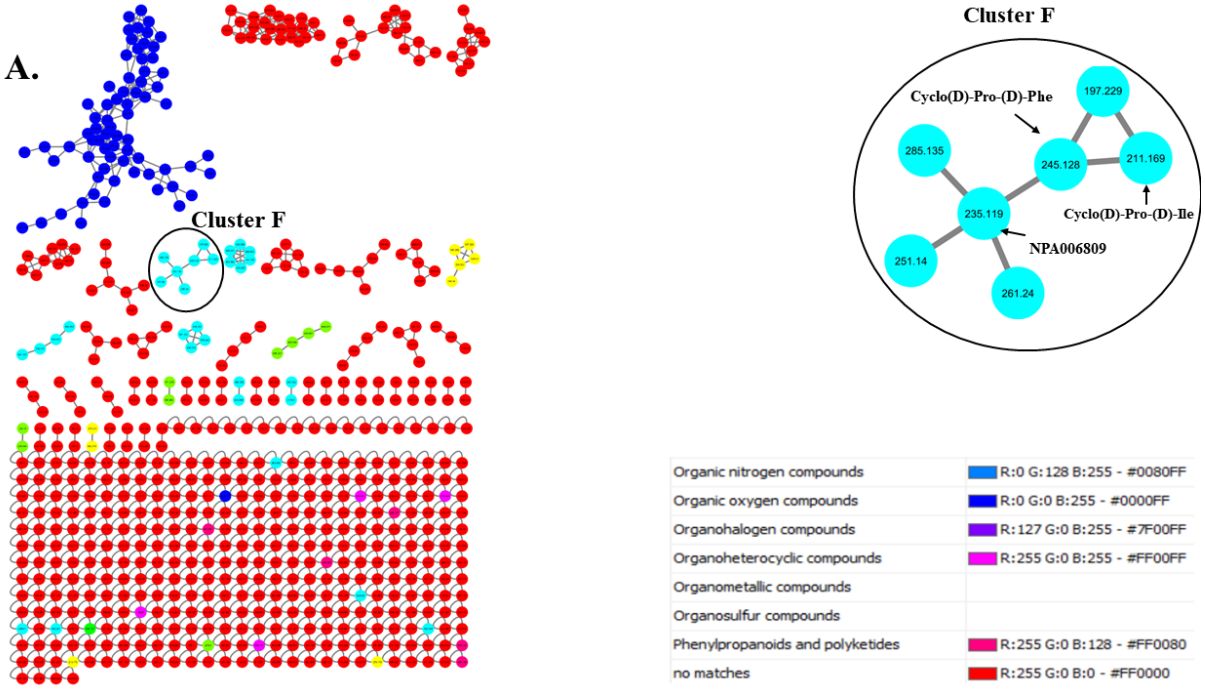


Figure 3.2. Molecular networking of bacterial crude extracts of **A**. *S. mycarofaciens* **B.** *B. subtilis* and **C.** *B. licheniformis.* Unknown compounds in red are not characterized. The molecular network annotation indicates that unknown compounds are abundant in *S. mycarofaciens*, *B. subtilis*, and *B. licheniformis.*

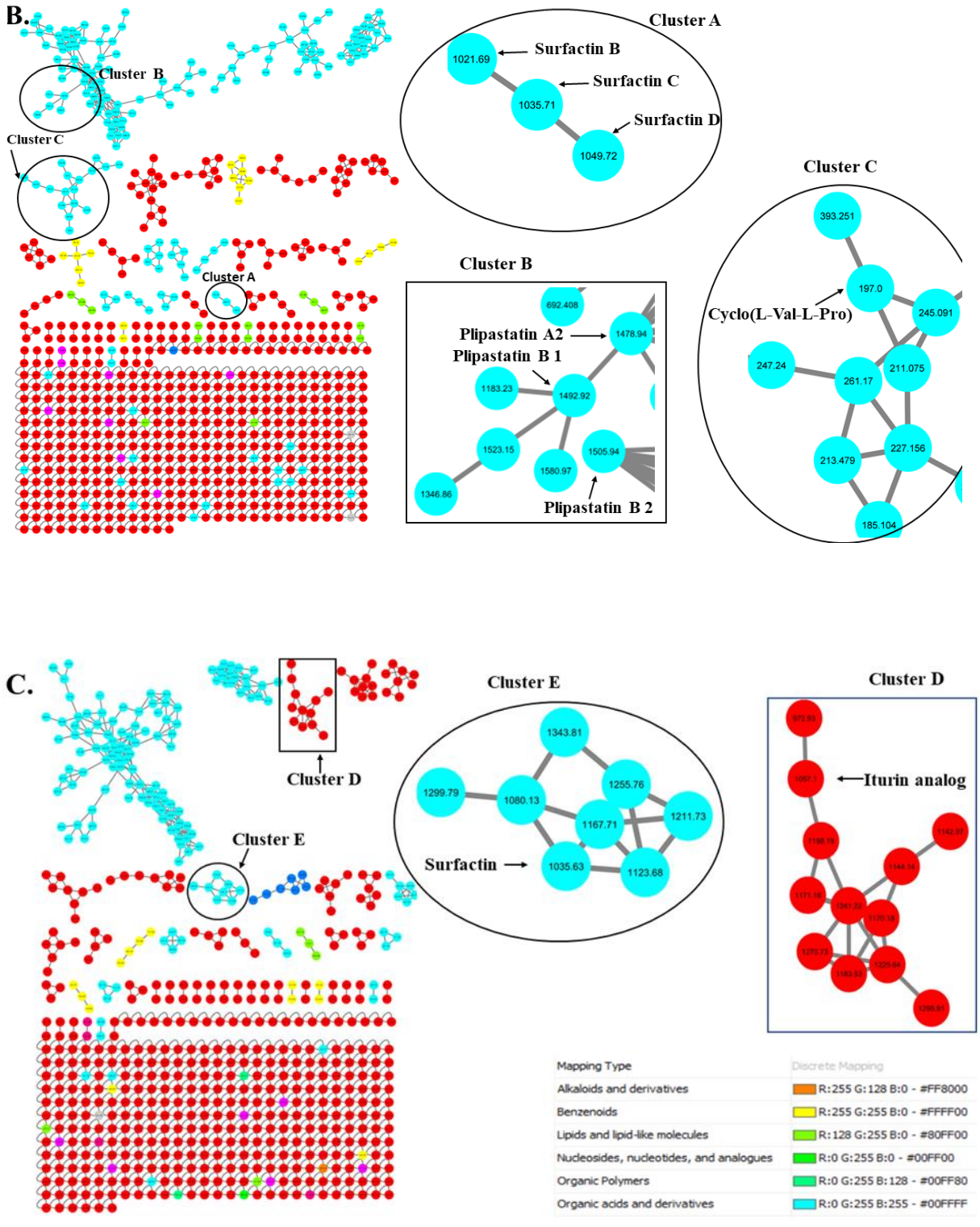


Figure 3.2. continued. *B. subtilis* and **C.** *B. licheniformis.*

3.4.3 Characterization and functional analysis of the key metabolic pathways

An investigation of the metabolic pathway of *B. subtilis*, *S. mycarofaciens*, and *B. licheniformis* which exhibited the strongest antimycobacterial activity was processed using MetaboAnalyst. Some of the pathways that were revealed from the pathway enrichment analysis; are the aminoacyl-tRNA biosynthesis, biosynthesis of secondary metabolites that have antimicrobial potential, and aminobenzoate degradation (Figure 3.3).

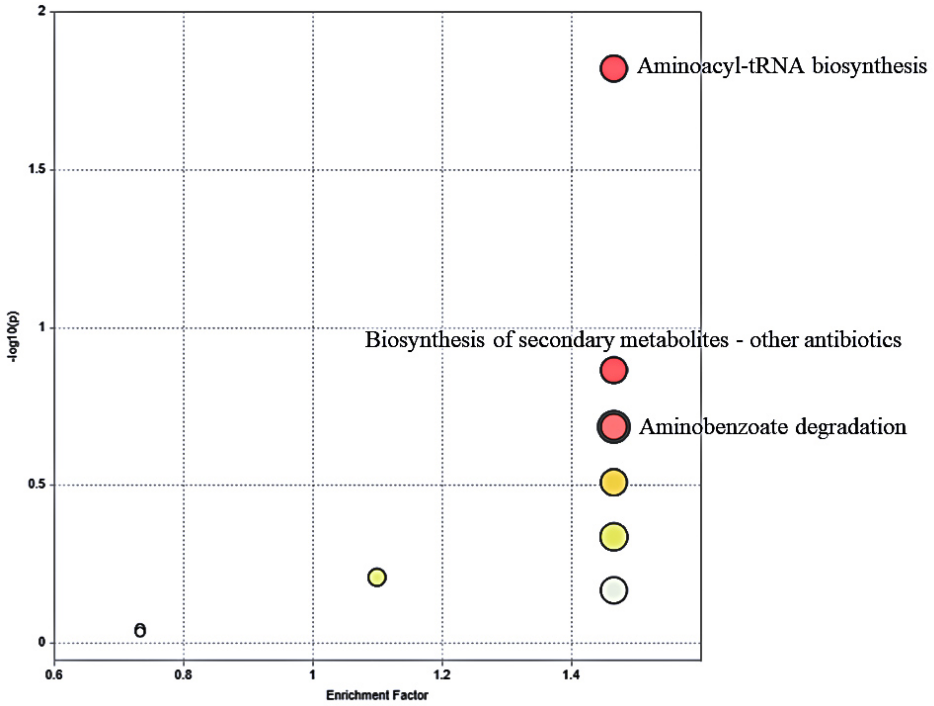
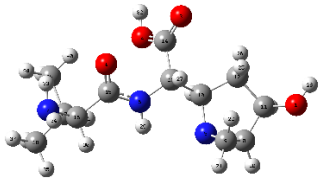


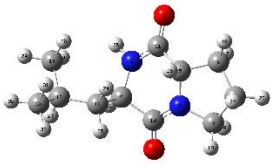
Figure 3.3. Metabolic pathway analysis generated with the MetaboAnalyst based on the metabolites identified from the crude extracts of *B. subtilis*, *S. mycarofaciens*, and *B. licheniformis*. The pathway enrichment analysis was based on the p-values on the Y-axis, to determine the significance of the metabolites. The range of colors on the plot, ranging from yellow to red, represents the varying levels of significance of the metabolites for the enrichment analysis.

3.4.4 Virtual screening and binding dynamics analysis

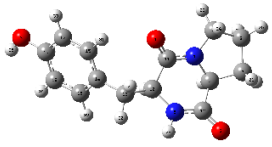
The RMSD of the co-crystalized and re-docked ligands of PknG and Pks13 was 0.61 and 1.19, respectively, as depicted in Figure A4. The tentatively identified bacterial compounds were subjected to VSW to identify compounds with the best binding postures and further explore their mode of interaction. From the Glide-XP docking outputs, eight compounds exhibited strong XP docking profiles to the two target proteins ranging from −8.8 kcal/mol to −11.9 kcal/mol. The optimised geometries of the identified bacterial compounds were derived through DFT calculations (as shown in Figure 3.4). To ascertain their stability, a frequency analysis was performed to confirm that these optimised scaffolds represent the lowest energy states, and in all cases, no negative frequencies were obtained. The quantum mechanics optimised scaffolds were re-docked and compared with the docking scores from ligands optimised using the OPLS4 force field. The docking scores of both quantum mechanics optimised, and OPLS4 force field optimised are similar, thus increasing the confidence of the docking protocol.



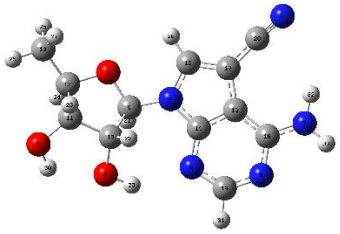
vazabitide A



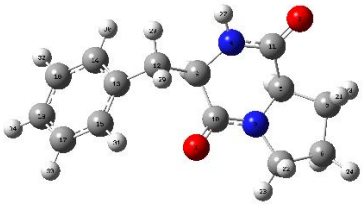
cyclo(Pro-Leu)



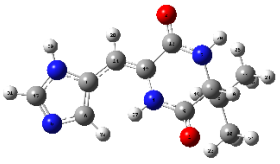
maculosin



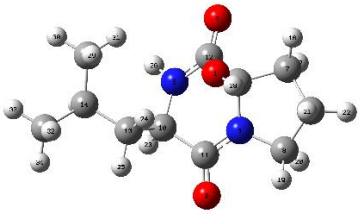
5'-Deoxytoyocamycin



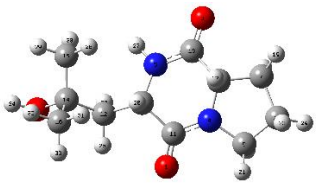
(3R,8aR)-3-benzyl-2,3,6,7,8,8a-hexahydropyrrolo[1,2-a]pyrazine-1,4-dione



NPA006809



cyclo(2-hydroxy-Pro-R-Leu)



cyclo-(L-Pro-4-OH-L-Leu)

Figure 3.4. DFT optimised geometric structures of the identified compounds returned from the virtual screening workflow.

The reactivity of the DFT-optimized compounds was computed using the M06-2X level of theory and 6-311++G (d,p) basis set. The global reactivity descriptors of the compounds calculated in this study, including chemical potential (μ), chemical hardness (ɲ), chemical softness (s), electronegativity (χ), and electrophilic index (ϖ), are shown in Table A2. Notably, the lowest unoccupied molecular orbital (LUMO) is expected to accept electrons, while the highest occupied molecular orbital (HOMO) is an electron donor. The difference between the HOMO and LUMO yields an energy gap (ΔE). Minus LUMO computes ionisation energy. A lower energy gap (ΔE) indicates higher reactivity and is associated with a soft compound. Conversely, a higher energy gap (ΔE) implies greater stability and lower reactivity. In this study, there were only small differences in the energy gap (ΔE), ranging from 0.23 to 0.31 eV. The compound NPA006809 exhibited the smallest energy gap (0.23 eV) and the lowest ionisation energy (0.27 eV). These results indicate that the compounds are soft and can interact as electron acceptors (Table A2 and Figure 3.5). The high docking scores also support that the compounds are reactive (Table 2). The LUMO surface was localized under regions where there are nitrogen rings.

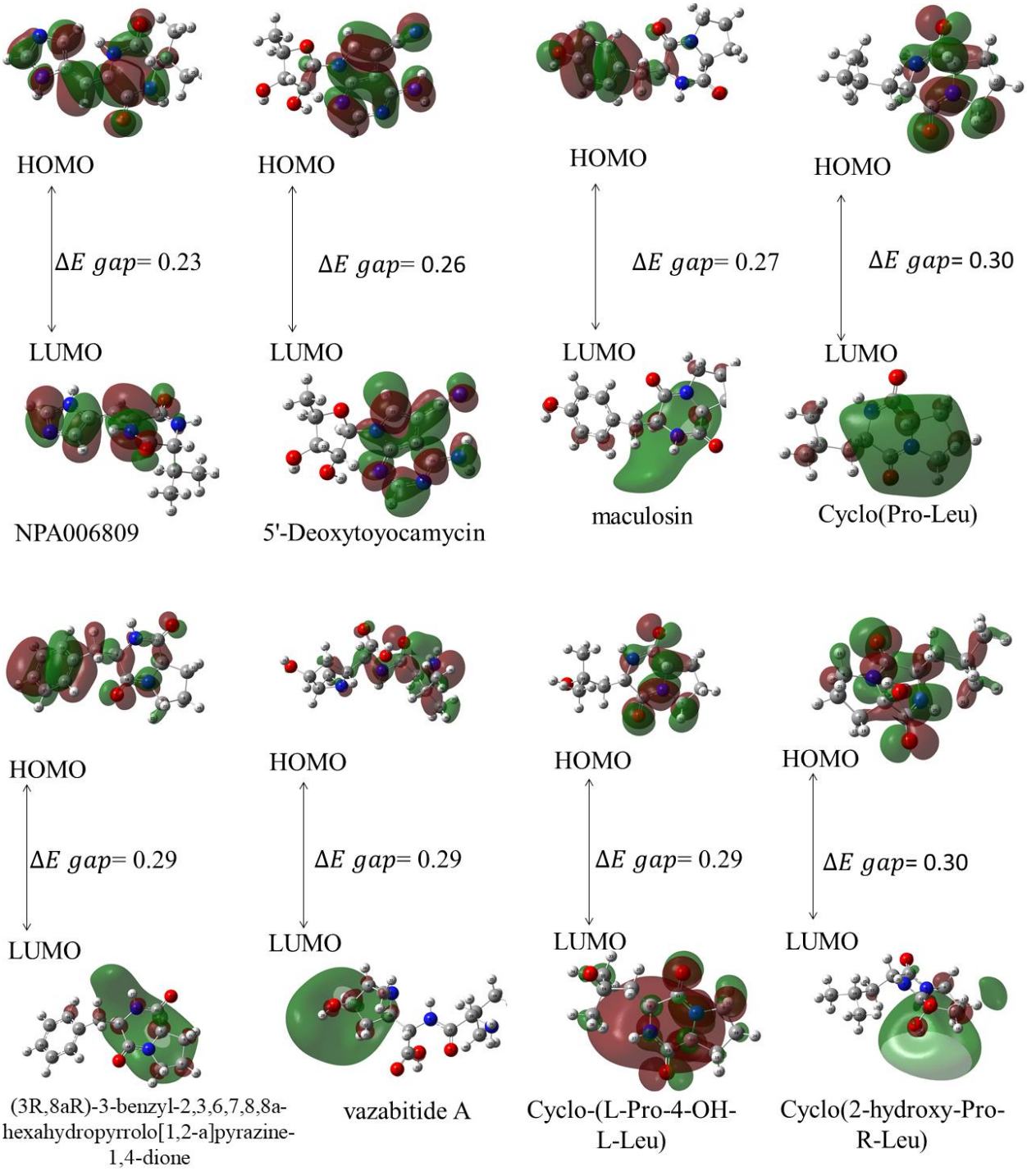
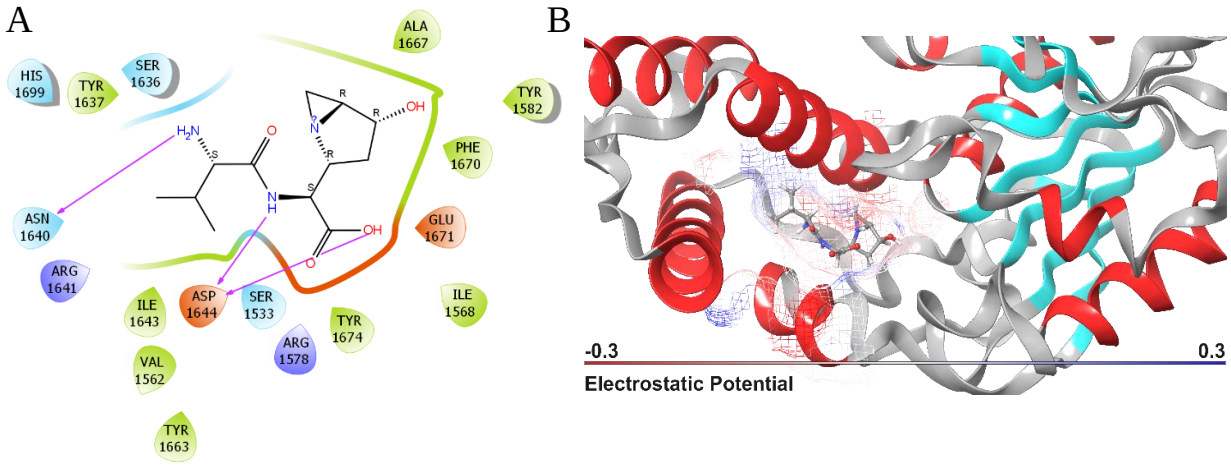


Figure 3.5. Depiction of the HOMO-LUMO surface maps computed with M06-2X/6-311 ++ (*d,p*). ΔE is measured in eV.

Notably, vazabitide A exhibited the strongest XP docking score of −11.92 kcal/mol against Pks13 and a ΔG_{bind} of −47.6 kcal/mol (Table 3.2). However, the percentage human oral absorption of vazabitide A was calculated to be relatively low at 27.76 %. The visualization of the binding interaction between vazabitide A and Pks13 showed that the compound was anchored in the hydrophobic pocket of Pks13 and interacted with the protein through multiple mechanisms. The interactions with proximal amino acid residues included hydrogen bonding with ASN1640, ASP1644, and TYR1663, pi-cation interaction with TYR1674, a salt bridge with ASP1644, positive charge interaction with ARG1578, and ARG164, polar interactions with SER1533, ASN1640, HIE1664, HIS169 (Figure 3.6). The interactions collectively contributed to the observed ΔG_{bind} −42.8 kcal/mol (Table 3.2). The control (7IJ) ligand exhibited a docking score of −8.2 kcal/mol and an MM-GBSA value of −42.0 kcal/mol. According to the HOMO-LUMO results, the interaction with the nitrogen rings was as expected.



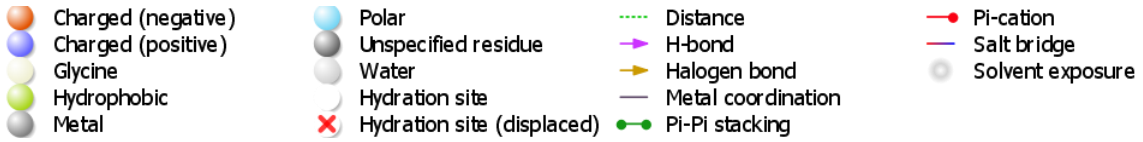


Figure 3.6. A concise overview of the interaction of vazabitide A with Pks13. A. 2D representation B. 3D representation.

For PknG, seven compounds exhibited strong XP docking scores, namely maculosin, NPA006809, 5'-Deoxytoyocamycin, (3R,8aR)-3-benzyl-2,3,6,7,8,8a-hexahydropyrrolo[1,2-a]pyrazine-1,4-dione, cyclo(Pro-Leu), cyclo-(L-Pro-4-OH-L-Leu), and cyclo(2-hydroxy-Pro-R-Leu) (Table 2). The pre-MD- MM-GBSA ΔG_{bind} values of these compounds docked against PknG ranged from −34.7 kcal/mol to −42.8 kcal/mol, with cyclo-(L-Pro-4-OH-L-Leu) exhibiting the highest affinity of −42.8 kcal/mol (Table 3.2). To further gain knowledge into the interactions that contributed to the highest ΔG_{bind} observed on PknG-cyclo-(L-Pro-4-OH-L-Leu) the complex was visualized in both two and three dimensions. In the active site of PknG, the hydrophobic interaction of cyclo-(L-Pro-4-OH-L-Leu) is depicted in Figure 3.7 and it contributed to the scaffold's high XP docking score. Cyclo-(L-Pro-4-OH-L-Leu) displayed multiple types of interactions with proximal amino acid residues, which include, hydrogen bonding with Val163 and Gly165, polar interactions with GLN166, negative charge interaction with GLU16, and positive interaction with LYS109 (Figure 3.7). To gain more comprehensive insights into the binding dynamics of the two selected PknG-cyclo-(L-Pro-4-OH-L-Leu) and Pks13- vazabitide A complexes, MD simulations and post-MD-ΔG_{bind} simulations were computed.

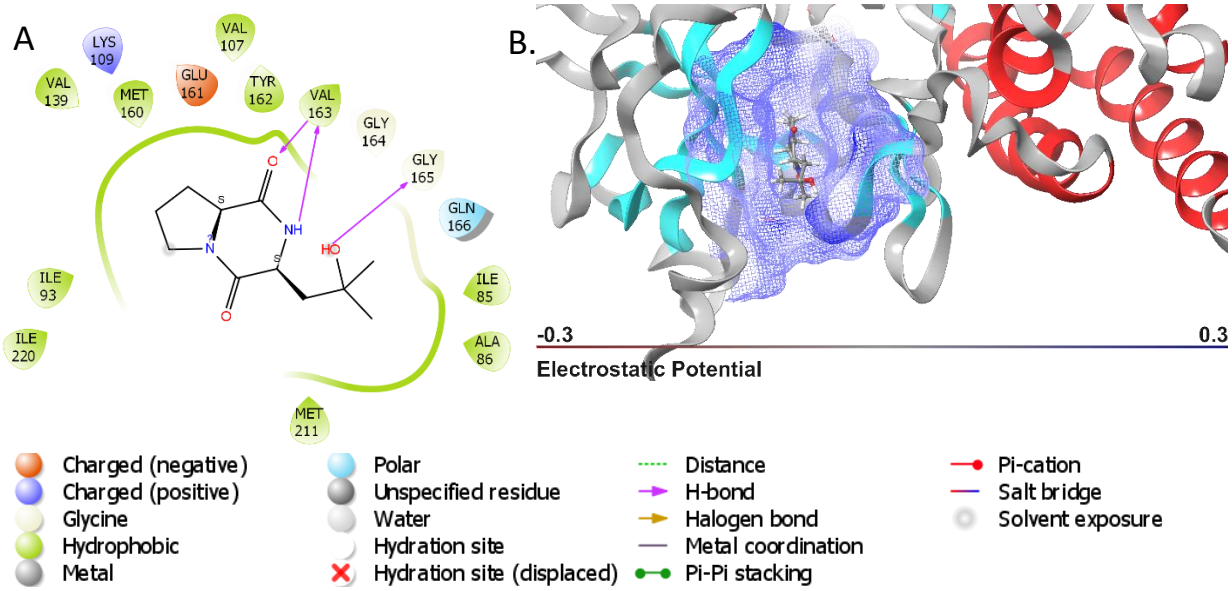


Figure 3.7. A concise superposition of the interaction of cyclo-(L-Pro-4-OH-L-Leu) with the binding pocket of PknG in two dimensions. A. 2D representation B. 3D representation.

Considering the pre-MD binding energies obtained from VSW, two ligands (6 and 8) were selected for further elucidation through MD simulations (Table 3.2). MD simulations were performed for the protein-ligand complexes, and proteins without ligands for 200 ns, and the obtained trajectories were analysed to gain insight into the complexes' interaction dynamics. The co-crystallized ligands were simulated for comparison. The MD simulations of the unbound PknG revealed that the protein's stability decreased during 200 ns simulation as shown by the gradual increase in RMSD from 1.8 Å to 3.6 Å (Figure A.1). While the RMSD of Pks13 increased from 1.2 Å from 0 ns to 25 ns and then stabilised at approximately 2.8 Å for the rest of the simulation as depicted by Figure A.2. During the 200 ns MD simulation of the PknG complexed with cyclo-(L-Pro-4-OH-L-Leu), the RMSD of PknG Cα-atoms remained relatively stable fluctuating below in the range 3.0-1.5 Å for the first 100 ns. However, after 100 ns, the RMSD increased slightly to 3.5 Å and remained constant for the rest of the simulation period (Figure 3.8A). The RMSD of cyclo-(L-Pro-4-OH-L-

Leu) showed that it was mostly stable throughout the simulation, except for a sudden spike fluctuation at 150 ns, in which the ligand slightly shifted from the designated active site and started interacting with the protein's loop (Figure 3.8.A). The stability of the PknG-co-crystallized ligand (8ZC) complex was investigated through a 200 ns MD simulation, and the RMSD of PknG Cα-atoms ranged from 5.6 to 3.2 Å (Figure 3.8.B), indicating that the complex was not stable during the simulation period. The RMSD of 8ZC ranged around 9 Å which shows that the ligand is not strong binder of PknG. A comparison of the RMSD values of PknG-cyclo-(L-Pro-4-OH-L-Leu) and the PknG-8ZC revealed that Cyclo-(L-Pro-4-OH-L-Leu) had a more favorable binding mode with PknG for most parts of the simulation. However, the RMSD profiles of two ligands show that the ligands are not strong binders of PknG. The RMSF revealed that residues 43-51, 147-152, 171-173, and 272-277 were relatively mobile, with RMSF values above 3.0, in both the PknG-cyclo-(L-Pro-4-OH-L-Leu) complex and PknG-8ZC complexes.

Table 3.2. Virtual screening of bacterial compounds against *M. tb* macromolecular targets (PknG and Pks13).

Complex	mol MW (170-725)	Dipole (1.0–12.5)	SASA	*Qplog S (− 6.5 to 0.5)	PSA (7.0–200.0)	Volume	%Human Oral Absorption	XP GScore[a] (kcal/mol)	XP Gscore[b]	Pre-MD-MM-GBSA[a] ΔG_{bind} (kcal/mol)	Pre-MD-MM-GBSA[b] ΔG_{bind} (kcal/mol)
PknG-ligand complex											
Maculosin	260.29	2.96	528.57	–1.09	92.29	875.01	59.48	−9.4	−9.4	−35.7	−35.7
NPA006809	234.26	3.85	466.23	–2.38	103.50	775.73	71.17	−9.4	−9.4	−39.8	−39.8
5'-Deoxytoyocamycin	275.27	6.02	487.25	–3.16	128.19	829.80	54.70	−9.3	−9.3	−38.0	−38.2
(3R,8aR)-3-benzyl-2,3,6,7,8,8a-hexahydropyrrolo[1,2-a]pyrazine-1,4-dione	244.29	1.80	516.11	–1.20	69.75	852.12	72.70	−9.1	−9.0	−41.1	−41.0
Cyclo(Pro-Leu)	210.28	1.90	458.39	–0.05	69.87	764.78	68.98	−8.9	−8.9	−42.6	−42.6
Cyclo-(L-Pro-4-OH-L-Leu)	226.28	3.50	473.85	–0.50	88.06	789.50	64.63	−8.8	−8.8	−42.8	−42.8
Cyclo(2-hydroxy-Pro-R-Leu)	226.26	2.14	330.80	0.11	87.68	781.66	61.59	−8.8	−8.5	−34.7	−35.1
Control **(8ZC)**								−9.5	−9.5	−48.4	−48.4
Pks13-ligand complex											
Vazabitide A	271.32	13.40	515.62	–0.90	133.66		27.76	−11.9	−8.218	−47.6	−47.6
Control (7IJ)								−8.2	−8.2	−42.0	−42.0

[a]Ligands docked using OPLS4 force field. [b]Ligands docked after M06-2X/6-311++G(*d,p*) optimization. SASA (solvent accessible surface area). MW (molecular weight). SASA: Solvent Accessible Surface Area; *QplogS: Predicted Log Solubility; PSA: polar surface area

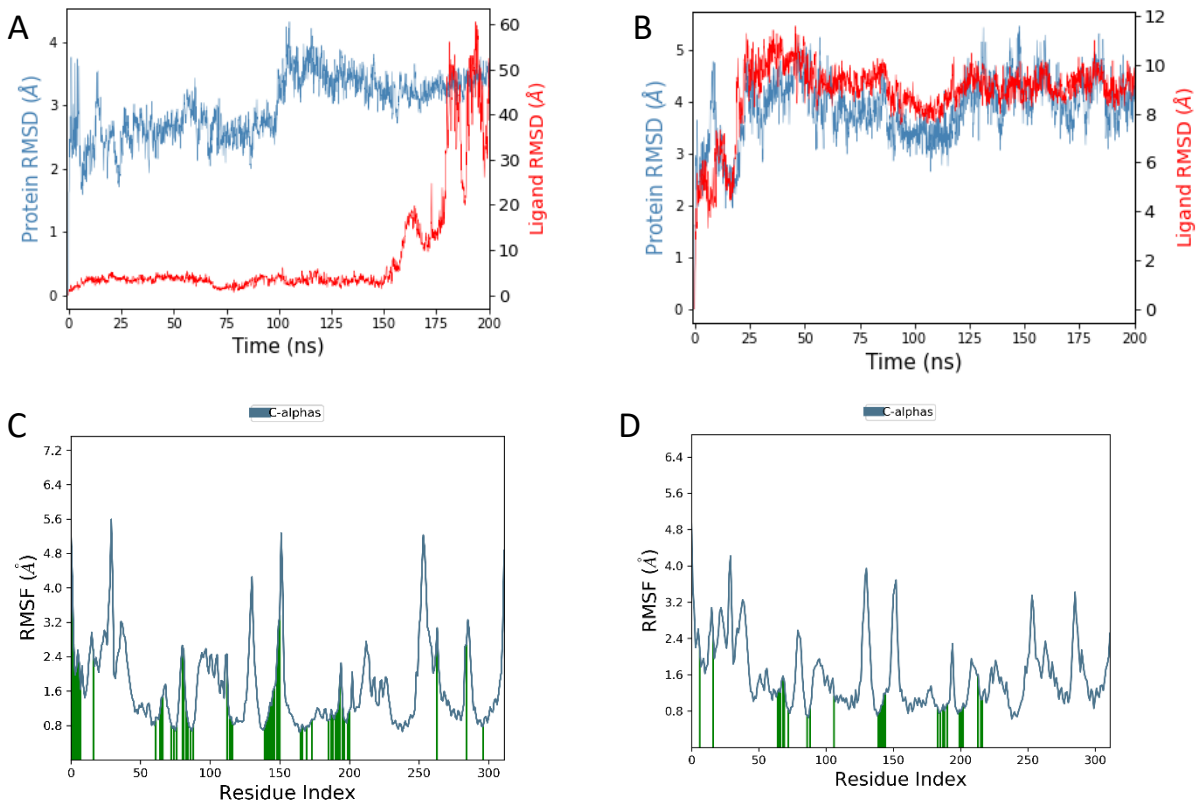


Figure 3.8. A. RMSD of PknG Cα-atoms and Cyclo-(L-Pro-4-OH-L-Leu) over a 200 ns simulation. **B.** RMSD for PknG Cα-atoms and the co-crystallized ligand (8ZC) over a 200 ns simulation. **C.** RMSF per residue of PknG in complex with Cyclo-(L-Pro-4-OH-L-Leu). **D.** RMSF per residue of PknG in complex with 8ZC ligand.

Throughout the 200 ns simulation, the partial stability of the PknG complex with cyclo-(L-Pro-4-OH-L-Leu) was mainly maintained by various non-covalent interactions, including hydrogen bonds involving LEU18, LEU21, H315, VA162, VAL163, GLY164, GLY165, LYS169, ARG170, and GLU214, as well as hydrophobic interactions involving LUE21, ILE85, ALA86, ILE93, MET160, TYR162, MET211, and ILE220. Additionally, water bridges involving ILE15, ASP16, PRO17, GLU19, ALA20, LEU21, ILE85, ASN102, ARG104, GLU161, TYR162, VAL163, GLY164, GLY165, GLU166, SER167, LYS169, ARG170, GLU187, GLU208, THR213, GLU214, and ILE220, as well as ionic interactions, were observed, as illustrated in

Figure 3.9.A. During the simulation, the ligand's carbonyl group located at position 3 formed interactions with protein residues through water bridges for 43 % of the 200 ns MD simulation time with GLU161, and through hydrogen bonding with VAL163 for 63 % of the 200 ns MD simulation time. Additionally, the amide group located at position 4 of the ligand formed hydrogen bonding with the protein residue VAL163 as depicted in Figure 3.9B.

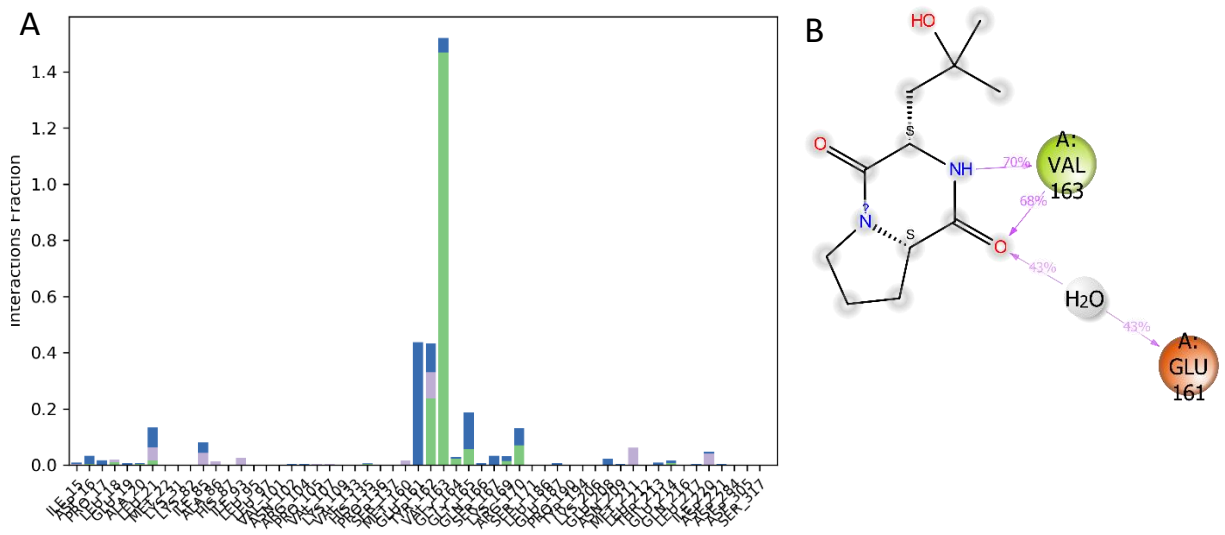


Figure 3.9. Interaction Fraction summary of PknG-cyclo-(L-Pro-4-OH-L-Leu) contacts. This graph is normalized by the total simulation time. A. Interaction fraction of PknG with the ligand cyclo-(L-Pro-4-OH-L-Leu). B. Interactions that occurred for more than 30 % of the 200 ns MD simulation.

The RMSD of Pks13 Cα-atoms complexed with vazabitide A gradually increased from 1.5 to 2.4 Å for the first 50 ns and then remained stable between 2.1 and 2.4 Å up to 200 ns as illustrated in Figure 3.10.A. The low RMSD value below 3 Å shows that the complex was stable. In contrast, the co-crystallized ligand (7IJ) also formed an excellent stable complex with the RMSD averaging around 1.5 Å, Figure 3.10.B. The RMSF of Pks13 Cα-atoms in all two systems showed that the residues did not fluctuate a lot except for the mobile residues (10-20 and 170-180).

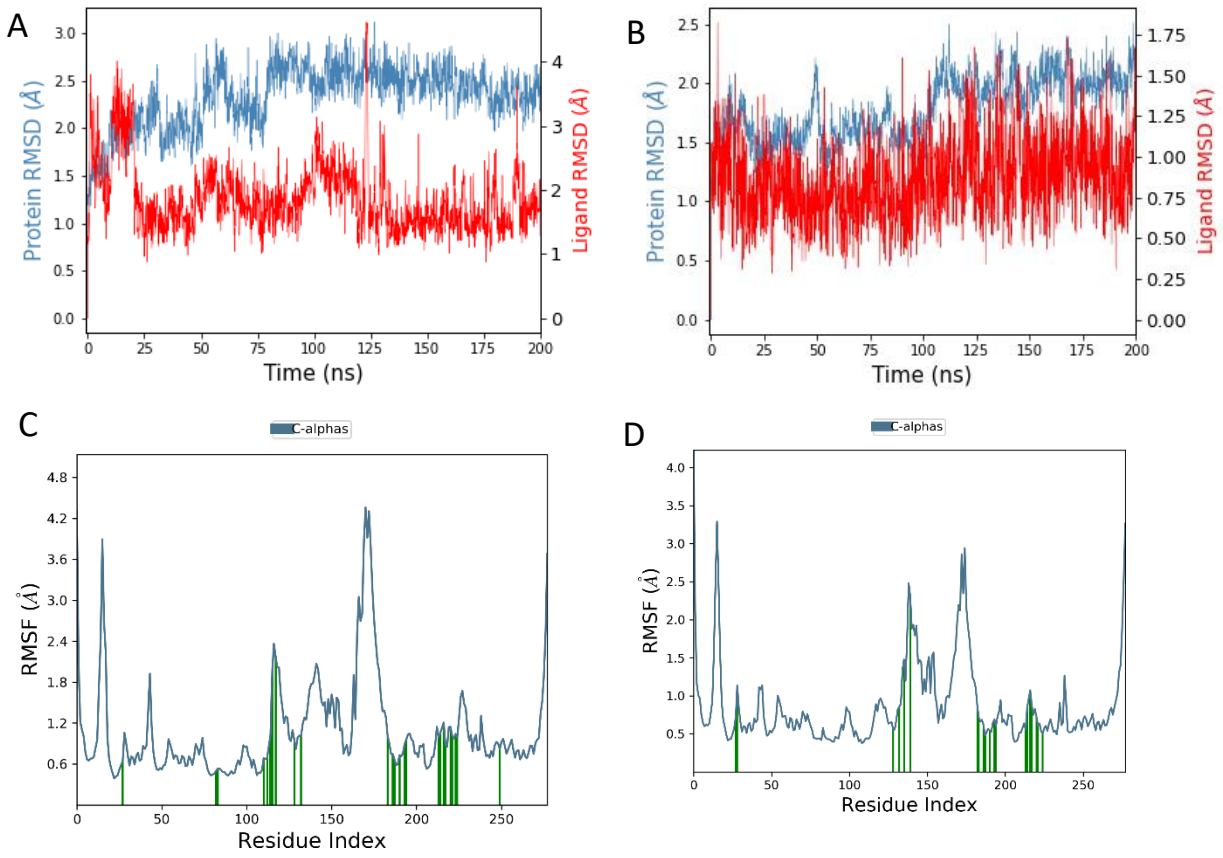


Figure 3.10. A. RMSD of Pks13 Cα-atoms and the vazabitide A over a 200 ns simulation. **B.** RMSD for Pks13 Cα-atoms and 7IJ over a 200 ns simulation. **C.** RMSF per residue of Pks13 in complex with the vazabitide A. **D.** RMSF per residue of Pks13 in complex with 7IJ.

The stability of Pks13- vazabitide A complex was significantly attributed to various non-covalent interactions, including hydrogen bonding with key protein residues such as ASP1562, ASN1640, ASP1644, HIS1664, GLU1671, and TYR1674. Water bridges also critically contributed to the stability, involving residues such as ALA1477, SER1533, ASP1560, ALA1564, GLU1567, TRY1582, GLN1633, SER1636, ASN1640, ASP1644, HIS1664, ASP1666, ALA1667, PHE1670, GLU1671, TRY1674, and HIS1699. Ionic interactions also cumulatively enhanced stability, involving residues such as ASP1560, GLU1567, ASP1644, and GLU167 as depicted in Figure

3.11.A. Figure 11B illustrates that the amino group of vazabitide A located on position 7 forms a hydrogen bond with GLU1671 and ASP1644, occurring for 34 % and 98 % of the 200 ns MD simulation time, respectively, and engaged in electrostatic interactions with the two protein residues. The amide group located on position 5 of vazabitide A also formed a hydrogen bond with ASN1640, occurring for 77% of the 200 ns MD simulation time. Vazabitide A also formed intramolecular hydrogen bonds between the hydroxyl group located at position 1 and a carboxylic acid group located at position 14.

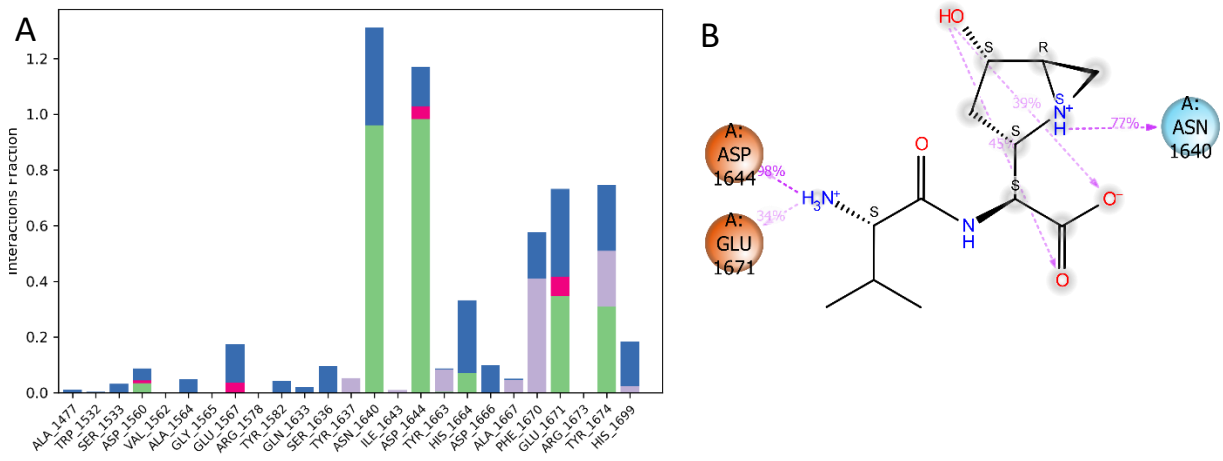


Figure 3.11. Interaction Fraction summary of Pks13- vazabitide A contacts. This graph is normalized by the total simulation time. **A**. Interaction fraction of the vazabitide A with Pks13. **B.** Interactions that occurred for more than 30 % of the 200 ns MD simulation.

Post-MD MM-GBSA binding free energy was used to determine and compare the binding affinities of the selected ligands against the co-crystallized ligands as controls as presented in Table 3.3. The binding free energy of Cyclo-(L-Pro-4-OH-L-Leu) for the PknG is−21.1 kcal/mol, while that of the control ligand 8ZC is −31.2 kcal/mol. The binding affinity of the co-crystallized ligand (7IJ) for Pks13 had a favorable more negative binding free energy (−83.4 kcal/mol) than for

vazabitide A (−37.2 kcal/mol). The ΔG_{bind} *vdW* interactions had a major contribution to the stability of the Pks13- vazabitide A complex. Some of the interactions that contributed to the stability include ΔG_{bind} *Lipophilicity,* ΔG_{bind} *Solvation GB,* ΔG_{bind}Hbond, and ΔG_{bind} Coulomb as shown in Table 3.3.

Table 3.3. Post-MD MM-GBSA binding free energy computation.

Complex	**MM-GBSA ΔG_{bind} kcal/mol**	**ΔG_{bind} Coulomb kcal/mol**	**ΔG_{bind} Covalent kcal/mol**	**ΔG_{bind} Hbond kcal/mol**	**ΔG_{bind} *Solv GB* kcal/mol**	**ΔG_{bind} *Lipo* kcal/mol**	**ΔG_{bind} *vdW* kcal/mol**
PknG-**(** cyclo-(L-Pro-4-OH-L-Leu)**)**	−21.1	−7.6	0.9	−0.6	8.2	−5.9	−16.16
PknG-co-crystallized (8ZC)	−31.2	−46.9	0.1	−1.5	54.6	−9.4	−28.15
Pks13-vazabitide A	−37.2	−11.4	1.8	−2.1	18.6	−10.0	−34.0
Pks13-co-crystallized (7IJ)	−83.4	−38.8	3.1	−1.9	49.3	−24.0	−57.2

ΔG_{bind}vdW = van der Waals contribution; Δ_{bind}Covalent = covalent bonding contribution; ΔG_{bind}Solv = polar contribution of solvation energy; ΔG_{bind}Lipophilicity = lipophilicity energy contribution; ΔG_{bind}Hbond = hydrogen bonding contribution; ΔG_{bind}Coulomb = electrostatic interaction; ΔG_{bind} = binding free energy.

3.4.5 *In-silico* evaluation of the modified compound

The modified compound identified by the following smiles notation (c1cc(Cl)c(Cl)cc1C[C@@H](N(C(=O)[C@@H]2[NH3+])C(=O)O[C@H](CC3)CCC23)CC(=O)NCc4c[nH+]cn4CCOCC) exhibited an improved profile such as the percentage of human oral absorption increased to 71.28, a lower docking score −13.2 kcal/mol than the score that was observed for vazabitide A which was −11.9 kcal/mol. Furthermore, the modified compound

showed a ΔG_{bind} of −81. 5 kcal/mol which is incredibly lower than that of vazabitide A (Table 3.4). All the other parameters were within the acceptable range.

Table 3.4. Virtual screening of the modified compound against *M. tb* Pks13.

Complex	mol MW (170-725)	Dipole (1.0–12.5)	SASA	*QplogS (− 6.5 to 0.5)	PSA (7.0–200.0)	Volume	%Human Oral Absorption	XP GScore (kcal/mol)	MM-GBSA ΔG_{bind} kcal/mol
7VJT 300405311 + 41735949 + 44455695	580.51	1.72	847.90	−4.52	126.21	1658.13	71.28	−13.2	−81.5

The position and chemical properties of atoms in the modified compound allow for a variety of interactions that contribute to stabilizing the Pks13-ligand complex as shown in Figure 3.12.B. The interaction of the modified compound with proximal amino acids residues of Pks13 included, hydrogen bond with ASN1640, and HE1644, hydrophobic interaction with ALA1477, PRO1476, TYR1663, ALA1667, PHE1670, TYR1674, ILE1700, TRP1532, ILE1643, TYR1582, and PHE1585, polar interactions with SER1533, HIS1699, ASN1640, SER1636, GLN1633, HE1632 and HIE1664, positive charges with ARG1641, and ARG1578, negative charges Arg1641, and Arg1578, pi-cation interaction with TYR1663, pi-pi stacking with PHE1670 , TYR1674 , and salt bridge with ASP1666 as shown in Figure 3.12.A.

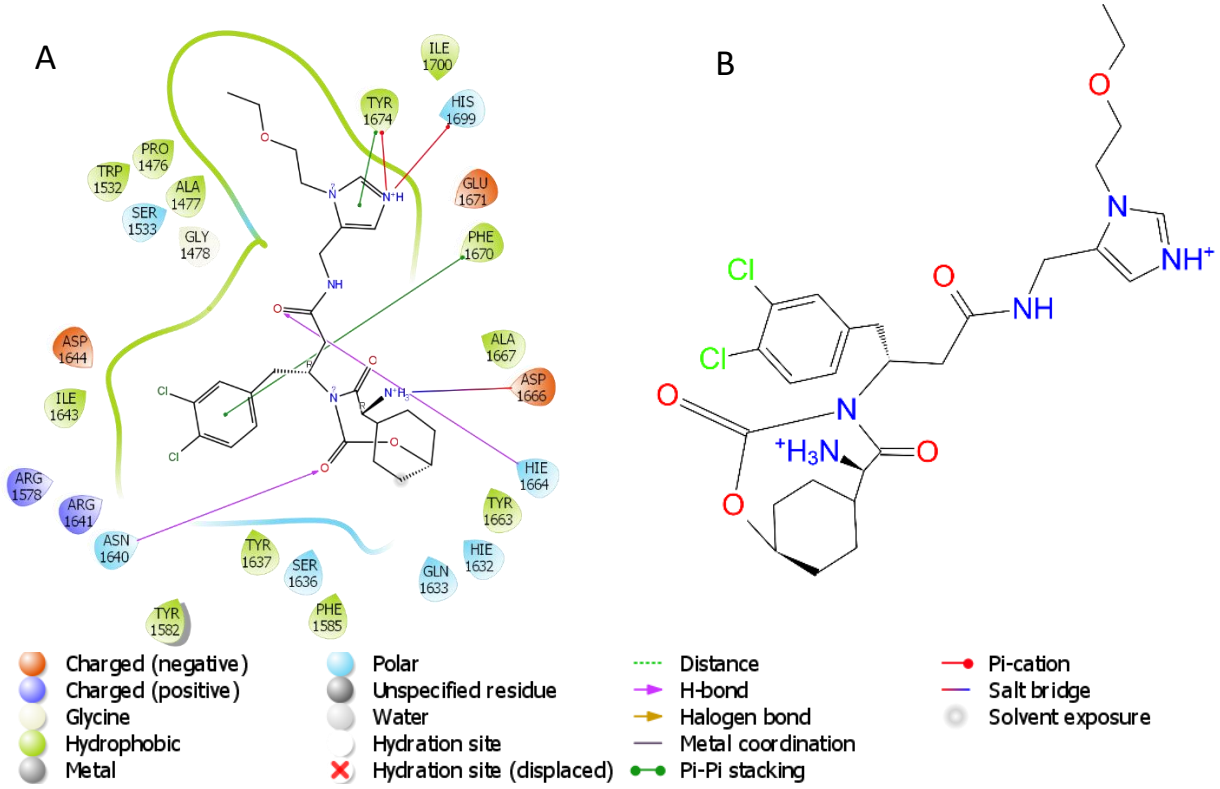


Figure 3.12. A concise overview of the interaction of the modified compound with Pks13 in two dimensions. A. Superposition of the interaction of the modified compound with the binding pocket of Pks13. B. Modified compound.

The RMSD of Pks13 Cα-atoms was computed during a 200 ns molecular dynamics (MD) simulation, and the results indicated that the protein Cα-atoms underwent a gradual increase in deviation from 0.9 Å to approximately 2.1 Å within the first 50 ns and then stabilized at approximately 2.1 Å up to 150 ns as shown in Figure 3.13. Subsequently, the RMSD began to gradually decrease to approximately 1.8 Å up to 200 ns, which revealed that the protein was stable throughout the MD simulation. Conversely, the ligand RMSD fluctuations were slightly above 3 Å during the MD simulations, indicating that the ligand was mobile within the active site of the

protein. The RMSF of Pks13 Cα-atoms showed that the residues did not fluctuate a lot except for the residues (10-20 and 170-180) which were constantly moving.

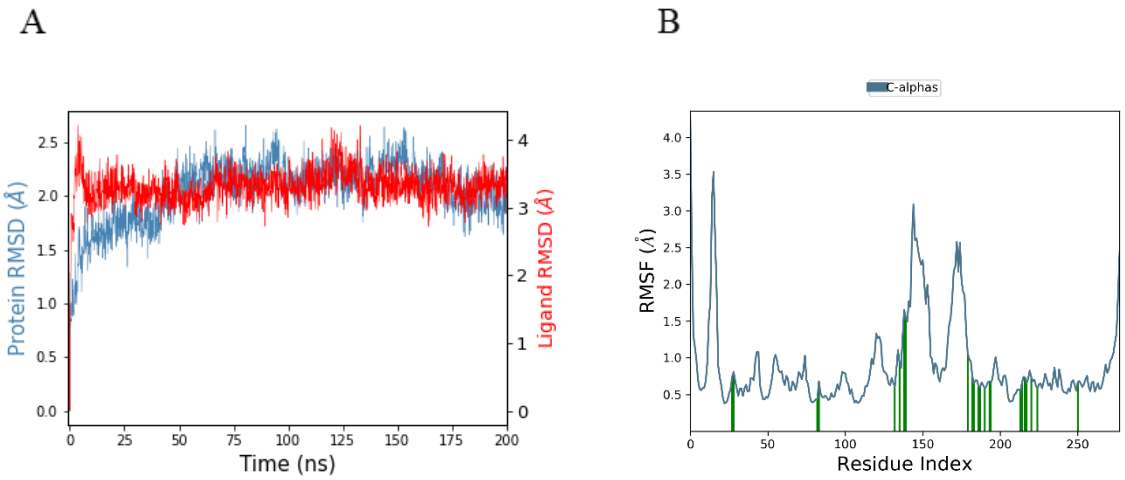


Figure 3.13. A. RMSD of Pks13 Cα-atoms and the modified compound over a 200 ns simulation. **B**. RMSF per residue of Pks13 in complex with the modified compound.

During the 200 ns molecular dynamics simulation, the carbonyl groups at positions 4 and 9 of the modified compound exhibited persistent hydrogen bonding interactions with polar amino acids ASN1640 and GLN1633, respectively, for a substantial fraction of the simulation time as shown in Figure 3.14.B. Furthermore, the nitrogen atom on position 32 formed water bridges with ASP1644 for 40 % of the simulation time. The substituted benzene ring, bearing two chlorine atoms at positions 31 and 28, contributed to hydrophobic interactions with nearby amino acid residues, including PHE1670, PHE1585, and TYR1582, throughout the simulation period. Notably, the substituted benzene ring also engaged in pi-pi stacking interactions with PHE1670 for 50 % and with PHE1585 for 45 % of the simulation duration, underscoring the aromatic nature

of the interaction. Overall, hydrogen bonds, hydrophobic interactions, ionic interactions, and water bridges influenced the stability of the protein-ligand complex.

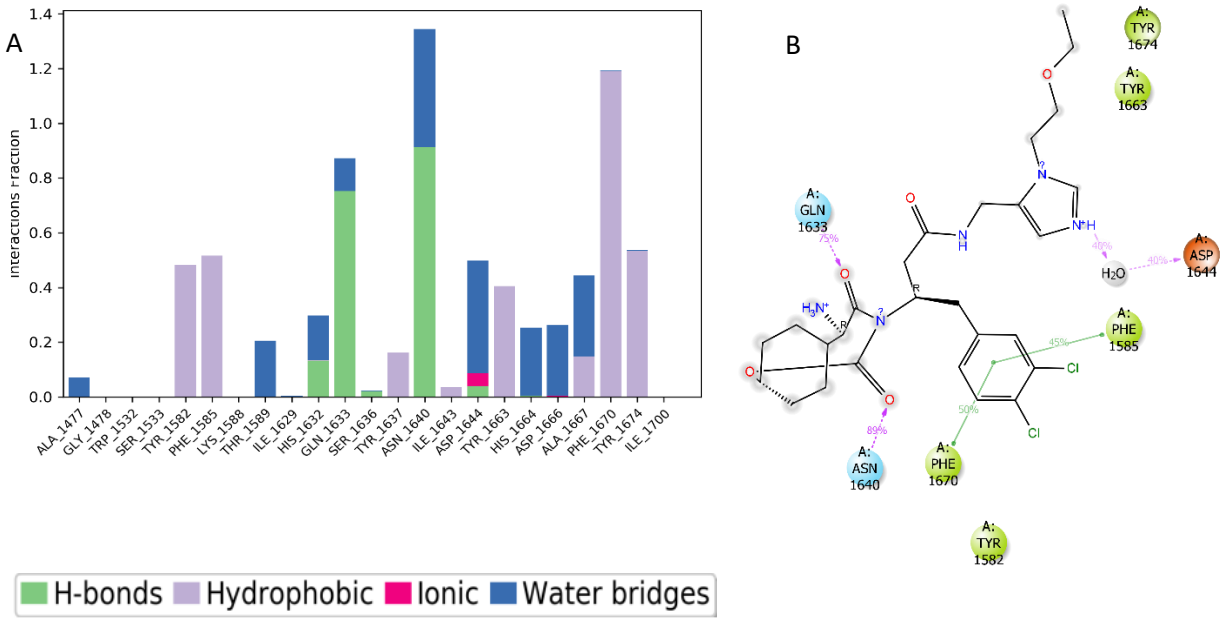


Figure 3.14. Interaction Fraction summary of Pks13- modified compound contacts. This graph is normalized by the total simulation time. **A.** Interaction fraction of the modified compound with Pks13. **B.** Interactions that occurred for more than 30 % of the 200 ns MD simulation.

The modified compound exhibited a higher binding affinity to Pks13 as indicated by the MM-GBSA ΔG_{bind} of −85.8 kcal/mol than co-crystallized (7IJ) and vazabitide A (Table 3.5, and Table 3.3). The ΔG_{bind} Coulomb and ΔG_{bind} *vdW* interactions significantly contributed to the favorable binding energy of the Pks13-vazabitide A complex. The ΔG_{bind} Lipophilicity of the modified compound increased and was more than that of the co-crystallized (7IJ) and vazabitide A.

Table 3.5. Post-MD MM-GBSA binding free energy computation.

Complex	MM-GBSA ΔG_{bind} kcal/mol	ΔG_{bind} Coulomb kcal/mol	ΔG_{bind} Covalent kcal/mol	ΔG_{bind} Hbond kcal/mol	ΔG_{bind} Solv GB kcal/mol	ΔG_{bind} Lipo kcal/mol	ΔG_{bind} vdW kcal/mol
Pks13-modified compound	-85.8	-69.0	0.4	-1.1	64.8	-22.5	-54.5

ΔG_{bind}vdW = van der Waals contribution; Δ_{bind}Covalent = covalent bonding contribution; ΔG_{bind}Solv = polar contribution of solvation energy; ΔG_{bind}Lipophilicity = lipophilicity energy contribution; ΔG_{bind}Hbond = hydrogen bonding contribution; ΔG_{bind}Coulomb = electrostatic interaction; ΔG_{bind} = binding free energy.

3.5 Discussion

Gold mine tailings are composed of various fractions of mineral species, for example, nonessential elements such as Pb, As, and essential elements such as Fe, Mg, Co, Mn, Cr, K, Cn [37]. Literature has shown that the acidic pH in gold tailings molecularly modifies the bioavailability and solubility of the heavy metals, creating an extreme environmental niche and, thus, selectively modulating microbial proliferation [38–40]. Therefore, the indigenous bacteria to the tailings must possess the molecular machinery that enables them to be intimately associated with heavy metals transformation and tolerate the oxidative stress caused by heavy metals [41]. Various reports have described the intimate interaction of *Bacillus sp*., *Acidithiobacillus sp*., *Arthrobacter sp*., *Pseudomonas sp*., *Microbacterium sp*., and *Sphingomonas sp* with heavy metals [38, 42, 43]. Similar to previous reports, the bacteria in this study are actinobacteria, which are commonly found in different soil environments [44, 45]. *Bacillus sp* was the most abundant culturable bacteria from South African gold mine tailings.

In accordance with the results from this report that showed that *Bacillus sp* isolated from the heavy metal-rich environment, produced multiple isomers of surfactins that possibly enhanced the solubilization of heavy metals (Table A1). The literature reveals that *Bacillus sp*., *Micrococcus*

sp., and *Pseudomonas sp*. are able to survive under abiotic stress [43, 46, 47]. Molecular mechanisms to alleviate heavy metal stress and striving in low pH in bacteria identified in this study are well characterize, suggesting their unique nature of metal acquisition and resistance determinants for inhabiting heavy metal-rich environments [41]. Fundamentally, the surfactins/siderophores are extracellularly secreted by *Bacillus sp* to facilitate the removal of heavy metals from surrounding environments, increase the bioavailability of water-insoluble nutrients, and communicate between bacterial cells as well as an antimicrobial agent [48].

A previous study illustrated that bacterial species that are closely related and with higher biosynthetic gene cluster homology have a lower chance of inhibiting each other, whilst distant species in the same genus are likely to suppress each other more fiercely [49]. In this study, the crude extracts from *B. licheniformis, S. mycarofaciens,* and *B. subtilis* showed potent antimicrobial efficacy against *M. smegmatis* MC^2155, and *M. aurum* A+., whilst *B. paralicheniformis* only showed activity against *M. smegmatis* MC^2155. Only *B. subtilis and B. licheniformis* exhibited activity against *M. tb. B. subtilis* and *B. licheniformis* are known for producing lethal broad-spectrum antimicrobials arsenal that inhibits Gram-positive and Gram-negative bacteria by downregulating peptidoglycan synthesis [50]. Our results are supported by reports from [51, 52] which demonstrated that *B. licheniformis* CG1 produced bioactive metabolites that inhibited the growth of *M. smegmatis.*

The current study showed promising antimycobacterial activity of crude extracts from *B. subtilis* and *B. licheniformis,* thus, metabolite profiling was performed to tentatively identify compounds responsible for activity (Figure 3.2; Table A.1). Microorganisms existing in extremophilic niches are known to have a higher probability of producing multifarious novel bioactive chemical classes

[28]. Through pathway enrichment analysis several pathways associated with bioactivity were observed in this study, including the biosynthesis of secondary metabolites and other antibiotics (Figure 3.3; Table A.1). These pathways play essential roles in the production of various bioactive metabolites. The identification of these pathways provides valuable insights into the metabolic skill of bacteria from gold mine tailings and may aid in the development of novel antimycobacterial agents through genome mining of biosynthetic functional gene clusters. Results from this investigation agree with the literature that revealed that microorganisms from extreme environments can produce unique metabolites as was revealed by many uncharacterized compounds (Figure 3.2). It has been reported that the Gram-positive *Bacillus* genus is strongly associated with the production of diverse secondary metabolites, for example, nonribosomal polyketides, nonribosomal lipopeptides, ribosomally synthesized and post-translationally modified peptides, and peptide-polyketide hybrid compound [49, 53]. Findings from this study are consistent with reports, showing that cyclic lipopeptides, and cyclic dipeptides, for instance, cyclo(proline-leucine), isomers of surfactin, and cyclo(L-Leu-L-4-Hyp), iturin derivatives are produced by *B. subtilis* and *B. licheniformis* [52]. Surfactins, cyclo (proline-leucine), and cyclo (L-Leu-L-4-Hyp) produced by various *Bacillus* sp have been revealed to possess pronounced permeabilization of microbial cell membranes and anti-microbial activity against Gram-positive and Gram-negative bacteria [52].

While there is limited research specifically on the activity of the isomers of surfactin, e.g., surfactin C, surfactin A, and surfactin D against *M. tb* H37Rv, some studies have investigated the activity of crude extracts containing surfactin against the *Mycobacteria* species [52]. The current study agrees with the findings that revealed that multiple isomers were tentatively identified in a *Bacillus*

sp. crude extract. In addition, the crude extract from the *Bacillus* sp exhibited potent activity against *M. tb* H37Rv. In the current study, strong inhibition exhibited by the crude extracts of *B. subtilis* and *B. licheniformis* may be due to increased pore formation of the *Mycobacterium* cell membrane caused by different surfactin isoforms and other cyclic lipopeptides present in the crude extracts. The synergistic effect of cyclic lipopeptides, and cyclic dipeptides may have ultimately contributed to the potent anti-mycobacterial efficacy against *M. smegmatis* MC^2155, *M. aurum* A+, and *M. tb* H37Rv as shown by MICs from *B. subtilis* and *B. licheniformis* (Table A1).

Interaction mapping from molecular docking and MD simulations is an important step in revealing the significance of ligand binding toward the stability of a protein-ligand complex and the inhibition of a protein target. The HOMO-LUMO results in the study showed that the compounds returned from the virtual screening are soft and highly reactive, as observed by the low docking scores. This study investigated the contribution of non-covalent interactions to the stability of PknG- cyclo-(L-Pro-4-OH-L-Leu) and Pks13- vazabitide A complexes during 200 ns MD simulations. The results revealed that various types of non-covalent interactions, including hydrogen bonding, water bridges, ionic interactions, and hydrophobic interactions played a critical role in restricting major conformational changes of the protein-ligand complexes. According to the HOMO-LUMO results, the interaction with the nitrogen rings was as expected. Specifically, key residues, including GLU167, ASN1640, and ASP1644 formed hydrogen bonds with vazabitide A which occurred for more than 30 % of the 200 ns simulation time and were crucial in stabilizing the Pks13- vazabitide A complex. The post-molecular dynamic simulation analysis was carried out by calculating the protein-ligand binding free energies based on MD simulation trajectories**.** A comparison of the MM/GBSA binding free energy of vazabitide A and cyclo-(L-Pro-4-OH-L-Leu)

is supporting the RMSD profiles from the two Pks13 complexes which showed excellent stability to the co-crystallized ligand (7IJ). Vazabitide A also interacted with the SER1533 of Pks13 thioesterase domain (TE) through water bridges. The TE domain is acyltransferase that is responsible for cleaving the thioester bond and forming an ester bond between the mycolic β-ketoester and the hydroxyl group of Ser1533 of the TE domain to form a trehalose monomycolate ketone. Studies have revealed that blocking the TE domain of Pks13 abolished the biogenesis of mycolic acids and consequently inhibits the growth of *M. tb* [19]. It is important to note that vazabitide A is a natural compound and further scaffold modification can enhance important properties, including affinity, toxicity, and ultimately activity. Furthermore, in this study, the RMSD generated from the trajectory of PknG Cα-atoms complexed with cyclo-(L-Pro-4-OH-L-Leu) during 200 ns simulation revealed that cyclo-(L-Pro-4-OH-L-Leu) had a relatively better binding dynamics as compared to the co-crystallized ligand for the first 100 ns. These results shed light on the relatively favorable binding mode of the cyclo-(L-Pro-4-OH-L-Leu) to PknG for the first 100 ns. However, the MM-GBSA free energy calculations show the control ligand (8ZC) had a more favorable energy profile because the binding free energy was calculated for the last 1000 frames. The interaction analysis suggests that some of the binding characteristics may be conserved while others are increased during scaffold modification against Pks13.

The carbonyl groups at positions 4 and 9 of the compound formed persistent hydrogen bonding interactions with polar amino acids ASN1640 and GLN1633, respectively, for a significant portion of the simulation time (Figure 3.12). This reveals that the electrostatic and polar interactions between the carbonyl groups and the amino acids were sufficiently strong and contributed to maintaining the stability of the Pks13-modified compound-complex. Furthermore, the substituted

benzene ring with two chlorine atoms contributed to the lipophilicity (Table 3.4) and binding affinity in the hydrophobic active site of Pks13 (Figure 3.12). The nitrogen on the imidazole ring at position 32 of the modified compound formed water bridges with ASP1644, highlighting the presence of polar interactions which played a crucial role in stabilizing the Pks13-modified compound complex. Literature has reported various activities of microbial activity of the structural motifs with an imidazole ring. These findings provide a different perspective on the developing of new antimycobacterial scaffolds with improved potency and selectivity [54–56].

3.6 Conclusion

The study suggests that the South African gold mine tailings niche may be a good source of fiercely antagonistic bacteria exhibiting antimycobacterial efficacy. As such, *B. licheniformis, S. mycarofaciens*, and *B. subtilis* isolated from a metal-rich niche might have the potential to inhibit *Mycobacterium* strains via the production of potent broad-spectrum functional secondary metabolites. Our results offer valuable insights into the metabolic pathways of *B. subtilis*, highlighting potential targets for antibiotic production. These findings may prove crucial for future studies in this area, advancing our understanding of this versatile microorganism and its antibiotic-producing capabilities. These findings provide valuable insights into the interaction mapping of protein-ligand complexes and may have implications for the development of novel therapeutic agents targeting PknG and Pks13. This could help reduce expenses and save time spent on processes such as extraction, purification, and retesting.

Author Contributions: Conceptualization, V.M., and K.N.; methodology, V.M., K.N., K.I.T., K.G., D.J.W., and M.N; formal analysis, K.N.; investigation, K.N.; resources, V.M., M.N., L.S., K.G., L.W., and I.SN.; writing original draft preparation, K.N.; writing review and editing, V.M.,

M.N., K.G., K.I.T., L.B., L.S., D.J.W., L.W., and I.S.N.; supervision, V.M., M.G., K.G., L.S, and I.S.N; project administration, L.S.; funding acquisition, V.M., and L.S. Molecular docking and Molecular dynamics simulations, K.N., K.G., K.I.T., D.J.W., and M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the South African Medical Research Association Council (SAMRC), the Centre of Excellence for Biomedical Tuberculosis Research (CBTBR), and the National Research Fund (NRF) [NRF GRANT UID129364]. K.N. is highly indebted to the CTR and CBTBR for financial support.

Ethical Approval: Ethical approval for this study was approved by the Research Ethics Committee: Biological and Environmental Safety (REC: BES) of Stellenbosch University with reference number **BEE-2022-3188**.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: We acknowledge the Open Access Fund of Stellenbosch University. We would also like to extend our appreciation to the Center for High-Performance Computing, South Africa, for providing us with access to the infrastructure used to conduct the simulations. The authors would like to acknowledge and thank Promolab (Pty) Ltd, trading as Separations, South Africa for their contribution to the X500R QTOF instrument.

3.7 References

1. Kanehiro, Y.; Tomioka, H.; Pieters, J.; Tatano, Y.; Kim, H.; Iizasa, H.; Yoshiyama, H. Identification of Novel Mycobacterial Inhibitors against Mycobacterial Protein Kinase G. *Front Microbiol* **2018**, *9*, 1517, doi:10.3389/FMICB.2018.01517/BIBTEX.

2. World TB Day 2023 Available online: https://www.who.int/campaigns/world-tb-day/2023 (accessed on 4 April 2023).

3. Guzman, J.D.; Mortazavi, P.N.; Munshi, T.; Evangelopoulos, D.; McHugh, T.D.; Gibbons, S.; Malkinson, J.; Bhakta, S. 2-Hydroxy-Substituted Cinnamic Acids and Acetanilides Are Selective Growth Inhibitors of Mycobacterium Tuberculosis. *Medchemcomm* **2013**, *5*, 47–50, doi:10.1039/C3MD00251A.

4. Barale, S.S.; Ghane, S.G.; Sonawane, K.D. Purification and Characterization of Antibacterial Surfactin Isoforms Produced by Bacillus Velezensis SK. *AMB Express* **2022**, *12*, 1–20.

5. Hamdache, A.; Lamarti, A.; Aleu, J.; Collado, I.G. Non-Peptide Metabolites from the Genus Bacillus. *J. Nat. Prod* **2011**, *74*, 893–899, doi:10.1021/np100853e.

6. Morelli, L.; Maris, S.; Rivera, R.; Rokana, N.; Panwar, H.; Elshaghabee, F.M.F.; Gulhane, R.D.; Sharma, C. Bacillus As Potential Probiotics: Status, Concerns, and Future Perspectives. **2017**, doi:10.3389/fmicb.2017.01490.

7. Etchegaray, A.; De Castro Bueno, C.; De Melo, I.S.; Tsai, S.M.; De Fátima Fiore, M.; Silva-Stenico, M.E.; De Moraes, L.A.B.; Teschke, O. Effect of a Highly Concentrated Lipopeptide Extract of Bacillus Subtilis on Fungal and Bacterial Cells. *Arch Microbiol* **2008**, *190*, 611–622.

8. Medeot, D.B.; Fernandez, M.; Morales, G.M.; Jofré, E. Fengycins From Bacillus Amyloliquefaciens MEP218 Exhibit Antibacterial Activity by Producing Alterations on the Cell Surface of the Pathogens Xanthomonas Axonopodis Pv. Vesicatoria and Pseudomonas Aeruginosa PA01. *Front Microbiol* **2020**, *10*, 3107, doi:10.3389/FMICB.2019.03107/XML/NLM.

9. Jones, S.E.; Elliot, M.A. Streptomyces Exploration: Competition, Volatile Communication and New Bacterial Behaviours. *Trends Microbiol.* **2017**, *25*, 522–531, doi:10.1016/j.tim.2017.02.001.

10. Fang, Q.; Maglangit, F.; Wu, L.; Ebel, R.; Kyeremeh, K.; Andersen, J.H.; Annang, F.; Pérez-Moreno, G.; Reyes, F.; Deng, H. Signalling and Bioactive Metabolites from Streptomyces Sp. RK44. *Molecules* **2020**, *25*, doi:10.3390/molecules25030460.

11. Bruning, J.B.; Murillo, A.C.; Chacon, O.; Barletta, R.G.; Sacchettini, J.C. Structure of the Mycobacterium Tuberculosis D-Alanine:D-Alanine Ligase, a Target of the Antituberculosis Drug D-Cycloserine. *Antimicrob Agents Chemother* **2011**, *55*, 291–301.

12. Hartkoorn, R.C.; Sala, C.; Neres, J.; Pojer, F.; Magnet, S.; Mukherjee, R.; Uplekar, S.; Boy-Röttger, S.; Altmann, K.H.; Cole, S.T. Towards a New Tuberculosis Drug: Pyridomycin – Nature's Isoniazid. *EMBO Mol Med* **2012**, *4*, 1032–1042, doi:10.1002/EMMM.201201689.

13. Fashola, M.O.; Ngole-Jeme, V.M.; Babalola, O.O. Heavy Metal Pollution from Gold Mines: Environmental Effects and Bacterial Strategies for Resistance. *Int J Environ Res Public Health* 2016, *13*.

14. Lin, Y.; Fan, H.; Hao, X.; Johnstone, L.; Hu, Y.; Wei, G.; Alwathnani, H.A.; Wang, G.; Rensing, C. Draft Genome Sequence of Halomonas Sp. Strain HAL1, a Moderately Halophilic Arsenite-Oxidizing Bacterium Isolated from Gold-Mine Soil. *J Bacteriol* 2012, *194*, 199–200.

15. Sanyal, S.K.; Reith, F.; Shuster, J. A Genomic Perspective of Metal-Resistant Bacteria from Gold Particles: Possible Survival Mechanisms during Gold Biogeochemical Cycling. *FEMS Microbiol Ecol* **2020**, *96*, 111, doi:10.1093/FEMSEC/FIAA111.

16. Chen, Y.Y.; Chen, L.Y.; Chen, P.J.; El-Shazly, M.; Peng, B.R.; Chen, Y.C.; Su, C.H.; Su, J.H.; Sung, P.J.; Yen, P.T.; et al. Probing Anti-Leukemic Metabolites from Marine-Derived Streptomyces Sp. LY1209. *Metabolites* **2022**, *12*, 320, doi:10.3390/METABO12040320/S1.

17. Milojevic, T.; Albu, M.; Kölbl, D.; Kothleitner, G.; Bruner, R.; Morgan, M.L. Chemolithotrophy on the Noachian Martian Breccia NWA 7034 via Experimental Microbial Biotransformation. *Commun Earth Environ* **2021**, *2*, doi:10.1038/s43247-021-00105-x.

18. Garcia-Vallve, S.; Palan, J.; Romeu, A. Horizontal Gene Transfer in Glycosyl Hydrolases Inferred from Codon Usage in Escherichia Coli and Bacillus Subtilis. *Molecular Biology and Evololution* **1999**, *16*, 1125–1134.

19. Santhi Sudha, S.; Aranganathan, · V Experimental Elucidation of an Antimycobacterial Bacteriocin Produced by Ethnomedicinal Plant-Derived Bacillus Subtilis (MK733983). **2021**, *203*, 1995–2006, doi:10.1007/s00203-020-02173-7.

20. Singh, V.; Chibale, K. Strategies to Combat Multi-Drug Resistance in Tuberculosis. *Acc Chem Res* **2021**, *54*, 2361–2376.

21. Basarab, G.S.; Ghorpade, S.; Gibhard, L.; Mueller, R.; Njoroge, M.; Peton, N.; Govender, P.; Massoudi, L.M.; Robertson, G.T.; Lenaerts, A.J.; et al. Spiropyrimidinetriones: A Class of DNA Gyrase Inhibitors with Activity against Mycobacterium Tuberculosis and without Cross-Resistance to Fluoroquinolones. *Antimicrob Agents Chemother* **2022**, *66*.

22. Schön, T.; Miotto, P.; Köser, C.U.; Viveiros, M.; Böttger, E.; Cambau, E. Mycobacterium Tuberculosis Drug-Resistance Testing: Challenges, Recent Developments and Perspectives. *Clinical Microbiology and Infection* 2017, *23*, 154–160.

23. Nimmo, C.; Millard, J.; van Dorp, L.; Brien, K.; Moodley, S.; Wolf, A.; Grant, A.D.; Padayatchi, N.; Pym, A.S.; Balloux, F.; et al. Population-Level Emergence of Bedaquiline and Clofazimine Resistance-Associated Variants among Patients with Drug-Resistant Tuberculosis in Southern Africa: A Phenotypic and Phylogenetic Analysis. *Lancet Microbe* **2020**, *1*, e165–e174, doi:10.1016/s2666-5247(20)30031-8.

24. Kidwai, S.; Bouzeyen, R.; Chakraborti, S.; Khare, N.; Das, S.; Gosain, T.P.; Behura, A.; Meena, C.L.; Dhiman, R.; Essafi, M.; et al. *NU-6027 Inhibits Growth of Mycobacterium Tuberculosis by Targeting Protein Kinase D and Protein Kinase G*; 2019;

25. Nagarajan, S.N.; Lenoir, C.; Grangeasse, C. Recent Advances in Bacterial Signaling by Serine/Threonine Protein Kinases. *Trends Microbiol* **2022**, *30*, 553–566, doi:10.1016/J.TIM.2021.11.005.

26. Zhao, G.; Tian, X.; Wang, J.; Cheng, M.; Zhang, T.; Wang, Z. The Structure-Based Virtual Screening of Non-Benzofuran Inhibitors against: M. Tuberculosis Pks13-TE for Anti-Tuberculosis Phenotypic Discovery. *New Journal of Chemistry* **2021**, *45*, 1286–1300, doi:10.1039/D0NJ03828H.

27. Purves, K.; Macintyre, L.; Brennan, D.; Hreggviðsson, G.; Kuttner, E.; Ásgeirsdóttir, M.E.; Young, L.C.; Green, D.H.; Edrada-Ebel, R.; Duncan, K.R. Using Molecular Networking for Microbial Secondary Metabolite Bioprospecting. *Metabolites* **2016**, *6*, doi:10.3390/METABO6010002.

28. Xu, L.; Ye, K.X.; Dai, W.H.; Sun, C.; Xu, L.H.; Han, B.N. Comparative Genomic Insights into Secondary Metabolism Biosynthetic Gene Cluster Distributions of Marine Streptomyces. *Mar Drugs* **2019**, *17*, doi:10.3390/MD17090498.

29. Lecina, D.; Gilabert, J.F.; Guallar, V. Adaptive Simulations, towards Interactive Protein-Ligand Modeling. *Sci Rep* **2017**, *7*, doi:10.1038/s41598-017-08445-5.

30. Wang, Q.; Wang, L.; Zhang, Y.; Zhang, X.L.; Zhang, L.; Shang, W.; Bai, F. Probing the Allosteric Inhibition Mechanism of a Spike Protein Using Molecular Dynamics Simulations and Active Compound Identifications. *J Med Chem* **2022**, *65*, 2827–2835.

31. Yang, P.; Bokros, N.; Debolt, S.; Zhao, Z.; Xia, Y. Genome Sequence Source of Bacillus Amyloliquefaciens Strain GD4a, a Bacterial Endophyte Associated with Switchgrass Plants. *PBIOMES* **2022**, *6*, 354–357, doi:10.1094/PBIOMES-09-21-0054-A.

32. Nyambo, K.; Sithole Niang, I. Bioprospecting of Endophytes Isolated from Selected Zimbabwean Medicinal Plants. *World Journal of Environmental Biosciences* **2020**, *9*, 1–12.

33. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **2016**, *33*, 1870–1874, doi:10.1093/MOLBEV/MSW054.

34. Uche-Okereafor, N.; Sebola, T.; Tapfuma, K.; Mekuto, L.; Green, E.; Mavumengwana, V. Antibacterial Activities of Crude Secondary Metabolite Extracts from Pantoea Species Obtained from the Stem of Solanum Mauritianum and Their Effects on Two Cancer Cell Lines. *International Journal of Environmental Research and Public Health 2019, Vol. 16, Page 602* **2019**, *16*, 602, doi:10.3390/IJERPH16040602.

35. Tapfuma, K.I.; Nyambo, K.; Adu-Amankwaah, F.; Baatjies, L.; Smith, L.; Allie, N.; Keyster, M.; Loxton, A.G.; Ngxande, M.; Malgas-Enus, R.; et al. Antimycobacterial Activity and Molecular Docking of Methanolic Extracts and Compounds of Marine Fungi from Saldanha and False Bays, South Africa. *Heliyon* **2022**, *8*, e12406, doi:10.1016/J.HELIYON.2022.E12406.

36. Venugopala, K.N.; Tratrat, C.; Pillay, M.; Mahomoodally, F.M.; Bhandary, S.; Chopra, D.; Morsy, M.A.; Haroun, M.; Aldhubiab, B.E.; Attimarad, M.; et al. Anti-Tubercular Activity of Substituted 7-Methyl and 7-Formylindolizines and in Silico Study for Prospective Molecular Target Identification. *Antibiotics* **2019**, *8*, doi:10.3390/antibiotics8040247.

37. Abdelrahman, S.M.; Dosoky, N.S.; Hanora, A.M.; Lopanik, N.B. Metabolomic Profiling and Molecular Networking of Nudibranch-Associated Streptomyces Sp. SCSIO 001680. *Molecules* **2022**, *27*, 4542, doi:10.3390/MOLECULES27144542/S1.

38. Wang, M.; Carver, J.J.; Phelan, V. V; Sanchez, L.M.; Garg, N.; Peng, Y.; Duy Nguyen, D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and Community Curation of Mass

Spectrometry Data with Global Natural Products Social Molecular Networking. **2016**, doi:10.1038/nbt.3597.

39. da Silva Oliveira, J.P.; Garrett, R.; Bello Koblitz, M.G.; Furtado Macedo, A. Vanilla Flavor: Species from the Atlantic Forest as Natural Alternatives. *Food Chem* **2022**, *375*, 131891, doi:10.1016/J.FOODCHEM.2021.131891.

40. Guo, Y.S.; Tao, J.Z. Metabolomics and Pathway Analyses to Characterize Metabolic Alterations in Pregnant Dairy Cows on D 17 and D 45 after AI. *Sci. Rep. 2018 8:1* **2018**, *8*, 1–8, doi:10.1038/s41598-018-23983-2.

41. Zong, K.; Xu, L.; Hou, Y.; Zhang, Q.; Che, J.; Zhao, L.; Li, X.; Brezovsky, J.; Sousa, S.F. Molecules Virtual Screening and Molecular Dynamics Simulation Study of Influenza Polymerase PB2 Inhibitors. **2021**, doi:10.3390/molecules26226944.

42. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J Med Chem* **2006**, *49*, 6177–6196.

43. Pandey, A.K.; Shukla, D.V.; Singh, V.; Narayan, V. Structural, IR Spectra NBO, TDDFT, AIM Calculation, Biological Activity and Docking Property of [1,2,4]-Triazolo[3,4- b][1,3,4] Thiadiazole . *Egyptian J. Basic Appl. Sci.* **2018**, *5*, 280–288, doi:10.1016/j.ejbas.2018.10.001.

44. Gumede, N.J. Pathfinder-Driven Chemical Space Exploration and Multiparameter Optimization in Tandem with Glide/IFD and QSAR-Based Active Learning Approach to Prioritize Design Ideas for FEP+ Calculations of SARS-CoV-2 PLpro Inhibitors. *Molecules* **2022**, *27*, 8569, doi:10.3390/MOLECULES27238569/S1.

45. Wang, M.; Carver, J.J.; Phelan, V. V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat Biotechnol 2016 34:8* **2016**, *34*, 828–837, doi:10.1038/nbt.3597.

46. Fu, P.; Li, Z.; Feng, J.; Bian, Z. Recovery of Gold and Iron from Cyanide Tailings with a Combined Direct Reduction Roasting and Leaching Process. *Metals (Basel)* **2018**, *8*, doi:10.3390/met8070561.

47. Qian, L.; Lin, H.; Li, B.; Dong, Y. Physicochemical Characteristics and Microbial Communities of Rhizosphere in Complex Amendment-Assisted Soilless Revegetation of Gold Mine Tailings. *Chemosphere* **2023**, *320*, 138052, doi:10.1016/J.CHEMOSPHERE.2023.138052.

48. Qi, R.; Xue, N.; Wang, S.; Zhou, X.; Zhao, L.; Song, W.; Yang, Y. Heavy Metal(Loid)s Shape the Soil Bacterial Community and Functional Genes of Desert Grassland in a Gold Mining Area in the Semi-Arid Region. *Environ Res* **2022**, *214*, 113749, doi:10.1016/J.ENVRES.2022.113749.

49. Santini, T.C.; Raudsepp, M.; Hamilton, J.; Nunn, J. Extreme Geochemical Conditions and Dispersal Limitation Retard Primary Succession of Microbial Communities in Gold Tailings. *Front Microbiol* **2018**, *9*, 2785, doi:10.3389/FMICB.2018.02785/BIBTEX.

50. Gadd, G.M. Metals, Minerals and Microbes: Geomicrobiology and Bioremediation. *Microbiol (N Y)* 2010, *156*, 609–643.

51. Ji, H.; Zhang, Y.; Bararunyeretse, P.; Li, H. Characterization of Microbial Communities of Soils from Gold Mine Tailings and Identification of Mercury-Resistant Strain. *Ecotoxicol Environ Saf* **2018**, *165*, 182–193, doi:10.1016/J.ECOENV.2018.09.011.

52. Chen, L.X.; Hu, M.; Huang, L.N.; Hua, Z.S.; Kuang, J.L.; Li, S.J.; Shu, W.S. Comparative Metagenomic and Metatranscriptomic Analyses of Microbial Communities in Acid Mine Drainage. *The ISME Journal 2015 9:7* **2014**, *9*, 1579–1592, doi:10.1038/ismej.2014.245.

53. Robertsen, H.L.; Musiol-Kroll, E.M. Actinomycete-Derived Polyketides as a Source of Antibiotics and Lead Structures for the Development of New Antimicrobial Drugs. *Antibiotics* **2019**, *8*, doi:10.3390/antibiotics8040157.

54. Mast, Y.; Stegmann, E. Actinomycetes: The Antibiotics Producers. *Antibiotics* **2019**, *8*, 10–13, doi:10.3390/antibiotics8030105.

55. Mrunalini, B.R.; Girisha, S.T. Molecular Phylogenetic Analysis of Soil Bacteria. *Res J Chem Environ* **2017**, *21*, 35–41.

56. Li, H.; Wang, G.; Liu, J.; Cui, X.; Liu, Z.; Guo, Z.; Yu, Z.; Yao, Q.; Sui, Y.; Jin, J.; et al. The Diversity and Geographic Distribution of Cultivable Bacillus-Like Bacteria Across Black Soils of Northeast China. **2019**, *10*, 1424, doi:10.3389/fmicb.2019.01424.

57. Dunbar, J.; Ticknor, L.O.; Kuske, C.R.; Icrobiol, A.P.P.L.E.N.M. Assessment of Microbial Diversity in Four Southwestern United States Soils by 16S RRNA Gene Terminal Restriction Fragment Analysis. **2000**, *66*, 2943–2950.

58. Zhang, J.; Bernat, P.; Nitschke, M.; Sharma, D.; Singh, S.S.; Baindara, P.; Sharma, S.; Khatri, N.; Grover, V.; Patil, P.B.; et al. Surfactin Like Broad Spectrum Antimicrobial Lipopeptide Co-Produced With Sublancin From Bacillus Subtilis Strain A52: Dual Reservoir of Bioactives. **2020**, doi:10.3389/fmicb.2020.01167.

59. Xia, L.; Miao, Y.; Cao, A.; Liu, Y.; Liu, Z.; Sun, X.; Xue, Y.; Xu, Z.; Xun, W.; Shen, Q.; et al. Biosynthetic Gene Cluster Profiling Predicts the Positive Association between Antagonism and Phylogeny in Bacillus. *Nat Commun* **2022**, *13*, doi:10.1038/s41467-022-28668-z.

60. Zhu, J.; Li, L.; Wu, F.; Wu, Y.; Wang, Z.; Chen, X.; Li, J.; Cai, D.; Chen, S. Metabolic Engineering of Aspartic Acid Supply Modules for Enhanced Production of Bacitracin in Bacillus Licheniformis. *ACS Synth Biol* **2021**, *10*, 2243–2251, doi:10.1021/acssynbio.1c00154.

61. Martín-González, D.; Bordel, S.; Solis, S.; Gutierrez-Merino, J.; Santos-Beneit, F. Characterization of Bacillus Strains from Natural Honeybee Products with High Keratinolytic Activity and Antimicrobial Potential. *Microorganisms* **2023**, *11*, 456, doi:10.3390/MICROORGANISMS11020456/S1.

62. Quintero, M.; Blandón, L.M.; Vidal, O.M.; Guzman, J.D.; Gómez-Marín, J.E.; Patiño, A.D.; Molina, D.A.; Puerto-Castro, G.M.; Gómez-León, J. In Vitro Biological Activity of Extracts from Marine Bacteria Cultures against Toxoplasma Gondii and Mycobacterium Tuberculosis. *J Appl Microbiol* **2022**, *132*, 2705–2720, doi:10.1111/JAM.15397.

63. Zhang, H.; Yang, Q.; Zhao, J.; Chen, J.; Wang, S.; Ma, M.; Liu, H.; Zhang, Q.; Zhao, H.; Zhou, D.; et al. Metabolites from Bacillus Subtilis J-15 Affect Seedling Growth of Arabidopsis Thaliana and Cotton Plants. *Plants* **2022**, *11*, 3205, doi:10.3390/PLANTS11233205/S1.

64. Desai, N.C.; Maheta, A.S.; Rajpara, K.M.; Joshi, V. V.; Vaghani, H. V.; Satodiya, H.M. Green Synthesis of Novel Quinoline Based Imidazole Derivatives and Evaluation of Their Antimicrobial Activity. *J Saudi Chem Soc* **2014**, *18*, 963–971, doi:10.1016/J.JSCS.2011.11.021.

65. Fan, Y.L.; Jin, X.H.; Huang, Z.P.; Yu, H.F.; Zeng, Z.G.; Gao, T.; Feng, L.S. Recent Advances of Imidazole-Containing Derivatives as Anti-Tubercular Agents. *Eur J Med Chem* **2018**, *150*, 347–365, doi:10.1016/J.EJMECH.2018.03.016.

66. Dhameliya, T.M.; Patel, K.I.; Tiwari, R.; Vagolu, S.K.; Panda, D.; Sriram, D.; Chakraborti, A.K. Design, Synthesis, and Biological Evaluation of Benzo[d]Imidazole-2-Carboxamides as New Anti-TB Agents. *Bioorg Chem* **2021**, *107*, 104538, doi:10.1016/J.BIOORG.2020.104538.

Chapter 4

In-silico Screening of Fungi Secondary Metabolites against *Mycobacterium tuberculosis* Ser/Thr kinases

Kudakwashe Nyambo[1], Kudzanai Ian Tapfuma[1], Francis Adu-Amankwaah[1], Lauren Julius[1], Lucinda Baatjies[1], Krishna Govender [4], [5], Idah Sithole Niang[2], Liezel Smith [1], Mkhuseli Ngxande [3], and Vuyo Mavumengwana [1], *

[1] *DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town; KN: knyambo@sun.ac.za; KT: kudzanait@sun.ac.za; LJ: laurenjulius@sun.ac.za; LB: lbaatjies@sun.ac.za; LS: liezels@sun.ac.za; VM: vuyom@sun.ac.za*

[2]*Department of Biochemistry, University of Zimbabwe, B064, Mount Pleasant, Harare, Zimbabwe; sitholeidah2015@gmail.com*

[3]*Computer Science Division, Department of Mathematical Sciences, Faculty of Science University of Stellenbosch, Matieland, South Africa; ngxandem@sun.ac.za*

[4]*Department of Chemical Sciences, University of Johannesburg, P. O. Box 17011, Doornfontein Campus, 2028, Johannesburg, South Africa; krishnag@uj.ac.za*
[5]*National Institute for Theoretical and Computational Sciences (NITheCS), South Africa*

**Corresponding authors.*

Email address: vuyom@sun.ac.za; Tel: +27 718502949

4.1 Abstract

The consistent development of resistance to current tuberculosis (TB) drugs poses a significant threat to human health, necessitating the exploration of natural products as alternative sources for the discovery of antimycobacterial agents. Fungi from gold mine tailings are a promising reservoir of bioactive compounds that may possess antimycobacterial activity. In this study, a total of 15

compounds tentatively identified from fungi isolated from gold mine tailings were virtually screened against *Mycobacterium tuberculosis* PknA, PknB, PknD, and PknE proteins using XP molecular docking and pre-molecular dynamics simulation MM-GBSA using Schrodinger. To evaluate protein-ligand interactions, molecular dynamics simulations were performed for the PknB-aurovertin D, PknA-aurovertin D, PknD-verticillin A, and PknE-roquefortine C complexes. Subsequently, the binding free energies of the complexes were computed. The compounds were ranked based on docking scores, which ranged from -3.7 kcal/mol to -7.4 kcal/mol. Notably, aurovertin D (-7.2 kcal/mol), aurovertin D (-6.7 kcal/mol), verticillin A (-6.5 kcal/mol), and roquefortine C (-6.7 kcal/mol) displayed XP docking scores for PknB, PknA, PknD, and PknE, respectively. Furthermore, the compounds with the lowest ΔGBind values were aurovertin D against PknA (-50.9 kcal/mol), aurovertin D against PknB (-50.7 kcal/mol), verticillin A against PknD (-36.8 kcal/mol), and roquefortine C against PknE (-53.4 kcal/mol). Notably, PknD-verticillin A exhibited the lowest binding free energy (-53.67 kcal/mol), followed by roquefortine C against PknE (-53.37 kcal/mol), aurovertin D against PknA (-45.34 kcal/mol), and aurovertin D against PknB (-42.87 kcal/mol). The results suggest that verticillin A is a potential lead compound for further modification, to develop a potent antimycobacterial inhibitor. These findings pave the way for future target-based TB drug design and highlight the importance of natural product-derived compounds as valuable resources for novel anti-TB agent development.

Keywords: Tuberculosis; antimicrobial resistance; molecular docking; molecular dynamics, MM-GBSA, binding free energy, secondary metabolites.

4.2 Introduction

Mycobacterium tuberculosis (*M. tb*) is a pathogenic bacterium responsible for causing the infectious disease tuberculosis (TB) [1, 2]. Over the past two decades, TB control regimen cocktails have consisted of combinations of four first-line drugs, including rifampicin, isoniazid, ethambutol, and pyrazinamide, to treat the disease. The effectiveness of these first-line drugs is, however, thwarted by the emergence of drug-resistant, multidrug-resistant, and extensively-drug resistant TB (DR-, MDR- and XDR-TB), which poses a significant challenge to the "End TB Strategy" because the cases are steadily rising [3, 4]. Genetic studies have illustrated that multidrug resistance in *M. tb* is principally due to point mutations, deletions, and insertions of drug-target genes. In addition, the selection of resistant mutants from patients with treatment failure, the rise in coinfection with HIV (Human Immunodeficiency Virus) infections, and the inefficiency of the laboratories to rapidly identify and perform susceptibility testing of *M. tb* isolates also accelerates the spread of resistant mutants. Therefore, new, and effective TB drugs must be discovered to control and end the TB pandemic [5–9].

Numerous drugs exert their life-saving effects by interacting with druggable macromolecular targets that play an essential role in the cells, of microorganisms, for example, nucleic acids and proteins. When an essential protein is inhibited by potent bioactive molecules, the downstream processes that are critical for the survival and virulence of the microbe are rapidly and negatively affected [10–12]. Fundamentally, for a viable *M. tb* cell to survive and maintain the genome integrity in a host, it constantly responds and adapts to the stressful oxidative environment. The powerful regulatory and adaptive mechanisms in *M. tb* are triggered by the post-translational modification of essential proteins, particularly the Serine/Threonine protein kinases (STPKs) [13,

14]. These STPKs function in a phosphorylation-dependent signal transduction manner by converting extracellular stimuli into a cellular response of multiple metabolic processes, for instance, transcription regulation, cell division, stress response, regulation of numerous metabolic pathways, and pathogenesis.

When *M. tb* invades the alveoli, it is phagocytosed by macrophages as an innate defense mechanism [15]. However, the bacilli efficiently evade the host's innate and adaptive immunity by secreting PknG into the host cytosol and phagosome lumen. PknG is an STPK that inhibits the exchange and recruitment of Rab7 endosomal markers, consequently preventing phagosome maturation and lysosome fusion. Some studies on two *M. tb* STPKs (PknA and PknB) have illustrated their roles in modulating cell division, morphogenesis, and virulence [14, 16]. *M. tb* responds to heat and oxidative stress via PknB phosphorylating RshA and SigH. Zeng et al. [17] revealed that the inhibition of PknB modulates SigH, thus, negatively impacting the bacterial transcription network. Furthermore, PknD has been associated with the regulation of osmotic stress via transcription. Reports of gene knockouts and inhibition of PknA further validated that depletion of the functions of PknA consequently results in decreased phosphorylation of *Wag31*, *FtsZ*, and *MurD*, which regulate cell division, morphology, and peptidoglycan biosynthesis [18, 19]. PknE has been implicated in inhibiting apoptosis in infected macrophages. In this regard, *M. tb* STPKs are undoubtedly valuable potential anti-TB targets [20].

Natural products have been a valuable source of mining major potent antimicrobials and chemotherapeutics. Microorganisms, particularly fungi, actinomycetes, and myxobacteria have been reported to be the principal sources of a wide array of bioactive secondary metabolites [21–23]. In this study, it is assumed that fungi inhabiting gold mining tailings have a unique metabolic

potential to synthesis lethal bio-arsenal which it uses in competing with other microorganisms residing in the gold mine tailings. The metabolite profile and anticancer activity of small molecules obtained from fungi isolated from old gold mine tailings collected from Johannesburg, South Africa (26º13'7.08" S, 28º29'8.64"), have been previously evaluated [24]. However, this fungus' small molecules have not been assessed for bioactivity against *M. tb'* essential proteins. In this study, the small molecules library produced by *Penicillum janthinellum* KTMT5, *Penicillium oxalicum* KTMT4, and *Acidiella americana* KTMT6 was therefore virtually screened against *M. tb* druggable proteins to discover potential scaffolds that serve as molecular starting points for further optimization. The binding interactions were evaluated based on docking the compounds against potential drug targets (*M. tb* PknA, PknE, PknB, and PknD). The dynamic behavior of the top scoring ligands conformation for each respective protein was further validated by molecular dynamics (MD) simulations and free energy calculation.

4.3 Materials and methods

4.3.1 Data collection

Tapfuma et al. [24], isolated, identified, fermented, and extracted the potato dextrose broth cultures of *P. janthinellum* KTMT5, *Penicillium oxalicum* KTMT4, and *Acidiella americana* KTMT6 using ethyl acetate. The crude extract was then analyzed using tandem liquid chromatography-quadrupole time of flight mass spectrometry (LC-QTOF-MS/MS), and secondary metabolites were identified as described by Tapfuma et al. [24]. The compounds depicted in Figure 4.1 were used for virtual screening.

4.3.2 Molecular docking

The *M. tb* H37Rv proteins' crystal structures available in the Protein Data Bank corresponding to PDB ID 6B2Q, 6B2P, 1RWL, and 2H34, for proteins PknA, PknB, PknD, and PknE, respectively, were retrieved for molecular docking. The protein preparation was performed as described by Baptista et al. [25] in Schrödinger Release 2021-1. Briefly, hydrogen atoms were added, hydrogen-bond assignments were optimized, the loop was refined, and the OPLS-4 (Optimized Potentials for Liquid Simulations 4) (Optimized Potentials for Liquid Simulations) force field was used for energy minimization. The structural coordinates of the co-crystalized ligand (CJJ) with 6B2Q and 6B2P were used for generating the binding site. For 1RWL, and 2H34, the Site Map module was used to predict the binding sites, which were then used for generating respective receptor grids using the Receptor Grid Generating module (Schrödinger Release 2021-1). A control compound 5-(6-chloro-4-((5-cyclopropyl-1H-pyrazol-3-yl) amino)quinazolin-2-yl) thiophene-2-sulfonamide, with dual inhibitory activity against PknA and PknB as described by [18] was downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/compound/134815875). The selected fungi compounds were also downloaded from PubChem. The compounds were prepared for docking using the LigPrep module (Schrödinger Release 2021-1) according to the following parameters: the energy was minimized by an OPLS4 force field, generate ionization states at pH 7.0 + 2.0, and generate multiple conformers to develop a library of compounds. The library created was subjected to extra-precision (XP) molecular docking calculations against the selected target proteins using Glide [26]. Thereafter, pre-MD MM-GBSA free energy calculations were computed. The compounds were then ranked according to a numerical score in kcal/mol. The ligands with the highest MM-GBSA score were later subjected to MD simulations.

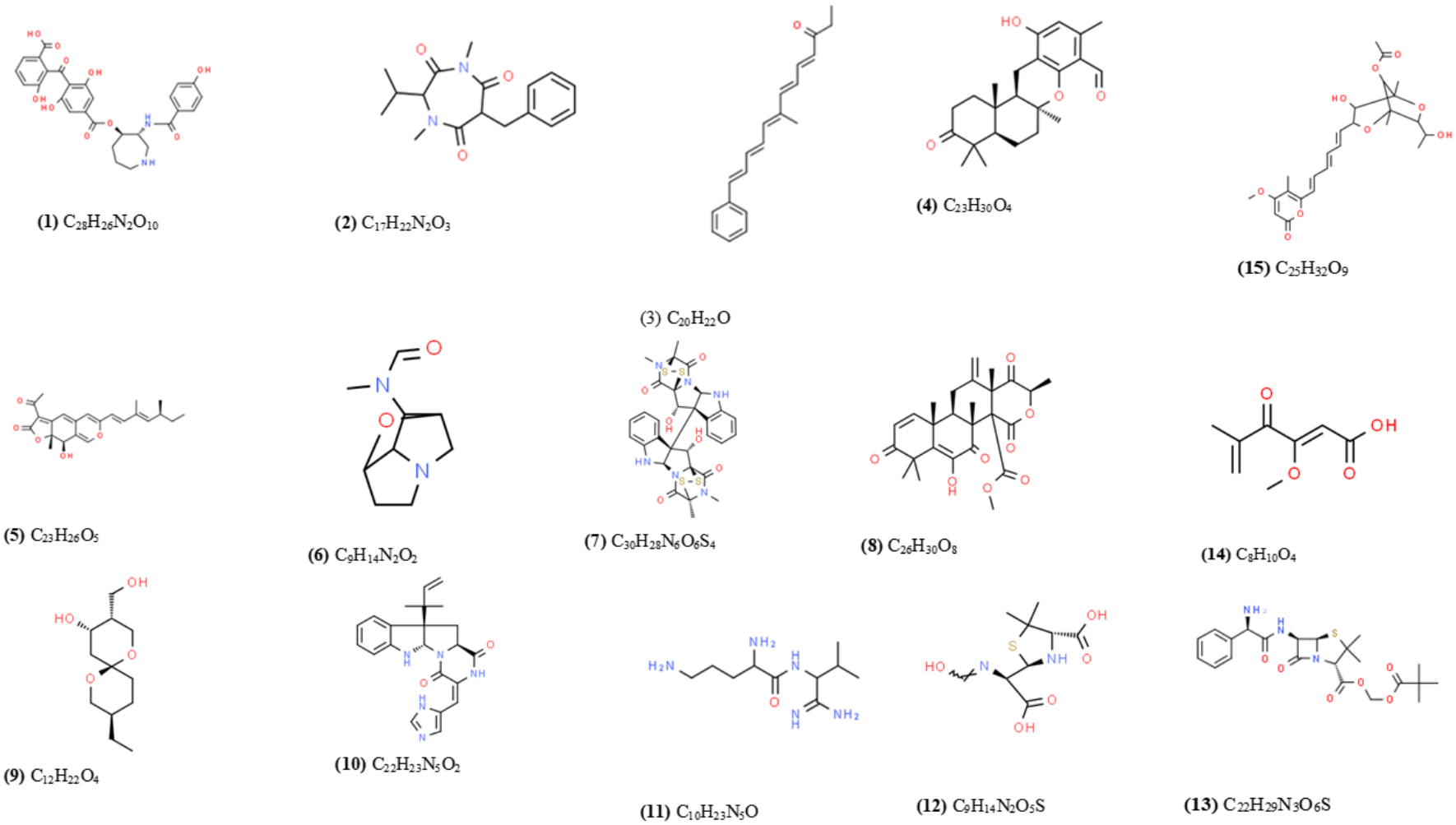


Figure 4.1. Fungi compounds that were docked against M. *tuberculosis* PknA, PknB, PknD and PknE**.**

4.3.3 Molecular Dynamics Simulation

The selected docked protein-ligand complexes with the highest docking scores were subjected to 200 nanoseconds (ns) MD simulations to further explore protein and ligand interactions, and stability using Desmond (Schrödinger Release 2021-1). A total of four MD systems were set up using Maestro (Schrödinger Release 2021-1) according to the following parameters: For each protein-ligand complex, an MD system was created by explicitly solvating the complex using TIP3P hydration model in an orthorhombic box with a buffer boundary dimension of (10 Å × 10 Å × 10 Å). Counter ions (Na+ and Cl-) 0.15 M were added to neutralize the system and to precisely predict the physical properties of a realistic system. For long electrostatic forces, periodic grid conditions were automatically generated for Particle-mesh Ewald FFT. The entire system was energy minimized and equilibrated at constant pressure (1. 01325 bar) and temperature (303.15 K). The MD simulations were performed in the NPT ensemble. The Nose-Hoover thermostat was used with a 1.0 ps interval and Martyna- Tobias-Klein as the default barostat with a 2.0 ps interval by applying an isotropic coupling style. The systems were subjected to MD simulations for 200 ns and the internal energy was stored for every 1000 ps of the actual frame. The structural changes and dynamic behavior of the protein-ligand complexes were calculated by the Simulation Interaction Diagram module in Maestro (Schrodinger) and represented as the root mean fluctuation (RMSF), and the root means square deviation (RMSD).

4.3.4 Post-molecular Dynamics Simulation Analysis

The post-molecular dynamic simulation analysis was carried out by calculating the protein-ligand binding free energies based on MD simulation trajectories. The free energy change (ΔG_{Bind}) for

the interaction between the receptor and the ligand to form a complex is described by Molecular mechanics/generalized Molecular Mechanics Generalized Born Surface Area MM/GBSA as the summation of different interactions according to the equation below:

$$\Delta G_{Bind} = E_{Complex} - [E_{Receptor} + E_{Ligand}]$$

Where ΔG_{Bind} is the calculated relative free energy which includes both ligand and receptor strain energy. The $E_{Complex}$ is the MM-GBSA energy of the minimized complex, and E_{Ligand} is the MM-GBSA energy of the ligand after removing it from the complex and allowing it to relax. $E_{Receptor}$ is the MM-GBSA energy of relaxed protein after separating it from the ligand. The MM-GBSA calculation was performed based on the clustering method for energy calculation.

4.4 Results

The molecular docking-based screening was performed to preliminarily gain an insight into the molecular interactions between the *M. tb* Ser/Thr protein kinases (PknA, PknE, PknB, and PknD) and the fungi metabolites as represented in Table 4.1. The Site-score and the DSCORE were used to rank the druggable regions of PknD, and PknE (Table B1). All the site scores for the proteins were above 0.9, which implies that the docking cavity regions in this study are fit enough to be used as docking regions (Table B1). The study revealed that all the investigated compounds could bind to the active sites of the PknA, PknE, PknB, and PknD and were ranked by docking scores (Table 4.1). The XP docking scores of all the compounds vary from -3.7 kcal/mol to -7.4 kcal/mol while the ΔG_{Bind} varies from 9.8 to -53.4 kcal/mol. The control ligand, 5-(6-chloro-4-((5-cyclopropyl-1H-pyrazol-3-yl) amino) quinazolin-2-yl) thiophene-2-sulfonamide (CJJ) used in the study had a docking score of –11.3 kcal/mol, -11.1 kcal/mol, -5.8 kcal/mol, and -4.3 kcal/mol for

the proteins PknA, PknB, PknD, and PknE respectively. Aurovertin D obtained a docking score of -7.2 kcal/mol and -6.7 kcal/mol against PknA and PknB respectively, while verticillin A had a docking score of -6.5 kcal/mol against PknD, and roquefortine C exhibited -6.7 kcal/mol against PknE. Noteworthy, the compounds that exhibited the lowest ΔG_{Bind} on each protein are aurovertin D (-50.9 kcal/mol) against PknA, aurovertin D (-50.7 kcal/mol) against PknB, Verticillin A (-36.8 kcal/mol) against PknD and roquefortine C (-53.4 kcal/mol) against PknE.

Table 4.1. XP docking of compounds against 5 Ser/Thr kinases (PknA, PknE, PknB, and PknD).

No.	Tentative Identification	Docking Scores (kcal/mol)							
		PknA 6B2Q		PknB (6B2P)		PknD (1rwl)		PknE (2H34)	
		XP	ΔG_{Bind}	XP	ΔG_{Bind}	XP	ΔG_{Bind}	XP	ΔG_{Bind}
1	Balanol	-6.4	-38.9	-4.0	-29.7	-6.7	-23.8	-4.5	-26.6
2	Terretrione B	-4.3	-28.8	-5.7	-32.2	-5.0	-24.4	-4.7	-36.3
3	8-methyl-13-phenyltrideca-4,6,8,10,12-pentaen-3-one	-6.7	-44.6	-6.9	-49.7	-2.7	-23.3	-4.1	-43.5
4	Phomoarcherin C	-4.9	-34.5	-5.6	-34.8	-4.1	-20.5	-4.5	-22.9
5	Rotiorinol A	-4.1	-49.0	-3.6	-43.8	-0.9	-10.4	-2.9	-33.7
6	N-Formylloline	-2.3	-24.4	-4.0	-22.4	-5.2	-21.4	-4.9	-32.1
7	Verticillin A	-7.0	-48.3	-3.7	-22.6	-6.5	-36.8	-6.5	-36.9
8	Terretonin F	-4.3	-28.8	-5.9	-10.3	-3.1	-18.3	-7.2	-37.5
9	Talaromycin A	-5.1	-19.5	-6.4	-37.5	-6.5	-28.0	-6.1	-35.7
10	Roquefortine C	-4.6	-23.6	-6.3	-32.4	-3.7	-13.6	-6.7	-53.4
11	2,5-diamino-N-(1-amino-1-imino-3-methylbutan-2-yl)pentanamide	-3.8	-10.4	-5.9	-29.7	-6.1	-33.2	-6.0	-31.2

Table 4.1. continued. XP docking of compounds against 5 Ser/Thr kinases (PknA, PknE, PknB, and PknD).

No.	Tentative Identification	Docking Scores (kcal/mol)							
		PknA (6B2Q)		PknB (6B2P)		PknD (1rwl)		PknE (2H34)	
		XP	ΔG_{Bind}	XP	ΔG_{Bind}	XP	ΔG_{Bind}	XP	ΔG_{Bind}
12	Penicilloic acid	-2.4	-2.5	-4.9	-10.6	-4.2	3.8	-3.5	4.0
13	Pivampicillin	-6.1	-45.6	-5.4	-39.6	-5.1	-33.27	-6.6	-37.9
14	Penicillic acid	-5.2	-6.27	-4.8	-20.4	-4.3	-15.8	-2.3	-14.8
15	Aurovertin D	-7.2	-50.9	-6.7	-50.7	-3.4	-23.8	-6.9	-49.0
16	Benzylpenicilloic acid	-6.0	-13.8	-6.0	9.8	-4.4	-18.4	-5.2	-16.6
	Control (CJJ)	-11.3	-61.9	-11.1	-59.1	-5.8	-35.3	-4.3	-41.5

The two-dimensional binding poses for the ligands that exhibited the lowest ΔG_{Bind} MM-GBSA scores (Table 4.1) were used to visually inspect the specific nature of the intermolecular interactions as depicted in Figure 4.2 and Figure 4.3. Aurovertin D formed hydrogen bonds with several residues of PknA, namely Arg17, Lys42, Arg38, Asn99, and Lys143. Additionally, hydrophobic interactions with Leu18, Leu19, Ala20, Val98, Leu97, Pro102, and Gly145. Furthermore, polar interactions were identified with Asn99, Asn146, and Thr158. Negative charges were found to interact with Asp159, Glu29, and Glu101, while positive charges interacted with Arg17, Arg38, Kys42, and Lys143. Glycine residues (Gly22, Gly23, Gly100, and Gly145) also contributed to the binding, as depicted in Figure 4.2.A. Aurovertin D also interacted with residues of PknB through hydrophobic interactions with Leu17, Phe19, Val25, Ala38, Tyr94, Val95, Val98, Ala142, Met145, and Ile177. Hydrogen bonds were formed with Val95, Lys140, and Asn143. Positive charges interacted with Asp156, Asp138, and Asp156, while polar interactions were observed with Ser23, Thr99, Asn143, and Thr179 as illustrated in Figure 4.2.B.

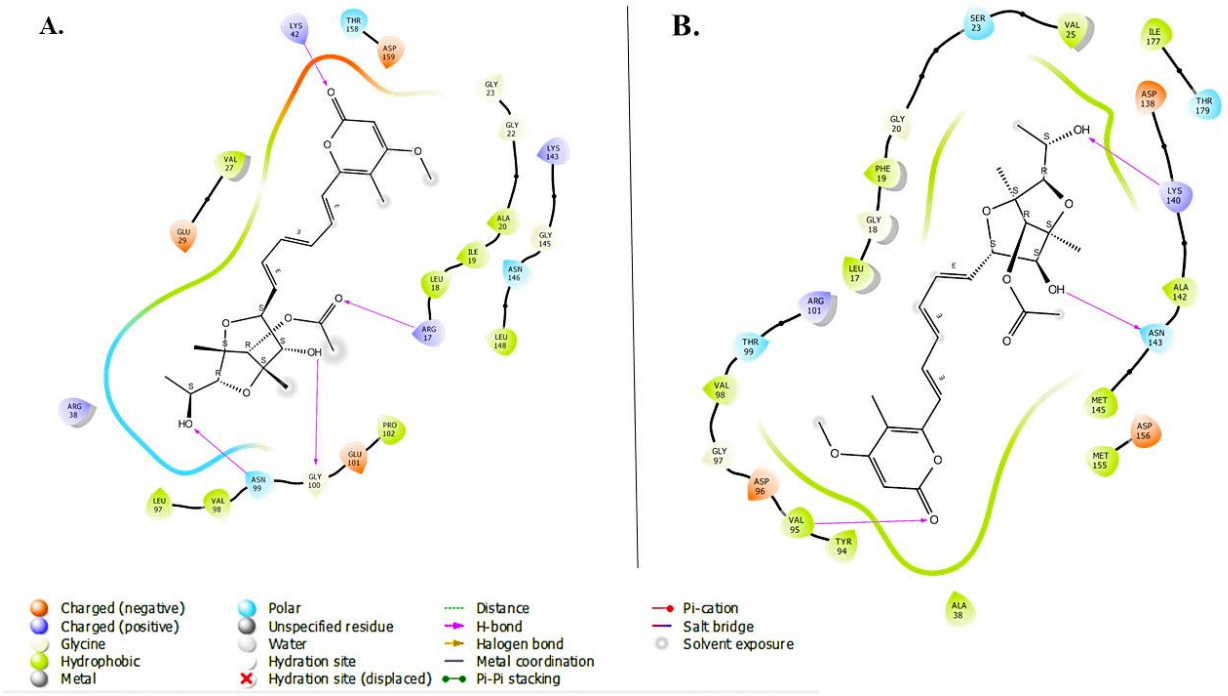


Figure 4.2. Two-dimensional depictions of the compounds with the lowest pre-MD ΔG_{Bind} MM-GBSA. A. PknA and Aurovertin D complex. B. PknB and Aurovertin D complex.

Verticillin A engaged the active domain of PknD through diverse types of interactions which include hydrogen bonds that were formed with Ser33, Gly76, and Glu201. Hydrophobic interactions were observed with Val31, Val74, Val115, and Val157. Positive charges were involved in interactions with Asp32, Asp75, Asp116, Asp158, and Glu201. Additionally, polar interactions were identified with Ser33, Thr117, Asn159, and Ser243. These intermolecular interactions are illustrated in Figure 4.3.A. Conversely, roquefortine C formed hydrogen bonds with PknE residues namely, Lys168, and Thr170 as shown in Figure 4.3.B. Furthermore, intermolecular interactions depicted in Figure 4.3.B include hydrophobic interactions through Leu21, Val22, Ala161, and Leu172, positive charges with Arg24, and Lys168, polar Ser162, Thr164, Thr165, and Thr170, negative charge interactions with Asp166 and Glu167 and glycine

Gly23. It is essential to fully understand molecular recognition of the ligands by PknA, PknB, PknD, and PknE. Thus, ligands that exhibited the lowest ΔG_{Bind} MM-GBSA for the PknA, PknB, PknD, and PknE (Table 4.1) were further subjected to molecular dynamics simulations to evaluate the structural dynamics and stability of the protein-ligand complexes during the 200 ns simulation.

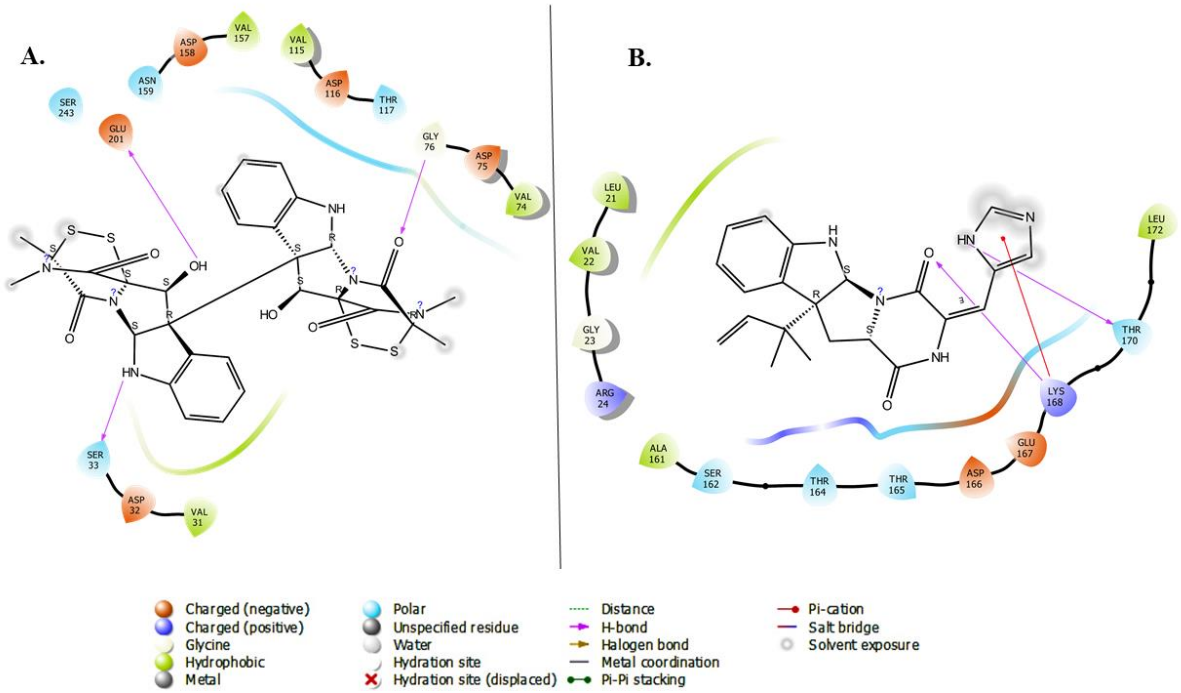


Figure 4.3. Two-dimension representation of A. Docked Verticillin A in binding pocket PknD B. Docked view of roquefortine C in the binding pocket of PknE.

The structural stability and fluctuations of the protein-ligand complexes and the unbound native proteins in this study were evaluated by performing MD simulations and measured by root-mean-square-deviation (RMSD) and root-mean-square-fluctuation (RMSF). The RMSD and RMSF per residue of the target proteins Cα-atoms - ligand atoms are depicted in Figure 4.4, Figure 4.6, Figure 4.8, and Figure 4.10. The RMSD represents the displacements of atoms at a given time during the simulation as compared to the reference structures obtained at the initiation of the simulation. During the 200 ns MD simulations, the RMSD of the unbound PknA Cα-atoms exhibited a gradual

increase from 1.2 to 2.4 Å within the first 30 ns, after which it stabilized (Figure A.1.A). For the unbound PknB, the RMSD trajectory fluctuated between 2.0 and 3.5 Å during the initial 75 ns, followed by a gradual reduction and stabilization around 2.8 Å (Figure A.1.C). The RMSD fluctuations of unbound PknD Cα-atoms remained relatively stable below 2.8 Å for the first 120 ns, followed by slight conformational changes with deviations ranging between 2.8 and 3.2 Å from 121 to 175 ns. Subsequently, the RMSD gradually decreased to 1.6 Å until the end of the simulation (Figure A.2.A). As for unbound PknE, its RMSD trajectory showed deviations ranging between 2.5 and 3.5 Å throughout the 200 ns simulation duration (Figure A.2.C)

The RMSD trajectory of PknA protein Cα-atoms in complex with aurovertin D exhibited a gradual increase from 1.5 Å to approximately 2.3 Å within the initial 50 ns as depicted in Figure 4.4.A. Subsequently, the RMSD stabilized at 2.3 Å indicating relatively stable until the end of the 200 ns simulation. However, the RMSD of aurovertin D showed a large fluctuation between 55 to 70 ns, suggesting that the ligand underwent conformational change within the designed binding site, thus, rendering it not stable (Figure 4.4.A). In contrast, the RSMD trajectory of PknA protein Cα-atoms in complex with control CJJ showed a relatively stable trajectory throughout the simulation, with minor confirmation changes occurring between 100 ns to 125 ns, where the RMSD increased from 2.0 Å to 2.8 Å. Subsequently, the RMSD decreased and stabilized at approximately 2.0 Å up to 200 ns (Figure 4.4.A). Further, RMSF evaluated the flexibility of the PknA residues in the two complexes. The RMSF profiles of PknA showed that the protein residues were stable and with minimal fluctuations above 2.5 Å (Figure 4.4.B and Figure 4.4.D). Large fluctuations were only observed on the mobile loop at residue Gly175.

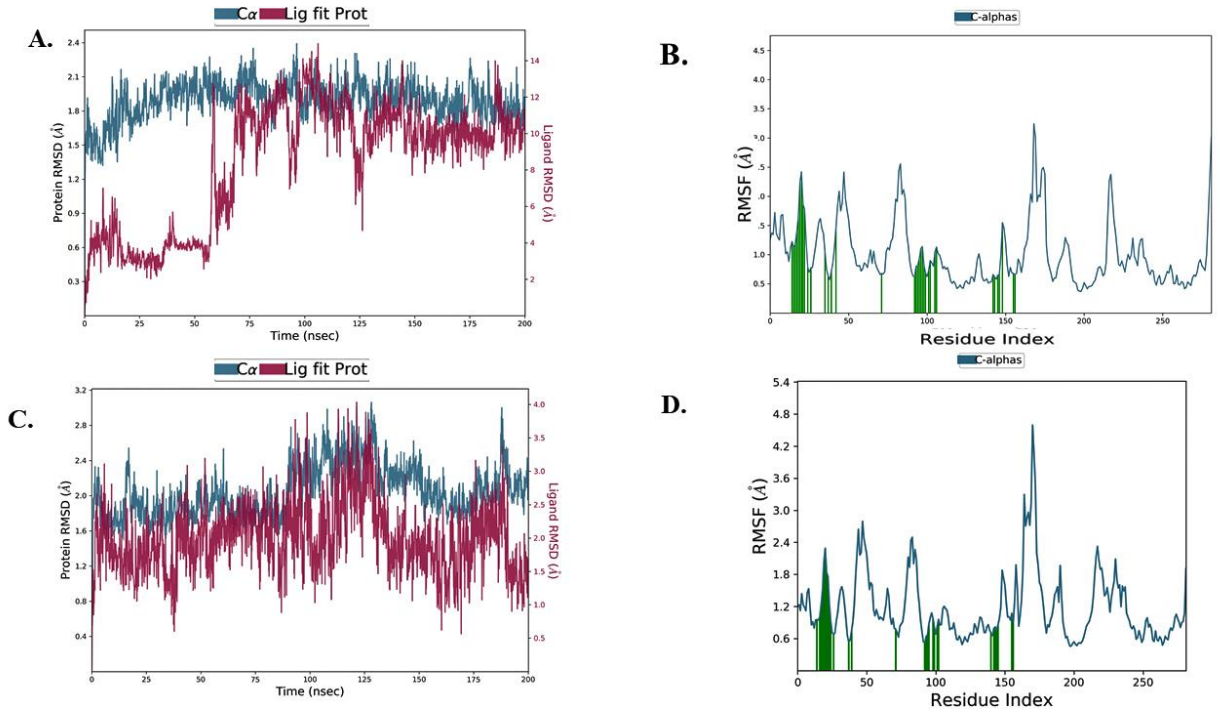


Figure 4.4. (**A**) Root-mean-square deviation (RMSD) of PknA Cα-atoms and Aurovertin D as a function of simulation time (200ns) (**B**) Root-mean-square fluctuation (RMSF) per residue of PknA in complex with Aurovertin D (**C**) RMSD of PknA Cα-atoms and control ligand (CJJ) as a function of simulation time (200ns) (**D**) RMSF per residue of PknA in complex with the control. Green lines show the residues in contact with the ligand.

Visual representations were used to further understand the non-covalent molecular interactions between the PknA-aurovertin D complex. Figure 4.5.A illustrates the contributions of hydrogen bonds, hydrophobic interactions, ionic interactions, and water bridges in stabilizing the PknA-aurovertin D complex. Noteworthy, only the hydrophobic interaction with Pro102 lasted for more than 30 % of the 200 ns simulation duration (Figure 4.5.B), explaining the instability observed in (Figure 4.4.A).

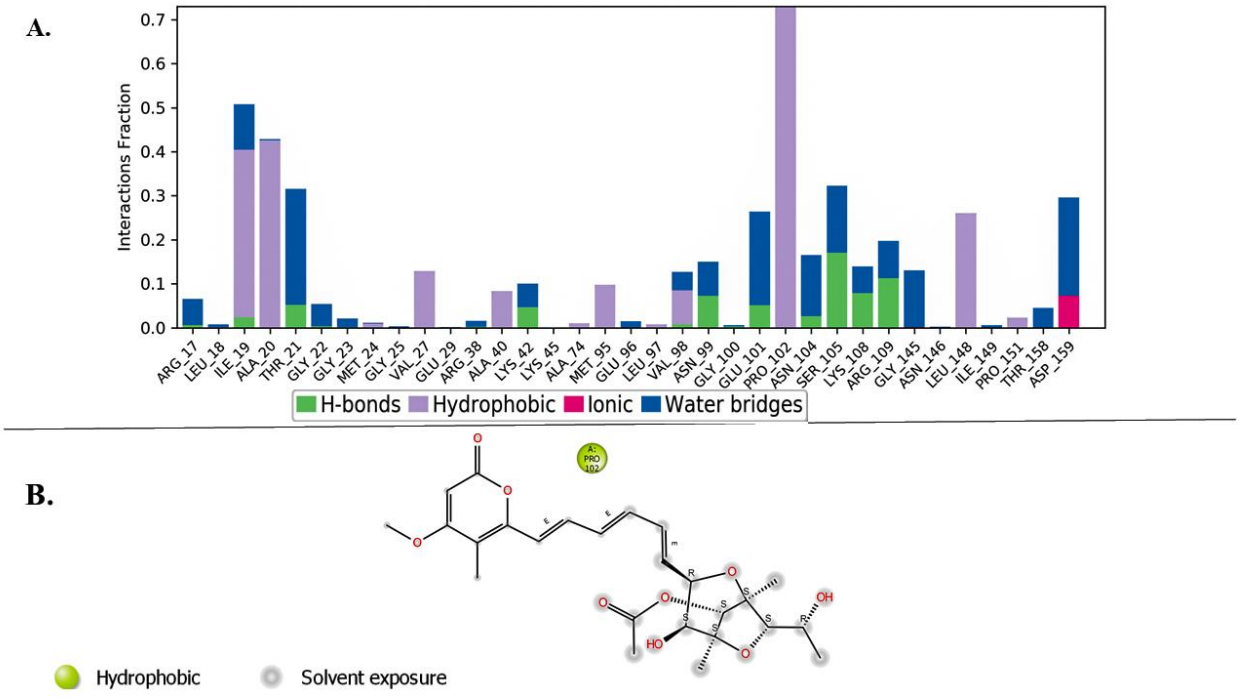


Figure 4.5. An overview of protein-ligand contacts over a 200 ns molecular dynamics simulation. **A**. Interaction types and fraction between PknA and Aurovertin D. **B.** Interaction types between PknA and that occurred for more than 30 % of the simulation duration.

The RSMD trajectory of PknB Cα-atoms complexed to aurovertin D displayed a sharp increase from 1.5 Å to 3.0 Å within the initial 25 ns. Subsequently, the RMSD stabilized around 3.0 Å from 25 to 100 ns. Notably, a large fluctuation was observed between 100 and 115 ns, followed by stabilization from 115 to 175 ns, during which the RMSD remained approximately 3.0 Å. At 175 ns, a reduction to 2.5 Å was observed, followed by an increase to 3.0 Å from 178 to 200 ns. The large fluctuations of PknB indicate conformation changes during the simulation period (Figure 4.6.A). Similarly, the RMSD of the aurovertin D revealed its structural instability, with fluctuations above 3.0 Å (Figure 4.6.A). A large fluctuation occurred at 100 ns, further indicating aurovertin D's instability in the designated binding site. The RMSD profile of the control showed large fluctuations between 25 and 50 ns indicating conformations changes during this period

(Figure 4.6.C). Comparing the RMSD profiles of the control (CJJ) and aurovertin D complexes, showed that the control maintained relatively stable throughout most of the simulation with an RSMD range of 2.4 Å. The RMSF shows that the protein Cα-atoms were relatively stable for most parts of the simulation period (Figure 4.6.B and Figure 4.6.D). However, the protein residue Val170 and Pro83 exhibited consistent large fluctuation in both complexes, representing a highly mobile loop region conformational flexibility (Figure 4.6.B and Figure 4.6.D).

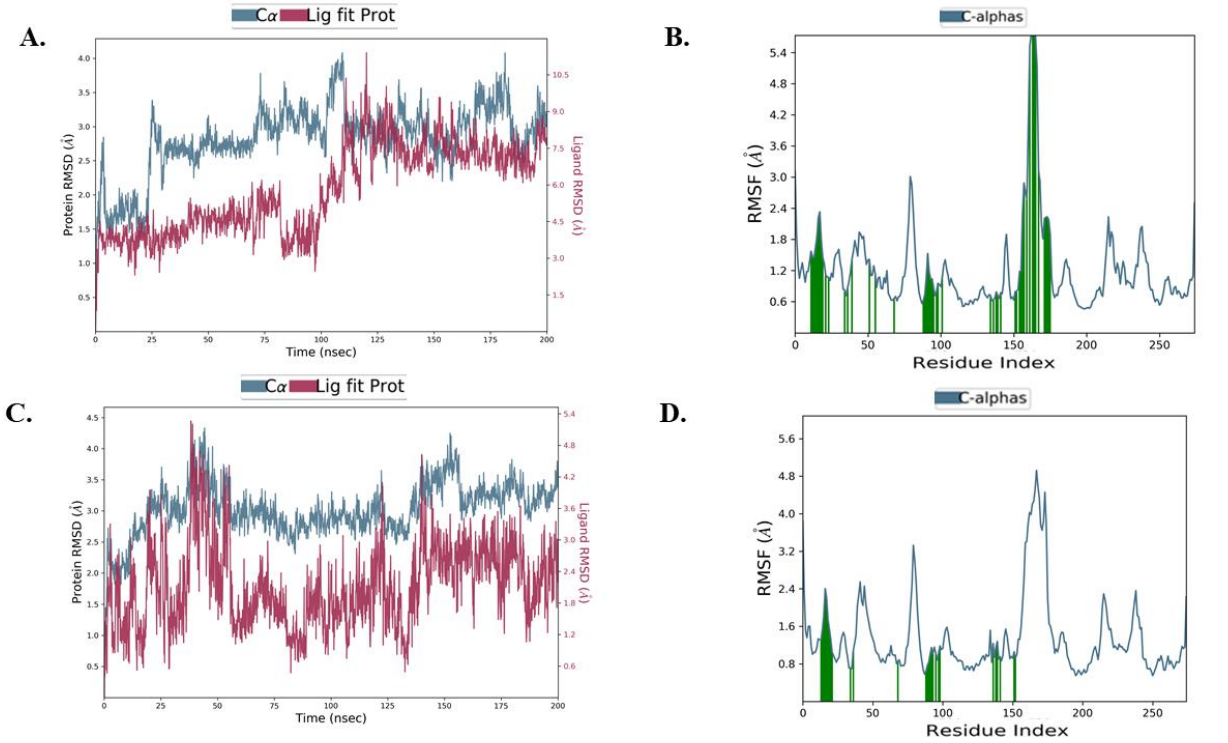


Figure 4.6. (**A**) RMSD PknB Cα-atoms and aurovertin D as a function of simulation time (200ns) (**B**) RMSF per residue of PknB in complex with aurovertin D (**C**) RSMD of PknB Cα-atoms and control ligand (CJJ) as a function of simulation time (200ns) (**D**) RMSF per residue of PknB in complex with the control. The green lines indicate the residues in contact with the ligand.

The bonds that contributed to the interaction of the PknB-aurovertin D are hydrogen bonds, hydrophobic interactions, ionic interactions, and water bridges (Figure 4.6.B). Aurovertin D engaged PknB via hydrogen bonds with Leu17, Phe19, Gly21, Ser23, Lys40, Val95, Asp96,

Gly97, Gly18, Gly21, Met22, Ser23, Lys40, Val95, Asp96, Gly97, Arg101, Lys140, Ala142, Asp156, Arg161, and Asn168. Hydrophobic contacts were observed with Leu17, Val25, Leu27, Ala38, Val72, Met92, Thr94, Ala142, Met145, Met155, Ile159, and Ile177. The formation of water bridges involved PknB residues Leu17, Phe19, Gly20, Gly21, Met22, Ser 23, Lys40, Arg43, Glu59, Glu93, Val95, Asp96, Gly97, Thr99, Arg101, Asp138, Lys140, Ala142, Asn143, Asp156, Ala160, Arg161, Ile177, and Gly178. Further, an ionic interaction occurred between aurovertin D and Asp96. Asp96 formed hydrogen bonds for approximately 39 % of the 200 ns simulation, indicating its significant contribution in stabilizing the PknB-aurovertin D complex (Figure 4.7.A). Notably, intramolecular hydrogen bond and hydrogen bonds mediated by water bridges (41 %) were also observed (Figure 4.7.A).

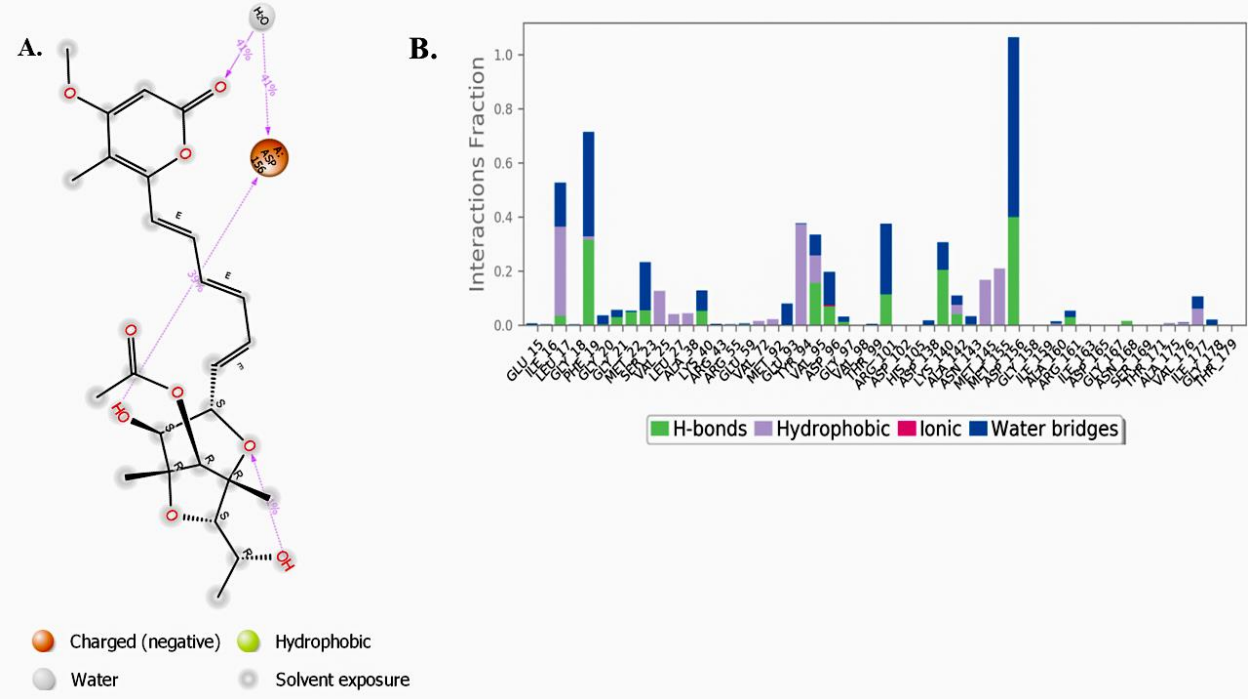


Figure 4.7. Protein-ligand contacts over a 200 ns molecular dynamics simulation. **A**. shows the interaction types between PknA that occurred for more than 30 % of the simulation duration. **B**. The types of interaction and fractions which occurred between Aurovertin D and PknB.

The RMSD trajectory of the PknD-verticillin A complex indicated that the complex was stable with RMSD values consistently below 2.1 Å throughout the 200 ns simulation period as shown in Figure 4.8.A. The RMSD of verticillin A exhibited minimal deviation maintaining a stable structural state at approximately 1.6 Å from 0 to 160 ns. A slight fluctuation occurred between 160-178 ns, suggesting a minor conformational change, and then followed by stabilization at 1.6 Å from 178 to 200 ns (Figure 4.8.A). The RMSF of PknD bound to verticillin A shows that the protein residues were stable throughout the 200 ns simulation as depicted in Figure 4.7.B. In contrast, the RMSD of the PknD Cα-atoms complex with CJJ (control ligand) showed distinct dynamics. Initially, the RMSD trajectory was stable at approximately 2.0 Å from 0 to 78 ns. Thereafter, a large fluctuation (2.0 to 3.6 Å) occurred between 80 and 100 ns, followed by a gradual increase to 2.8 Å from 100 to 200 ns. In contrast, the control ligand (CJJ) exhibited a large fluctuation at 50 ns, resulting in the RMSD increasing from 2 Å to approximately 6 Å (Figure 4.8.C). The RMSD revealed that the control ligand was not stable within the binding site of PknD as summarized in Figure 4.8.C.

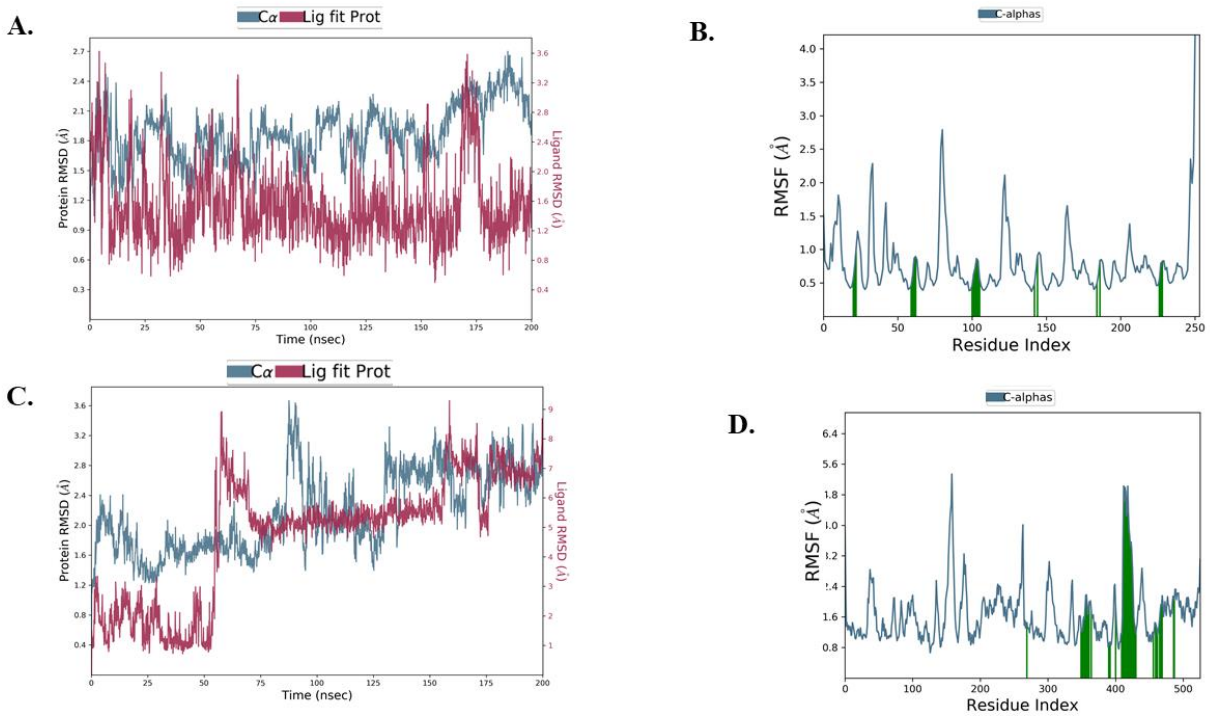


Figure 4.8.A. RMSD of PknD Cα-atoms and verticillin A over a 200 ns simulation. B. RSMF per residue of PknD in complex with verticillin A. C. RSMD of PknD Cα-atoms and control ligand as a function of 200 ns simulation time. D. The RSMF per residue of PknD in complex with control ligand. The green lines indicate the residues in contact with the ligand**.**

The stability of the PknD-verticillin A complex was due to a combination of hydrogen bonds, hydrophobic interactions, ionic interactions, and water bridges. Notably, verticillin A interacted with Asn159, Glu201, and Ser243.Water bridges with Val31, Asp32, Ser33, Val74, Asp75, Gly76, Ala77, Val115, Asp116, Thr117, Asn159, Glu201, Val241, and Ser243; and ionic interactions with Glu201 further enhanced the stability as shown in Figure 4.9.B. Furthermore, the most prominent interactions that occurred for more than 30 % of the 200 ns simulation duration between PknD-verticillin A include, hydrogen bond interactions with ASP116 (74 %), Asn159 (33 %), Ser243 (30 %), and Gly76 (89 %) as depicted in Figure 4.9.A.

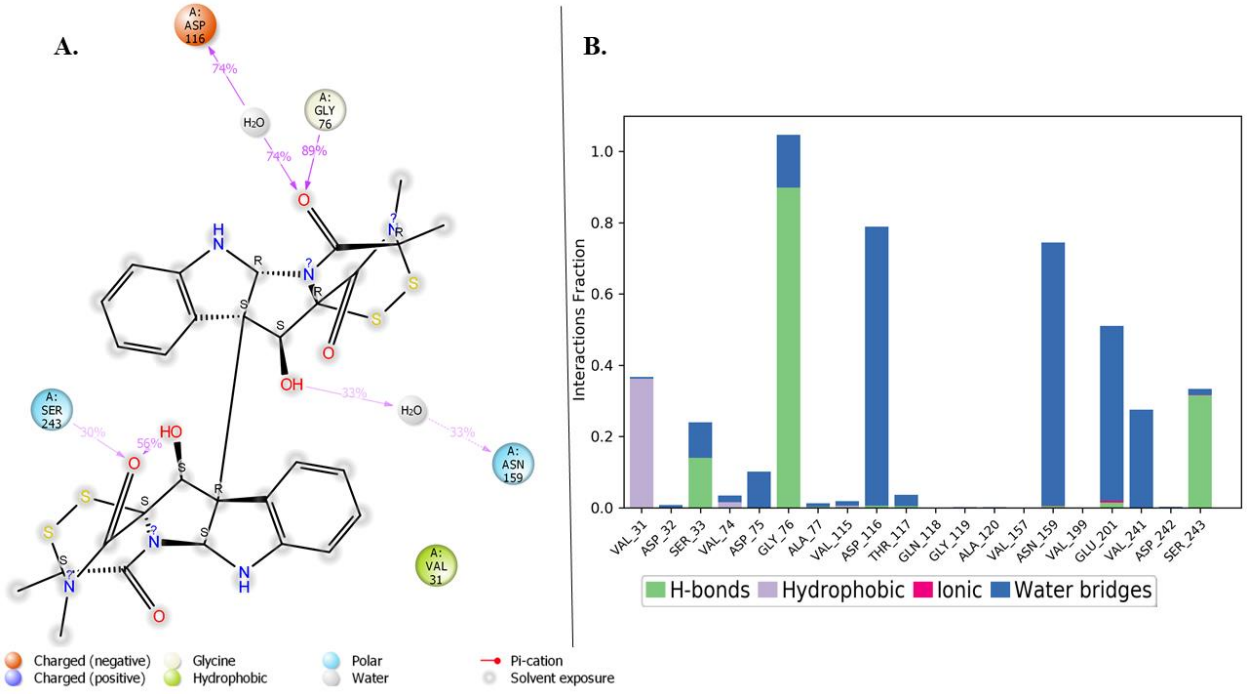


Figure 4.9. An overview of the protein-ligand contacts over a 200 ns molecular dynamics simulation. **A.** The bonds that contributed to the PknD-Verticillin A. Water bridges are hydrogen bonds between the protein and ligand facilitated by a water molecule. **B.** Contacts between PknD and Verticillin A.

The RMSD trajectory of PknE showed relative stability with an RMSD value of 2.5 Å within the initial 125 ns (Figure 4.10.A). Subsequently, the RMSD gradually increased to approximately 3.8 Å between 125 to 150 ns, then stabilized at around 3.8 Å from 150 to 200 ns. Similarly, the RMSD trajectory of roquefortine C displayed a comparable pattern, with initial RMSD values of 4 Å within the initial 125 ns, followed by large fluctuations ranging from 4 to 8 Å between 125 and 150 ns. The RMSD trajectory then reduced to 5 Å before increasing again to 8 Å until the end of the simulation (Figure 4.10.A). Analyzing the RMSF of the PknE complex with roquefortine C depicted in Figure 4.10., showed that most protein residues exhibited stability throughout the simulation. Notably, a mobile loop at residue 160 exhibited a significant fluctuation, suggesting a region of conformational flexibility within the complex. In contrast, the RMSD of the control

ligand, CJJ, exhibited a large fluctuation between 15 and 30 ns, followed by continued RMSD fluctuation ranging from 12 to 21 Å (Figure 4.10.C). This indicates that the control ligand was not stable within the designed binding pocket. Nevertheless, the RMSF of the PknE complex to control ligand was relatively stable as depicted in Figure 4.10.D.

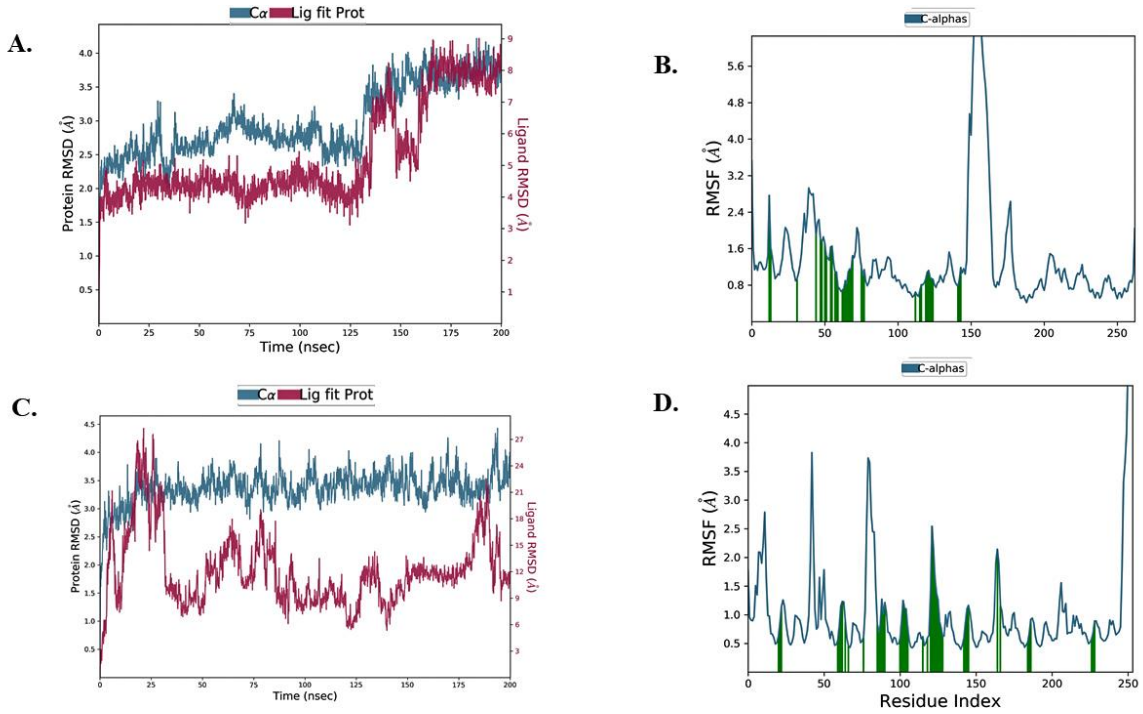


Figure 4.10. A. RSMD of PknE Cα-atoms and roquefortine C as a function of simulation time (200ns) **. B.** RMSF for PknE residues in complex with roquefortine C. **C.** RSMD of PknE Cα-atoms and control as a function of simulation time (50ns). **D.** RSMF of the residues PknE in complex with control. The green lines indicate the residues in contact with the ligand.

The stability of the PknE-roquefortine C complex which lasted for 125 ns was due to a combination of hydrogen bonds, hydrophobic interactions, ionic interactions, and water bridges. Roquefortine C formed hydrogen bonds with Glu64 and Ala65. Hydrophobic contacts were observed with Lys45, Ala65, Ala68, Leu71, Val76, Val77, Pro78, Phe82, Val91, Ala130, Ala131, Ala135 and Leu155. Water bridges with Lys45, Arg58, MSE61, Gln62, Glu64, Ala65, Gly69, Gln72, Glu73, Val77, Ile79, His80, Gly83, Try90, Ala133, Ala135, His137, Arg138 and Asp157; and ionic

interactions with Glu64, Ala135, Asp157 further enhanced the stability (Figure 4.11.B). Furthermore, the most prominent interactions that occurred for more than 30 % of the 200 ns simulation duration between PknE-roquefortine C include hydrogen bond interactions with Glu64 (50 %), hydrophobic interactions with Ala65, Val76, Phe82, Val91, and Tyr890 as depicted in Figure 4.9.A.

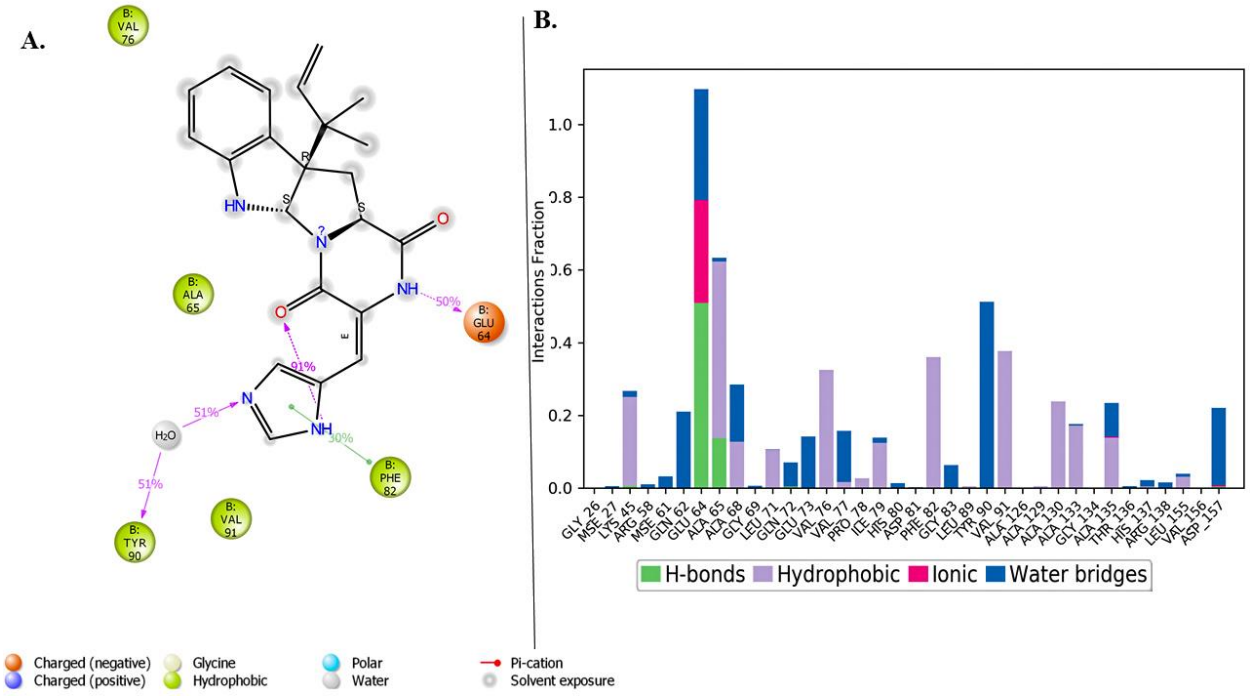


Figure 4.11. Protein-ligand contacts over a 200 ns molecular dynamics simulation. **A.** Interaction types between PknA and that occurred for more than 30 % of the simulation duration. **B.** The types of interaction and fractions which occurred between PknE and roquefortine C.

Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) ΔG_{Bind} enables the accurate description of the thermodynamics of protein-ligand interaction. As observed from Table 4.2 of verticillin A had the lowest ΔG_{Bind} against PknD of –53.67 kcal/mol, followed by roquefortine C against PknE (-53.37 kcal/mol), followed by aurovertin D against PknA (-45.34 kcal/mol) and

lastly aurovertin D against PknB (-42.87 kcal/mol). Noteworthy, verticillin A and roquefortine C exhibited higher affinity to PknD and PknE, respectively, more than the control ligand (CJJ).

Table 4.2. Binding free energy (ΔG_{Bind}) values computed by MM/GBSA.

Complex	**MM-GBSA ΔG_{bind} kcal/mol**	**ΔG_{Bind} Coulomb kcal/mol**	**ΔG_{Bind} Covalent kcal/mol**	**ΔG_{Bind} H_{bond} kcal/mol**	**ΔG_{Bind} *Solv GB* kcal/mol**	**ΔG_{Bind} *Lipo* kcal/mol**	**ΔG_{Bind} *vdW* kcal/mol**
PknA-aurovertin D	-45.34	-10.86	1.50	-0.69	18.60	-14.02	-39.87
PknA-control (CJJ)	-57.69	-25.40	1.58	-2.45	27.96	-15.96	-43.41
PknB-aurovertin D	-42.87	-13.30	2.40	-0.83	23.89	-13.29	-41.72
PknB-control (CJJ)	-60.43	-24.28	2.25	-2.61	29.38	-16.91	-48.14
PknD-verticillin A	-53.67	-14.32	1.76	-0.68	25.29	-13.82	-51.35
PknD-control (CJJ)	-35.33	-13.84	3.14	-1.59	18.68	-8.18	-32.77
PknE-roquefortine C	-53.37	-5.86	1.58	-0.68	18.24	-19.20	-47.27
PknE-control (CJJ)	-35.19	-14.76	2.25	-1.01	19.86	-9.15	-30.51

4.5 Discussion

In targeted structure-based drug design, computational modeling and the prediction of binding free energy play a critical role in molecular recognition and can be harnessed to innovate new lead design molecules with a new mode of interaction, increased efficacy, and metabolic stability. The knowledge derived from both a known active ligand and macromolecular drug targets may be used for filtering libraries of diverse compounds based on estimating the likelihood of protein-ligand interacting and the dynamic stability of the complex formed [27–29]. Most of the TB drugs that

are currently in clinical use were either mined from or derived from natural products, for example, capreomycin and kanamycin are from natural sources while rifampicin is a derivative of a rifamycin [14]. This study used a targeted virtual screening approach including molecular docking, MD simulations, and free-energy calculation to estimate the dynamic interaction of small molecules from *P. janthinellum* against *M. tb* SER/Thr protein kinases and to identify possible lead compounds that can potentially be subject to further modifications.

Small bioactive molecules are known to reversibly bind to macromolecular drug targets in non-covalent interactions with side chains of binding domain residues of the protein targets. The interaction between a small molecule and the protein target modulates essential biomolecular machinery, for instance, gene regulation and molecular recognition [30–32]. In this study, molecular docking was used as a preliminary predictive tool for screening different conformations of fungal compounds against PknA, PknE, PknB, and PknD. Aurovertin D, aurovertin D, verticillin A, and roquefortine C exhibited the highest pre-MD ΔG_{Bind} for PknA, PknB, PknD, and PknE respectively (Table 4.1). In molecular docking, the protein is rigid and does not account for all the entropic and enthalpic factors governing the formation of the protein-ligand complex. Thus, to further evaluate the interactions and dynamic behavior of the protein-ligand complexes MD simulations were performed.

Early reports revealed that PknA and PknB are essential fragility points in *M. tb* proteome and if they are inhibited cellular arrest consequently occurs, thus, they are attractive macromolecular druggable targets [14, 16, 33]. Furthermore, Carette et al. (2018) [18] illustrated the dual inhibitory efficacy of CHEMBL4166160 (CJJ) against PknA and PknB. In this regard, CHEMBL4166160 (CJJ) was selected in this research as a control ligand for comparison purposes. Aurovertins are

reported to be potent inhibitors of ATP hydrolysis and oxidative phosphorylation [34, 35]. This study investigated the stability and binding dynamics of the PknA and PknB proteins when complexed with aurovertin D. The RMSD trajectory result indicated the stability of the PknA complexed with aurovertin D, as the RMSD fluctuations were below 3 Å throughout the simulation (Figure 4.4.A). However, the RMSD values of aurovertin D exceeded 3 Å, suggesting the instability of the ligand in the designated binding site of PknA (Figure 4.4.A). In the case, of the PknB-aurovertin D complex, the RMSD trajectory revealed that both the protein and the ligand exhibited fluctuations above 3 Å, indicating conformational changes occurring during the 200 ns MD simulation (Figure 3.6.A). Previous studies highlighted that non-covalent forces, which include, hydrophobic interactions with Leu17, Val25, Ala38, and Met92, as well as van der Waals interactions with Gly18, Leu17, Ala38, Glu93, Val95, Try94, Met145, and Met155 were critical in the inhibition of PknB [14, 16]. In this study, the binding affinity of aurovertin D to PknB was driven by hydrophobic interactions with Leu17, Val25, Leu27, Ala38, Val72, Met92, Thr94, Ala142, Met145, Met155, Ile159, and Ile177. Additionally, hydrogen bonds were identified between aurovertin D and specific residues including Leu17, Phe19, Gly21, Ser23, Lys40, Val95, Asp96, Gly97, Gly18, Gly21, Met22, Ser23, Lys40, Val95, Asp96, Gly97, Arg101, Lys140, Ala142, Asp156, Arg161, and Asn168. Water bridges involving Glu93, Gly18, Tyr94, and Met155 were also observed (Figure 4.7.A). To further explore and understand the free energy differences that govern the protein-ligand interactions, the binding free energy (ΔG_{Bind}) was computed using the MM-GBSA approach. In this study, a binding affinity of -45.34 kcal/mol for aurovertin D with PknA and -42.87 kcal/mol with PknB (Table 4.2) was observed were both higher than the for the control ligand (CJJ).

In the same fashion as before for PknB and PknA, the RMSD, and binding free energy were computed for PknD-verticillin A and PknE roquefortine C complexes. The MD simulation revealed that the PknD-verticillin A complex maintained remarkable stability throughout the 200 ns, as shown by low RMSD fluctuations of PknD Cα-atoms (> 2.1 Å) (Figure 4.8.A). Notably, verticillin A exhibited minor fluctuations in its RMSD trajectory, signifying a stable binding pose in the active site of PknD (Figure 4.8.A). The computed ΔG_{Bind} value of –53.67 kcal/mol of PknD-verticillin A further elaborated the attractive forces between the protein and ligand, which were mediated by a combination of van der Waals forces, Coulomb, lipophilicity, and hydrogen bonds (Table 4.2). In contrast, the control ligand was not a good binder to PknD, as shown by higher RMSD fluctuations of PknD Cα-atoms in Figure 4.8.C.

The MD simulation of the PknE-Roquefortine C complex revealed that the complex was stable for the first 125 ns with RMSD of PknE ranging around 2.5 Å (Figure 4.10.A). However, after 125 ns, the RMSD trajectory increased to approximately 3.8 Å, indicating the protein's instability and conformational change during the last 75 ns. In comparison, both the PknE and the control ligand, CJJ, displayed relative instability as revealed by the RMSD trajectory (Figure 4.10.C). Roquefortine C exhibited a ΔG_{Bind} of -53.37 kcal/mol against PknE, while the control had a lower ΔG_{Bind} of -35.19 kcal/mol, suggesting that roquefortine C has a higher affinity to the designated binding site as compared to the CJJ.

4.6 Conclusions

In conclusion, this study employed molecular docking, molecular dynamics simulations, and binding free energy computations to explore the prospects of the potential use of natural products for targeted drug discovery against TB. Detailed results of the binding interactions and ΔG_{bind}

revealed that verticillin A is a potential lead scaffold that can further be modified to enhance its affinity, potency, and solubility, thus, establishing a foundation for more structure-based drug design probes targeting PknD Ser/Thr. Integrating computational approaches in the sampling of natural product libraries, this study contributes to the accelerated target-based drug discovery and design in TB treatment.

Author Contributions: Conceptualization, V.M., and K.N.; methodology, V.M., K.N., K.G., and M.N; formal analysis, K.N.; investigation, K.N.; resources, V.M., M.N., L.S., K.G., and I.SN.; writing original draft preparation, K.N.; writing review and editing, V.M., M.N., K.G., K.T., L.B., L.S., F.A., and L.J; supervision, V.M., M.G., K.G., L.S and I.S.N; project administration, L.S.; funding acquisition, V.M., and L.S. Molecular docking and Molecular dynamics simulations, K.N., K.G., and M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the South African Medical Research Association (SAMRC), the Centre of Excellence for Biomedical Tuberculosis Research (CBTBR), and the National Research Fund (NRF) [NRF GRANT UID129364].

Data Availability Statement: Raw data are stored among the authors and available upon request.

Acknowledgments: We acknowledge the Open Access Fund of Stellenbosch University. We would also like to extend our appreciation to the Center for High-Performance Computing, South Africa, for providing us with access to the infrastructure used to conduct the simulations.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of fungi and compounds are available from the authors.

4.7 References

1. Maertzdorf, J.; Ota, M.; Repsilber, D.; Mollenkopf, H.J.; Weiner, J.; Hill, P.C.; Kaufmann, S.H.E. Functional Correlations of Pathogenesis-Driven Gene Expression Signatures in Tuberculosis. *PLoS One* **2011**, *6*, doi:10.1371/journal.pone.0026938.

2. Dagne, B.; Desta, K.; Fekade, R.; Amare, M.; Tadesse, M.; Diriba, G.; Zerihun, B.; Getu, M.; Sinshaw, W.; Seid, G.; et al. The Epidemiology of First and Second-Line Drug-Resistance Mycobacterium Tuberculosis Complex Common Species: Evidence from Selected TB Treatment Initiating Centers in Ethiopia. *PLoS One* **2021**, *16*, doi:10.1371/journal.pone.0245687.

3. Cambau, E.; Viveiros, M.; Machado, D.; Raskine, L.; Ritter, C.; Tortoli, E.; Matthys, V.; Hoffner, S.; Richter, E.; Perez Del Molino, M.L.; et al. Revisiting Susceptibility Testing in MDR-TB by a Standardized Quantitative Phenotypic Assessment in a European Multicentre Study. *J Antimicrob Chemother* **2015**, *70*, 686–696, doi:10.1093/jac/dku438.

4. Sotgiu, G.; Ferrara, G.; Matteelli, A.; Richardson, M.D.; Centis, R.; Ruesch-Gerdes, S.; Toungoussova, O.; Zellweger, J.P.; Spanevello, A.; Cirillo, D.; et al. Epidemiology and Clinical Management of XDR-TB: A Systematic Review by TBNET. *Eur Respir J* 2009, *33*, 871–881.

5. Gajdács, M.; Albericio, F. Antibiotic Resistance: From the Bench to Patients. *Antibiotics* **2019**, *8*, 8–11, doi:10.3390/antibiotics8030129.

6. Nimmo, C.; Millard, J.; van Dorp, L.; Brien, K.; Moodley, S.; Wolf, A.; Grant, A.D.; Padayatchi, N.; Pym, A.S.; Balloux, F.; et al. Population-Level Emergence of Bedaquiline and Clofazimine Resistance-Associated Variants among Patients with Drug-Resistant Tuberculosis in

Southern Africa: A Phenotypic and Phylogenetic Analysis. *Lancet Microbe* **2020**, *1*, e165–e174, doi:10.1016/s2666-5247(20)30031-8.

7. Lonsdale, D.O.; Lipman, J. Antimicrobial Resistance: We Must Pursue a Collaborative, Global Approach and Use a "One Health" Approach. *Antibiotics* **2019**, *8*, 10–12, doi:10.3390/antibiotics8040237.

8. Martinez, J.L.; Muniesa, M.; Dionisio, F.; Kaur, P.; Peterson, E. Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. **2018**, doi:10.3389/fmicb.2018.02928.

9. Riccardi, G.; Pasca, M.R.; Buroni, S. Mycobacterium Tuberculosis: Drug Resistance and Future Perspectives. *Future Microbiol* 2009, *4*, 597–614.

10. Öztürk, H.; Ozkirimli, E.; Özgür, A. Classification of Beta-Lactamases and Penicillin Binding Proteins Using Ligand-Centric Network Models. *PLoS One* **2015**, *10*, 1–23, doi:10.1371/journal.pone.0117874.

11. Gazi, M.A.; Kibria, M.G.; Mahfuz, M.; Islam, M.R.; Ghosh, P.; Afsar, M.N.A.; Khan, M.A.; Ahmed, T. Functional, Structural and Epitopic Prediction of Hypothetical Proteins of Mycobacterium Tuberculosis H37Rv: An in Silico Approach for Prioritizing the Targets. *Gene* **2016**, *591*, 442–455, doi:10.1016/j.gene.2016.06.057.

12. Marrer, E.; Schad, K.; Satoh, A.T.; Page, M.G.P.; Johnson, M.M.; Piddock, L.J.V. Involvement of the Putative ATP-Dependent Efflux Proteins PatA and PatB in Fluoroquinolone Resistance of a Multidrug-Resistant Mutant of Streptococcus Pneumoniae. *Antimicrob Agents Chemother* **2006**, *50*, 685–693, doi:10.1128/AAC.50.2.685-693.2006.

13. Sundar, S.; Thangamani, L.; Manivel, G.; Kumar, P.; Piramanayagam, S. Molecular Docking, Molecular Dynamics and MM/PBSA Studies of FDA Approved Drugs for Protein Kinase a of Mycobacterium Tuberculosis; Application Insights of Drug Repurposing. *Inform Med Unlocked* **2019**, *16*, doi:10.1016/J.IMU.2019.100210.

14. Antunes, S.S.; Won-Held Rabelo, V.; Romeiro, N.C. Natural Products from Brazilian Biodiversity Identified as Potential Inhibitors of PknA and PknB of M. Tuberculosis Using Molecular Modeling Tools. *Comput Biol Med* **2021**, *136*, doi:10.1016/J.COMPBIOMED.2021.104694.

15. Giraud-Gatineau, A.; Coya, J.M.; Maure, A.; Biton, A.; Thomson, M.; Bernard, E.M.; Marrec, J.; Gutierrez, M.G.; Larrouy-Maumus, G.; Brosch, R.; et al. The Antibiotic Bedaquiline Activates Host Macrophage Innate Immune Resistance to Bacterial Infection. *Elife* **2020**, *9*, 1–29, doi:10.7554/eLife.55692.

16. England, P.; Wehenkel, A.; Bellinzoni, M.; Graña, M.; Duran, R.; Villarino, A.; Fernandez, P.; Andre-Leroux, G.; Takiff, H.; Cerveñansky, C.; et al. Mycobacterial Ser/Thr Protein Kinases and Phosphatases: Physiological Roles and Therapeutic Potential. **2007**, doi:10.1016/j.bbapap.2007.08.006.

17. Zeng, J.; Platig, J.; Cheng, T.Y.; Ahmed, S.; Skaf, Y.; Potluri, L.P.; Schwartz, D.; Steen, H.; Branch Moody, D.; Husson, R.N. Protein Kinases PknA and PknB Independently and Coordinately Regulate Essential Mycobacterium Tuberculosis Physiologies and Antimicrobial Susceptibility. *PLoS Pathog* **2020**, *16*, doi:10.1371/journal.ppat.1008452.

18. Carette, X.; Platig, J.; Young, D.C.; Helmel, M.; Young, A.T.; Wang, Z.; Potluri, L.P.; Moody, C.S.; Zeng, J.; Prisic, S.; et al. Multisystem Analysis of Mycobacterium Tuberculosis Reveals Kinase-Dependent Remodeling of the Pathogen-Environment Interface. *mBio* **2018**, *9*, doi:10.1128/mBio.02333-17.

19. Lougheed, K.E.A.; Osborne, S.A.; Saxty, B.; Whalley, D.; Chapman, T.; Bouloc, N.; Chugh, J.; Nott, T.J.; Patel, D.; Spivey, V.L.; et al. Effective Inhibitors of the Essential Kinase PknB and Their Potential as Anti-Mycobacterial Agents. *Tuberculosis* **2011**, *91*, 277–286, doi:10.1016/j.tube.2011.03.005.

20. Abdulhamid, A.; Awad, T.A.; Ahmed, A.E.; Koua, F.H.M.; Ismail, A.M. Acetyleugenol from Acacia Nilotica (L.) Exhibits a Strong Antibacterial Activity and Its Phenyl and Indole Analogues Show a Promising Anti-TB Potential Targeting PknE/B Protein Kinases. *Microbiol Res (Pavia)* **2021**, *12*, 1–15, doi:10.3390/microbiolres12010001.

21. Grigalunas, M.; Burhop, A.; Zinken, S.; Pahl, A.; Gally, J.M.; Wild, N.; Mantel, Y.; Sievers, S.; Foley, D.J.; Scheel, R.; et al. Natural Product Fragment Combination to Performance-Diverse Pseudo-Natural Products. *Nat Commun* **2021**, *12*, doi:10.1038/s41467-021-22174-4.

22. Gurnani, N.; Mehta, D.; Gupta, M.; Mehta, B.K. Natural Products: Source of Potential Drugs. *Afri J Basic Appl Sci* **2014**, *6*, 171–186, doi:10.5829/idosi.ajbas.2014.6.6.21983.

23. Green, J.; Loesgen, S.; Stone, J. Antibacterial Natural Products from Fungal Endophytes. **2019**.

24. Tapfuma, K.I.; Sebola, T.E.; Uche-Okereafor, N.; Koopman, J.; Hussan, R.; Makatini, M.M.; Mekuto, L.; Mavumengwana, V. Anticancer Activity and Metabolite Profiling Data of Penicillium Janthinellum KTMT5. *Data Brief* **2020**, *28*, doi:10.1016/j.dib.2019.104959.

25. Baptista, R.; Bhowmick, S.; Shen, J.; Mur, L.A.J. Molecular Docking Suggests the Targets of Anti-Mycobacterial Natural Products. *Molecules* **2021**, *26*, 475, doi:10.3390/molecules26020475.

26. Zong, K.; Xu, L.; Hou, Y.; Zhang, Q.; Che, J.; Zhao, L.; Li, X.; Brezovsky, J.; Sousa, S.F. Molecules Virtual Screening and Molecular Dynamics Simulation Study of Influenza Polymerase PB2 Inhibitors. **2021**, doi:10.3390/molecules26226944.

27. Schuster, D.; Li, H. *Virtual Drug Design*; 2020; ISBN 9782889633593.

28. Cerón-Carrasco, J.P. When Virtual Screening Yields Inactive Drugs: Dealing with False Theoretical Friends. **2022**, doi:10.1002/cmdc.202200278.

29. Grigalunas, M.; Burhop, A.; Zinken, S.; Pahl, A.; Gally, J.M.; Wild, N.; Mantel, Y.; Sievers, S.; Foley, D.J.; Scheel, R.; et al. Natural Product Fragment Combination to Performance-Diverse Pseudo-Natural Products. *Nat Commun* **2021**, *12*, doi:10.1038/s41467-021-22174-4.

30. Atanasov, A.G.; Zotchev, S.B.; Dirsch, V.M.; Orhan, I.E.; Banach, M.; Rollinger, J.M.; Barreca, D.; Weckwerth, W.; Bauer, R.; Bayer, E.A.; et al. Natural Products in Drug Discovery: Advances and Opportunities. *Nat Rev Drug Discov* 2021, *20*, 200–216.

31. Pant, S.; Singh, M.; Ravichandiran, V.; Murty, U.S.N. Peptide-like and Small-Molecule Inhibitors Against. *J Biomol Struct Dyn* **2020**, *0*, 1–10, doi:10.1080/07391102.2020.1757510.

32. Jiménez, A.; García, P.; de La Puente, S.; Madrona, A.; Camarasa, M.J.; Pérez-Pérez, M.J.; Quintela, J.C.; García-del Portillo, F.; San-Félix, A. A Novel Class of Cationic and Non-Peptidic Small Molecules as Hits for the Development of Antimicrobial Agents. *Molecules* **2018**, *23*, doi:10.3390/molecules23071513.

33. Salvi, A.; Amrine, C.S.M.; Austin, J.R.; Kilpatrick, K.A.; Russo, A.; Lantvit, D.; Calderon-Gierszal, E.; Mattes, Z.; Pearce, C.J.; Grinstaff, M.W.; et al. Verticillin A Causes Apoptosis and Reduces Tumor Burden in High-Grade Serous Ovarian Cancer by Inducing DNA Damage. *Mol Cancer Ther* **2020**, *19*, 89–100, doi:10.1158/1535-7163.MCT-19-0205.

34. Li, W.; Ma, Z.; Chen, L.; Yin, W.B. Synthesis and Production of the Antitumor Polyketide Aurovertins and Structurally Related Compounds. *Appl Microbiol Biotechnol* **2018**, *102*, 6373–6381, doi:10.1007/S00253-018-9123-1/FIGURES/3.

35. Wang, F.; Luo, D.Q.; Liu, J.K. Aurovertin E, a New Polyene Pyrone from the Basidiomycete Albatrellus Confluens. *J Antibiot 2005 58:6* **2005**, *58*, 412–415, doi:10.1038/ja.2005.53.

Chapter 5

Exploring the Metabolic Potential of *Bacillus licheniformis* for the Production of Antimycobacterial Secondary Metabolites

Kudakwashe Nyambo [1], Kudzanai Ian Tapfuma [1], Francis Adu-Amankwa [1], Lucinda Baatjies [1], Idah Sithole Niang [2], Liezel Smith [1], Mkhuseli Ngxande [3], and Vuyo Mavumengwana [1, *]

[1] *DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town; KN: knyambo@sun.ac.za;FN:* fa@sun.ac.za; *K.I.T: kudzanait@su.ac.za; LB: lbaatjies@sun.ac.za; LS: liezels@sun.ac.za; VM: vuyom@sun.ac.za*

[2]*Department of Biotechnology and Biochemistry, University of Zimbabwe, B064, Mount Pleasant, Harare, Zimbabwe; sitholeidah2015@gmail.com*

[3]*Computer Science Division, Department of Mathematical Sciences, Faculty of Science University of Stellenbosch, Matieland, South Africa; ngxandem@sun.ac.za*

***** Corresponding author.

Email address: *vuyom@sun.ac.za;* Tel: +27 718502949

5.1 Abstract

Mycobacterium tuberculosis (M. tb) presents a significant complication to disease control globally due to its resistance to multiple antibiotics. Therefore, the current study aimed to comprehensively analyze the genomic features of *Bacillus licheniformis* SAMN36381076, a metagenome-assembled genome (MAG) isolated from gold mine tailings. Herein, the primary objective was to evaluate functional capacity for synthesizing bioactive scaffolds with potential to combat *M. tb*. The bacterial genomic DNA was isolated and sequenced using Illumina's NextSeq platform. A

KBase metagenomic pipeline was used for processing, assembling the identified genomes, constructing a phylogenetic tree, and predicting the presence of genes associated with carbohydrate-active enzymes, genes associated with biochemical cycles, and nitrogen metabolism. The presence of antibiotic-resistance genes was evaluated using a module on PATRIC. The antiSMASH platform was used to determine the biosynthetic gene clusters (BGCs) present in the metagenome-assembled genomes (MAG). Two MAGs were identified as *B. licheniformis SAMN36381076* and *Peribacillus simplex* SAMN36381075. The *B. licheniformis* SAMN36381076 MAG had an estimated genome length of 4.213156 Mb, with a G+C content of 46.08%, comprising 58 contigs and exhibiting an N50 length of 165,033 bp. Genomic annotation revealed 4,613 protein-coding sequences (CDS), encompassing 1,008 hypothetical proteins and 3,605 proteins with assigned functions, as well as 58 tRNAs. The *B. licheniformis* SAMN36381076 MAG had six diverse antibiotic resistance mechanisms, with efflux pumps as the predominant mechanism of resistance. A total of 13 BGCs responsible for encoding synthesis of diverse secondary metabolites, including fengycin, butirosin A, butirosin B, schizokinen, pulcherriminic acid, bacillibactin, bacillibactin E, bacillibactin F, lichenicidin VK21 A1, Lichenicidin VK21 A2, and thermoactinoamide A. Additionally, the *B. licheniformis* SAMN36381076 MAG harbored genes associated with encoding a wide repertoire of carbohydrate-active enzymes (CAZymes) involved in carbohydrate degradation and assembly processes. Further, the annotations revealed genes that encode enzymes responsible for oxyanion biogeochemical cycles and nitrogen metabolism highlighting the bacterium's adaptive capacity to the unstable gold mine tailings environmental conditions. This genomic-based study comprehensively uncovers the genomic features and functional capabilities of *B. licheniformis*

SAMN36381076 MAG, thus, providing essential knowledge to pave the way for the discovery of anti-*M. tb* natural products.

Keywords: *Bacillus licheniformis*, genome analysis, functional annotation, protein families, enzymatic activities, ecological adaptability.

5.2 Introduction

Gold mine tailings are host to complex microbial communities that play an essential role in modulating nutrient cycling, heavy metal mineralization, and stabilization [1]. These below ground microbial communities, influence microbe-microbe and host-microbe interactions by synthesising a variety of specialized metabolites known as natural products [2, 3]. Earlier research has focused on the biogeographic patterns of microbial distribution and biogeochemical transformation of mineral elements [4]. A study by Courchesne et al. [5] explored the bacterial diversity in mine tailings and identified a microbial community comprised of *Acidoferrobacterals*, *Bacillalles*, *Rhizobiales*, *Betaproteocterials*, *Sphingomonadales*, *Gaiellales*, *Pyrinomodales*, and *Solirubrobacterales*. While other metagenomics investigations revealed that the *Halothiobacillaceae*, *Acidithiobacillus*, *Leptospirillium*, *Ferrimicrobium*, *Actinobacteria*, and *Acidibacter* family was the most abundant in mine tailings [6–8]. However, most ecological studies to date have focused more on microbial community succession in gold mine tailings and bioremediation, neglecting the metabolic potential of microorganisms that drive the colonization for drug discovery [3, 9].

Bacterial natural products are widely recognized as one of the key drivers of microbial diversity and composition in microenvironments (Begley et al., 2009; Parkinson et al., 2018). All the possible secondary metabolites encoded by microbes make up the biosynthetic chemical space. Exploration of the biosynthetic chemical space by using genome mining techniques will reveal the potential bioactive natural products that contribute to the fierce competition between different microorganisms. The secondary metabolites are synthesized by multidomain enzymes encoded by genes that tend to cluster within a genome and are known as biosynthetic gene clusters (BGCs) [12, 13]. Recent efforts to harness the metabolic potential of actinobacteria through microbial genome mining revealed the presence of 20-80 different BGCs, with many remaining silent under standard laboratory culture conditions [14, 15]. These untapped genes present opportunities to discover architecturally diverse compounds which have strong affinity towards microbial targets. Knowledge of the evolutionary connections, biosynthetic routes, and bioactive secondary metabolites guides the exploration of the microbial resources in drug discovery [16, 17]. Target-directed genome mining, based on the notion that a bacterium must resist the self-produced arsenal to avoid suicide, enables the prediction of a compound's mode of action, regardless of the unknown chemical structure [18, 19]. However, the major drawback lies in selecting and prioritizing the bacterial strains harboring the highest potential to synthesize novel bioactive molecules.

Diverse bacterial natural products have been proven to be a reference point for the discovery of important chemotherapeutics used for fighting infectious diseases such as tuberculosis (TB) [4]. Genome sequencing data have revealed that the *Bacillus subtilis* group is known to produce a wide array of bioactive secondary metabolites in the classes which include, cyclic lipopeptides, non-ribosomal dipeptides, terpenes, polyketides, bacteriocins, ribosomally synthesized and post-

translationally modified lantibiotics with a wide range of pharmacological functionalities [20–22]. Cyclic lipopeptides are amongst the most lethal antimicrobial agents produced by *B. subtilis* group. Earlier studies reported the activity of surfactin, iturin, and fengycin analogs against antimicrobial-resistant strains [23, 24]. A previous study exploring the anti-TB bioactive compounds reported that the secondary metabolites from *Bacillus* sp. in the *B. subtilis* group have shown inhibition of *M. tb* H37Rv *in-vitro* [22, 25]. Further, in our previous study *B. licheniformis* was identified using 16s rRNA-produced crude extracts that exhibited activity against *M. tb* H37Rv (Chapter 3). The aim of this study was to explore the potential metabolic of an enriched culture through genome-guided mining. This step expands knowledge of the bacteria's natural products and is an important step to harness the novel products which are much needed in antimycobacterial drug development.

5.3 Materials and methods

5.3.1 Bacteria samples

The gold mine tailings samples were collected from five sites located around the Germiston area, Johannesburg, South Africa (26º13'7.08" S, 28º29'8.64" E). At each site, 0.5 kg of sample material was collected at a depth of 12 cm. Samples were stored in polyethylene bags at 4 ℃ until further processing. Bacteria cultures were isolated and enriched as described previously in chapter 3. Briefly, the enriched bacteria were cultivated on Luria-Bertani agar at 37 ℃ [26].

5.3.2 Genomic DNA isolation and Sequencing

Genomic DNA isolation and purification were performed using the Promega Wizard® Genomic DNA Purification Kit (Promega, WI, USA) according to the manufacturer's instructions. The quality of the genomic DNA was quantified for concentrations 230/260- and 260/280 using a

Microplate Spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA). The whole-genome sequencing was performed using Illumina's NextSeq platform at Inqaba Biotech. Briefly, the genomic DNA sample was fragmented using an enzymatic approach (NEB Ultra II FS kit), according to the manufacturer's instructions. The resulting DNA fragments were selected according to size (200 – 500bp), using AMPure XP beads. Thereafter, the fragments were end-repaired, and Illumina-specific adapter sequences were ligated to each fragment. Each sample was individually indexed and subjected to a second size selection step. Samples were then quantified, using a fluorometric method, diluted to a standard concentration (4nM), and then sequenced on Illumina's NextSeq platform, using a NextSeq 300 cycle kit, following a standard protocol as described by the manufacturer. 1Gb of data (2x150bp paired-end reads) were produced for each sample.

5.3.3 Metagenomic assembly, binning, taxonomic and functional annotation

The sequence data from the bacteria were processed as described previously [27, 28]. Briefly, the initial quality check of the raw paired-end Illumina reads in (.fastq) format was evaluated using FastQC (v0.11.9) [29]. To improve the overall quality of the raw reads; low-quality bases and adapter sequences were trimmed from the Illumina raw reads, using Trimmomatic (v0.36) [30]. The trimmed reads were subjected to further quality checks using FastQC (v0.11.9) [29]. MetaSPADES (v3.15) [31] was utilized to assemble high-quality reads into contigs. High-quality contigs were binned into metagenome-assembled genomes (MAGs) using MaxBin2 (v2.2.4) [32] and MetaBAT2 (v1.7) . The DAS tool (v.1.1.1) [33] was used to integrate the MAGs. The quality of the bins was estimated based on 5 % contamination and completeness >90 % by CheckM (v1.0.18) [34]. This step ensured the selection of high-quality genomic bins for downstream

analysis. The evolutionary and phylogenetic relationships of the assembled genomes were determined by SpeciesTree (v2.2.0) [35]. The assembled genomes were functionally annotated by Rapid Annotations using Subsystems Technology toolkit (RASTtk) (v1.073) on The PathoSystems Resource Integration Center (PATRIC) (http://www.patricbrc.org/) [36]. Antimicrobial resistance genes and mutations conferring resistance to antibiotics were determined by Comprehensive Antibiotic Resistance Database (CARD v3.2.4) [37] in PATRIC (v3.6.9) (http://www.patricbrc.org/). Additionally, HMMER (v3.2.1) [38] was used to search for environmental bio-element cycling families enzymes. The genes that encode carbohydrate-active enzymes (CAZymes) in the MAGs were identified and classified using CAZymes (CAZyDB: http://www.cazy.org/) [39]. The MAGs were deposited in NCBI and accession numbers were allocated.

5.3.4 Identification of BGCs, PKS KS domain, NRPS C domain.

The BGCs and core genes in the genome were annotated using antiSMASH v5.0.0 (https:// antismash.secondarymetabolites.org, accessed on 6/05/2023) [40] using relaxed detection strictness.

5.4 Results

The metagenome-assembled genomes (MAGs) of the enriched *Bacillus* strains were characterized using Illumina sequencing platforms. A total of two MAGs belonging to the genus *Bacillus* had high completeness (100 %), and low contamination, thus, a comprehensive analysis and comparison could be carried out (Table 1). *B. simplex* SAMN36381075 MAG (bin.001.) and *B. licheniformis* SAMN36381076 MAG (bin.002) were identified (Figure 5.1) with 5.90 Mb and 4.21

Mb genome sizes, respectively. *B. licheniformis* SAMN36381076 MAG was selected and extracted as an assembly for further analysis.

Table 5.1 Characteristics of the identified MAGs

Feature	**Species**	**Species**
Genome Name	*Bacillus licheniformis* SAMN36381076	*Peribacillus simplex* SAMN36381075
Fine Consistency (%)	99.3	92.2
Completeness (%)	100	100
Contamination (%)	0.7	1.4
Contig count	58	183
DNA size (Mb)	4.21	5.90
Contigs N50 (bp)	165033	70335
Mean Coverage (%)	69.98	67.78

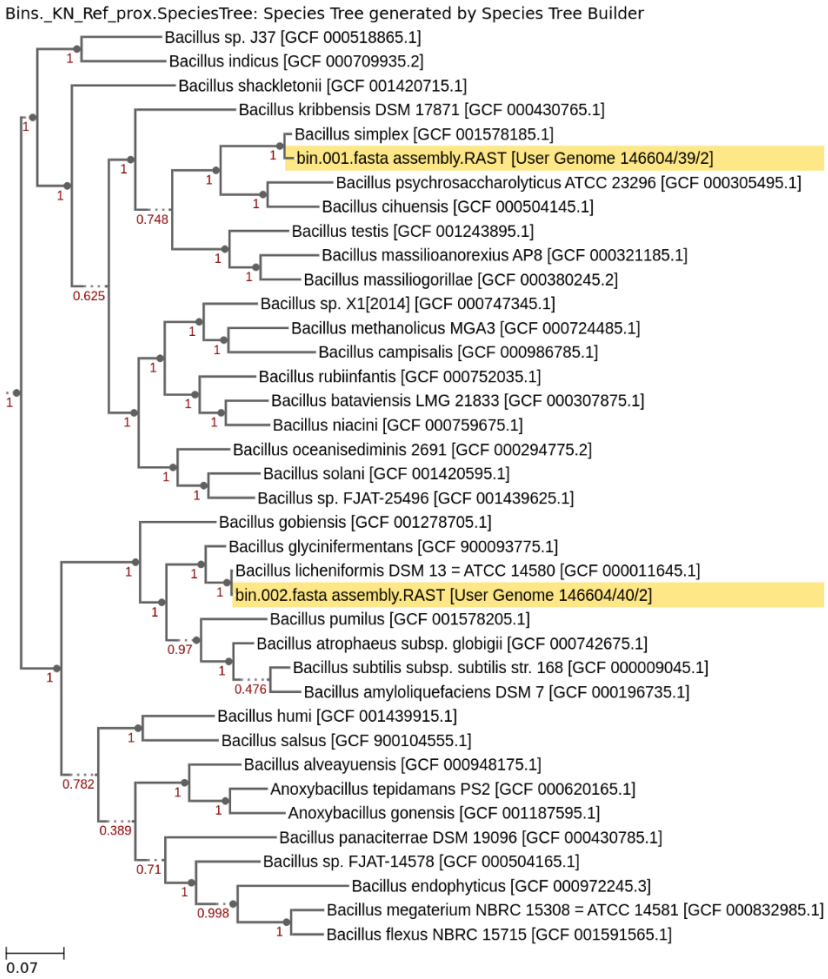


Figure 5.1. The phylogenetic positions of *B. licheniformis* SAMN36381076 MAG (bin.002) and *B. simplex* SAMN36381075 MAG (bin.001), along with their closest neighbors within the Bacillus genus, were determined.

The characteristics of *B. licheniformis* SAMN36381076 MAG are summarized in Table 5.2 and Figure 5.3. Briefly, the estimated genome length of *B. licheniformis* SAMN36381076 was 4.213156 Mb, with an average G+C content of 46.08 %. The N50 length, representing the shortest sequence length at 50 % of the genome, was 165,033 bp. The genome assembly consisted of 58 contigs, with an L50 count of 8. Further genomic annotation identified 4,613 protein-coding sequences (CDS), and 58 transcribed RNA (tRNA)s. Among the CDS, 1,008 were predicted as hypothetical proteins, and 3,605 proteins were assigned specific functions. Functional assignments

included 1,070 proteins with Enzyme Commission (EC) numbers, 898 with Gene Ontology (GO) assignments, and 789 proteins mapped to KEGG pathways. Further, the annotation showed the presence of 4,398 proteins belonging to genus-specific protein families (PLFams) and 4,408 proteins categorized under cross-genus protein families (PGFams). A circular graphical representation of the genome annotation is depicted in Figure 5.2.A. A total of 96 proteins were identified to be responsible for metabolism, 43 for protein processing, 34 for stress response, defense, and virulence, 30 for cellular processes, 27 for energy, 17 for DNA processing, 16 for membrane transport, 13 for RNA processing, four for cell envelope, three for regulation and cell signaling, and lastly three miscellaneous Figure 5.2.B.

Table 5.2. Genome features of *B. licheniformis* SAMN36381076

Features	**Chromosome**
Contigs	58
GC content	46.08
Plasmids	0
Contig L50	8
Genome Length	4.21 Mb
Contig N50	165033
Chromosomes	0
CDS	4613
tRNA	58
Partial CDS	0
Miscellaneous RNA	0
Repeat Regions	0
Hypothetical proteins	1008
Proteins with functional assignments	3605
Proteins with EC number assignments	1070
Proteins with GO assignments	898
Proteins with Pathway assignments	789
Proteins with Subsystem assignments	1298
Proteins with PARTIC cross-genus family (PGfam) assignments	4408

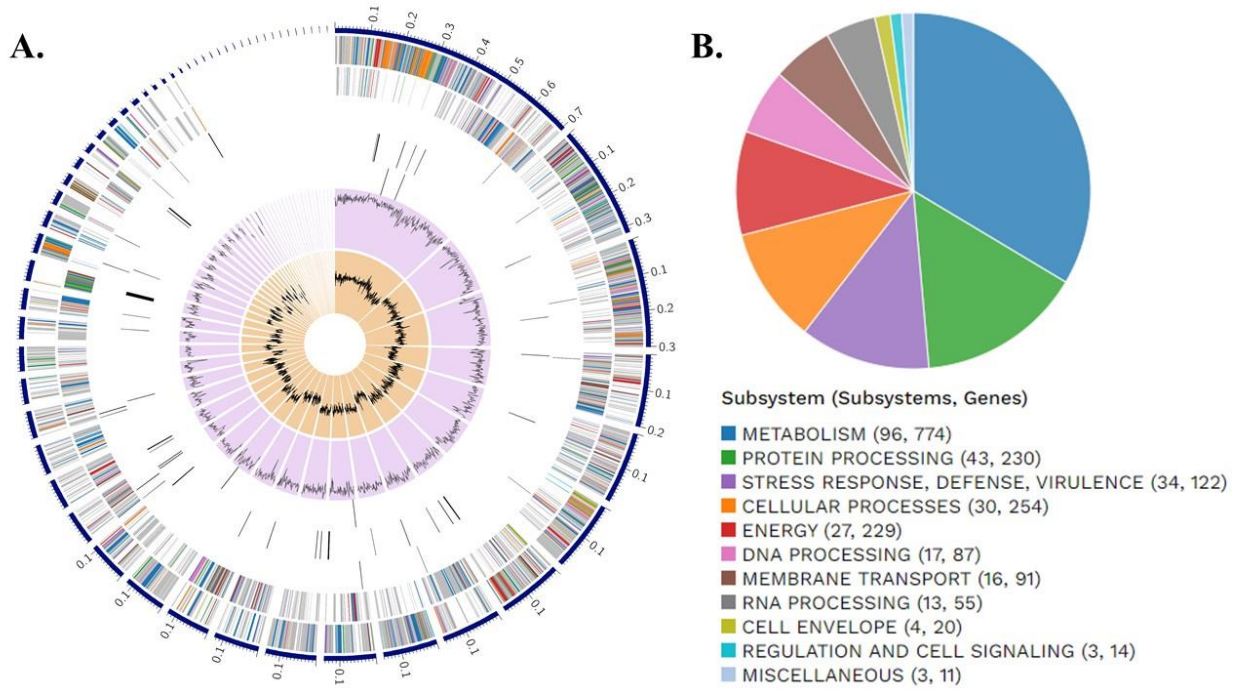


Figure 5.2. Circular genome of *Bacillus licheniformis* SAMN36381076 MAG. This includes, from outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to known virulence factors, GC content, and GC skew. The colors of the CDS on the forward and reverse strands represent the subsystem to which the genes belong. **B**. Distribution of annotated subsystems categories observed in *B. licheniformis* SAMN36381076 MAG.

A total of 48 genes were annotated and found to have homology to the genes that are responsible for antibiotic resistance, 35 for a drug target, 117 for transporter, and 4 for virulence (Table 5.3). The blast result of antibiotic resistance genes was based on three databases, including CARD, NDARO, and PATRIC.

Table 5.3. Specialty genes

Features	Source	Genes
Antibiotic resistance	CARD	2
Antibiotic Resistance	NDARO	1
Antibiotic Resistance	PATRIC	45

Drug target	DrugBank	33
Drug target	TTD	2
Transporter	TCDB	117
Virulence Factor	PATRIC_VF	2
Virulence	Victros	2

The assembled genome of *B. licheniformis* SAMN36381076 was annotated to discover the AMR genes and corresponding AMR mechanisms (Table 5.4). In this study, a total of six distinct resistance mechanisms were classified under antibiotic inactivation enzyme, antibiotic target modifying enzyme, antibiotic target protection protein, antibiotic target replacement protein, efflux pump, absence of a gene, and protein-altering cell wall charge were identified (Table 5.4). These mechanisms of action confer resistance to antimicrobial classes, including macrolides, lincosamides, aminoglycosides, peptide antibiotics, multiple antibiotic resistance, tetracyclines, aminoglycosides, and phenicol antibiotics. Among the mechanisms, antibiotic efflux pumps were identified as the prominent mode of resistance. Further, several genes encoding for efflux pumps, such as *BceA*, *BceB*, *EbrA*, *EbrB*, and *YkkCD* were identified in the *B. licheniformis* SAMN36381076 MAG. The presence of these genes suggests co-resistance for antibiotics and putative toxic elements produced by *B. licheniformis* SAMN36381076, for instance, bacitracin. Notably, the resistance to bacitracin indicates the ability of *B. licheniformis* SAMN36381076 to synthesize the antimicrobial peptide.

BGCs are responsible for the synthesis of bioactive secondary metabolites which are involved in facilitating the competition amongst bacteria within a niche. In this study, the genome of *B. licheniformis* SAMN36381076 was analyzed and 13 BGCs associated with the biosynthesis of various secondary metabolites classes were identified, as depicted in Figure 5.3. These BGCs

encompassed diverse metabolite classes, including betalactone, terpene, thiopeptide, lasso peptides, RiPP recognition element (RRE)-containing, Type III Polyketide Synthase (T3PKS), lanthipeptide (class II), NI-siderophore, tRNA-dependent cyclodipeptide synthase (CDPS), and non-ribosomal peptides (NRPS). Notably, gene clusters 12.1, and 47.1 displayed 100 % amino acid homology to gene clusters that are known to produce Lichenicidin VK21 A1/ Lichenicidin VK21 A2, and thermoactinoamide A, respectively. Furthermore, gene clusters 1.1, 5.1, 13.1, 15.1, 23.1 and 36.1 exhibited homology 53 %, 7 %, 50 %, 60 %, and 66 %, with fengycin, butirosin A/butirosin B, schizokinen, pulcherriminic acid, and bacillibactin/bacillibactin E/bacillibactin F, respectively. Gene clusters 11.1 and 18.1 exhibited 50 % and 57 % homology with lichenysin, respectively. Notably, gene clusters 1.2, 7.1, and 9.1 showed no homology with known clusters in the MIBiG (Minimum Information about a Biosynthetic Gene cluster) database. These findings suggest that these clusters may encode novel metabolites within the *B. licheniformis* SAMN36381076 genome.

Table 5.4. Antimicrobial Resistance Genes

AMR Mechanism	Genes	Product	Antibiotics class
Antibiotic inactivation enzyme	CatA family	Chloramphenicol O-acetyltransferase (EC 2.3.1.28) CatA superfamily	Phenicol antibiotics
	FosB,	Fosfomycin resistance protein FosB	Fosfomycin
Antibiotic target modifying enzyme	RlmA(II)	23S rRNA (guanine(748)-N(1))-methyltransferase (EC 2.1.1.188)	Macrolides, Lincosamides
Antibiotic target protection protein	BcrC	Undecaprenyl-diphosphatase BcrC (EC 3.6.1.27),	Peptide antibiotics conveys bacitracin
Antibiotic target replacement protein	fabL	Enoyl-[acyl-carrier-protein] reductase [NADPH] (EC 1.3.1.104), FabL	Triclosan
Efflux pump conferring antibiotic resistance	BceA,	Bacitracin export ATP-binding protein BceA	Peptide antibiotics (bacitracin)
	BceB	Bacitracin export permease protein BceB	Peptide antibiotics (bacitracin)
	EbrA	Multidrug resistance protein EbrA	Multiple antibiotic resistance
	EbrB	Multidrug resistance protein EbrB	Multiple antibiotic resistance
	YkkCD	Broad-specificity multidrug efflux pump YkkC	Tetracyclines, aminoglycosides, phenicol antibiotics
Gene conferring resistance via absence	gidB	16S rRNA (guanine(527)-N(7))-methyltransferase (EC 2.1.1.170)	Aminoglycosides (streptomycin)
Protein altering cell wall charge conferring antibiotic resistance	GdpD	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)	Peptide antibiotics (daptomycin)
	MprF	L-O lysylphosphatidylglycerol synthase (EC 2.3.2.3)	Peptide antibiotics (daptomycin, defensin)
	PgsA	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5)	Peptide antibiotics (daptomycin)

However, they were predicted in terpene biosynthesis (cluster 1.2), lasso peptide (gene cluster 7.1), and T3PKS (gene cluster 9.1).

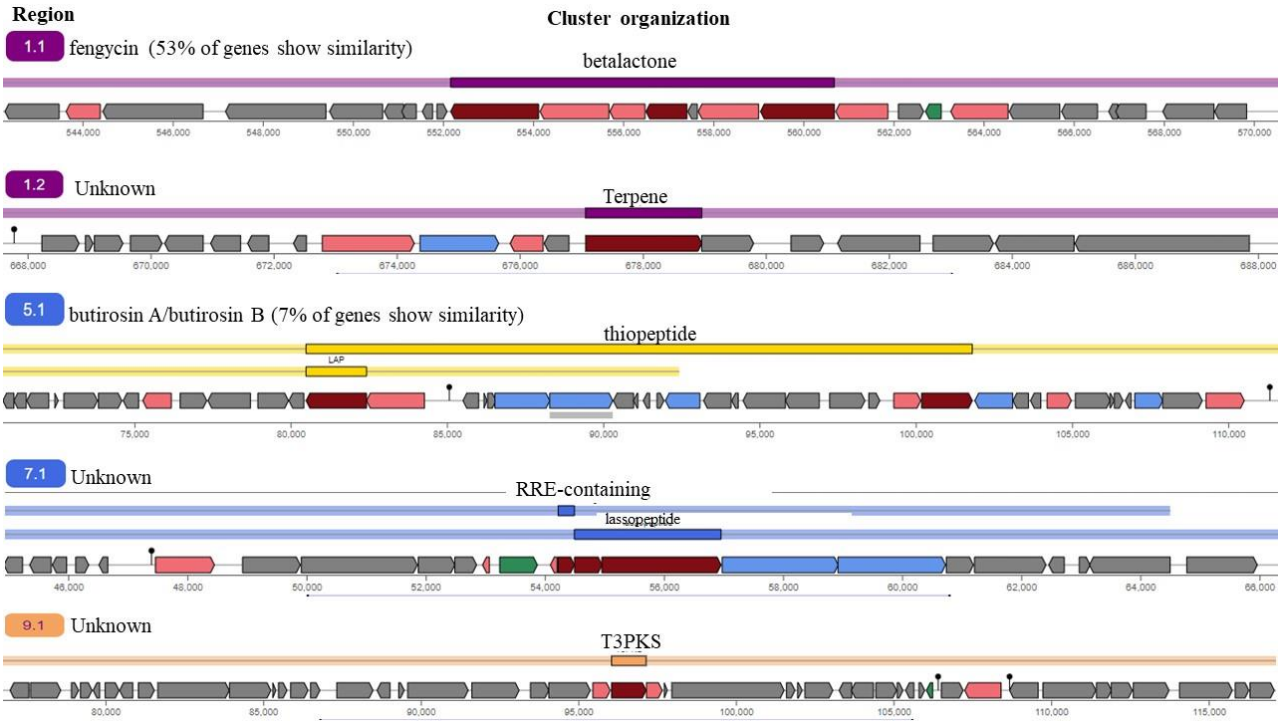


Figure 5.3. Predicted biosynthetic gene clusters of strain *B. licheniformis* SAMN36381076 MAG by antiSMASH. Fengycin, butirosin A, and 3 unknown metabolites were predicted with core biosynthetic genes.

Figure 5.3. continued. Predicted biosynthetic gene clusters of strain *B. licheniformis* SAMN36381076 MAG by antiSMASH. Lichenysin, bacillibactin/bacillibactin E/bacillibactin F,

thermoactinoamide A, Lichenicidin VK21 A1/ Lichenicidin VK21 A2, Schizokinen and pulcherriminic acid were predicted with core biosynthetic genes.

In the present study, the *B. licheniformis* SAMN36381076 genome was found to possess a wide array of genes responsible for encoding numerous Carbohydrate-active enzymes (CAZy) that are involved in the breakdown and assembly of carbohydrate complexes. Further, the coexistence of genes that encode for (auxiliary activities; AA) may suggest the importance of the bacteria in initial lignin degradation. A total of 41 CAZyme genes were annotated in the *B. licheniformis* SAMN36381076 assembled genome (Figure 5.4). The identified genes are different families, which include encoding polysaccharide lyases (18), carbohydrate esterases (6), glycosyltransferases (10), auxiliary activities (3), and carbohydrate-binding modules (4) (Figure 5.4.B, and Figure 5.4.C). The genome of *B. licheniformis* SAMN36381076 harbors a range of genes that play critical roles in the biogeochemical cycle of oxyanions (nitrate, selenate, and sulfate) (Figure 5.4.A). Notably, *cysN* and *cysC* genes are responsible for sulfate reduction, while *phsA* gene is involved in polysulfide reduction/thiosulfate disproportionation. The genes *sat*, *sqr*, and *sulfur_dioxygenase* are associated with sulfide oxidation, and the gene *soxB* is involved in thiosulfate oxidation. Additionally, the genome of *B. licheniformis* SAMN36381076 also contains several nitrogen metabolism genes, including *narG* which encodes nitrate reductase, *norB* which encodes nitric oxide reductase, and *nirB* and *nirD* which encodes nitrite reductase. Evidence of arsenite oxidation is shown by the presence of gene encoding arsenite oxidation.

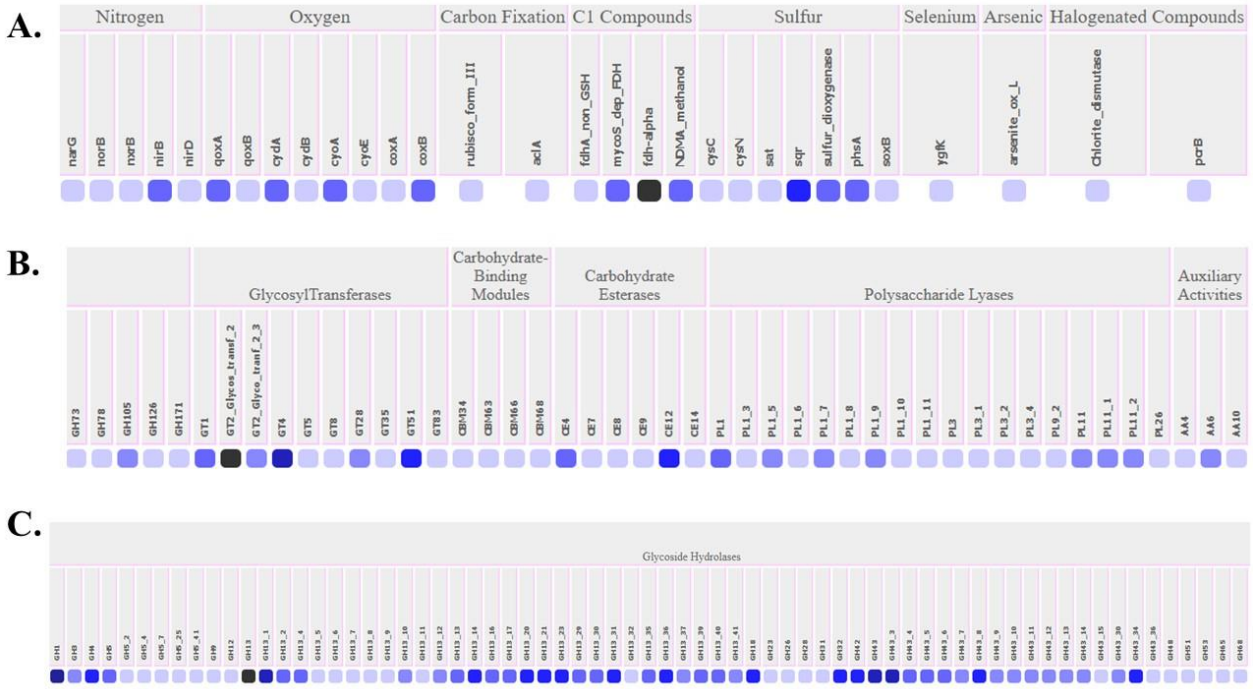


Figure 5.4.A. Genes responsible for encoding environmental bio-element cycling families. **B.** Annotated CAYZmes in *B. licheniformis* SAMN36381076 MAG. **C.** Annotated CAYZmes in *B. licheniformis* SAMN36381076 MAG.

5.5 Discussion

In the quest to combat multi-drug-resistant pathogens, natural products have gained increased attention in biomedical research due to their inhibitory activity against various drug-resistant pathogens. *Bacillus sp.* are known to natively synthesize a diverse array of bioactive secondary metabolites encoded by BGCs. These products have demonstrated utility in medicinal, agricultural, and bioremediation applications [41]. Earlier reports have highlighted the conservation of BGCs within the *Bacillus* genus that are responsible for producing clade-specific bioactive metabolites [42]. In this study, some secondary metabolites encoded by the BGCs were identified as clade-specific, for instance, gene clusters 12.1, 47.1, 1.1, 11.1, 18.1, and 36.1 were associated with the synthesis of lichenicidin VK21 A1/ Lichenicidin VK21 A, fengycin,

lichenysin, and bacillibactin/bacillibactin E/bacillibactin F, respectively, are known to be produced by *B. licheniformis.* The presence of these BGCs was also observed in several other *B. licheniformis* related genome studies [19, 20, 43, 44]. Interestingly, the BGC 47.1 displayed 100 % amino acid homology to gene clusters associated with the biosynthesis of thermoactinoamide A. This result aligns with studies demonstrating the acquisition of BGCs through horizontal gene transfer and the transferability of functional BGCs between different bacterial genomes, contributing to enhanced fitness [44].

Studies have shown that *B. licheniformis* is a reservoir for producing peptides including RiPPs and nonribosomal peptides and ribosomally synthesized [19, 20, 44]. In this study, we annotated the *B. licheniformis* SAMN36381076 MAG to decode the potential anti-mycobacterial roles of the secondary metabolites. Genome annotation results in this study identified anti-microbial peptide-related genes from the annotated BGCs, including fengycin, bacillibactin, lichenicidin, lichenysin, thermoactinoamide A, and pulcherriminic acid. A recent study reported that bacillibactin analogs which play a pivotal role in chelating ferric ions exhibited inhibitory activity against *Mycobacterium smegmatis* MTCC6 (MIC 22.15 µM) [22]. Lichenysin is a lipopeptide produced by *B. licheniformis* and is structurally similar to surfactins from *B. subtilis,* thus, has a similar mode antimicrobial mode of action against gram-positive and gram-negative bacteria [10, 45]. An earlier report revealed that lichenicidin exhibited inhibitory activity in gram-positive bacteria, including *M. smegmatis* [44]. The mode of action of lichenicidin is forming pores in the cell wall and membrane and inhibiting peptidoglycan biosynthesis [46].

Microbial antagonism within an ecological niche is facilitated by a wide array of toxins, for instance, strains-specific bacteriocins and broad-spectrum antibiotics [42, 47]. In this regard, the

bacteria producing lethal antimicrobial agents have evolved self-protection mechanisms to avoid self-destruction, thus, elucidation of the mechanisms of resistance guides the discovery of putative bioactive products and their mode of action [48]. In this study, the antibiotic resistance mechanisms of *B. licheniformis* SAMN36381076 were classified into six categories, including antibiotic inactivation, antibiotic target modifying enzyme, antibiotic target protection protein, antibiotic target replacement protein, efflux pump, absence of a gene, and protein-altering cell wall charge (Table 5.5). These mechanisms are depended on the scaffolds of natural products, and their molecular targets. In the present study, three genes *BcrC*, *BceA*, and *BceB* present in *B. licheniformis* SAMN36381076 genome confer resistance to bacitracin [49]. Bacitracin is a cyclic dodecyl peptide known to be synthesized by *B. licheniformis* and is known to exhibit wide-spectrum antimicrobial activity [50]. An earlier study by Wu et al. [20] reported that *B. licheniformis* produced bacitracin A, which possesses antimicrobial activity against gram-positive bacteria. Interestingly, bacitracin was reported to exhibit inhibitory activity against mycothiol (MSH) and ergothioneine (ERG) thiol-deficient *M. tb* mutants [51]. Further, the results from this investigation identified the presence of genes that confer resistance to peptides that possess a mode of activity that is similar to daptomycin, and defensins. The BGCs identified showed the presence of genes that encode lipopeptides that also have mechanisms of action that target the cell membrane similar to daptomycin, for instance, fengycin, and lichenysin.

Further, the results of this study showed the presence of a BGC in *B. licheniformis* SAMN36381076 with seven percent homology to genes associated with the synthesis of aminoglycosides in the neomycin family (butirosin A and butirosin B) (Figure 5.3), hence the erythromycin resistance (Table 5.5). Aminoglycoside antibiotics are medicinally essential agents

that have been reported to inhibit *M. tb* by binding to the 30S subunit of the mycobacterial ribosome, disrupting the reading of genetic code and subsequently leading to death [18]. Numerous micro-stressors within an environmental niche like gold mine tailings enhance the selection of bacteria with multiple defense mechanisms [36]. In this study, multi-drug resistance genes were detected indicating a general resilience to chemical warfare within the mine tailings and also the ability to produce various antimicrobials with different mechanisms of action.

Earlier studies reported high concentrations of Cu, Cr, Cd, Pb, Sb, Ni, Zn, V, As, and sulfate in gold mine tailings and acid mine drainage [8, 52, 53]. Previous research revealed that *B. licheniformis* A6 was able to tolerate high concentrations of As, Co, Cd, Cr, Hg, Mn, Se, Pb, and Zn. Further, they demonstrated that the *B. licheniformis* A6 was able to oxidize arsenite into arsenate [54]. This study showed the presence of an *arsenite oxidation gene* which may be responsible for the oxidation of arsenite. Selenium (Se) is widespread in various rocks and can form complexes with iron (Fe) and manganese (Mn) oxides or hydroxides [55]. A previous study has reported that microbial reductive processes play an essential role in the removal of selenate from natural sediments. In the present study, the presence of a *ygfk* gene, which codes for ygfk reductase was identified suggesting the potential Se reduction by *B. licheniformis* SAMN36381076. In addition, nitrate reductases primarily have a high preference for the reduction of nitrate and have also been reported to exhibit the reduction of selenate.

Actinobacteria, proteobacteria, and Firmicutes are known to play critical roles in sulfate reduction and sulfur oxidation in various environmental niches [56]. Previous studies revealed the involvement of sulfate-reducing bacteria (SRB) in hydrocarbon bioremediation and the precipitation of heavy metals [53]. The *B. licheniformis* SAMN36381076 MAG contained sulfite

oxidation genes such as *sat*, *sqr*, and *sulfur_dioxygenase*. While *cysN* and *cysC* genes are responsible for assimilatory sulfate reduction, while *phsA* gene is involved in polysulfide reduction/thiosulfate disproportionation. Further, *soxB* was found in the genome, thus, indicating the potential of the bacteria in thiosulfate oxidation.

Denitrification and nitrification are essential processes in the nitrogen cycle and are facilitated by distinct enzymes encoded by functional genes within a specific group of microorganisms [57]. This study revealed the potential contribution of *B. licheniformis* SAMN36381076 MAG to nitrogen cycling in gold mine tailings. The *B. licheniformis* SAMN36381076 MAG contains several nitrogen metabolism genes, for instance, *narG*, which encodes nitrate reductase which catalyzes the conversion of nitrate (NO^{3-}) to nitrite (NO^{2-}). Further, the presence of *norB*, encoding nitric oxide reductase, highlights the ability of *B. licheniformis* SAMN36381076 to reduce NO^{2-} to nitric oxide (NO). In addition, *nirB* and *nirD* were present in the *B. licheniformis* SAMN36381076 MAG, and they encode nitrite reductase which catalyzes the conversion of NO^{2-} to nitrous oxide (N_2O) or N_2.

This study also investigated the metabolic potential of *B. licheniformis* SAMN36381076 MAG by analyzing the genes encoding carbohydrate-active enzymes (CAZYmes). In total, 41 CAZYme genes were identified, and categorized into different enzyme classes. These classes include polysaccharide lyases (18), carbohydrate esterases (6), glycosyltransferases (10), auxiliary activities (3) associated with polysaccharide and lignin degradation, and carbohydrate-binding modules (4). The presence of these CAZYmes suggests that *B. licheniformis* SAMN36381076 possesses a potential metabolic capacity for the biotransformation and recombination of lignin

derivatives, leading to the production of bioactive aromatic compounds. These compounds may have applications in the pharmaceutical industry [58, 59].

Conclusion:

The whole genome showed the global biosynthetic potential of *B. licheniformis* SAMN36381076. The bacteria is capable of producing potent broad-spectrum bioactive secondary metabolites that may be explore for their potential antimycobacterial activity.In addition, the bacteria may also be used for bioprocessing of lignin to produce novel aromatic compounds that may have a potential antimycobacterium activity.

Author Contributions: Conceptualization, V.M., and K.N.; methodology, K.N., V.M., M.N., K.G., L.S, and I.S.N.; formal analysis, K.N.; investigation, K.N.; resources, V.M., M.N., L.S., K.G., and I.SN.; writing original draft preparation, K.N.; writing review and editing, V.M., I.S.N., M.N., K.T., L.B., L.S., and FA.; supervision, V.M., M.N., K.G., L.S, and I.S.N; project administration, L.S.; funding acquisition, V.M., and L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the South African Medical Research Association Council (SAMRC), the Centre of Excellence for Biomedical Tuberculosis Research (CBTBR), and the National Research Fund (NRF) [NRF GRANT UID129364]. K.N. is highly indebted to the CTR and CBTBR for financial support.

Ethical Approval: Ethical approval for this study was approved by the Research Ethics Committee: Biological and Environmental Safety (REC: BES) of Stellenbosch University with reference number **BEE-2022-3188**.

5.6 References

1. Qian, L.; Lin, H.; Li, B.; Dong, Y. Physicochemical Characteristics and Microbial Communities of Rhizosphere in Complex Amendment-Assisted Soilless Revegetation of Gold Mine Tailings. *Chemosphere* **2023**, *320*, 138052, doi:10.1016/J.CHEMOSPHERE.2023.138052.

2. Santini, T.C.; Raudsepp, M.; Hamilton, J.; Nunn, J. Extreme Geochemical Conditions and Dispersal Limitation Retard Primary Succession of Microbial Communities in Gold Tailings. *Front Microbiol* **2018**, *9*, 2785, doi:10.3389/FMICB.2018.02785/BIBTEX.

3. Gagnon, V.; Gagnon, V.; Rodrigue-Morin, M.; Tremblay, J.; Wasserscheid, J.; Champagne, J.; Bellenger, J.P.; Greer, C.W.; Roy, S. Life in Mine Tailings: Microbial Population Structure across the Bulk Soil, Rhizosphere, and Roots of Boreal Species Colonizing Mine Tailings in Northwestern Québec. *Ann Microbiol* **2020**, *70*, 1–18, doi:10.1186/S13213-020-01582-9/FIGURES/8.

4. Waschulin, V.; Borsetto, C.; James, R.; Newsham, K.K.; Donadio, S.; Corre, C.; Wellington, E. Biosynthetic Potential of Uncultured Antarctic Soil Bacteria Revealed through Long-Read Metagenomic Sequencing. *The ISME Journal 2021 16:1* **2021**, *16*, 101–111, doi:10.1038/s41396-021-01052-3.

5. Courchesne, B.; Schindler, M.; Mykytczuk, N.C.S. Relationships Between the Microbial Composition and the Geochemistry and Mineralogy of the Cobalt-Bearing Legacy Mine Tailings in Northeastern Ontario. *Front Microbiol* **2021**, *12*, 1227, doi:10.3389/FMICB.2021.660190/BIBTEX.

6. Whaley-Martin, K.; Jessen, G.L.; Nelson, T.C.; Mori, J.F.; Apte, S.; Jarolimek, C.; Warren, L.A. The Potential Role of Halothiobacillus Spp. In Sulfur Oxidation and Acid Generation in Circum-Neutral Mine Tailings Reservoirs. *Front Microbiol* **2019**, *10*, 297, doi:10.3389/FMICB.2019.00297/BIBTEX.

7. Mardanov, A. V.; Beletsky, A. V.; Ivasenko, D.A.; Karnachuk, O. V.; Ravin, N. V. Metagenome Sequence of a Microbial Community from the Gold Mine Tailings in the Kuzbass Area, Russia. *Genome Announc* **2017**, *5*, doi:10.1128/GENOMEA.01355-17.

8. Qi, R.; Xue, N.; Wang, S.; Zhou, X.; Zhao, L.; Song, W.; Yang, Y. Heavy Metal(Loid)s Shape the Soil Bacterial Community and Functional Genes of Desert Grassland in a Gold Mining Area in the Semi-Arid Region. *Environ Res* **2022**, *214*, 113749, doi:10.1016/J.ENVRES.2022.113749.

9. Ji, H.; Zhang, Y.; Bararunyeretse, P.; Li, H. Characterization of Microbial Communities of Soils from Gold Mine Tailings and Identification of Mercury-Resistant Strain. *Ecotoxicol Environ Saf* **2018**, *165*, 182–193, doi:10.1016/J.ECOENV.2018.09.011.

10. Begley, M.; Cotter, P.D.; Hill, C.; Ross, R.P. Identification of a Novel Two-Peptide Lantibiotic, Lichenicidin, Following Rational Genome Mining for LanM Proteins. *Appl Environ Microbiol* **2009**, *75*, 5451–5460, doi:10.1128/AEM.00730-09/ASSET/643DCCBA-2523-494A-830C-F77265B4B7FC/ASSETS/GRAPHIC/ZAM0170902170003.JPEG.

11. Parkinson, E.I.; Tryon, J.H.; Goering, A.W.; Ju, K.S.; McClure, R.A.; Kemball, J.D.; Zhukovsky, S.; Labeda, D.P.; Thomson, R.J.; Kelleher, N.L.; et al. Discovery of the Tyrobetaine

Natural Products and Their Biosynthetic Gene Cluster via Metabologenomics. *ACS Chem. Biol.* **2018**, *13*, 1029–1037, doi:10.1021/acschembio.7b01089.

12. Miethke, M.; Pieroni, M.; Weber, T.; Brönstrup, M.; Hammann, P.; Halby, L.; Arimondo, P.B.; Glaser, P.; Aigle, B.; Bode, H.B.; et al. Towards the Sustainable Discovery and Development of New Antibiotics. *Nat Rev Chem 2021 5:10* **2021**, *5*, 726–749, doi:10.1038/s41570-021-00313-1.

13. Yao, T.; Liu, J.; Liu, Z.; Li, T.; Li, H.; Che, Q.; Zhu, T.; Li, D.; Gu, Q.; Li, W. Genome Mining of Cyclodipeptide Synthases Unravels Unusual TRNA-Dependent Diketopiperazine-Terpene Biosynthetic Machinery. *Nat Commun* **2018**, *9*, doi:10.1038/S41467-018-06411-X.

14. Zhou, L.; Song, C.; Li, Z.; Kuipers, O.P. Antimicrobial Activity Screening of Rhizosphere Soil Bacteria from Tomato and Genome-Based Analysis of Their Antimicrobial Biosynthetic Potential. *BMC Genomics* **2021**, *22*, 1–14, doi:10.1186/S12864-020-07346-8/FIGURES/6.

15. Jackson, S.A.; Crossman, L.; Almeida, E.L.; Margassery, L.M.; Kennedy, J.; Dobson, A.D.W. Diverse and Abundant Secondary Metabolism Biosynthetic Gene Clusters in the Genomes of Marine Sponge Derived Streptomyces Spp. Isolates. *Mar Drugs* **2018**, *16*, doi:10.3390/MD16020067.

16. Doroghazi, J.R.; Metcalf, W.W. Comparative Genomics of Actinomycetes with a Focus on Natural Product Biosynthetic Genes. *BMC Genomics* **2013**, *14*, 611, doi:10.1186/1471-2164-14-611.

17. Belknap, K.C.; Park, C.J.; Barth, B.M.; Andam, C.P. Genome Mining of Biosynthetic and Chemotherapeutic Gene Clusters in Streptomyces Bacteria. *Sci. Rep.* **2020**, *10*, 2003, doi:10.1038/s41598-020-58904-9.

18. Ogawara, H. Comparison of Antibiotic Resistance Mechanisms in Antibiotic-Producing and Pathogenic Bacteria. *Molecules 2019, Vol. 24, Page 3430* **2019**, *24*, 3430, doi:10.3390/MOLECULES24193430.

19. Agersø, Y.; Bjerre, K.; Brockmann, E.; Johansen, E.; Nielsen, B.; Siezen, R.; Stuer-Lauridsen, B.; Wels, M.; Zeidan, A.A. Putative Antibiotic Resistance Genes Present in Extant Bacillus Licheniformis and Bacillus Paralicheniformis Strains Are Probably Intrinsic and Part of the Ancient Resistome. *PLoS One* **2019**, *14*, e0210363, doi:10.1371/JOURNAL.PONE.0210363.

20. Wu, Z.; Li, Y.; Xu, Y.; Zhang, Y.; Tao, G.; Zhang, L.; Shi, G. Transcriptome Analysis of Bacillus Licheniformis for Improving Bacitracin Production. *ACS Synth Biol* **2022**, *11*, 1325–1335, doi:10.1021/ACSSYNBIO.1C00593/SUPPL_FILE/SB1C00593_SI_001.PDF.

21. Barale, S.S.; Ghane, S.G.; Sonawane, K.D. Purification and Characterization of Antibacterial Surfactin Isoforms Produced by Bacillus Velezensis SK. *AMB Express* **2022**, *12*, 1–20, doi:10.1186/S13568-022-01348-3/TABLES/3.

22. Nalli, Y.; Singh, S.; Gajjar, A.; Mahizhaveni, B.; Dusthackeer, V.N.A.; Shinde, P.B. Bacillibactin Class Siderophores Produced by the Endophyte Bacillus Subtilis NPROOT3 as Antimycobacterial Agents. *Lett Appl Microbiol* **2023**, *76*, doi:10.1093/LAMBIO/OVAC026.

23. Dimkic, I.; Stankovic, S.; Nišavic, M.; Petkovic, M.; Ristivojevic, P.; Fira, D.; Beric, T. The Profile and Antimicrobial Activity of Bacillus Lipopeptide Extracts of Five Potential Biocontrol Strains. *Front Microbiol* **2017**, *8*, 259438, doi:10.3389/FMICB.2017.00925/BIBTEX.

24. Lin, L.Z.; Zheng, Q.W.; Wei, T.; Zhang, Z.Q.; Zhao, C.F.; Zhong, H.; Xu, Q.Y.; Lin, J.F.; Guo, L.Q. Isolation and Characterization of Fengycins Produced by Bacillus Amyloliquefaciens JFL21 and Its Broad-Spectrum Antimicrobial Potential Against Multidrug-Resistant Foodborne Pathogens. *Front Microbiol* **2020**, *11*, 579621, doi:10.3389/FMICB.2020.579621/BIBTEX.

25. Quintero, M.; Blandón, L.M.; Vidal, O.M.; Guzman, J.D.; Gómez-Marín, J.E.; Patiño, A.D.; Molina, D.A.; Puerto-Castro, G.M.; Gómez-León, J. In Vitro Biological Activity of Extracts from Marine Bacteria Cultures against Toxoplasma Gondii and Mycobacterium Tuberculosis. *J Appl Microbiol* **2022**, *132*, 2705–2720, doi:10.1111/JAM.15397.

26. Su, Z.; Chen, X.; Liu, X.; Guo, Q.; Li, S.; Lu, X.; Zhang, X.; Wang, P.; Dong, L.; Zhao, W.; et al. Genome Mining and UHPLC–QTOF–MS/MS to Identify the Potential Antimicrobial Compounds and Determine the Specificity of Biosynthetic Gene Clusters in Bacillus Subtilis NCD-2. *BMC Genomics* **2020**, *21*, 1–16, doi:10.1186/S12864-020-07160-2/FIGURES/8.

27. Arkin, A.P.; Cottingham, R.W.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol 2018 36:7* **2018**, *36*, 566–569, doi:10.1038/nbt.4163.

28. Wattam, A.R.; Abraham, D.; Dalay, O.; Disz, T.L.; Driscoll, T.; Gabbard, J.L.; Gillespie, J.J.; Gough, R.; Hix, D.; Kenyon, R.; et al. PATRIC, the Bacterial Bioinformatics Database and Analysis Resource. *Nucleic Acids Res* **2014**, *42*, D581–D591, doi:10.1093/NAR/GKT1099.

29. Brown, J.; Pirrung, M.; Mccue, L.A. FQC Dashboard: Integrates FastQC Results into a Web-Based, Interactive, and Extensible FASTQ Quality Control Tool. *Bioinformatics* **2017**, *33*, 3137–3139, doi:10.1093/BIOINFORMATICS/BTX373.

30. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120, doi:10.1093/BIOINFORMATICS/BTU170.

31. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. MetaSPAdes: A New Versatile Metagenomic Assembler. *Genome Res* **2017**, *27*, 824–834, doi:10.1101/GR.213959.116.

32. Wu, Y.W.; Simmons, B.A.; Singer, S.W. MaxBin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets. *Bioinformatics* **2016**, *32*, 605–607, doi:10.1093/BIOINFORMATICS/BTV638.

33. Sieber, C.M.K.; Probst, A.J.; Sharrar, A.; Thomas, B.C.; Hess, M.; Tringe, S.G.; Banfield, J.F. Recovery of Genomes from Metagenomes via a Dereplication, Aggregation and Scoring Strategy. *Nat Microbiol 2018 3:7* **2018**, *3*, 836–843, doi:10.1038/s41564-018-0171-1.

34. Chklovski, A.; Parks, D.H.; Woodcroft, B.J.; Tyson, G.W. CheckM2: A Rapid, Scalable and Accurate Tool for Assessing Microbial Genome Quality Using Machine Learning. *bioRxiv* **2022**, 2022.07.11.499243, doi:10.1101/2022.07.11.499243.

35. Parks, D.H.; Chuvochina, M.; Waite, D.W.; Rinke, C.; Skarshewski, A.; Chaumeil, P.A.; Hugenholtz, P. A Standardized Bacterial Taxonomy Based on Genome Phylogeny Substantially Revises the Tree of Life. *Nat Biotechnol 2018 36:10* **2018**, *36*, 996–1004, doi:10.1038/nbt.4229.

36. Noman, S.M.; Shafiq, M.; Bibi, S.; Mittal, B.; Yuan, Y.; Zeng, M.; Li, X.; Olawale, O.A.; Jiao, X.; Irshad, M. Exploring Antibiotic Resistance Genes, Mobile Gene Elements, and Virulence Gene Factors in an Urban Freshwater Samples Using Metagenomic Analysis. *Environ Sci Pollut Res* **2023**, *30*, 2977–2990, doi:10.1007/S11356-022-22197-4/FIGURES/8.

37. Alcock, B.P.; Huynh, W.; Chalil, R.; Smith, K.W.; Raphenya, A.R.; Wlodarski, M.A.; Edalatmand, A.; Petkau, A.; Syed, S.A.; Tsang, K.K.; et al. CARD 2023: Expanded Curation, Support for Machine Learning, and Resistome Prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* **2023**, *51*, D690–D699, doi:10.1093/NAR/GKAC920.

38. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions. *Nucleic Acids Res* **2013**, *41*, e121–e121, doi:10.1093/NAR/GKT263.

39. Li, Y.; Lei, L.; Zheng, L.; Xiao, X.; Tang, H.; Luo, C. Genome Sequencing of Gut Symbiotic Bacillus Velezensis LC1 for Bioethanol Production from Bamboo Shoots. *Biotechnol Biofuels* **2020**, *13*, 1–12, doi:10.1186/S13068-020-1671-9/FIGURES/5.

40. Su, Z.; Liu, G.; Liu, X.; Li, S.; Lu, X.; Wang, P.; Zhao, W.; Zhang, X.; Dong, L.; Qu, Y.; et al. Functional Analyses of the Bacillus Velezensis HMB26553 Genome Provide Evidence That Its Genes Are Potentially Related to the Promotion of Plant Growth and Prevention of Cotton Rhizoctonia Damping-Off. *Cells* **2023**, *12*, 1301, doi:10.3390/CELLS12091301/S1.

41. Uche-Okereafor, N.; Sebola, T.; Tapfuma, K.; Mekuto, L.; Green, E.; Mavumengwana, V. Antibacterial Activities of Crude Secondary Metabolite Extracts from Pantoea Species Obtained from the Stem of Solanum Mauritianum and Their Effects on Two Cancer Cell Lines. *Int J Environ Res Public Health 2019, Vol. 16, Page 602* **2019**, *16*, 602, doi:10.3390/IJERPH16040602.

42. Tapfuma, K.I.; Nyambo, K.; Adu-Amankwaah, F.; Baatjies, L.; Smith, L.; Allie, N.; Keyster, M.; Loxton, A.G.; Ngxande, M.; Malgas-Enus, R.; et al. Antimycobacterial Activity and Molecular Docking of Methanolic Extracts and Compounds of Marine Fungi from Saldanha and False Bays, South Africa. *Heliyon* **2022**, *8*, e12406, doi:10.1016/J.HELIYON.2022.E12406.

43. Venugopala, K.N.; Tratrat, C.; Pillay, M.; Mahomoodally, F.M.; Bhandary, S.; Chopra, D.; Morsy, M.A.; Haroun, M.; Aldhubiab, B.E.; Attimarad, M.; et al. Anti-Tubercular Activity of Substituted 7-Methyl and 7-Formylindolizines and in Silico Study for Prospective Molecular Target Identification. *Antibiotics* **2019**, *8*, doi:10.3390/antibiotics8040247.

44. Abdelrahman, S.M.; Dosoky, N.S.; Hanora, A.M.; Lopanik, N.B. Metabolomic Profiling and Molecular Networking of Nudibranch-Associated Streptomyces Sp. SCSIO 001680. *Molecules* **2022**, *27*, 4542, doi:10.3390/MOLECULES27144542/S1.

45. Wang, M.; Carver, J.J.; Phelan, V. V; Sanchez, L.M.; Garg, N.; Peng, Y.; Duy Nguyen, D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. **2016**, doi:10.1038/nbt.3597.

46. da Silva Oliveira, J.P.; Garrett, R.; Bello Koblitz, M.G.; Furtado Macedo, A. Vanilla Flavor: Species from the Atlantic Forest as Natural Alternatives. *Food Chem* **2022**, *375*, 131891, doi:10.1016/J.FOODCHEM.2021.131891.

47. Guo, Y.S.; Tao, J.Z. Metabolomics and Pathway Analyses to Characterize Metabolic Alterations in Pregnant Dairy Cows on D 17 and D 45 after AI. *Sci Rep 2018 8:1* **2018**, *8*, 1–8, doi:10.1038/s41598-018-23983-2.

48. Kudo, F.; Eguchi, T. Biosynthetic Genes for Aminoglycoside Antibiotics. *J Antibio 2009 62:9* **2009**, *62*, 471–481, doi:10.1038/ja.2009.76.

49. Xia, L.; Miao, Y.; Cao, A.; Liu, Y.; Liu, Z.; Sun, X.; Xue, Y.; Xu, Z.; Xun, W.; Shen, Q.; et al. Biosynthetic Gene Cluster Profiling Predicts the Positive Association between Antagonism and Phylogeny in Bacillus. *Nat Commun 2022 13:1* **2022**, *13*, 1–11, doi:10.1038/s41467-022-28668-z.

50. Zhu, J.; Li, L.; Wu, F.; Wu, Y.; Wang, Z.; Chen, X.; Li, J.; Cai, D.; Chen, S. Metabolic Engineering of Aspartic Acid Supply Modules for Enhanced Production of Bacitracin in *Bacillus Licheniformis*. *ACS Synth Biol* **2021**, *10*, 2243–2251, doi:10.1021/acssynbio.1c00154.

51. Shenkarev, Z.O.; Finkina, E.I.; Nurmukhamedova, E.K.; Balandin, S. V.; Mineev, K.S.; Nadezhdin, K.D.; Yakimenko, Z.A.; Tagaev, A.A.; Temirov, Y. V.; Arseniev, A.S.; et al. Isolation, Structure Elucidation, and Synergistic Antibacterial Activity of a Novel Two-Component Lantibiotic Lichenicidin from *Bacillus Licheniformis* VK21. *Biochemistry* **2010**, *49*, 6462–6472, doi:10.1021/BI100871B/SUPPL_FILE/BI100871B_SI_001.PDF.

52. Théatre, A.; Cano-Prieto, C.; Bartolini, M.; Laurin, Y.; Deleu, M.; Niehren, J.; Fida, T.; Gerbinet, S.; Alanjary, M.; Medema, M.H.; et al. The Surfactin-Like Lipopeptides From Bacillus Spp.: Natural Biodiversity and Synthetic Biology for a Broader Application Range. *Front Bioeng Biotechnol* **2021**, *9*, 118, doi:10.3389/FBIOE.2021.623701/XML/NLM.

53. Barbosa, J.C.; Gonçalves, S.; Makowski, M.; Silva, Í.C.; Caetano, T.; Schneider, T.; Mösker, E.; Süssmuth, R.D.; Santos, N.C.; Mendo, S. Insights into the Mode of Action of the Two-Peptide Lantibiotic Lichenicidin. *Colloids Surf B Biointerfaces* **2022**, *211*, 112308, doi:10.1016/J.COLSURFB.2021.112308.

54. Han, J.; Liu, X.; Zhang, L.; Quinn, R.J.; Feng, Y. Anti-Mycobacterial Natural Products and Mechanisms of Action. *Nat Prod Rep* **2022**, *39*, 77–89, doi:10.1039/D1NP00011J.

55. Alanjary, M.; Kronmiller, B.; Adamek, M.; Blin, K.; Weber, T.; Huson, D.; Philmus, B.; Ziemert, N. The Antibiotic Resistant Target Seeker (ARTS), an Exploration Engine for Antibiotic Cluster Prioritization and Novel Drug Target Discovery. *Nucleic Acids Res* **2017**, *45*, W42–W48, doi:10.1093/NAR/GKX360.

56. Birošová, L.; Koščová, J.; Zhang, Y.; Lv, X.; Cao, W.; Zhang, H.; Shi, L.; Bai, W.; Ye, L. Survey of Colistin Resistance in Commensal Bacteria from Penaeus Vannamei Farms in China. *Foods 2023, Vol. 12, Page 2143* **2023**, *12*, 2143, doi:10.3390/FOODS12112143.

57. Cai, D.; Zhu, J.; Zhu, S.; Lu, Y.; Zhang, B.; Lu, K.; Li, J.; Ma, X.; Chen, S. Metabolic Engineering of Main Transcription Factors in Carbon, Nitrogen, and Phosphorus Metabolisms for Enhanced Production of Bacitracin in *Bacillus Licheniformis*. *ACS Synth Biol* **2019**, *8*, 866–875, doi:10.1021/ACSSYNBIO.9B00005/ASSET/IMAGES/LARGE/SB-2019-00005F_0007.JPEG.

58. Sao Emani, C.; Williams, M.J.; Wiid, I.J.; Baker, B.; Carolis, C. Compounds with Potential Activity against Mycobacterium Tuberculosis. *Antimicrob Agents Chemother* **2018**, *62*, doi:10.1128/AAC.02236-17.

59. Kiventerä, J.; Sreenivasan, H.; Cheeseman, C.; Kinnunen, P.; Illikainen, M. Immobilization of Sulfates and Heavy Metals in Gold Mine Tailings by Sodium Silicate and Hydrated Lime. *J Environ Chem Eng* **2018**, *6*, 6530–6536, doi:10.1016/J.JECE.2018.10.012.

60. Allard, O.; Lopez, M.; Demers, I.; Coudert, L. Gold Recovery from Sulfide Concentrates Produced by Environmental Desulfurization of Mine Tailings. *Minerals 2022, Vol. 12, Page 1011* **2022**, *12*, 1011, doi:10.3390/MIN12081011.

61. Sher, S.; Sultan, S.; Rehman, A. Characterization of Multiple Metal Resistant *Bacillus Licheniformis* and Its Potential Use in Arsenic Contaminated Industrial Wastewater. *Appl Water Sci* **2021**, *11*, 1–7, doi:10.1007/S13201-021-01407-3/FIGURES/3.

62. Hendry, M.J.; Biswas, A.; Essilfie-Dughan, J.; Chen, N.; Day, S.J.; Barbour, S.L. Reservoirs of Selenium in Coal Waste Rock: Elk Valley, British Columbia, Canada. *Environ Sci Technol* **2015**, *49*, 8228–8236, doi:10.1021/ACS.EST.5B01246/SUPPL_FILE/ES5B01246_SI_001.PDF.

63. Ayala-Muñoz, D.; Macalady, J.L.; Sánchez-España, J.; Falagán, C.; Couradeau, E.; Burgos, W.D. Microbial Carbon, Sulfur, Iron, and Nitrogen Cycling Linked to the Potential Remediation of a Meromictic Acidic Pit Lake. *The ISME Journal 2022 16:12* **2022**, *16*, 2666–2679, doi:10.1038/s41396-022-01320-w.

64. Luo, G.; Xue, C.; Jiang, Q.; Xiao, Y.; Zhang, F.; Guo, S.; Shen, Q.; Ling, N. Soil Carbon, Nitrogen, and Phosphorus Cycling Microbial Populations and Their Resistance to Global Change Depend on Soil C:N:P Stoichiometry. *mSystems* **2020**, *5*, doi:10.1128/MSYSTEMS.00162-20/SUPPL_FILE/MSYSTEMS.00162-20-SF005.TIF.

65. Varman, A.M.; He, L.; Follenfant, R.; Wu, W.; Wemmer, S.; Wrobel, S.A.; Tang, Y.J.; Singh, S. Decoding How a Soil Bacterium Extracts Building Blocks and Metabolic Energy from Ligninolysis Provides Road Map for Lignin Valorization. *Proc Natl Acad Sci U S A* **2016**, *113*, E5802–E5811, doi:10.1073/pnas.1606043113.

66. Vinardell, M.P.; Mitjans, M. Lignins and Their Derivatives with Beneficial Effects on Human Health. *Int J Mol Sci* 2017, *18*.

Chapter 6

Comparison of predictive machine learning-based quantitative structure-activity relationship models for targeting InhA, Pks13 and MptpA

Kudakwashe Nyambo[1], Precious Chiwira[6], Kudzanai Ian Tapfuma[1], Lucinda Baatjies[1], Idah Sithole Niang[2], Liezel Smith[1], Krishna Govender[4,5], Mkhuseli Ngxande[3], and Vuyo Mavumengwana[1,*]

[1] *DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town; KN: knyambo@sun.ac.za; K.I.T: kudzanait@su.ac.za; LB: lbaatjies@sun.ac.za; LS: liezels@sun.ac.za; VM: vuyom@sun.ac.za*

[2]*Department of Biotechnology and Biochemistry, University of Zimbabwe, B064, Mount Pleasant, Harare, Zimbabwe; sitholeidah2015@gmail.com*

[3]*Computer Science Division, Department of Mathematical Sciences, Faculty of Science University of Stellenbosch, Matieland, South Africa; ngxandem@sun.ac.za*

[4]*Department of Chemical Sciences, University of Johannesburg, P. O. Box 17011, Doornfontein Campus, 2028, Johannesburg, South Africa; krishnag@uj.ac.za*

[5]*National Institute for Theoretical and Computational Sciences (NITheCS), South Africa*

[6] *Mathematics Division, Department of Mathematical Sciences, Faculty of Science, University of Stellenbosch, Matieland South Africa, chiwirap@gmail.com*

***** Corresponding author.

Email address: *vuyom@sun.ac.za;* Tel: +27 718502949

6.1 Abstract

Predictive quantitative structure activity relationship (QSAR) models are increasingly employed in the drug discovery process, as they facilitate the screening of large libraries of compounds. These computational models complement traditional *in-vitro* or *in-vivo* assays, significantly accelerating the drug discovery process by increasing the probability of identifying promising lead

compounds. In this study, the focus was on developing and comparing models using both molecular descriptors and an alternative approach without handcrafted molecular descriptors. Predictive classification models based on pIC50 values were trained using molecular descriptors. Additionally, Long Short-Term Memory (LSTM) neural networks to learn the intrinsic features from the SMILES notation of compounds with pIC50 as the dependent variable. The results showed that logistic regression (LR) and multi-layer perceptron (MLP) emerged as the most promising models, exhibiting better performance metrics including recall, AUC ROC, RMSE, and accuracy. However, the LSTM model performed suboptimal (negative R^2 values), suggesting the necessity for hyperparameter tuning to enhance its predictive capabilities. In conclusion, machine learning-based QSAR architectures have significant potential to accelerate the discovery of anti-TB agents.

6.1 Introduction

Traditional experimental drug discovery is the most reliable way to identify therapeutic agents which satisfy the optimum biological and chemical properties, but it is tedious and cost-ineffective since the chemical space is vast [1]. On the other hand, the integration of computer-aided drug design (CADD), notably recognized as the fourth industrial revolution, has been paramount in cost-effectively accelerating drug discovery [2]. Machine learning based computational modules are gaining much attention in biomolecular drug discovery fields, for instance, prediction of molecular energy, transcriptomics, elucidation of reaction mechanisms, and molecular docking [3]. Fundamentally, the modules developed by machine learning carefully and efficiently address the challenges faced in conventional drug discovery by virtually screening large ensembles of

biological data, integrating heterogeneous biological data sets, using numerous statistical measures, and finally selecting unbiased significant predictions [4, 5].

Some supervised machine learning-guided algorithms are currently under development and deployment with the goal of identifying hit compounds from large and diverse datasets [6]. The primary objective of these algorithms is to predict well-defined endpoints that enable the discovery of compounds that target essential drug targets involved in bacterial and viral pathogenicity [7]. QSAR is a multidiscipline chemometrics technique, that aims to correlate compounds' properties with biological activities such as Ki, pIC50, pEC50, MIC, etc., or categorical biological properties [8]. This approach has been widely adopted by pharmaceutical companies to cost-effectively accelerate drug discovery [5]. For instance, Shahbaaz et al. [9] employed molecular dynamics simulations for the identification of inhibitors against the MmpS5-MmpL5 efflux pump, while Sadawi et al. [10] developed a multi-task QSAR (quantitative structure-activity relationship) tool for screening active compounds against different drug targets.

Numerous classical machine learning algorithms have been successfully implemented for developing QSAR models, for instance, support vector machine (SVMs), Naïve Bayesian classifiers, K-Nearest neighbors (KNN), Random Forest (RF), e.t.c [7, 11, 12]. Earlier studies have reported the development of models based on fixed molecular descriptors and fingerprints, that encode the chemical and physical properties of compounds [13]. Molecular descriptors are the "fingerprint" of one molecular on the micro-level, including theoretical and experimental parameters such as atom numbers; chemical bonds numbers; MLI (molecular connectivity index); ionization constant, pKa; electric dipole moment μ; MR (molecular molar refractive index), logP (lipophilic parameter), etc [13, 14]. Molecular descriptors play an essential role in the formulation

of robust and intuitive models that may be significant in drug discovery. Another study revealed the occurrence of false negative results in a multitarget QSAR investigation [14]. This highlighted a major limitation in using encoded handcrafted features and descriptors, as they often fail to capture some molecular structural properties that contribute to the antimicrobial activity of compounds. Deep learning-based architectures, for instance, Recurrent Neural Networks (RNN), Long Short-Term Memory, Convolutional Neural Networks (CNN), and Transformers are currently being developed and implemented to address that drawback. A recent study has demonstrated the efficiency of Deep Neural Networks at predicting hit compounds with desired chemical properties [7]. A significant advantage posed by deep learning-based architectures in QSAR modeling is the extraction of statistical dependencies directly from chemical scaffolds represented as SMILES notations, 3D images, or chemical graphs [3, 15].

Despite the rapid attention and influence of virtual screening tools in *Mycobacterium tuberculosis* (*M. tb*) research, several limitations persist. Some essential target proteins that confer virulence of the *M. tb*, such as MptpA and Pks13, have relatively small datasets, and some datasets lack crucial physiochemical information needed for designing novel and potent *M. tb* inhibitors. In this study, machine learning classifiers were utilized to develop single-target predictive models using the open-source Scikit-learn ML Python library (for training and prediction). The dataset's SMILES representation was augmented, and Long Short-Term Memory Networks (LSTMs) were employed to learn intrinsic features directly from the SMILES without relying on fixed descriptors and fingerprints (for training and prediction) [16].

6.3 Materials and methods

6.3.1 Experimental Dataset Retrieving

To build a meaningful predictive model, the datasets should have more than five samples representing both inactive and active compounds. The datasets for virtual screening were retrieved from the ChEMBL bioassay protocol, which comprises a comprehensive list of HTS bioassays involving the target *M. tb* H37Rv strain. The concentration required to cause 50% inhibition (IC_{50}) values were assigned as the criterion to classify the bioactivity of chemical scaffolds against desired target proteins. A python script was used to extract all the compounds and inhibition data from ChEMBL that satisfy the following criteria: Bioassays reporting experimental results for *M. tb* target proteins (InhA, mPtpA, and Pks13), in the form of IC_{50} values [1, 13]. The dataset was extracted and saved into a CSV file format using a proprietary script. The retrieved datasets were curated as described by Fourches et al. [17]. Briefly, the datasets were curated by removing compounds with missing features such as SMILES, IC_{50} (unit of activity), and duplicates.. The compounds in the dataset were annotated as either active if the IC_{50} value is $\leq$ 1000 nM or inactive if $\geq$ 10, 000 nM for all the selected protein targets. The criteria used for selecting the cut-off values are as follows: 1. Sub-micromolar range which ensures that the potent hits are rigorously searched 2. Prevent any excessive imbalance between the number of molecules assigned as active against those labeled as inactive [1]. The compounds' structures were standardized using MolStandardize SMARTS-based functionality in RDKit, to ensure rigorous deduplication, performance measures, consistent descriptor generation, and preserving stereochemistry. Some of the filters implemented include removing salt/solvent components; neutralizing compounds, removal of duplicate compounds, and Canonical SMILES were then formalized into specific tautomers using RDKit.

The IC_{50} values were converted to PIC_{50}. The SMILES were processed by standardizing them into canonical SMILES strings and those that failed to convert were discarded. RDKit Salt Stripper was used to desalt input compounds (that is, remove the salt/solvent components). The input molecules were neutralized by addition/ removal of hydrogen atoms. Only one stereoisomers in the class was kept because they were regarded as duplicates. Canonical SMILES were then formalized into specific tautomers using RDKit. The IC_{50} values were converted to PIC_{50}.

6.3.2 Feature engineering and selection

In building QSAR models, chemicals are encoded into a set of molecular descriptors. The descriptors of each compound were computed as described by [13]. Briefly, 12 different classes of molecular descriptors were considered, which include AtomPairs 2D count, AtomPairs2D, CDK graph only, Chemistry Development Kit (CDK) fingerprinter, CDK extended E-state, Klekota-Roth count, Klekota-Roth, Molecular ACCess System (MACCS), PubChem, substructure count, and substructure. A total of 16,095 molecular features were obtained from the InhA library and underwent pre-processing, which involved removing 15 % of the features with less variance. In the resultant dataset, highly correlated features with an r^2 greater than 0.8 were grouped together, and only the first feature from each group was retained for model development. Significant features were selected and noise reduction was applied during feature selection using the recursive feature elimination with cross-validation (RFECV) and a logistic regression estimator. The RFECV method with logistic regression as the estimator and 5-fold cross-validation. A total of 18 features were selected to be significant for model development.

6.3.3 Model development

A total of 18 types of features including (FP93, FP97, FP182, FP250, FP464, FP536, FP579, FP868, ExtFP33, ExtFP99, ExtFP405, ExtFP511, ExtFP566, ExtFP630, ExtFP638, ExtFP710, GraphFP787, and MACCSFP133) were selected as input for all the machine learning classifiers. The dataset was divided into a 75 % training set and a 25 % test set. Machine learning models were trained using six different supervised classifiers Random Forest (RF), decision tree (DT), support vector machines (SVC), KNeighbors (KNN), logistic regression (LR), and multi-layer perceptron (MLP) implemented in the scikit-learn package. The machine learning models were developed in Anaconda Navigator software (v 2022.05) in Jupyter Notebook IDE (v 6.4.8) with Python 3.10.4 64-bit. The Python modules which were used include pandas (v 1.5.2). Given the setbacks posed by the imbalance between the number of active and inactive compounds in mPtpA and Pks13 datasets, a virtual library augmentation was performed as described previously [18] using Python script available on https://github.com/Ebjerrum/SMILES-enumeration. Thereafter LSTM was used to develop a predictive regression model using the augmented datasets using the following parameters: LSTM layer with 32 neurons, dropout rate of 0.5, L1 and L2 regularization of 0.005 and 0.01, respectively, linear activation for the Dense layer, MSE is the loss function used, RMSprop optimizer with a learning rate of 0.005 was used for optimizing the model, and ReduceLROnPlateau callback for reducing the learning rate of the optimizer [18].

6.3.4 Model validation

The predictivity of the models was evaluated using the confusion matrix, which comprises the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) samples. The accuracy of the models was then computed for the test and train sets (**1**). The receiver

operating curve (ROC) AUC was also estimated for the classification models. The standard deviation of the residuals (*RMSE*) was also computed (**1**). The correlations (R^2 values) were computed for the LSTM regression model.

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(\widehat{Yi} - Yi)^2}\ /n \quad (\mathbf{1})$$

$$\text{Accuracy} = (TP + TN)/ (TP + TN + FP + FN) \quad (\mathbf{2})$$

$$\text{F1–score} = 2P/ (2T + FP + FN \frac{TP.TN - TP.FN}{\sqrt{[(TP+FP)(TP+FN)(TN+FP)(TN+FN)]}} \quad (\mathbf{3})$$

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives}) \quad (\mathbf{4})$$

6.4 Results and Discussion

6.4.1 Chemical space analysis

QSAR models have gained a lot of attention in the development of therapeutic drugs with the aim of predicting the bioactivity of compounds against macromolecular druggable targets. These models are known to reliably predict measured endpoints, thus, are useful in virtual screening and optimizing the pharmacokinetics of compounds in drug design [19]. The distribution of active and inactive compounds in InhA, Pks13, and MptpA was visualized before the model because the distribution influences the model's predictability. The two datasets Pks13 and MptpA had a highly unbalanced ratio of active to inactive compounds Figure 6.1. Thus, only InhA dataset was used for the development of classification models.

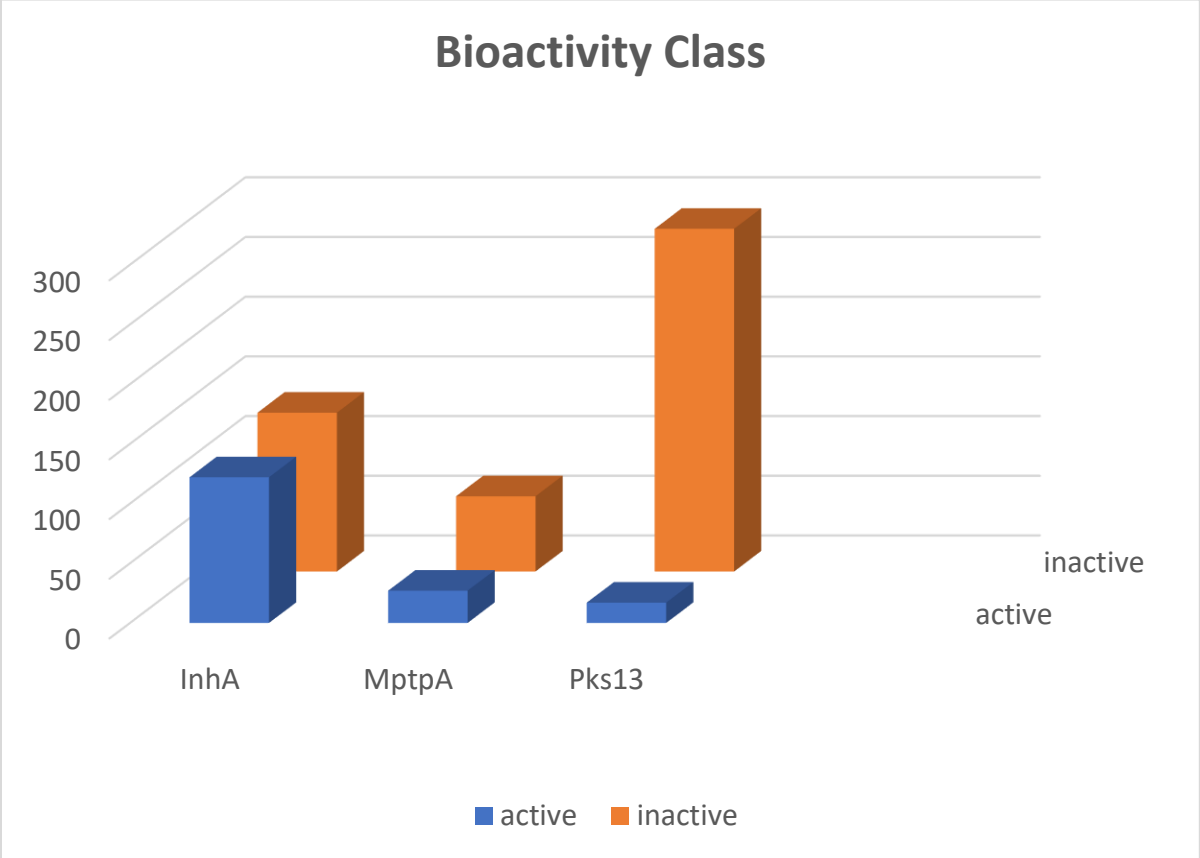


Figure 6.1. Exploration data analysis of the distribution of the chemical space of: **A.** Pks13 **B.** MptpA

Lipinski's rule-of-five (Ro5) descriptors are widely used in drug discovery to describe the drug-likeness based on the following parameters: octanol-water partition coefficient (log P; <5), number of hydrogen bond acceptors (<10), number of hydrogen bond donors (>5) and molecular weight (<500). In this study, the activity class (pIC50) was plotted against the Ro5 descriptors to explore the chemical space of the entire InhA. The results in this study agree with studies that have demonstrated that the Ro5 does not significantly contribute to differentiating between active and inactive, that is macromolecular target-ligand relationship (Figure 6.2.A, Figure 6.2.B, Figure 6.2.D). Ro5 may be used as one of the filtering parameters as it has been successfully implemented in describing most of the drugs in clinical use.

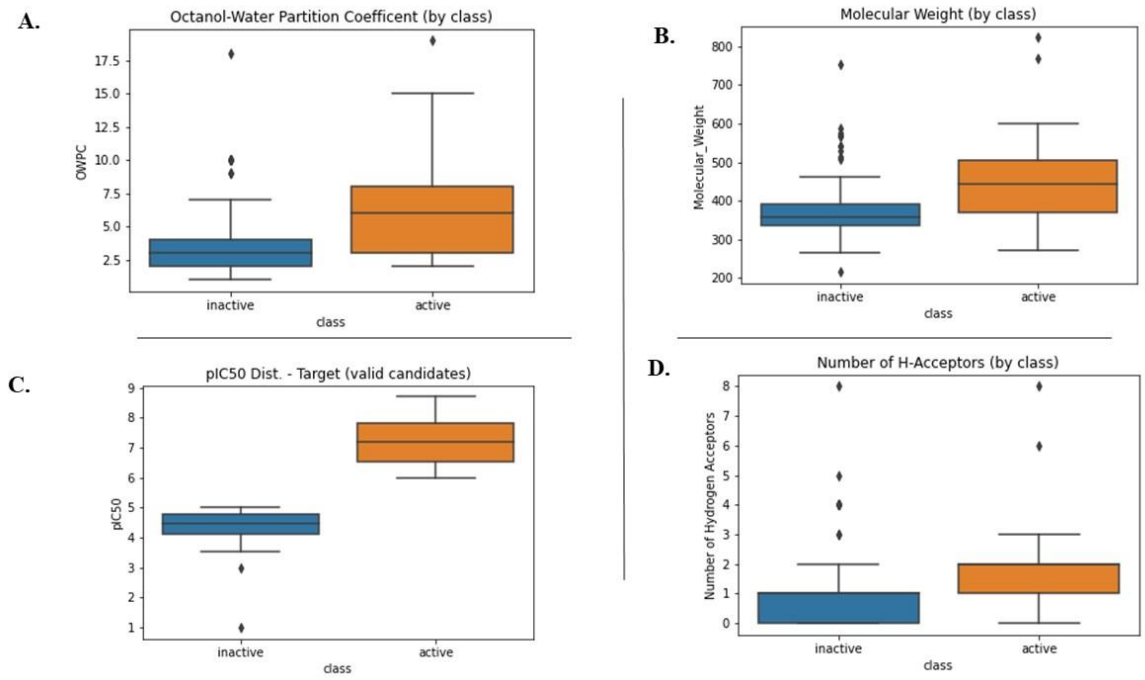


Figure 6.2. Exploratory data analysis of the chemical space of the target InhA. **A.** octanol–water partition coefficient (log P; < 5). **B.** molecular weight (<500). **C.** Distribution of pIC_{50} values for the InhA dataset. **D**. number of hydrogen bond acceptors (<10).

6.4.2 Model evaluation

A comparison of machine learning algorithms reveals that the RF and DT models had a consistent performance on both the train and test sets, which suggests that they are generalizing well and not overfitting. Likewise, the SCV showed the same RMSE as RF and RT on the test. The LR and MLP exhibited the lowest RMSE for the test, implying that they are good at classifying between active and inactive values.

Table 6.1. Comparison of the RMSE the classification models

Models Structure

Set	RF	DT	SVC	KNN	LR	MLP
Train	0.0340	0.0340	0.0454	0.0681	0.0511	-
Test	0.0339	0.0339	0.0339	0.0508	0.0169	0.0169

The performance and generalization of the classification models were evaluated on the train set and test. A summary of the statistical performance of the six machine learning models based on Recall, F1-score, Precision, AUC ROC, and Accuracy is shown in Table 6.2. Based on these results, it was observed that all the models performed well on both the train and test sets. It is important to note that the test set results are relatively close to the train set results, indicating that the models are generalizing well and are not overfitting. The Random Forest, Decision Tree, and Logistic Regression models exhibit high accuracy and F1-score, indicating good overall performance. The Multi-Layer Perceptron exhibited high accuracy, F1-score, and AUC ROC, implying that it can strongly classify between active and inactive capabilities (Table 6.2). In addition, LR is the best model based on the trade-off of all metrics used including Recall, precision, AUC ROC, and accuracy (Table 6.2).

Table 6.2. Comparison of the performance of six models based on AUC-ROC, precision, accuracy, F1-score, and Recall.

	Models	Recall	F1-score	precision	AUC-ROC	accuracy
RF						
	Train set	0.9780	0.9673	0.95698	0.9655	0.9659
	Test set	0.97	0.97	0.97	0.9660	0.9661
DT						
	Train set	0.956	0.9663	0.988505	0.9666	0.9659

	Test set	0.97	0.97	0.97	0.9660	0.9661
KNN						
	Train set	0.956	0.9355	0.9158	0.93096	0.9318
	Test set	0.95	0.95	0.95	0.9464	0.9491
LR						
	Train set	0.945	0.9503	0.9556	0.9489	0.9488
	Test set	0.98	0.98	0.98	0.9821	0.9830
MLP	-	-	-	-	-	-
	Test set	0.9354	0.9508	0.9667	0.9942	0.9830
SCV						
	Train set	0.9450	0.9556	0.9663	0.9548	0.9545
	Test set	0.97	0.97	0.97	0.9642	0.9661

In this study, all the datasets, particularly those for Pks13 and MptpA, are small in size and skewed, where active compounds are heavily underrepresented compared to inactive ones, as depicted in Figure 6.1. Thus, if models are developed from these two datasets they will lack robustness and may be biased toward the majority class (inactive compounds). To address this challenge, the SMILES representation of compounds within all datasets was augmented. An LSTM architecture was employed to develop a predictive regression model using the augmented datasets. However, the performance of the model was found to be unsatisfactory, as shown by the negative R^2 values (Table 6.3 and Figure 6.3). The poor performance indicates that the model needs extensive hyperparameter tuning to obtain satisfactory results. A study by Bjerrum (2017) reported the initial failure of a similar model before extensive hyperparameter tuning [18].

Table 6.3. Evaluation metrics of the multitarget LSTM models

	Models	**RMSE**	**R^2**
InhA			
	Train set	0.2013	-0.3140

	Test set	0.1863	-0.4529
Pks13			
	Train set	0.3449	-0.6664
	Test set	0.2896	-0.9904
MptpA			
	Train set	0.1131	-0.0080
	Test set	0.0982	-0.0707

6.6. Additional challenges and future applications

Small and skewed datasets for specific unexplored drug targets are common and they have a negative impact on the performance of machine learning models, especially deep learning models. LSTM and similar deep learning architectures are well-reported in capturing intrinsic patterns in data, but they require large amounts of data (compounds), to perform optimally. In this study, data augmentation did not yield meaningful results. This may be because the model requires extensive hyperparameter tuning. It is important to note that this study serves as a baseline from which further optimization, improvement, domain applicability assessment, and external validation of the models can be performed. However, the LR and MLP were promising as they showed good performance metrics.

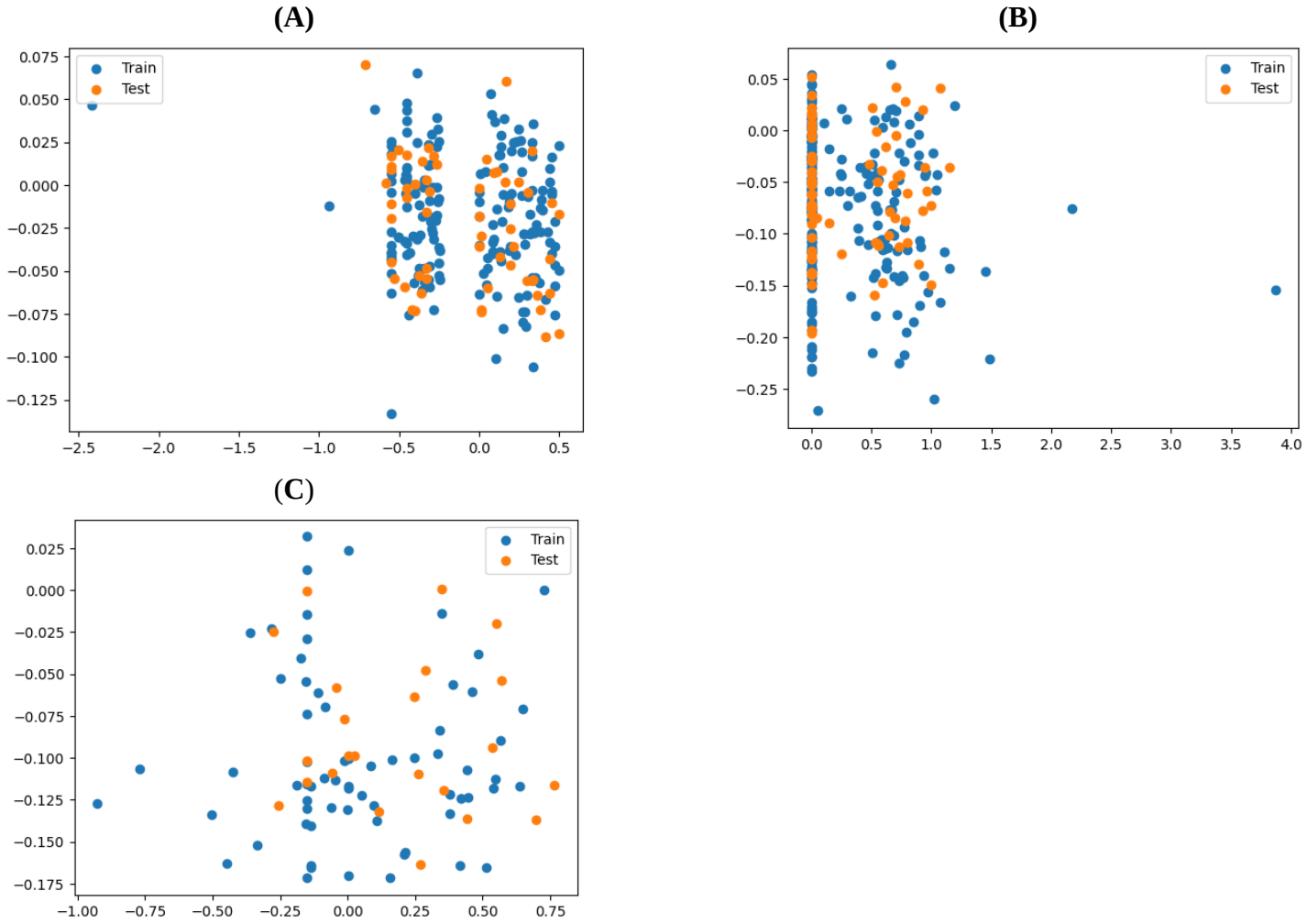


Figure 6.3. Regression plots for the augmented datasets (A) InhA, (B) shows Pks13, and (C) shows MptpA.

6.5 References

1. Kleandrova, V. v.; Scotti, L.; Bezerra Mendonça Junior, F.J.; Muratov, E.; Scotti, M.T.; Speck-Planche, A. QSAR Modeling for Multi-Target Drug Discovery: Designing Simultaneous Inhibitors of Proteins in Diverse Pathogenic Parasites. *Front Chem* **2021**, *9*, doi:10.3389/fchem.2021.634663.

2. Medina-Franco, J.L. Grand Challenges of Computer-Aided Drug Design: The Road Ahead. *Front Drug Discov* **2021**, *1*, 728551, doi:10.3389/FDDSV.2021.728551.

3. Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for de Novo Drug Design. **2018**.

4. Soufan, O.; Ba-Alawi, W.; Magana-Mora, A.; Essack, M.; Bajic, V.B. DPubChem: A Web Tool for QSAR Modeling and High-Throughput Virtual Screening. *Sci Rep* **2018**, *8*, doi:10.1038/s41598-018-27495-x.

5. Plisson, F.; Ramírez-Sánchez, O.; Martínez-Hernández, C. Machine Learning-Guided Discovery and Design of Non-Hemolytic Peptides. *Sci Rep* **2020**, *10*, doi:10.1038/s41598-020-73644-6.

6. Carrizosa, E.; Romero Morales, D. Supervised Classification and Mathematical Optimization. *Comput Oper Res* **2013**, *40*, 150–165, doi:10.1016/J.COR.2012.05.015.

7. Tsou, L.K.; Yeh, S.H.; Ueng, S.H.; Chang, C.P.; Song, J.S.; Wu, M.H.; Chang, H.F.; Chen, S.R.; Shih, C.; Chen, C.T.; et al. Comparative Study between Deep Learning and QSAR Classifications for TNBC Inhibitors and Novel GPCR Agonist Discovery. *Sci Rep* **2020**, *10*, doi:10.1038/s41598-020-73681-1.

8. Neves, B.J.; Braga, R.C.; Melo-Filho, C.C.; Moreira-Filho, J.T.; Muratov, E.N.; Andrade, C.H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front Pharmacol* **2018**, *9*, 418940, doi:10.3389/FPHAR.2018.01275/BIBTEX.

9. Shahbaaz, M.; Nkaule, A.; Christoffels, A. Designing Novel Possible Kinase Inhibitor Derivatives as Therapeutics against Mycobacterium Tuberculosis: An in Silico Study. *Sci Rep* **2019**, *9*, doi:10.1038/s41598-019-40621-7.

10. Sadawi, N.; Olier, I.; Vanschoren, J.; Van Rijn, J.N.; Besnard, J.; Bickerton, R.; Grosan, C.; Soldatova, L.; King, R.D. Multi-Task Learning with a Natural Metric for Quantitative Structure Activity Relationship Learning. *J Cheminform* **2019**, *11*, doi:10.1186/s13321-019-0392-1.

11. Chaube, S.; Goverapet Srinivasan, S.; Rai, B. Applied Machine Learning for Predicting the Lanthanide-Ligand Binding Affinities. *Sci Rep* **2020**, *10*, doi:10.1038/s41598-020-71255-9.

12. Ortiz, C.L.D.; Completo, G.C.; Nacario, R.C.; Nellas, R.B. Potential Inhibitors of Galactofuranosyltransferase 2 (GlfT2): Molecular Docking, 3D-QSAR, and In Silico ADMETox Studies. *Sci Rep* **2019**, *9*, doi:10.1038/s41598-019-52764-8.

13. Suvannang, N.; Preeyanon, L.; Malik, A.A.; Schaduangrat, N.; Shoombuatong, W.; Worachartcheewan, A.; Tantimongcolwat, T.; Nantasenamat, C. Probing the Origin of Estrogen Receptor Alpha Inhibition via Large-Scale QSAR Study †. **2018**, doi:10.1039/c7ra10979b.

14. Wei, Y.; Li, W.; Du, T.; Hong, Z.; Lin, J. Targeting HIV/HCV Coinfection Using a Machine Learning-Based Multiple Quantitative Structure-Activity Relationships (Multiple QSAR) Method. *International Journal of Molecular Sciences Article Int. J. Mol. Sci* **2019**, *20*, 3572, doi:10.3390/ijms20143572.

15. Anwar, M.U.; Adnan, F.; Abro, A.; Khan, M.R.A.; Rehman, A.U.; Osama, M.; Javed, S.; Baig, A.; Shabbir, M.R.; Assir, M.Z. Combined Deep Learning and Molecular Docking Simulations Approach Identifies Potentially Effective FDA Approved Drugs for Repurposing against SARS-CoV-2. *ChemRxiv* 2020.

16. Chakravarti, S.K.; Alla, S.R.M. Descriptor Free QSAR Modeling Using Deep Learning With Long Short-Term Memory Neural Networks. *Front Artif Intell* **2019**, *2*, 468526, doi:10.3389/FRAI.2019.00017/BIBTEX.

17. Fourches, D.; Muratov, E.; Tropsha, A. Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research., doi:10.1021/ci100176x.

18. Bjerrum, E.J. SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules.

19. Kelleci, F.; Ç, ¸; Karaduman, G.L. Machine Learning-Based Prediction of Drug-Induced Hepatotoxicity: An OvA-QSTR Approach. *J Chem Inf Model* **2023**, doi:10.1021/ACS.JCIM.3C00687.

Chapter 7

General conclusion, limitations, and future directions

7.1 Conclusion from this study

This research aimed to integrate multi-disciplinary strategies to accelerate the discovery of agents for addressing the challenges posed by *Mycobacterium tuberculosis* (*M. tb*). Many studies have demonstrated that natural products from plants, bacteria, and fungi possess antimicrobial and anticancer activities. However, the traditional top-down approach of screening and identifying compounds is time-consuming. In this study, crude extracts from bacteria and plants showing *in-vitro* antimycobacterial activity were tentatively identified. Subsequently, the compounds were virtually screened against essential *M. tb* targets.

Notably, the compounds verticillin A were identified as potential molecular starting points for developing inhibitors and PknD, respectively. Additionally, vazabitide A, tentatively identified from bacteria found in gold mine tailings, was structurally modified *in-silico* using multi-synthetic routes to create a potent analog with a strong affinity to PknG. These findings demonstrate the efficient use of *in-silico* approaches to tailor and discover new analogs for further *in-vitro* and *in-vivo* evaluation.

The study also focused on the metabolic capabilities of *Bacillus licheniformis* (*B. licheniformis*) SAMN36381076, which exhibited antimycobacterial activity, in an attempt to expand the chemical space. The results revealed that *B. licheniformis* SAMN36381076 possesses biosynthetic gene

clusters associated with the production of broad-spectrum antibiotics. Furthermore, the metabolic potential of *B. licheniformis* SAMN36381076 can be harnessed in the biotransformation of lignin to produce bioactive compounds with therapeutic potential.

In addition to natural products, a machine-learning-based predictive quantitative structure-activity relationship (QSAR) model was developed to screen compounds with activity against InhA, Pks13, and MptpA. The results showed that classic machine learning algorithms like logistic regression and multi-layer perceptron exhibited good performance and strong discriminatory capabilities. Machine learning based QSAR models can significantly accelerate targeted drug discovery.

Limitations

The linking of natural products synthesised by proteins encoded by biosynthetic gene clusters, such as Type II PKSs and terpene synthases, is a hurdle due to the lack of collinearity between the genes and the products. Another limitation is some compounds are produced at very low concentrations, which cannot be detected by the analytic methods. A total of 8 bacterial samples had 16 S rRNA sequences that had good quality, thus were used for constructing a phylogenetic tree. The datasets from Pks13 and MptpA were small, thus creating a machine learning model based on that data will be biased towards the predominant class. In this study, enumeration of the two datasets was performed but the dataset was still small for deep learning architecture.

Future directions

The study demonstrated efficacy and affinity of compounds from bacteria and plants in a crude mixture. *In-vitro* whole cell and protein validation of the pure compounds identified using virtual screening methods will be performed. The logistic regression and multi-layer perceptron models

constructed in the study may be further validation and used for predictive quantitative structure activity relationship classification tasks. When antimicrobial agents attack *M.tb*, it attempts to use multiple compensatory mechanisms to maintain fitness. When these compensatory mechanisms are insufficient, the tubercule only experiences growth retardation or death. Thus, incorporating DualSeq technologies will accelerate drug discovery by enabling researchers to precisely disentangle the genome-wide transcriptome dynamics between the host and bacteria pathogens simultaneously after treatment. Integrating data from transcriptome and functional profiles of *M.tb* after exposure to drugs with *in-silico* approaches such as molecular dynamics simulations and free-energy perturbation (FEP+) simulations increases the chances of discovering novel targets and new mechanisms of action. The combination of metabologenomic data from *B. licheniformis* SAMN36381076 will increases the number of novel bioactive secondary metabolites that can be identified.

Appendix A

Molecular docking, molecular dynamics simulations and binding free energy studies of interactions between *Mycobacterium tuberculosis* Pks13, PknG and bioactive constituents of extremophilic bacteria

Table A1. Tentatively identified metabolites present in bacterial extracts.

Sample	Classification	Formula	Tentative Identification	Precursor m/z $[M+H]^+$	m/z Error (ppm)
S. mycarofaciens					
	Alpha amino acids and derivatives	$C_7H_{10}N_2O_2$	cyclo-L-Prolylglycine	155.0811	0.000404
	Alpha amino acids and derivatives	$C_{14}H_{16}N_2O_3$	Maculosin	261.1234	0.8334
	Aminocyclitol glycosides	$C_{19}H_{37}N_5O6$	Istamycin C1	432.2802	0.7802
	Alpha amino acids and derivatives	$C_{11}H_{18}N_2O_2$	L,L-Cyclo(leucylprolyl)	211.1437	0.8737
	Alpha amino acids and derivatives	$C_{14}H_{16}N_2O_2$	cyclo-(L-Phe-L-Pro)	245.1283	
		$C_9H_{11}NO_3$	unknown	182.0815	-1.8241
	Alpha amino acids and derivatives	$C_{11}H_{14}N_4O_2$	NPA006809	235.1191	-0.6317
		$C_{11}H_{16}N_2O_2$	Unknown	209.1278	3.1434
	Organic acids and derivatives	$C_{12}H_{21}N_3O_4$	Vazabitide A	272.1614	-3.3833
	Glycosides	$C_{12}H_{13}N_5O_3$	5’- deoxytoyocamycin	276.1082	3.3286
	Organoheterocyclic	$C_{15}H_{21}N_3O_2$	Physostigmine	276.1714	-2.7133
		$C_{14}H_{25}NO_6$	unknown	304.1762	-2.4277
	Organic acids and derivatives	$C_{10}H_{18}N_2O_3$	Geralcin E	215.1393	-1.3127
	Lipids and lipid-like molecules	$C_{11}H_{20}N_2O_3$	α-Methyldethiobiotin	229.1558	-4.9573
	Alpha amino acids and derivatives	$C_{10}H_{16}N_2O_2$	cyclo(L-Pro-L-Val)	197.1294	-4.8225

	Alpha amino acids and derivatives	$C_{11}H_{18}N_2O_3$	Cyclo(2-hydroxy-Pro-R-Leu)	227.14	-4.3386
	Alpha amino acids and derivatives	$C_{14}H_{16}N_2O_3$	Cyclo(D-Pro-L-Tyr)	261.1239	-2.0422

Table A1. Tentatively identified metabolites present in bacterial extracts. Continued.

Sample	Classification	Formula	Tentative Identification	Precursor m/z $[M+H]^+$	m/z Error (ppm)
S. mycarofaciens					
	Alpha amino acids and derivatives	$C_{11}H_{18}N_2O_2$	Cyclo(D)-Pro-(D)-Ile	211.1447	-2.8348
	Alpha amino acids and derivatives	$C_{14}H_{16}N_2O_2$	Cyclo(D)-Pro-(D)-Phe	245.1285	-0.1876
	Alpha amino acids and derivatives	$C_{11}H_{14}N_4O_2$	JBIR-75	235.1191	0.8691
	Alpha amino acids and derivatives	$C_{19}H_{36}N_2O_5$	Lipoxamycin	373.2704	0.7704
	Alpha amino acids and derivatives	$C_{10}H_{16}N_2O_2$	2-amino-N-(2'-(cyclohex-2''-enyl)acetyl)acetimide	197.1291	0.8791
B. subtilis					
	Alpha amino acids and derivatives	$C_{11}H_{18}N_2O_3$	Cyclo-(L-Pro-4-OH-L-Leu)	227.1387	1.4102
	Alpha amino acids and derivatives	$C_{11}H_{18}N_2O_2$	Cyclo(proline-leucine)	211.1436	0.0005043
	Alpha amino acids and derivatives	$C_7H_{10}N_2O_2$	cyclo-L-Prolylglycine	155.0812	0.000304
	Indoles	$C_{11}H_9NO_2$	Indole-3-acrylic acid	188.0704	0.000205
	Alpha amino acids and derivatives	$C_{10}H_{16}N_2O_2$	cyclo(L-Pro-L-Val)	197.128	0.0004542
	Alpha amino acids and derivatives	$C_{14}H_{16}N_2O_2$	cyclo-(L-Phe-L-Pro)	245.128	1.8606

	Cyclic lipopeptide	$C_{76}H_{117}N_{11}O_{20}$	Plipastatin	1505.94	4.5387
	Cyclic depsipeptides	$C_{51}H_{89}N_7O_{13}$	Surfactin A	1008.658	3.9654

Table A1. Tentatively identified metabolites present in bacterial extracts. Continued.

Sample	Classification	Formula	Tentative Identification	Precursor m/z $[M+H]^+$	m/z Error (ppm)
B. subtilis					
	Cyclic lipopeptide	$C_{74}H_{114}N_{12}O_{20}$	Plipastatin	1492.4058	0.6058
	Cyclic depsipeptides	$C_{51}H_{89}N_7O_{13}$	Surfactin A	1008.658	3.9654
	Cyclic depsipeptides	$C_{53}H_{95}N_7O_{14}$	Gageostatin B	1054.704	-2.8688
	Cyclic depsipeptides	$C_{52}H_{91}N_7O_{13}$	Surfactin B	1022.676	4.1201
	Cyclic depsipeptides	$C_{53}H_{94}N_8O_{12}$	Surfactin C1	1036.69	0.3245
	Cyclic depsipeptides	$C_{53}H_{89}N_{13}O_8$	Surfactin D	1036.703	-0.0161
(***B. licheniformis***)					
	Valine and derivatives	$C_5H_{11}NO_2$	L-Valine	118.0858	3.8863
	Alpha amino acids and derivatives	$C_{11}H_{19}N_3O_2$	Unknown	226.1555	-2.2061
		$C_{12}H_{18}N_4O_2$	Cis-cyclo-(His,Leu)	251.1511	-3.3889
		$C_{12}H_{23}N_3O_2$	Unknown	242.1862	3.0860
		$C_{32}H_{43}N_5O_8$	Unknown	626.3157	4.3815
	Alpha amino acids and derivatives	$C_{10}H_{16}N_2O_2$	Cyclo(L-Pro-L-Val)	197.1286	-0.7434

		$C_{20}H_{33}N_5O_7$	unknown	456.2437	3.4593
	Alpha amino acids and derivatives	$C_{11}H_{18}N_2O_3$	Cyclo(4-hydroxy-R-Pro-S-Leu)	227.1396	-2.5697
	Alpha amino acids and derivatives	$C_{14}H_{16}N_2O_3$	Cyclo(D-6-Hyp-L-Phe)	261.1232	0.6489

Table A1. Tentatively identified metabolites present in bacterial extracts. Continued.

Sample	Formulae	Classification	Tentative Identity	Precursor m/z $[M+H]^+$	m/z Error (ppm)
(***B. licheniformis***)					
	$C_{11}H_{18}N_2O_2$	Alpha amino acids and derivatives	Cyclo(Pro-Leu)	211.1446	-2.3589
	$C_{12}H_{14}N_2O$		Unknown	203.1175	1.9272
	$C_{10}H_{16}N_2O_4$		Unknown	229.1181	0.8040
	$C_{16}H_{29}N_3O_6$		Unknown	360.2137	-2.1935
	$C_{29}H_{82}N_{34}O_8$		Unknown	1035.713	-0.2300
	$C_{11}H_{19}N_5O_2$	Alpha amino acids and derivatives	Cyclo(D-Arg-L-Pro)	254.1624	-4.9326
	$C_{11}H_{19}N_3O_2$		Unknown	226.1555	-2.2061
	$C_{12}H_{18}N_4O_2$	Alpha amino acids and derivatives	Cis-cyclo-(His,Leu)	251.1511	-3.3889
	$C_{53}H_{89}N_{13}O_8$	Cyclic depsipeptides	Surfactin D	1036.703	-0.0161
	$C_{53}H_{94}N_8O_{12}$	Cyclic depsipeptides	Surfactin_C1	1036.69	0.3245
	$C_{52}H_{91}N_7O_{13}$	Cyclic depsipeptides	Surfactin B	1022.676	4.1201
	$C_8H_{16}N_4O_3$	N-acyl-L-alpha-amino acids	(2S)-2-acetamido-5-guanidino-valeric acid	217.129	2.3912
	$C_{14}H_{16}N_2O_2$	Dipeptides	(3s,8ar)-3-benzyl-1-hydroxy-3h,6h,7h,8h,8ah-pyrrolo[1,2-a]pyrazin-4-one	245.1282	1.0413
	$C_{16}H_{22}O_4$	Sesquiterpenoids	mochiquinone	279.1594	-1.1303

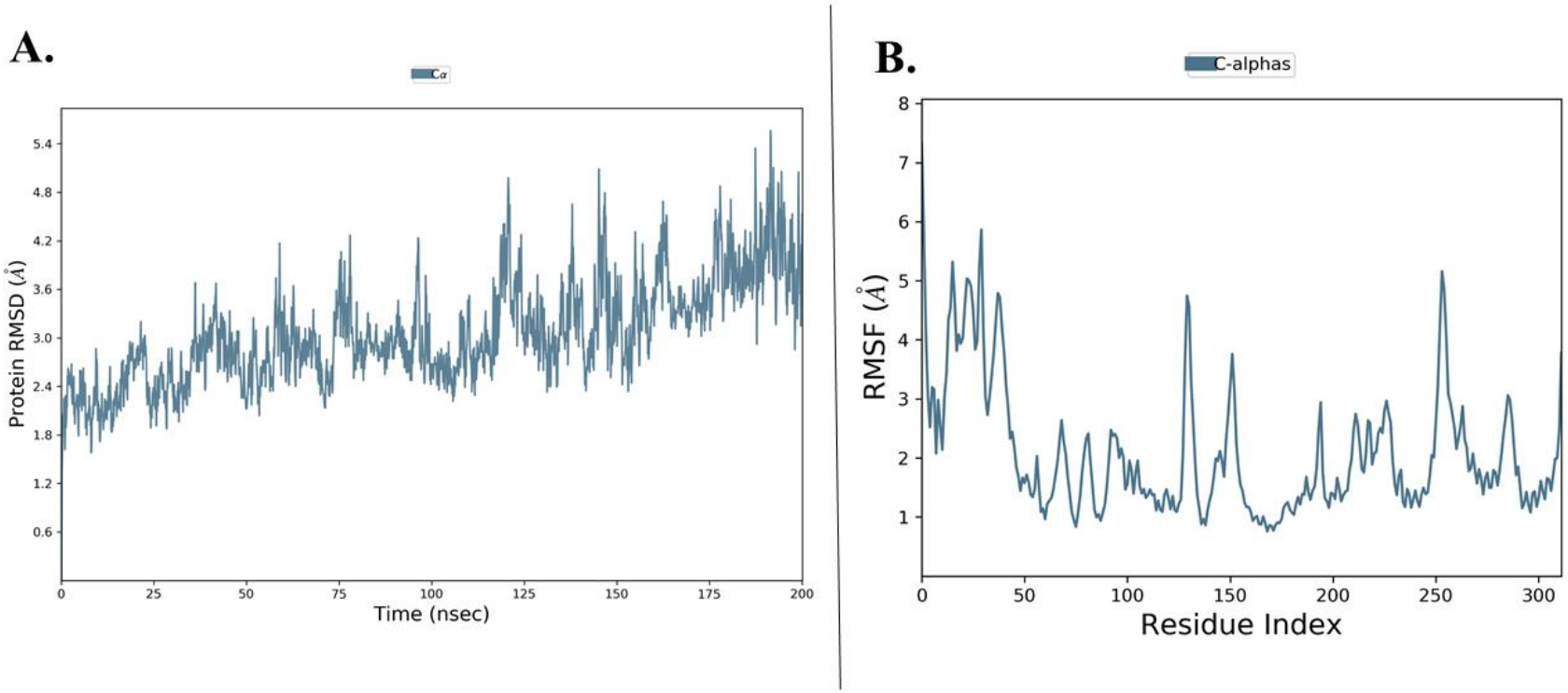


Figure A1. A RMSD PknG Cα-atoms over a period of 200 ns MD simulation and B RMSD for PknG Cα-atoms.

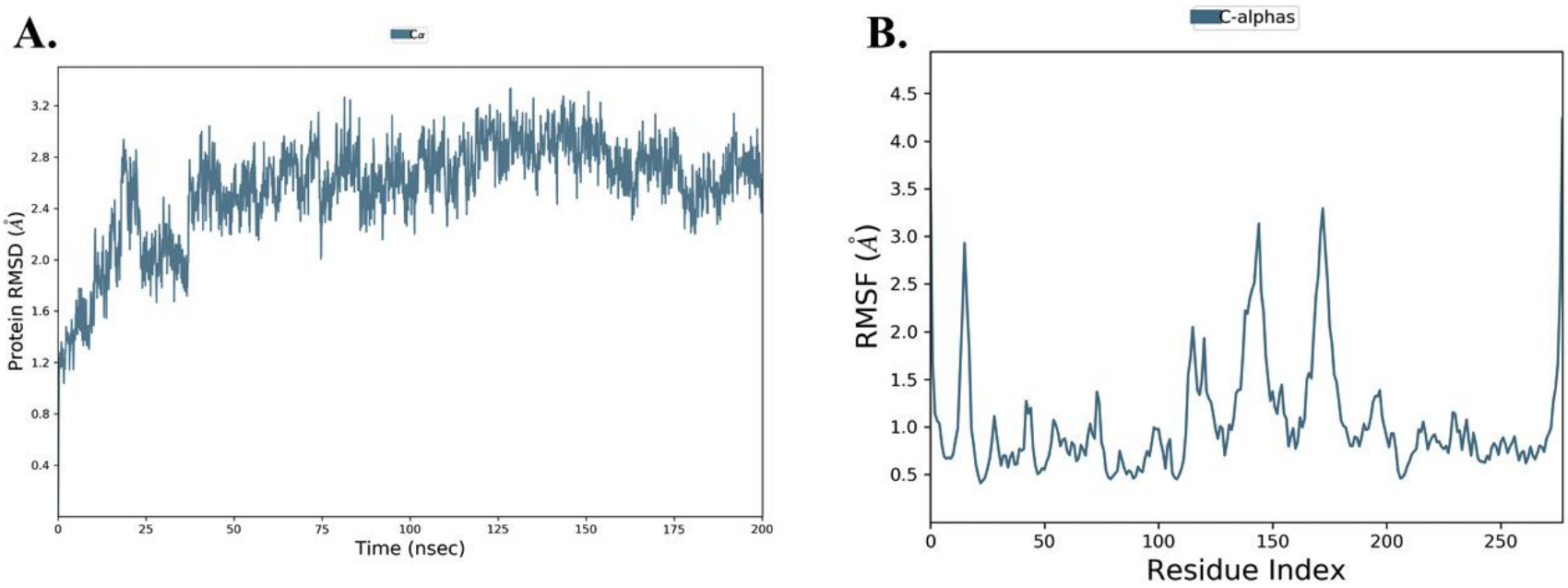


Figure A2. RMSD Pks13 Cα-atoms over a period of 200 ns MD simulation and B. RMSF of Pks13 Cα-atoms.

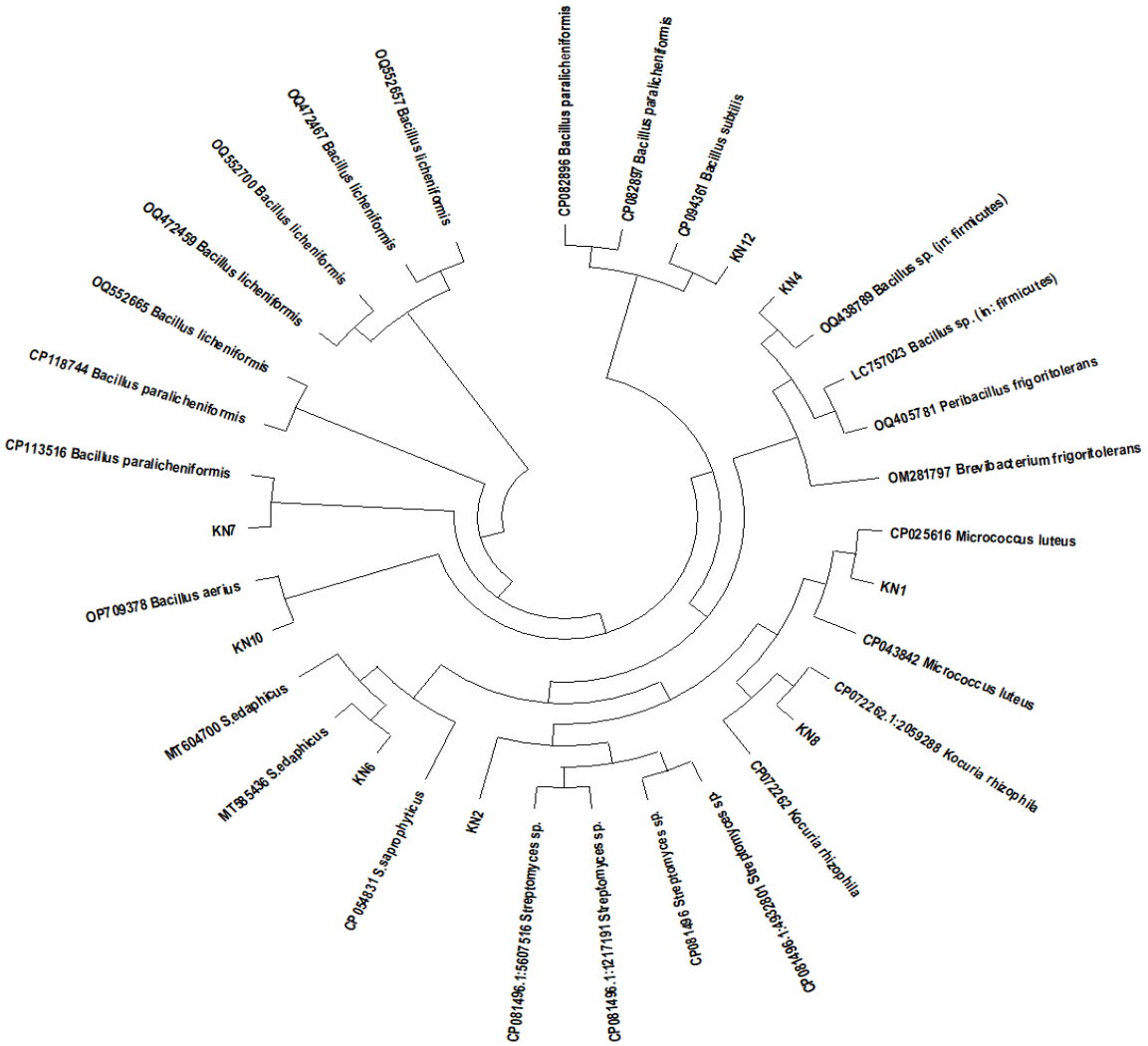


Figure A3. Phylogenetic position of 8 isolated bacterial strains. A maximum likelihood (ML) tree was constructed based on 16 S rRNA sequences of good quality.

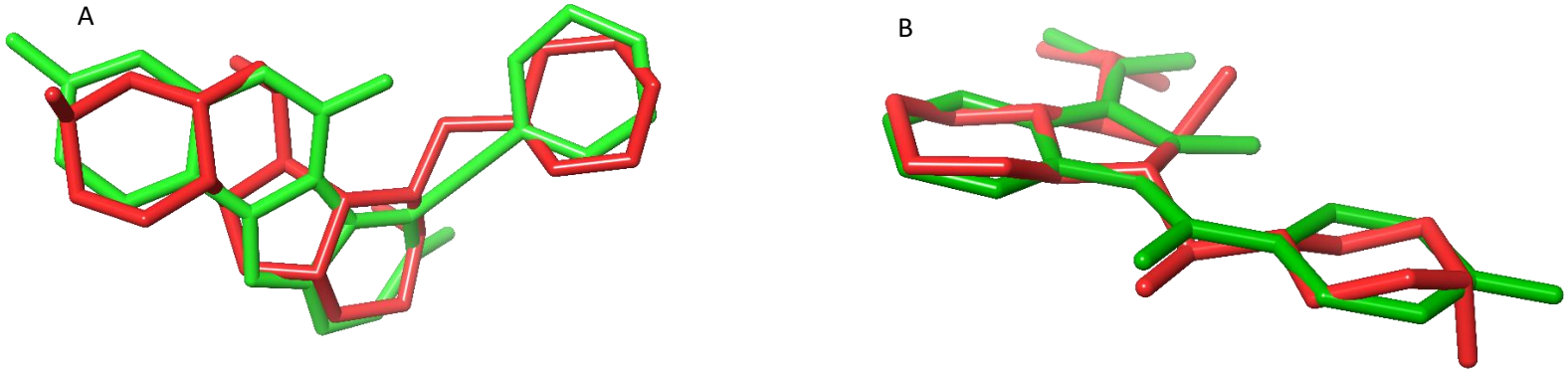


Figure A4: A. The RMSD of the co-crystalized ligand (7IJ) vs the docked co-crystalized ligand. B. The RMSD of the native co-crystalized ligand (8ZC) vs the docked co-crystalized ligand (8ZC). Green represents the native co-crystalized ligand while red represents docked co-crystalized ligand.

Table S2. Computed orbital energies of the tentatively identified bacterial compounds.

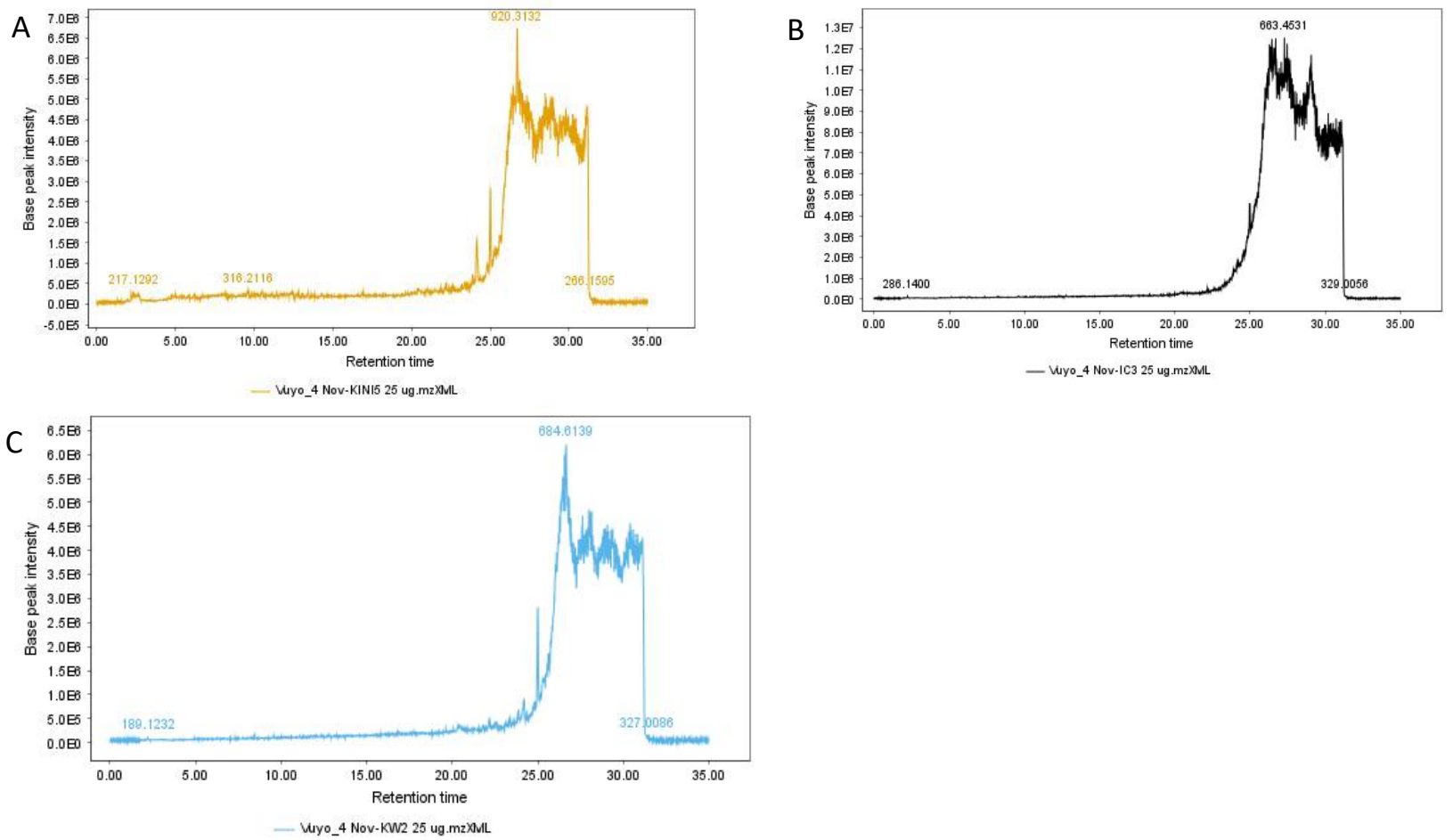
Compounds	E_{HOMO}	E_{LUMO}	Energy gap ΔE	Ionisation energy[a]	Electron Affinity[b]	μ	ɲ	s	χ	ω
vazabitide A	−0.31	−0.02	0.29	0.31	0.02	-0.16	0.15	6.83	0.16	0.09
Cyclo(Pro-Leu)	−0.31	−0.01	0.30	0.31	0.01	-0.16	0.15	6.72	0.16	0.09
maculosin	−0.28	−0.01	0.27	0.28	0.01	-0.15	0.13	7.51	0.15	0.08
5'-Deoxytoyocamycin	−0.29	−0.03	0.26	0.29	0.03	-0.15	0.13	7.65	0.16	0.10
(3R,8aR)-3-benzyl-2,3,6,7,8,8a-hexahydropyrrolo[1,2-a]pyrazine-1,4-dione	−0.30	−0.01	0.29	0.30	0.01	-0.16	0.14	6.94	0.16	0.08
NPA006809	−0.27	−0.04	0.23	0.27	0.04	-0.16	0.12	8.61	0.16	0.10
Cyclo(2-hydroxy-Pro-R-Leu)	−0.32	−0.01	0.30	0.32	0.01	-0.17	0.15	6.59	0.17	0.09
cyclo-(L-Pro-4-OH-L-Leu)	−0.31	−0.01	0.29	0.31	0.01	-0.16	0.15	6.79	0.1	0.09

chemical potential (μ), chemical hardness (ɲ), chemical softness (s), electronegativity (χ) and electrophilic index (ω).

$$\chi = -\mu \quad (1)$$

$$\omega = \frac{\mu^2}{2\eta} \quad (2)$$

Figure A5. LC-MS-QTOF detection of natural products. A. Base peak intensity corresponding to *B. subtilis.* B. Base peak intensity corresponding to *S. mycarofaciens*. C. Base peak intensity corresponding *B. licheniformis.*



Appendix B

In-silico screening of fungi secondary metabolites against *Mycobacterium tuberculosis* Ser/Thr kinases

Table B1. Site-map scores PknD and PknE.

Protein	**Site score**	**DSCORE**	**Size**	**Volume**
PknD	0.997	0.991	98	342.657
PknE	1.101	1.126	154	414.001

DSCORE-druggability score

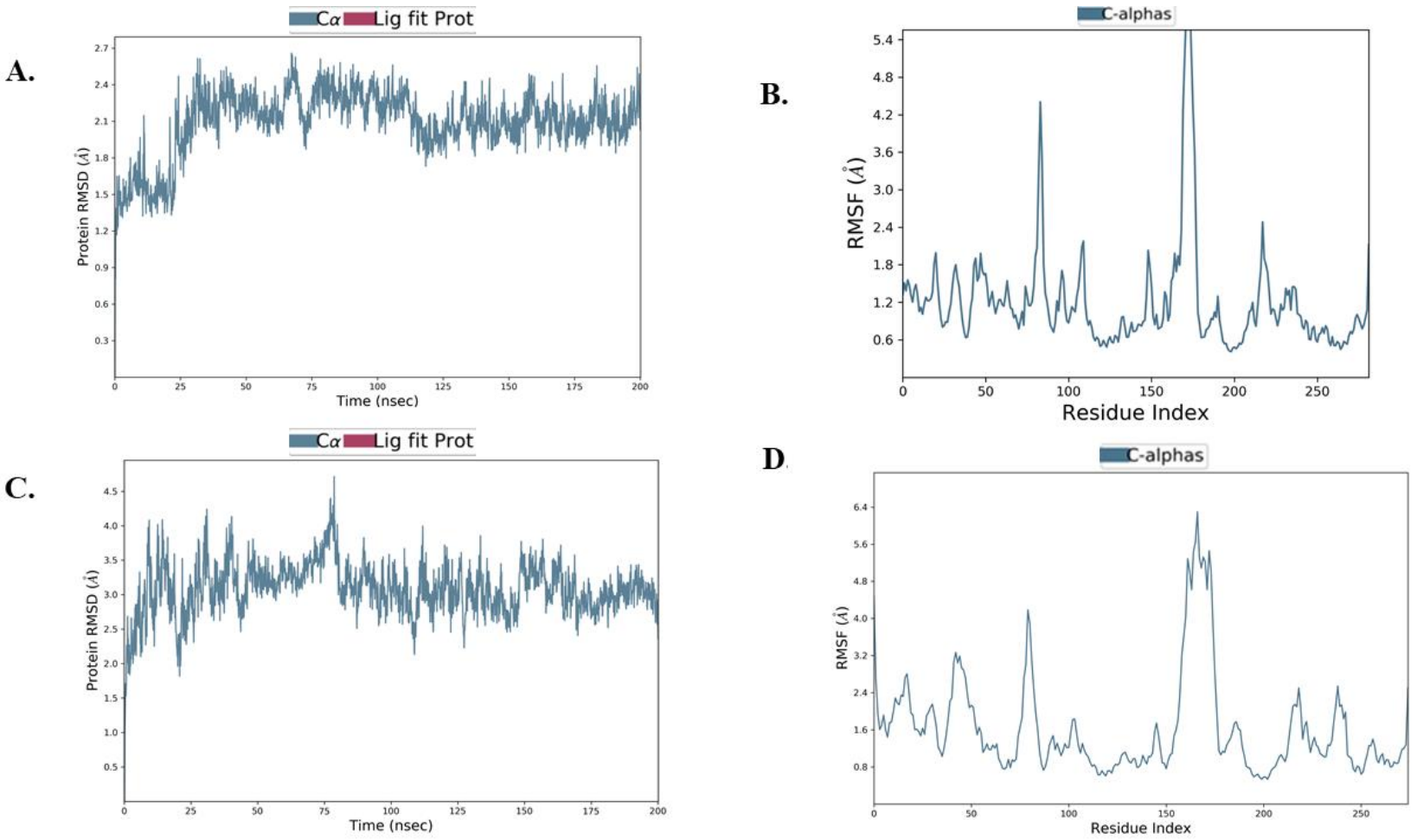


Figure B1. Root-mean-square deviation (RMSD) of the native PknA Cα-atom as a function of simulation time (200ns). **B.** Root-mean-square fluctuation (RMSF) per residue of PknA. **C.** RMSD of PknB Cα-atoms as a function of simulation time (200ns). **D.** RMSF per residue of the native PknB.

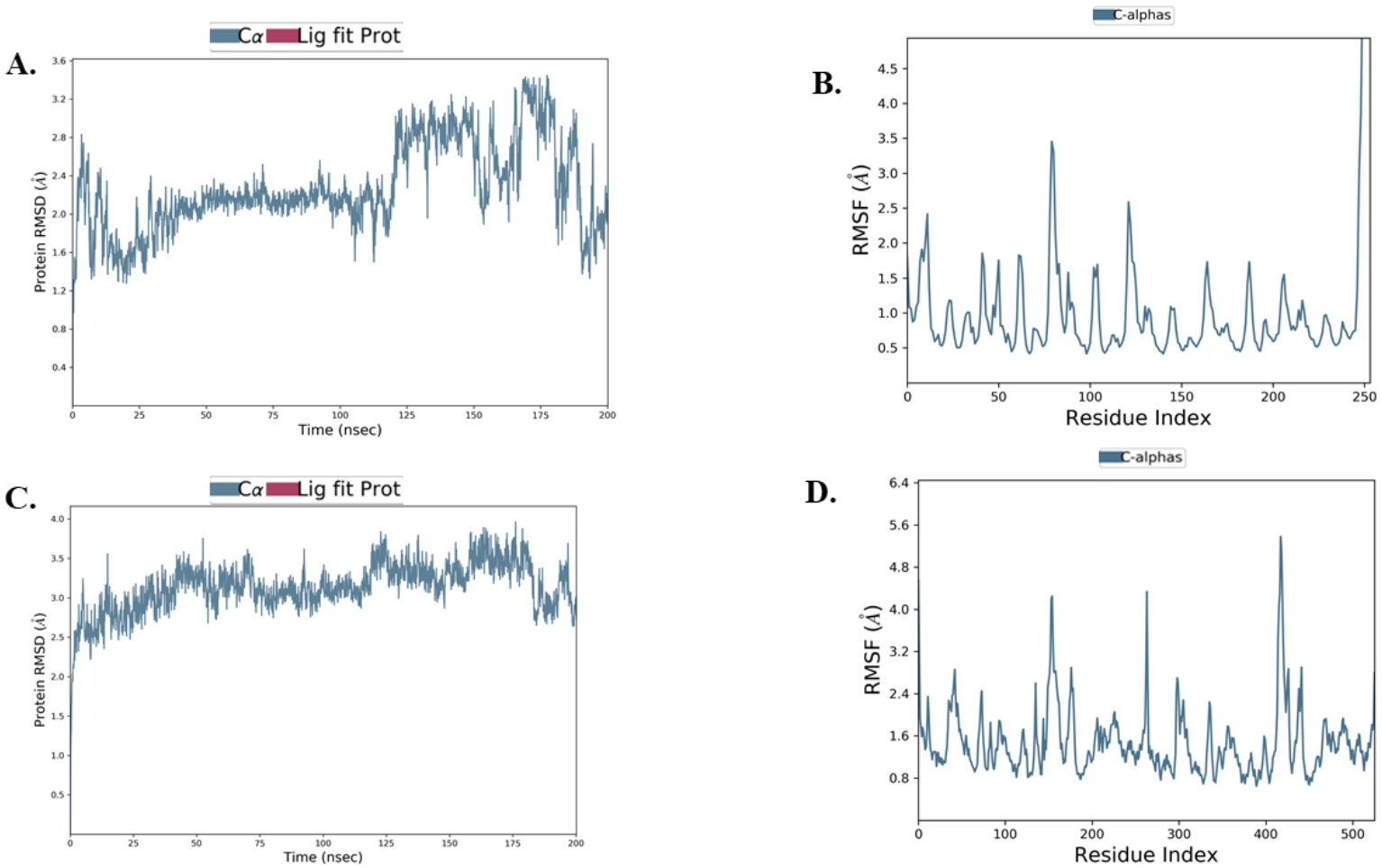


Figure B2. A. Root-mean-square deviation (RMSD) of PknD Cα-atoms as a function of simulation time (200ns). B. Root-mean-square fluctuation (RMSF) per residue of PknD. C. RMSD of PknE Cα-atoms as a function of simulation time (200ns). D. RMSF per residue of PknE.

Appendix C

In-silico and *in-vitro* assessments of some *Fabaceae*, *Rhamnaceae*, *Apocynaceae*, and *Anacardiaceae* species against *Mycobacterium tuberculosis* H37Rv and triple-negative breast cancer cells

Kudakwashe Nyambo[1*], Francis Adu-Amankwaah[1*], Kudzanai Ian Tapfuma[1], Lucinda Baatjies[1], Lauren Julius[1], Liezel Smith[1], Mkhuseli Ngxande[2], Krishna Govender[3,4], Lawrence Mabasa[5], Afsatou Traore[6], Maano Valerie Masiphephethu[6], Idah Sithole Niang[7] and Vuyo Mavumengwana[1†]

[1] *DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town; KN: knyambo@sun.ac.za; K.I.T: kudzanait@sun.ac.za, F.A: fa@sun.ac.za, LB: lbaatjies@sun.ac.za; LJ: laurenjulius@sun.ac.za; LS: liezels@sun.ac.za; VM: vuyom@sun.ac.za*

[2] *Computer Science Division, Department of Mathematical Sciences, Faculty of Science University of Stellenbosch, Matieland, South Africa; ngxandem@sun.ac.za*

[3] *Department of Chemical Sciences, University of Johannesburg, P. O. Box 17011, Doornfontein Campus, 2028, Johannesburg, South Africa; krishnag@uj.ac.za*

[4] *National Institute for Theoretical and Computational Sciences (NITheCS), South Africa*

[5] *Biomedical Research and Innovation Platform (BRIP), South African Medical Research Council (SAMRC), Tygerberg 7505, South Africa; Lawrence.Mabasa@mrc.ac.za*

[6]*Department of Biochemistry & Microbiology, University of Venda; Afsatou.Traore@univen.ac.za; masiphemv@gmail.com*

[7]*Department of Biochemistry, University of Zimbabwe, B064, Mount Pleasant, Harare, Zimbabwe; sitholeidah2015@gmail.com*

** These authors contributed equally to this paper.*

[†]*Corresponding author email address: vuyom@sun.ac.za*

C.1 Abstract

Medicinal plants play a huge role in the treatment of various diseases in the Limpopo province (South Africa). Traditionally, concoctions used for treating tuberculosis and cancer are sometimes prepared from plant parts naturally occurring in the region, these include (but not limited to)

Schotia brachypetala, Rauvolfia caffra, Schinus molle, Ziziphus mucronate, and *Senna petersiana*. In this study, the aim was to evaluate the potential antimycobacterial activity of the five medicinal plants against *Mycobacterium smegmatis* MC^2155, *Mycobacterium aurum* A+, and *Mycobacterium tuberculosis* H37Rv, and cytotoxic activity against MDA-MB 231 triple-negative breast cancer cells. Phytochemical constituents present in *R. caffra* and *S. molle* were tentatively identified by LC-TOF-MS/MS as these extracts showed antimycobacterial and cytotoxic activity. A rigorous Virtual Screening Workflow (VSW) of the tentatively identified phytocompounds was then employed to identify potential inhibitor/s of *M. tb* pantothenate kinase (PanK). Molecular dynamics simulations and post-MM-GBSA free energy calculations were used to determine the potential mode of action and selectivity of selected phytocompounds. The results showed that plant crude extracts exhibited poor antimycobacterial activity, except for *R. caffra* and *S. molle* which exhibited average efficacy against *M. tb* H37Rv with minimum inhibitory concentrations between 0.25-0.125 mg/mL. Only one compound with a favorable ADME profile, namely, norajmaline was returned from the VSW. Norajmaline exhibited a docking score of -7.47 kcal/mol, while, pre-MM-GBSA calculation revealed binding free energy to be -37.64 kcal/mol. All plant extracts exhibited a 50% inhibitory concentration (IC_{50}) of < 30 mg/mL against MDA-MB 231 cells. Flow cytometry analysis of treated MDA-MB 231 cells showed that the dichloromethane extracts from *S. petersiana, Z. mucronate,* and ethyl acetate extracts from *R. caffra* and *S. molle* induced higher levels of apoptosis than cisplatin. It was concluded that norajmaline could emerge as a potential antimycobacterial lead compound. Validation of the antimycobacterial activity of norajmaline will need to be performed *in vitro* and *in vivo* before chemical modifications to enhance potency and efficacy are done. *S. petersiana*, *Z. mucronate*, *R.caffra* and *S. molle* possess strong potential as

key contributors in developing new and effective treatments for triple-negative breast cancer in light of the urgent requirement for innovative therapeutic solutions.

Keywords

Mycobacterium tuberculosis, Schinus molle, Rauvolfia caffra, LC-MS-QTOF, virtual screening, molecular dynamics simulations, MM-GBSA, triple-negative breast cancer, flow cytometry, antioxidants,

Abbreviations

MM-GBSA - molecular mechanics generalized born surface area; RMSD - root mean square deviation; ADME - absorption, distribution, metabolism, and excretion; RMSF - root mean square fluctuation; LC-TOF-MS/MS - liquid chromatography-tandem quadrupole time-of-flight mass spectrometry; OPLS4 - optimized potentials for liquid simulations 4; NPT- isothermal isobaric ensemble; TIP3P – transferable intermolecular potential 3P

C.2 Introduction

Drug resistance in breast cancer, and the escalating spread of multidrug-resistant *Mycobacterium tuberculosis* (*M. tb*) strains is a major concern because it is straining the healthcare systems, especially that of developing countries. Current tuberculosis (TB) and cancer chemotherapies in clinical use have severe side effects that often result in the development of other health-related complications. Southern Africa is one of the heavily affected regions, due to a combination of various factors which include a plethora of factors usually associated with Low and Income-Middle-Class Countries (LIMCC) including, inadequate health facilities, HIV-TB co-infection, and socioeconomic factors [1–3]. It is therefore critical to develop new effective communicable

and noncommunicable chemotherapeutic agents that will be easily accessible to marginalized communities [4].

For most people in rural areas, the traditional pharmaceutical system is complemented by modern treatment procedures thus, broadening the scope of healthcare solutions usually available to individuals in urban areas [2, 5, 6]. Comprehensive knowledge of the diverse botanical landscape provides a baseline for prescribing complex concoctions for treating and curing various ailments. Due to strong cultural beliefs in communities embracing indigenous knowledge systems, it is not surprising to find that ethnobotanical treatment modalities are strongly adhered to [7–9]. As such, various extracts derived from plants such as *Schotia brachypetala, Senna petersiana, Ziziphus mucronata, Rauvolfia caffra,* and *Schinus molle,* are reported to be widely used as medicinal remedies in these communities. Traditional therapies constituting *R. caffra* have been reported as prescriptions for the treatment of ailments such as microbial infections, malaria, diabetes, diarrhea, skin infections, worm infections, and coughs [10, 11]. Oils extracted from *S. molle* have been associated with nutritional, antimicrobial, anti-inflammatory, anti-depressant, astringent, stimulant, and anti-cancer activity [12, 13]. Extracts from the bark of *Z. mucronata* were shown to exhibit broad-spectrum antimicrobial activity [14]. Ethnobotanical knowledge can be used to harness the prominent plant arsenal by selecting the species of plants that are prescribed as antibacterial remedies.

With the increase of drug-resistant *M. tb* strains, the adoption of *in-silico* techniques enables efficient and cost-effective identification of potential lead compounds that can further be developed into potent drugs [15]. Targeting *M. tb* enzymes that participate in essential biosynthetic pathways with bioactive phytocompounds may lead to the discovery of novel

scaffolds with novel mechanisms of action. *M. tb* pantothenate kinase (PanK) is a critical regulatory target that catalyzes the first and rate-limiting step of the biosynthesis of the CoA pathway. CoA is a crucial cofactor for the survival of the bacilli because it is vital for enzymes involved in lipid biosynthesis and catabolism. Lipids are essential building blocks for the cell envelope and serve as *M. tb* virulence factors [16, 17]. From this perspective, a targeted *in-silico* exploration of the ethnopharmacological derived compounds present in crude extracts against *M. tb* PanK may identify promising lead scaffolds. Herein, the study aims to evaluate the antimycobacterial activity of the crude extracts of *S. brachypetala, S. petersiana, Z. mucronata*, *R. caffra,* and *S. molle* and provide a detailed insight into compounds that could have complimentary conformational features required for binding in the PanK domain. Virtual screening workflow, molecular dynamics (MD) simulations, and MM-GBSA binding free energy were performed to reveal a new dimension on the dynamics of targeting the PanK with plant-derived ligands.

Breast cancer, particularly the aggressive triple-negative breast cancer subtype, remains a major worldwide health concern [18]. It is critical to develop innovative treatment techniques against this subtype. The plant species selected in this study are abundant in structurally diverse secondary metabolites, that is, indole alkaloids, phenols, terpenoids, and flavonoids. These secondary metabolites are well known for possessing anticancer activity [19, 20], thus, the second aim of the study was to explore the antiproliferative activity of the crude extracts of *S. brachypetala, S. petersiana, Z. mucronata*, *R. caffra,* and *S. molle* against MDA-MB 231, a triple-negative breast cancer cell line. Understanding their antiproliferative effects could pave the way for the development of new therapeutic interventions for breast cancer.

C.3 Materials and methods

C.3.1 Plant collection and preparation

In this study medicinal plants were collected in Tshififi, Siambe, and Lufule villages, Vhembe district, Limpopo province, South Africa. A voucher specimen was identified and authenticated by Professor P. Tshisikhawe at the UNIVEN herbarium, (Botany Department, University of Venda). Plant samples were separated into leaves, bark, and roots. The different parts of the plants were then dried at ambient temperature in the lab for two weeks. Thereafter, they were separately ground into a fine powder and kept in airtight containers in the dark until use. The crude ingredients present in all the plant material were exhaustively extracted by dissolving ten grams of each plant material in 100 mL of hexane, chloroform, dichloromethane, ethyl acetate, acetone, ethanol, and methanol (Merck, Kenilworth, NJ; USA). The solution was shaken for 1 hour at 200 rpm. The supernatant was then filtered into pre-weighed bottles and the process was repeated three times. Thereafter, crude plant extracts were combined and dried under a fume hood at room temperature and then stored at 4 °C until further analysis.

C.3.2 Antimycobacterial minimum inhibitory concentration assay

The antimycobacterial activity was performed to evaluate the activity of the crude extracts against *Mycobacterium smegmatis* MC^2155, *Mycobacterium aurum* A+, and *M. tb* H37Rv as described by [21]. Briefly, the Mycobacteria were cultured in Middlebrook 7H9 (Fluka M7H9) broth supplemented with 0.2% glycerol, 0.05% Tween 80, and 10% Middlebrook growth supplement OADC (oleic acid, albumin, dextrose, and catalase) at 37°C. The minimum inhibitory concentration (MIC) to obstruct *M. smegmatis*, *M. aurum*, and *M. tb* growth were determined following a procedure described by Eloff (1998), with a modest change for Mycobacteria. Dried

plant extracts were redissolved in dimethyl sulfoxide (DMSO) to a final concentration of 4 mg/mL which was followed by a twofold serial dilution in 96-well microtiter plates to achieve a series of concentrations ranging from 0.2-2 mg/mL. DMSO (negative control) and isoniazid (positive control). *M. smegmatis* and *M. aurum* plates were incubated at 37°C for 72 hours, while *M. tb* plates were incubated for seven days before adding 20 µL of 0.02% resazurin. The non-pathogenic strains were incubated for a further four hours and *M. tb* for a further 24 hours. Growth inhibition was indicated by a constant blue resazurin color while a pink color indicated the inactivity of extracts against Mycobacteria. All extracts were tested in triplicate.

C.3.3 Tentative identification of phytochemicals

A liquid chromatography connected to quadrupole time-of-flight with tandem mass spectrometry (LC-QTOF-MS/MS) was employed to identify the plant crude extracts as described Tapfuma et al., (2022) [21]. The system consists of a Waters Acquity ultra-performance liquid chromatography (U-PLC) coupled to Water Synapt G2 quadrupole time-of-flight mass spectrometer (Milford, MA, USA) and an Acquity photo-diode array (PDA) detector. The plant metabolites were chromatographically separated using a Waters UPLC BEH C18 column (1.7-µm particle size, 2.1× 100 mm, Waters Corp) with the following elution gradient., 95 % of formic acid (0.1 % (v/v)) as eluent A and 5 % acetonitrile as eluent B. The spectral data were acquired at 150 to 1500 m/z in positive centroid mode. Ionization was achieved with an electrospray source using a cone voltage of 15 V and capillary voltage of 2.5 kV. Nitrogen was used as the desolvation gas at 650 L/hr and the desolvation temperature was set to 275 ℃. The raw data containing spectral data were converted to .abf format. The .abf files were processed by MS-Dial module (version 4.24) and MS-Finder for tentative identification of compounds using the following parameters: error ppm < 7.0, $[M+H]^+$ adducts ions, KNapSacK (http://www.knapsackfamily.com/KNApSAcK_Family/),

Metfrag (https://msbi.ipb-halle.de/MetFrag/), and online published databases were used for tentatively identifying compounds.

C.3.4 Virtual Screening of Tentatively identified compounds

C.3.4.1 Receptor preparation

The raw crystal structure of *M. tb* pantothenate kinase (PanK) PDB:4BFX was downloaded from the protein data bank (https://www.rcsb.org/structure/4BFX). The structure was prepared as described [22] in Schrödinger (Release 2021-1) using the Protein Preparation Wizard module. Briefly, hydrogen atoms were added, the loop region was refined, H-bond assignments were optimized, and energy was minimized by an OPLS-4 force field. The coordinates of the co-crystallized ligand (1f) [23] were used for the generation of the binding domain using the Receptor Grid Generating module (Schrödinger Release 2021-1).

C.3.4.2 Ligand preparation

The compounds were prepared as previously described [21]. Briefly, the LigPrep module [24] (Schrödinger Release 2021-1) was used to prepare the compounds following these parameters: energy minimized by an OPLS4 force field, generate ionization states at pH 7.0 +2.0, and generate at most 32 conformers per each ligand to develop a new set of 640 possible compounds. The prepared library was subjected to a molecular docking-based virtual screening.

C.3.4.3 Structure-based Virtual Screening of Compounds

A Virtual screening Workflow (VSW) comprised of the following modules (Schrödinger Release 2021-1) [25]: QickProp, Lipinski's Rule of 5 filters, high-throughput virtual screening (HTVS), standard precision (SP), and lastly extra-precision, which were used for screening the library of phytocompounds to obtain a hit list. QickProp module filtered the phytocompounds based on

features of ADME (absorption, distribution, metabolism, and excretion). The obtained compound list was further subjected to Lipinski's rule of five filters. Using the Glide module, the returned compounds were subjected to three-step docking regimes with increasing precision. Briefly, the compounds were docked against PanK using high-throughput virtual screening (HTVS), standard precision (SP), and lastly extra-precision (XP). Only 20 % of the HTVS docking hits were applied to SP docking. Only 20% of SP docking outputs were subjected to XP docking, from which 30 % were retrieved as described by [26]. The pre-MM-GBSA (Molecular Mechanics, the Generalized Born model, and Solvent Accessibility) was performed to evaluate the free binding energy (ΔG_{bind}) of the protein-ligand complex/es as described by [25].

C.3.5 Molecular Dynamics Simulation

The selected poses for the PanK-phytocompound complex, PanK-control ligand (1f) complex, and native unbound PanK conformations were subjected to molecular dynamics (MD) simulations using Desmond (Schrödinger Release 2021-1) to evaluate the stability of the docked complex. A total of three separate MD systems were created according to the same parameters as described by [21]. Briefly, the protein-ligand complex was explicitly solvated by enclosing it in an orthorhombic TIP3P water box with the protein surface atoms 10 Å away from the box boundary. The system was neutralized by adding 0.15 M counter ions (Na^+ and Cl^-). All systems had implemented periodic grid conditions, long-range electrostatic interactions were generated for the particle-mesh Ewald method with a non-bonding cut-off distance of 12 Å. The systems were energy minimized and equilibrated at constant pressure and temperature (1.01325 bar and 303.15K, respectively) with Nose-Hoover thermostat, and Martyna-Tobias-Klein as the default barostat with a 2.0 ps interval by applying an isotropic coupling style. The internal energy was stored for every 1000 ps of the actual frame. The NPT ensemble MD simulations were performed for a duration of 50ns.

The stability for each complex was evaluated by computing the root-mean-square deviation (RMSD), and root-mean-square fluctuations (RMSF). The molecular mechanics generalized Born surface area (MM-GBSA) (ΔG_{Bind}) (kcal/mol) binding free energies were computed based on Molecular Mechanics + Implicit Solvent Energy Function [27].

C.3.6 Determination of cytotoxic effects of plant crude extracts

The cytotoxic effect of plants against Triple-negative breast cancer cells (MDA-MB-231 was conducted as described by [28] with modifications. Triple-negative breast cancer cells (MDA-MB-231, passage number 43) were donated by Prof Anna-Mart Engelbrecht, Stellenbosch University, South Africa. Briefly, cells were seeded in a 96-well plate at a density of 6 000 cells/well and left to attach for 24 hours. The plant crude extracts were dissolved in dimethyl sulfoxide (DMSO) to form a 100 mg/mL stock solution. MDA-MB-231 cells were treated with plant crude extracts (62.5-250 µg/mL) and cisplatin (reference drug at 3 µg/mL, Sigma Aldrich, USA) for 48 hours. Spent Dulbecco's Modified Eagle Media (DMEM), supplemented 10% foetal bovine serum (DMEM complete media) was removed and replaced with 0.5 mg/mL 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (Sigma Aldrich, USA) dissolved in DMEM complete media. After four hours of incubation, MTT solution was removed, and the formazan product dissolved in 100 µL DMSO. Absorbance was measured at 540 nm using a microtiter plate reader (FLUOstar Omega, BMG Labtech, Germany) [28]. All incubations were done in a humidified incubator (ESCO, Vivid Air) with 5% CO_2 at 37 °C.

C.3.7 Cytotoxic effects of plant crude extracts against HepG2/C3A and Vero cell lines

The cytotoxic effect of HepG2/C3A and Vero cells were studied as described by [29, 30] with modifications. In brief, Hep G2 clonal derivatives (C3A) with passage number 14 and Vero,

normal monkey kidney cells with passage number 11 were used to evaluate the cytotoxicity of plant extracts. These cells were purchased from Cellonex, South Africa by Prof Maryna van de Venter, Nelson Mandela University South Africa. Cells were cultured at 37°C in a humidified incubator with 5% CO_2 in 10 cm culture dishes.

The complete growth medium consisted of Eagles Minimal Essential Media (EMEM) supplemented with 10% FBS, 10% penicillin-streptomycin (penstrep), and 1x Non-Essential Amino Acid (NEAA), all purchased from GE Healthcare Life Science (Logan, UT, USA), were used to grow the HepG2 cells, a human hepatoma cell line, and DMEM supplemented with 10% FBS and 10% penstrep for Vero cell lines. Cells were seeded into 96-well microtiter plates at a density of 4 000 cells per well using their respective medium and incubated overnight at 37°C, 5% CO_2, and 100% relative humidity to allow for cell attachment. Thereafter, cells were treated with 100 µL aliquots of extracts at 50, 100, and 200 µg/mL concentration and 10, 20, and 40 µM melphalan (positive control) and incubated for a further 48 hours. After incubation, the treatment medium was aspirated from all the wells and 100 µL of Hoechst 33342 nuclear dye (5 µg/mL) was added to each well and incubated for 20 minutes at room temperature. Cells were stained with propidium iodide (PI) at 100 µg/mL to enumerate the proportion of dead cells within the population. Cells were imaged immediately after adding PI, using the ImageXpress Micro XLS Widefield Microscope (Molecular Devices) with a 10x Plan Fluor objective and DAPI and Texas Red filter cubes [29, 30].

C.3.8 Annexin V-FITC/PI apoptosis assay

MDA-MB-231 cells were seeded at a density of 2.5 x 10^5 cells/well in a 24-well plate and incubated overnight at 37 °C in a humidified incubator with 5% CO_2. Cisplatin (10 µM/3 µg/mL)

and plant crude extracts at their respective IC_{50} values were used to treat the cells for 48 hours at 37 °C in a humidified incubator (ESCO, Vivid Air) with 5% CO_2. Following the incubation period, the cells were detached by adding 80 µL of Accutase™ for 10 minutes or until cells were detached. One milliliter complete media were added to each well and incubated at 37 °C in a humidified incubator (ESCO, Vivid Air) with 5% CO_2 for an hour to allow cells to recover. The cells transferred to polypropylene flow cytometry tubes and harvested by centrifugation (1500 rpm) for 5 minutes at 4°C. The pellets were washed with ice-cold DMEM complete media and centrifuged (1500 rpm) for 5 minutes at 4°C. The Annexin V FITC/PI apoptosis detection kit was used to stain the cells as per manufacturer's instructions (Invitrogen, ThermoFisher Scientific). The pellets were redissolved in ice-cold 1x binding buffer. To each tube, 1 µL of Annexin V FITC and 5 µL of PI were added. Control tubes with single stains were also added and incubated in the dark for 15 minutes. After incubation, 400 µL of 1x annexin-binding buffer was added and gently mixed. The samples were read on a BC DxFlex flow cytometer (Beckman Coulter, USA) [31].

C.3.9 Statistical analysis

The statistical analysis of the behavioural data was conducted using the student t-test with GraphPad Prism (GraphPad Software Inc., San Diego, CA) and Microsoft Excel. The mean values ± standard deviation (SD) was reported for all data. Statistical significance was determined at a significance level of $p \leq 0.05$, indicating that differences with this level of probability or lower were considered statistically significant.

C.4 Results

C.4.1 Antimycobacterial activity

Crude extracts were extracted using seven different solvents of varying polarity. A total of 30 extracts obtained from *S. brachypetala, R. caffra, S. molle, Z. mucronata,* and *S. petersiana* were evaluated for antimycobacterial activity against *M. smegmatis* (MC^2155), *M. aurum* (A+), and *M. tb* (H37Rv). The MIC value of >2 mg/mL was selected as a cutoff for all *Mycobacteria* strains susceptibility. The crude extracts exhibited varying antimycobacterial activity (Table C.1, Table C.2, and Table C.3). All the extracts exhibited poor inhibition against *M. smegmatis*. In contrast, the control (isoniazid) showed an MIC of 0.03 mg/mL against *M. smegmatis* (Table C.1).

Table C.1. Minimum inhibitory concentration (MIC) of crude plant extracts against *M. smegmatis* MC^2155.

Plant species	**Extracts (mg/mL)**							
	Hexane	Chloro-form	Dichloro-methane	Ethyl acetate	Acetone	Ethanol	Methanol	INH
S. brachypetala	>2	>2	>2	>2	>2	>2	>2	-
R. caffra	>2	>2	>2	>2	>2	>2	>2	-
S. molle	>2	>2	>2	>2	>2	>2	>2	-
Z. mucronata	>2	>2	>2	>2	>2	>2	>2	-
S. petersiana	>2	>2	>2	>2	>2	>2	>2	-
***Control**	-	-	-	-	-	-	-	0.03

* INH (isoniazid) was used as a positive control.

All the crude extracts from *R. caffra* showed strong efficacy against *M. aurum* A+ (Table 7.2). While, hexane, chloroform, dichloromethane, ethyl acetate, acetone, and methanol extracts of *S.*

molle strongly exhibited *M. aurum* A+, dichloromethane and methanol crude extracts from *Z. mucronata* showed potent efficacy against *M. aurum* A+. While for *S. petersiana* only the dichloromethane extract strongly inhibited *M. aurum* A+, *S. brachypetala* exhibited poor inhibitory activity against *M. aurum* A+.

Table C.2. Minimum inhibitory concentration (MIC) of crude plant extracts against *M. aurum* A+.

Plant species	Extracts (mg/mL)							
	Hexane	Chloro-form	Dichloro-methane	Ethyl acetate	Acetone	Ethanol	Methanol	INH
S. brachypetala	>2	2	>2	>2	>2	>2	>2	-
R. caffra	0.13	0.04	0.07	0.07	0.13	0.5	0.25	-
S. molle	0.04	0.02	0.25	0.02	0.04	>2	0.04	-
Z. mucronata	2	2	0.04	>2	2	>2	0.25	-
S. petersiana	>2	2	0.04	>2	>2	>2	>2	-
***Control**	-	-	-	-	-	-	-	0.03

* INH (isoniazid) was used as a control.

Crude extracts from *R. caffra* and *S. molle* were observed to possess antimycobacterial activity for *M. tb* (Table C.3). *S. molle* exhibited higher anti-*M. tb* activity (MIC = 0.125 mg/mL) than *R. caffra* (MIC = 0.25 mg/mL). *R. caffra* and *S. molle* crude extracts may possess useful bioactive constituents that may have the potential to serve as drug leads. Plants are undoubtedly an invaluable bio-factory comprised of numerous diverse bio-active ingredients. Therefore, the constituents present in crude extracts of *R. caffra* and *S. molle* were tentatively identified by untargeted LC-QTOF-MS/MS.

Table C.3. Minimum inhibitory concentration (MIC) of crude plant extracts against *M. tb* H37Rv.

Plant species	Extracts (mg/mL)							
	Hexane	Chloroform	Dichloro-methane	Ethyl acetate	Acetone	Ethanol	Methanol	INH
S. brachypetala	>2	>2	>2	>2	>2	>2	>2	-

R. caffra	>2	>2	0.25	>2	>2	>2	>2	-
S. molle	>2	>2	0.125	>2	>2	>2	>2	-
Z. mucronata	>2	>2	>2	>2	>2	>2	>2	-
S. petersiana	>2	>2	>2	>2	>2	>2	>2	-
***Control**	-	-	-	-	-	-	-	<0.31

* INH (isoniazid) was used as a control.

C.4.2 Tentatively Identification of phytocompounds

The active phytoconstituents present in the *R. caffra* and *S. molle* crude extracts were tentatively identified using LC-MS-QTOF. Basically, the tentatively identified phytocompounds exhibited varied mass-to-charge ratio (m/z) values ranging from 117.1031 to 513.2248. The tentatively identified compounds are represented in Table C.4. The classes of constituents present in *R. caffra* were mostly alkaloids, terpenoids, indole alkaloids, and glycoalkaloids (Table C.4), while for *S. molle* the compounds present were mostly terpenoids, terpenes, sesquiterpenes, and triterpenoid saponins.

Table C.4. Tentatively identified compounds present in *R. caffra* crude extract.

RT (min)	Peak height	Precursor m/z	Molecular formula	Error ppm	Compound	Class
2.8152	1644.42	171.1031	$C_9H_{14}O_3$	8.9362	Boonein	Terpenoid
3.4693	7986.07	313.1922	$C_{19}H_{24}N_2O_2$	3.6591	Norajmaline	Indole Alkaloid
3.9888	8118.74	513.2248	$C_{27}H_{32}N_2O_8$	3.2305	Raucaffricine	Glucoalkaloid
4.4711	12250.83	351.1713	$C_{21}H_{22}N_2O_3$	2.7935	Raucaffrine	Alkaloid
4.0898	5624.79	327.2076	$C_{20}H_{26}N_2O_2$	2.7352	Ajmaline	Alkaloid
4.2318	2396.07	355.2032	$C_{21}H_{26}N_2O_3$	4.4509	Acetylnorajmaline	Alkaloid

4.3590	2533.79	367.1664	$C_{21}H_{22}N_2O_4$	3.1756	Apodine	Alkaloid
4.7589	3064.10	339.1703	$C_{20}H_{22}N_2O_3$	-0.056	Akuammicine N-oxide	Alkaloid
4.8187	13412.6	353.1863	$C_{21}H_{24}N_2O_3$	0.9371	Raucaffrinoline	Indole Alkaloid
4.8710	1292.01	323.1758	$C_{20}H_{22}N_2O_2$	1.2253	Norpurpeline	Indole Alkaloid
5.0131	2402.98	349.1563	$C_{21}H_{20}N_2O_3$	4.6712	Alstonine	Indole Alkaloid
5.4280	2063.06	383.1618	$C_{21}H_{22}N_2O_5$	4.3115	Apodinine	Alkaloid
7.1101	1719.26	357.1816	$C_{20}H_{24}N_2O_4$	2.0045	Compactinervinete	Alkaloid

Table C.5. Tentatively identified compounds present in *S. molle* crude extract.

RT (min)	**Peak height**	**Precursor m/z**	**Molecular formula**	**Error ppm**	**Compound**	**Class**
5.4580	4281.99	237.1853	$C_{15}H_{24}O_2$	1.6611	Aubergenone	Sesquiterpene
6.1084	3686.87	203.18	$C_{15}H_{22}$	2.8201	Beta-Spathulene	Sesquiterpene
10.6350	6452.70	471.3477	$C_{30}H_{46}O_4$	1.7269	Semialatic acid	Triterpene
9.3080	3843.11	205.1953	$C_{15}H_{24}$	1.0867	Beta-Caryophyllene	Terpene
9.9397	11675.21	453.3369	$C_{30}H_{44}O_3$	1.2749	Pistacigerrimone	Triterpenoid
11.6107	1499.63	455.3534	$C_{30}H_{46}O_3$	3.1360	Isomasticadienonic acid	Triterpenoid
12.0929	1943.07	457.3684	$C_{30}H_{48}O_3$	1.7010	Oleanolic acid	Triterpenoid saponin
13.2405	1172.57	441.372	$C_{30}H_{48}O_2$	-1.6018	28-Hydroxy-beta-Amyrone	Triterpenoid

C.4.3 *In-silico* screening of the tentatively identified compounds

The library of the tentatively identified compounds was screened by a Virtual Screening Workflow (VSW) (Qikprop, Lipinski's rule of 5, HVTS, SP, and XP docking) to filter and reduce false positive hit compounds (Table C.6). The XP docking was performed to precisely search for the best protein-ligand complementarity conformation. Norajmaline was returned as a potential hit

from the extensive filtering stages and exhibited an XP docking score of –7.465 kcal/mol (Table C.7). The best-hit compound, norajmaline, returned from the rigorous VSW, exhibited zero violations for Lipinski's rule of five, the percentage human oral absorption was 63.99%, Van der Waals surface area of polar nitrogen and oxygen atoms (PSA) was 61.53, QPlogS was –0.81 , Solvent accessible surface area (SASA) was 532.97, and the dipole value was 1.59. Overall, the ADME values were promising as they were all in the recommended ranges. The pre-MD simulation binding energy (ΔG_{Bind}) of norajmaline-PanK was -37.64 kcal/mol.

Table C.6. The predicted ADME features (SASA, dipole, Qplogs, % Human Oral Absorption and PSA), and the molecular docking XP score and Pre-MM-GBSA (ΔG_{Bind}) values of norajmaline against 4BFX.

Compound ID	mol MW (170-725)	Dipole (1.0–12.5)	SASA	*QplogS (– 6.5 to 0.5)	PSA (7.0–200.0)	Volume	%Human Oral Absorption	Rule of Five	XP GScore (kcal/mol)	ΔG_{Bind} (kcal/mol)
Norajmaline	318.46	1.59	532.97	-0.81	61.53	991.26	63.99	0	-7.47	-37.64

***QPlogS**-(Predicted aqueous solubility, log S. S in mol dm^{-3} is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.)

Analyses of the best XP docked configuration depicted in Figure C.1, revealed that norajmaline is buried in the hydrophobic internal cavity of the protein. The main driving forces involved in the binding of norajmaline against amino acid residues of PanK were predicted to be predominantly hydrophobic interactions (Try257, Met242, Phe239, Tyr235, Ala100, Val99, Ile276, Met144, Ile272, Phe254, Try177, and Tyr1820). In contrast, ASN277 was involved in hydrogen bonding, while (Arg238, His179, and Lys147) were involved in positively charged interactions and polar interactions (Asn280, Asn277, and Hie145). PanK-norajmaline complex, unbound PanK, and PanK-control ligand (**1f**) complex were further subjected to molecular dynamics simulations and free-binding energy calculation for the complex.

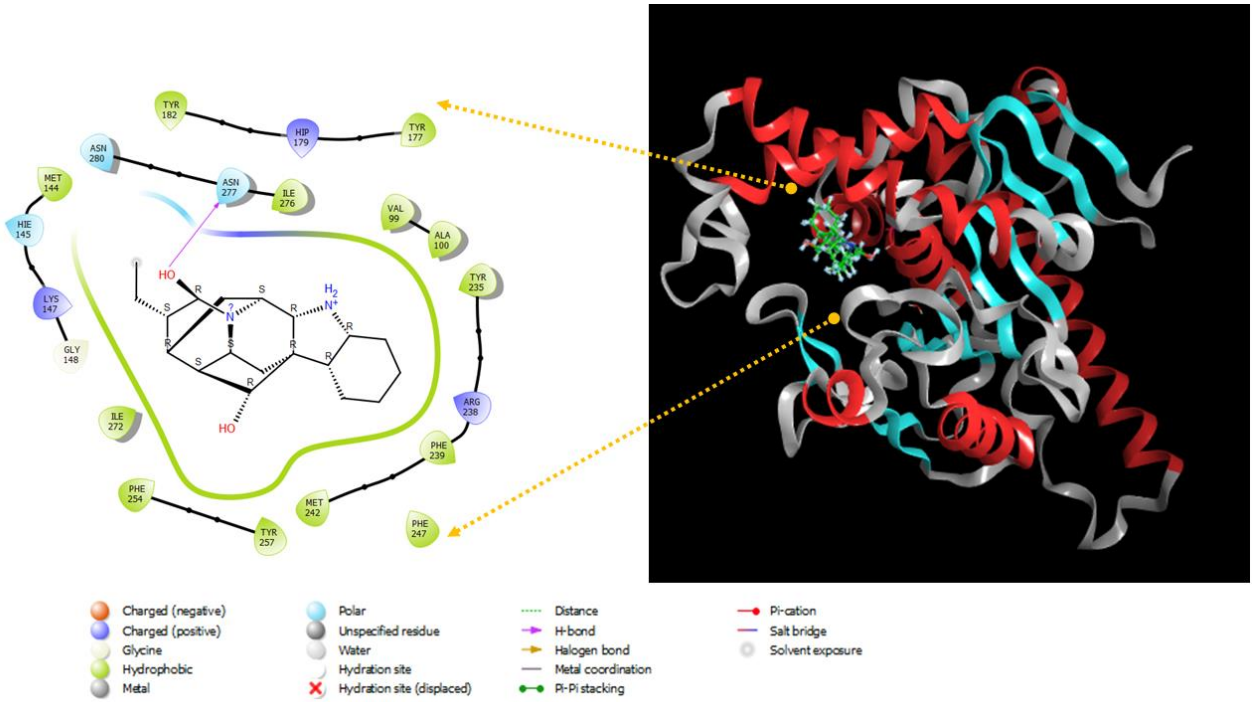


Figure C.1. Docked orientation and interaction of norajmaline with PanK residues in the binding site.

MD simulations were performed to provide a comprehensive insight into the structural dynamics of the binding of norajmaline in the hydrophobic cavity of the PanK. Root Mean Square Deviation (RMSD) of the PanK-norajmaline complex and the unbound cα atoms were

performed to evaluate the stability of the protein-ligand complex. The RMSD profile of bound PanK Cα-atoms shows a steep increase in deviation from 1.6 Å at 0 ns to approximately 2.6 Å at 8 ns, was then maintained between 2.4 Å and 2.6 Å up to 30 ns and then gradually decrease to 2.3 Å up to 50 ns as depicted in Figure C.2.A. The sharp increase observed from 0 ns to 8 ns indicates a change in PanK confirmation as it interacts with the norajmaline (Figure C.2.A). The Norajmaline RMSD (Figure C.2.A) was maintained at approximately 3.0 Å during the 50 ns simulation, which illustrates the ligand was fairly undergoing slight conformational changes. The RMSF of PanK illustrates a large fluctuation at residues between 0 and 100, while smaller fluctuations were observed from residues 100-300 which were participating in interacting with the ligand (Figure C.2.B). According to the RMSD, the protein-ligand complex was observed to be stable during the 50 ns simulation (Figure C.2.A).

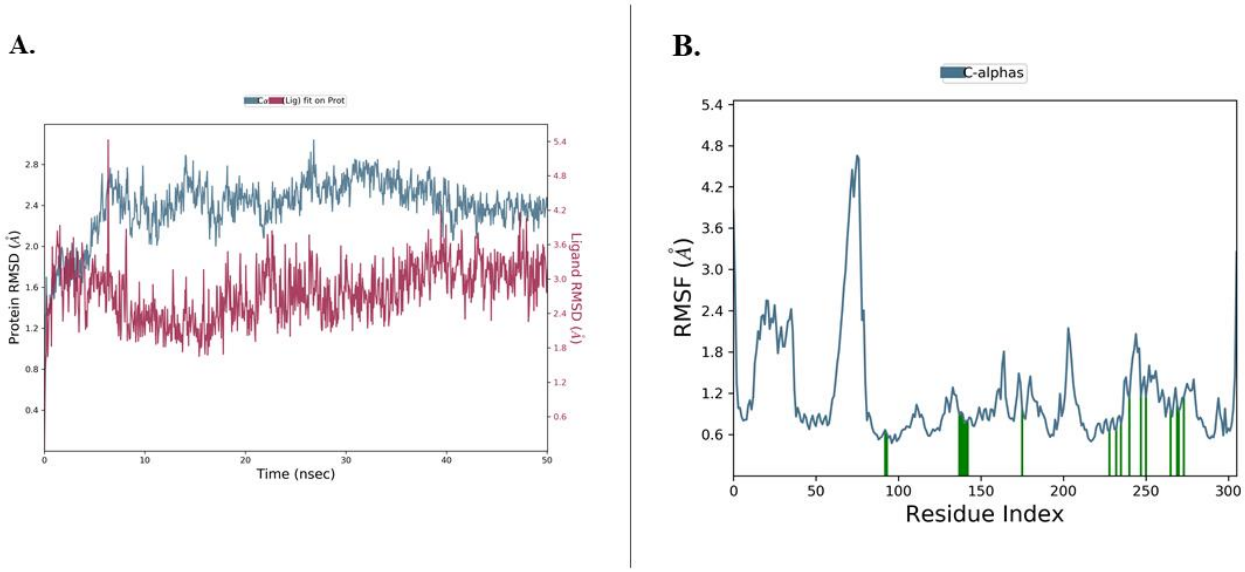


Figure C.2. Molecular dynamics simulation of PanK complexed with norajmaline. In the figure, **A.** shows the RSMD of C-α-residues of PanK observed during a 50 ns simulation. **B.** shows the RMSF of C-α-residues of PanK, where the green lines indicate the residues of 4BFX in contact with the ligand during the simulation.

The RMSD plots of native PanK Cα-atoms without a bound ligand were constant between 1.25 Å and 2.00 Å (Figure C.3.B). The RMSF of the unbound PanK residues was below 2.5 Å

(Figure C.3.B). A comparison of the RMSDs of the PanK-norajmaline complex and that of the unbound native PanK indicated that the binding of the ligand results in changes in slight protein confirmation. Likewise, the RMSF of the two systems showed that smaller fluctuations were observed on PanK C-α-residues that interacted with the ligand's atoms (Figure C.2.B, and Figure C.3.B).

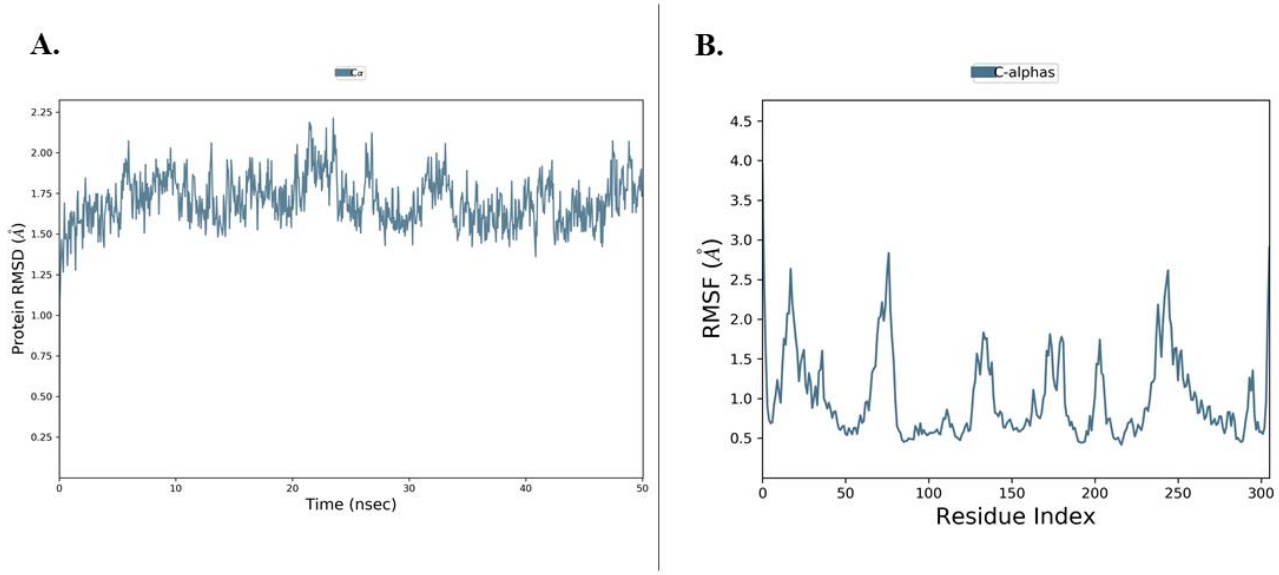


Figure C.3. Simulated native unbound PanK. In **A.** the RMSD of PanK C-α-residues observed during a 50 ns simulation is shown. In **B**. the RMSF of PanK C-α-residues is shown.

The interaction of PanK-norajmaline was mainly due to hydrogen bonds (His145, Tyr235, and Asn277), hydrophobic contacts (Val99, Try235, Phe239, Met242, Phe247, Phe254, Try257, Ile272, Ile276), ionic interactions (Tyr 182) and water bridges (His145, Lys174, Try182, Try257, and Asn277) (Figure C.4.A). The contacts which occurred for more than 30 % of the simulation time are charged (occurrence= 39%), polar interactions (occurrence= 64%), and hydrophobic interactions were also predominant since the ligand was in a hydrophobic pocket of PanK as depicted in Figure C.4.B.

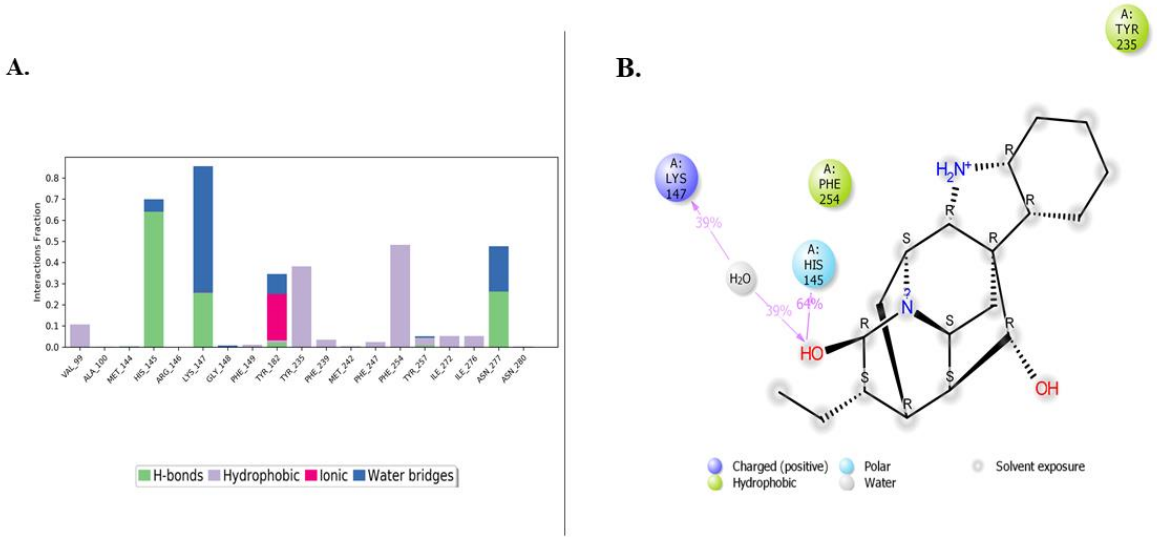


Figure C.4 In **A.** the interaction fraction of residues with norajmaline is displayed. In **B,**the occurrence of the between the ligand and the PanK residues is displayed.

C.4.4 Post-MD simulations MMGBSA (molecular mechanics generalized born surface area) binding energy calculations

The post-MD simulations binding free energy (ΔG_{Bind}) of the PanK-norajmaline complex were evaluated using the MMGBSA method. PanK-norajmaline complex exhibited an MMGBSA ΔG_{Bind} energy of -58.73 kcal/mol (Table C.7), while the control ligand exhibited -67.70 kcal/mol (Table C.7). The MMGBSA supports the stability of the complex as shown in RMSD plots (Figure C.2.).

Table C.7. MM-GBSA (molecular mechanics generalized born surface area) ΔG_{Bind} calculations

Compound ID	**MM-GBSA (kcal/mol)** ΔG_{Bind}
Control ligand (1f)	-67.70
Norajmaline	-58.73

C.4.5 Cytotoxicity effects of crude plant extracts against MDA-MB 231 cells

The crude extracts from medicinal plants, *S. brachypetala, R. caffra, S. molle, Z. mucronata,* and *S. petersiana* (62.5, 125 and 250 µg/mL) were studied for their anticancer effects on MDA-MB-321 cells, as shown in Figure C.5. Some of the intermediate polarity extracts from *R. caffra* (H2, D2 and EA2), *S. molle* (H3 and EA3), *Z. mucronata* (H4 and C4), and *S. petersiana* (C5 and D5) showed over 50% inhibition at the lowest concentration (62.5 µg/mL). *R. caffra* (H2, EA2) and *S. molle* (H3) extracts showed inhibition of over 80% at all concentrations. All the medicinal plant crude extracts showed they could slow the growth of MDA-MB-231 triple negative breast cancer cells. The extracts that were effective (as shown in Figure C.5.) were studied further to find the lowest concentration where they could inhibit 50% of growth.

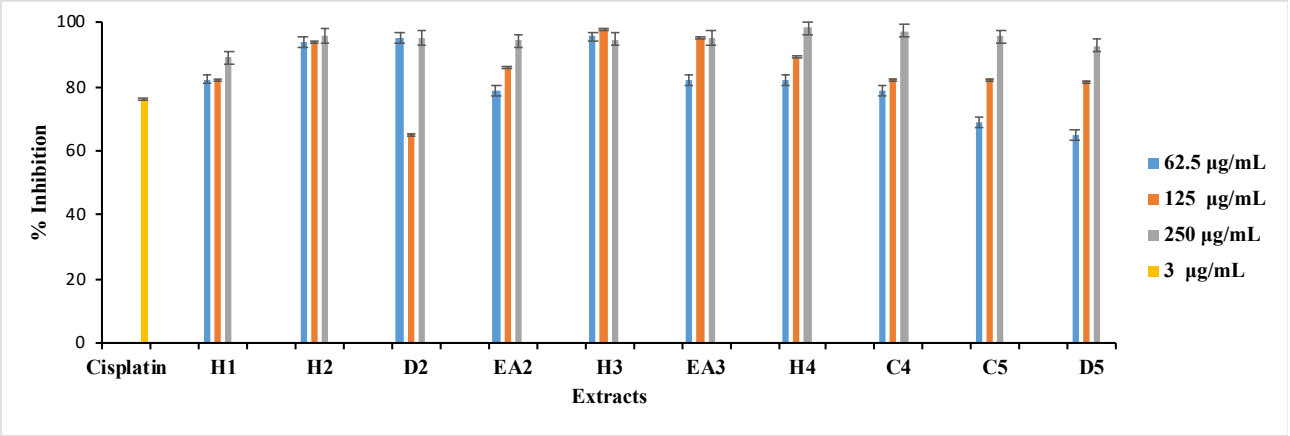


Figure C.5. Cytotoxicity activity of *R. caffra* (H2, D2 and EA2), *S. molle* (H3 and EA3), *Z. mucronata* (H4 and C4), and *S. petersiana* (C5 and D5) (62.5, 125 and 250 µg/mL) and cisplatin (3 µg/mL) as a control drug against MDA-MB 231 triple-negative breast cancer. Results represent the mean ± Standard deviation of triplicate determinations.

C.4.6 Cytotoxic effects of plant extracts against HepG2/C3A liver and Vero monkey kidney cell lines.

The crude extracts of each plant species were tested for potential cytotoxicity against HepG2/C3A and Vero cell lines. The distinction between these cell lines lies in the fact that HepG2/C3A represents a liver cell model, while Vero serves as a model for normal monkey kidney cells. In this study, the cytotoxicity test measured the number of live cells after treatment

of the two cell lines with the plant extracts. The black horizontal lines in Figure C.6.B and Figure C.7.B indicate half (50%) of the untreated control cells and extracts exhibiting cytotoxic potential are shown in Figure C.6.B and Figure C.7.B as having viable (live) cell numbers below the black line.

Figure C.6. shows that *R. caffra* is the least toxic, followed by *Z, mucronata*, *S. petersiana*, *S. molle* and *S. brachypetala*, the most toxic against HepG2/C3A cells at 50 and 200 µg/mL. Cytotoxicity against Vero cells was also determined, and the results depict that all plant crude extracts appear to be less toxic against the normal cell line (Vero cell) at the lowest concentration (50 µg/mL) tested and *S. petersiana* exhibited no harmful effect at all concentrations tested (50-200 µg/mL). However, *S. brachypetala* and *S. molle* were toxic to the Vero cells at a higher concentration (100 and 200 µg/mL) (Figure C.7). All crude plant extracts that showed lower cytotoxicity against the HepG2/C3A and the Vero cell lines were further investigated by determining their IC_{50} values on MDA-MB 231 triple-negative breast cancer.

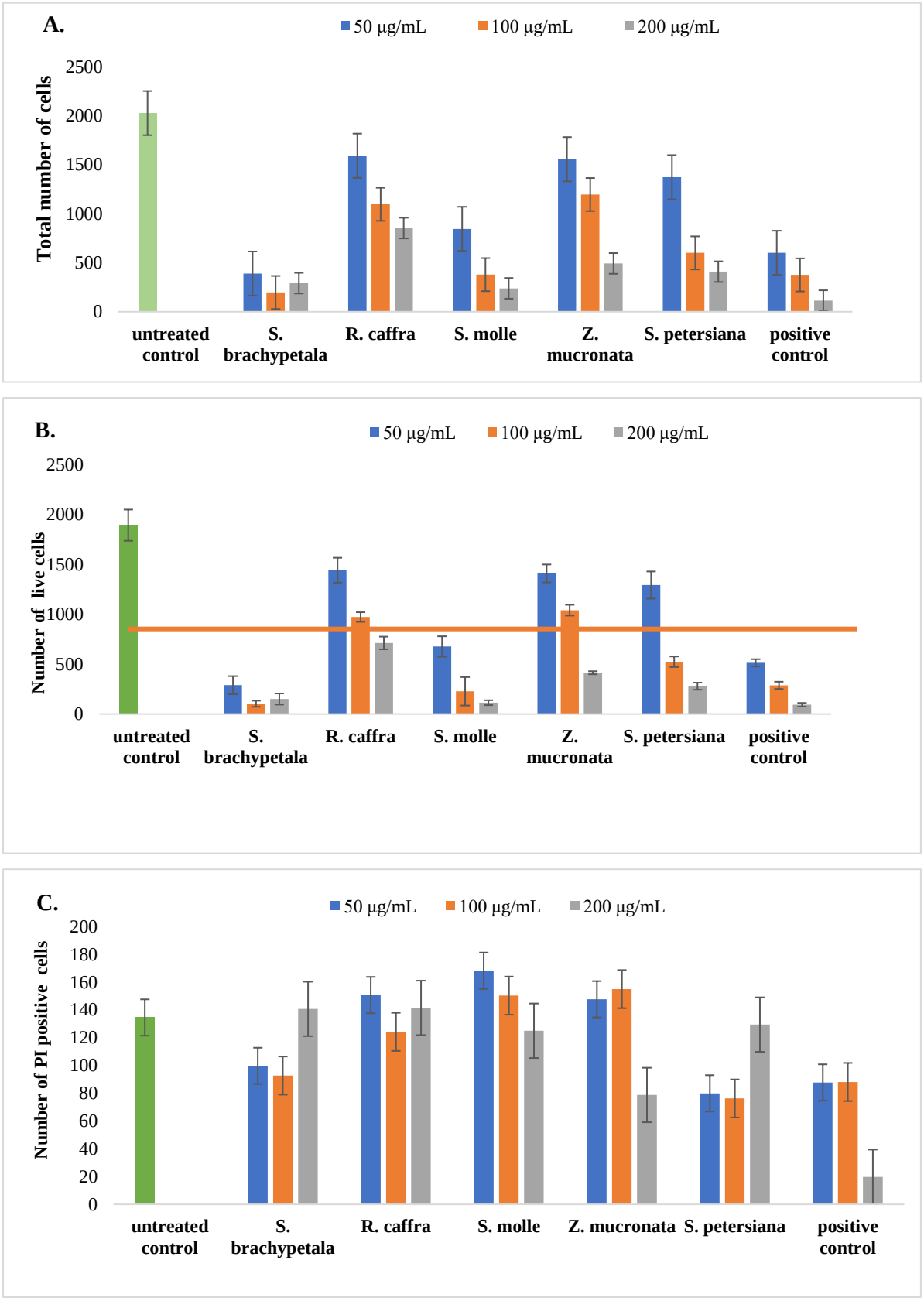


Figure C.6. Cytotoxicity of 5 extracts and Melphalan (10, 20, and 40 µM) as the reference drug against HepG2A/C3A after 48 hours of exposure. Results displayed as (**A**) total number of cells, (**B**) number of cells stained with Hoechst 33342 only and (**C**) Hoechst 33342 and PI.

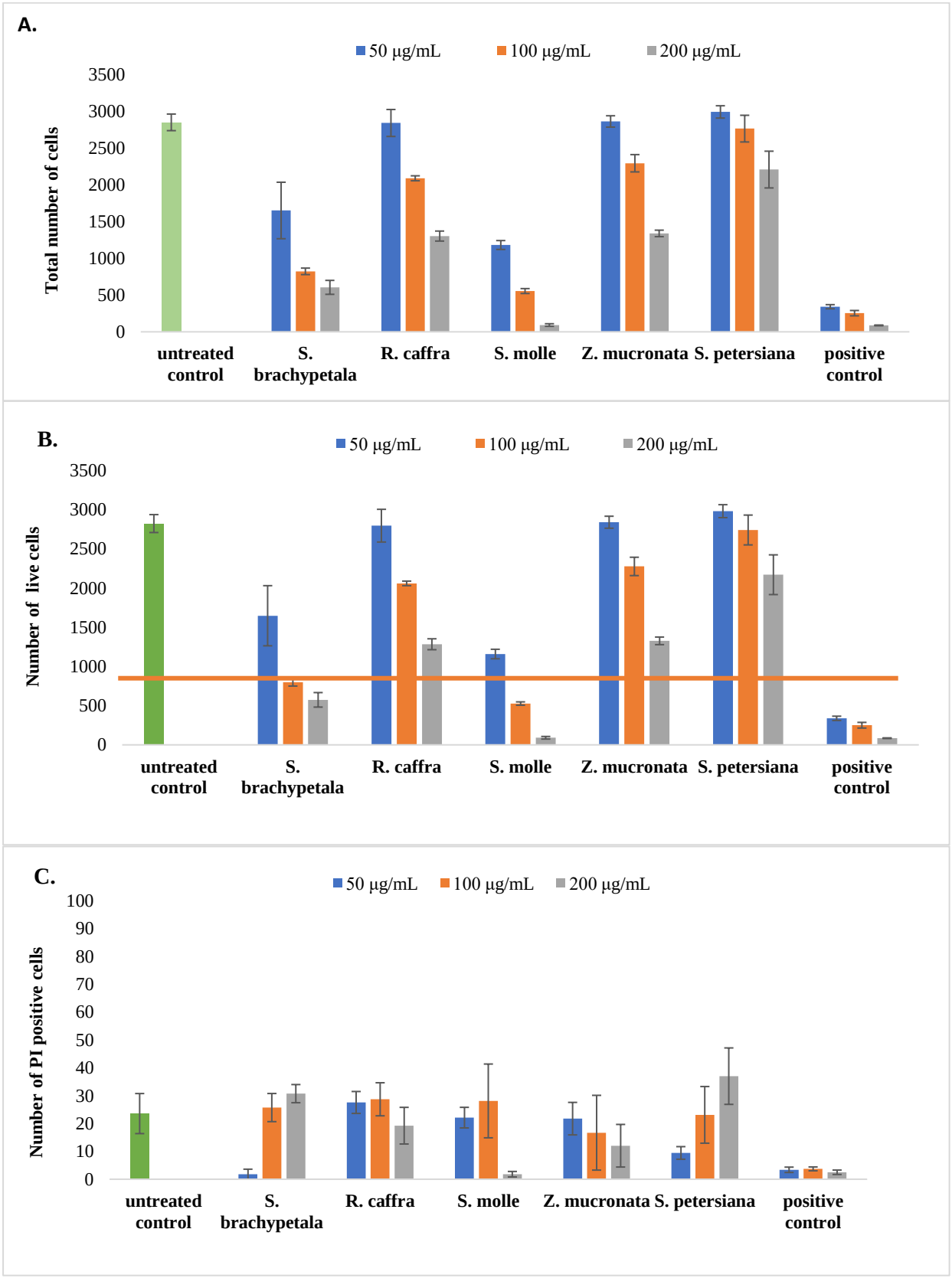


Figure C.7. Cytotoxicity of 5 extracts and Melphalan (10, 20, and 40 µM) as the reference drug against Vero cells after 48 hours of exposure. Results displayed as (A) total number of cells, (B) number of cells stained with Hoechst 33342 only (C) Hoechst 33342 and PI.

C.4.7 Anti-proliferation activity of selected plant extracts against MDA-MB 231 cell line.

Active crude plant extracts against MDA-MB 231 were chosen to be evaluated further to determine their 50% inhibitory concentration (IC_{50}). Cisplatin was used as the reference drug and the concentrations at which the crude plant extracts were tested are as follows: 3.906, 7.8125, 15.625, 31.25, 62.5, 125 and 250 µg/mL. The chloroform crude extract of *S. petersiana* (C5) showed the highest IC_{50} of 26.26 ±2.325, followed by the hexane crude extract of *R. caffra* (H2) at 8.625 ±0.163. However, the dichloromethane crude extract of *S. petersiana* (D5) showed the lowest IC_{50}, 1.525 ±0.458, even lower than the reference drug, cisplatin (2.017 ±0.09) (Figure C.8 and Table C.8).

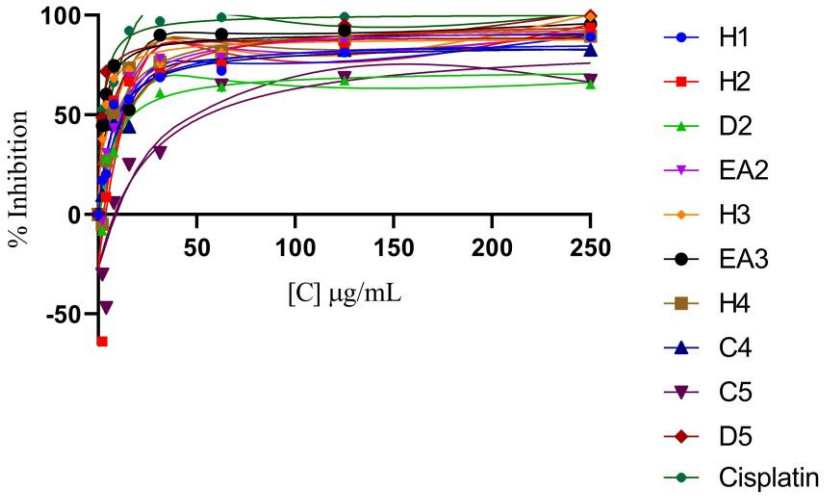


Figure C.8. Dose-response curve of the cytotoxicity of the medicinal plants (H1, H2, D2, EA2, H3, EA3, H4, C4, C5, D5) and the reference control drug (Cisplatin) against MDA-MB 231 triple-negative breast cancer cell line. Cells were treated with incubated varying concentrations of the selected crude extract for 48 hours, after which an MTT assay was performed. The data points shown represent the mean ± the standard deviation technical and biological triplicate repeats. The data were analysed using GraphPad Prism 8 software to obtain the IC_{50} concentrations.

Table C.8. Summary of IC50 values of the cytotoxic effect of cisplatin and H1, H2, D2, EA2 ,H3, EA3, H4, C4, C5 and D5 crude extracts against MDA-MB 231 triple negative cancer cell line.

Sample	IC_{50} (µg/mL)	*p* value
Cisplatin	2.017 ±0.09	0.00435
H1	7.425 ±0.911	0.0138
H2	8.625 ±0.37	0.0068
D2	6.829 ±0.37	0.0203
EA2	7.617 ±0.755	0.0266
H3	2.646 ±0.725	0.6516
EA3	2. 426 ±0.141	0.7672
H4	7.019 ±1.21	0.0375
C4	7.227 ±0.213	0.0157
C5	26.26 ±2.325	0.0157
D5	1.525 ±0.458	0.7225

C.4.8 Determination of mode of cell death using Annexin-V and PI staining

The plant crude extracts that showed growth inhibitory activity against MDA-MB 231 triple-negative breast cancer cell line were further investigated to determine the mode of cell death, using the Annexin V and Propidium Iodide (PI) kit (Invitrogen, ThermoFisher Scientific). Annexin V binds to phosphatidylserine (PS) which translocated from the inner cell membrane to the outer cell membrane during early apoptosis. Propidium Iodide enters dead cells via their compromised cell membranes and stains the nucleus of dead cells [32]. Flow cytometry was used to analyze the mode of cell death (Figure C.10-C.11).

The untreated control showed a high background apoptosis of 9.26% apoptosis. Cisplatin (control drug) induced 16.15% and 1.12% early apoptosis and necrosis respectively, (Figure C.9.A), while the untreated control showed 9.26% and 0.99% early apoptosis and necrosis, respectively (Figure C.9.A). The plant crude extracts of *R. caffra* (EA2, 19.75%), *S. molle*

(EA3, 23.89%), *Z. mucronata* (C4, 24.76%) and *S. petersiana* (D5, 41.97) induced higher percentage early apoptosis than cisplatin. D5 showed the best activity against the MDA-MB 231 cancer cells, with 41.97% early apoptosis. All extracts induced less than 2.5% necrosis under the condition screened (Figure C.9-C.10).

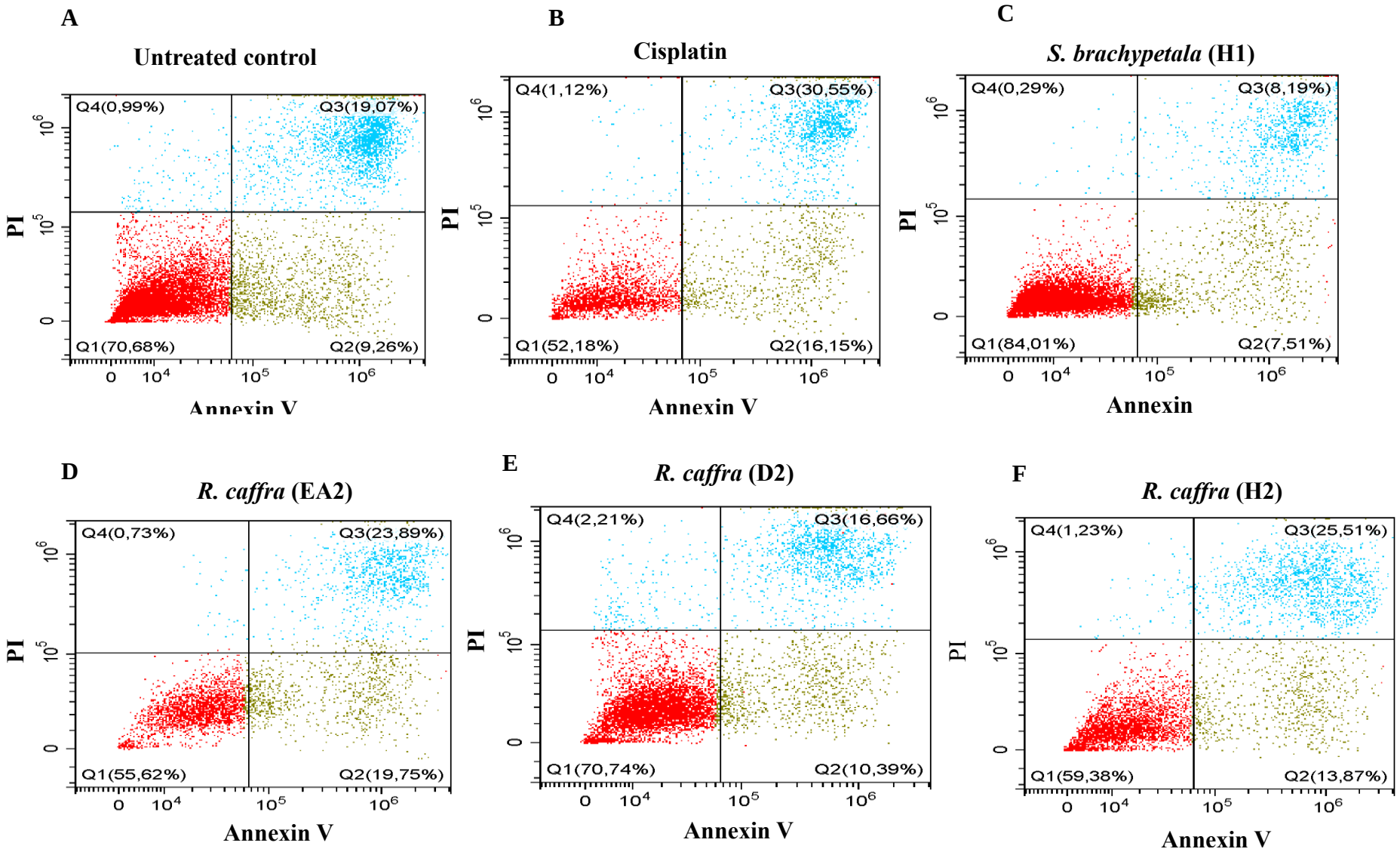


Figure C.9A-F. Effect of crude extracts from *S. brachypetala* (H1), and *R. caffra* (H2, D2, EA2) compared to untreated control and cisplatin (positive control), on stained MDA-MB 231 to identify mode of cell death using flow cytometry.

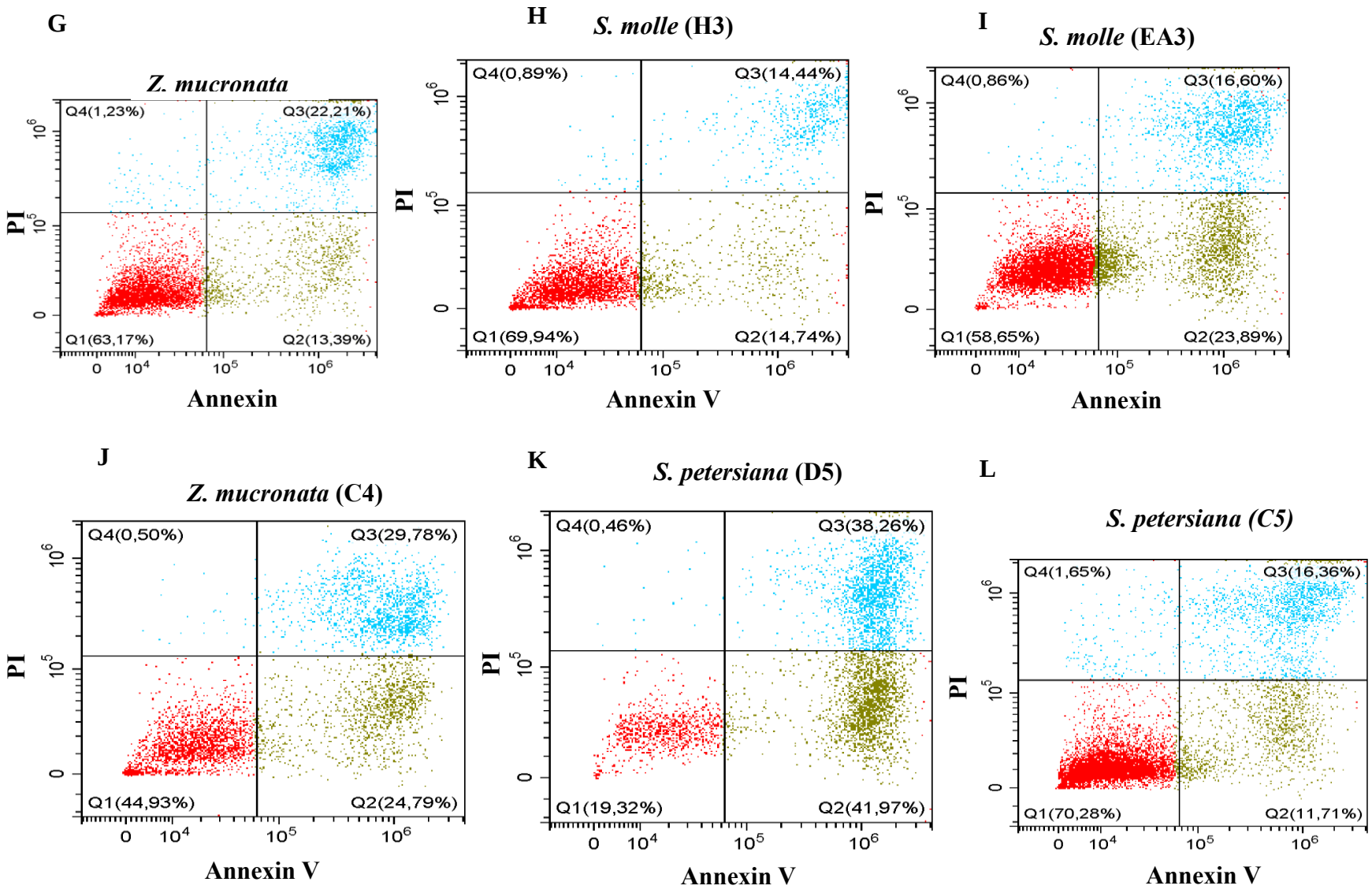


Figure C.9.G -L. Effect of crude extracts from *S. molle* (H3, EA3), *Z. mucronata* (H4, C4), and *S. petersiana* (C5, D5) compared to untreated control and cisplatin (positive control), on stained MDA-MB to identify mode of cell death using flow cytometry.

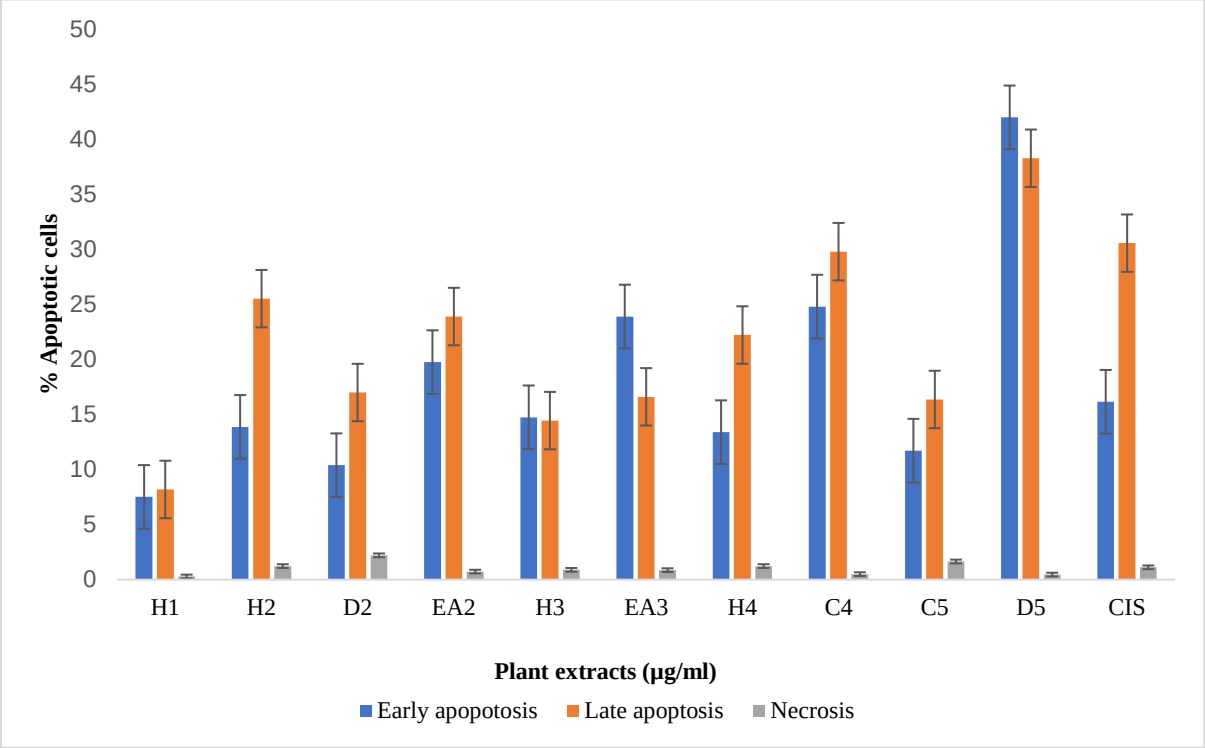


Figure C.10. Percentage of MDA-MB cells at early, late apoptotic and necrosis mode of death after being treated with *S. brachypetala* (H1), *R. caffra* (H2, D2, EA2), *S. molle* (H3, EA3,), *Z. mucronata* (H4, C4), *S. petersiana* (C5, D5), Cisplatin (CIS) on stained MDA-MB cells.

C.5 Discussion

Plants efficiently deploy sophisticated defense mechanisms to fight infections, which renders them immune to numerous pathogenic microorganisms. Phytocompounds are among the arsenal used by plants for combating microbial infections. This study explored the phytocompounds present in crude extracts from *S. brachypetala, S. petersiana, Z. mucronate, R. caffra,* and *S. molle* to decipher their antimycobacterial activity. The crude extracts from *R. caffra* exhibited strong growth inhibitory activity against *M. aurum* A+ (MIC ranging from 0.02-0.5 mg/mL) (Table C.2). Furthermore, *R. caffra* dichloromethane extracts exhibited potent growth inhibition against *M. tb* H37Rv (MIC range of 0.25-0.125 mg/mL.) (Table C.3). Traditional practitioners in Limpopo use

R. caffra concoctions to treat a wide range of diseases. Findings in this study confirm that extracts from *R. caffra* strongly inhibit *M. tb* as previously reported [10]. In this study, the chemical constituents from *R. caffra* extracts that might have synergistically contributed to the potent antimycobacterial activity were tentatively identified. Results from this study confirm the literature that showed that *R. caffra* is rich in alkaloids (Table C.4) [33]. Our study also supports a report by [34] which identified and isolated raucaffricine from *R. caffra* extracts. Extracts constituting alkaloids have previously been investigated for pharmacological activities and have been discovered to exhibit potent antimicrobial efficacy [10, 35].

In this study, crude extracts from *S. molle* presented strong antimycobacterial efficacy, resulting in high susceptibility patterns with low MIC values ranging from 0.02-0.5 mg/mL for *M. aurum* A+ and 0.25-0.125 mg/mL for *M. tb* H37Rv) shown in Table C.2 and Table C.3. A study by [36] demonstrated the inhibitory activity of extracts from *S. molle* against gram-positive bacteria (*Bacillus subtilis*). Another report by Bernardes et al. (2014) [37] demonstrated the significant susceptibility of *Mycobacterium bovis* BCG to the methanolic extract of *S. terebinthifolius* which belongs to the *Schinus* genus. To the best of our knowledge, this study is the first to report *in-vitro* anti-*M. tb* H37Rv activity of *S. molle.* In this study, tentative identification of phytocompounds constituting *S. molle* extracts showed the presence of sesquiterpene, triterpene, terpene, triterpenoid, and triterpenoid saponin. Our study agrees with the phytochemical profiling of *S. molle* by various studies which showed the presence of a wide array of potentially bioactive compounds from classes of compounds including sesquiterpenes, terpenes, and triterpenes [12, 13, 36]. Tannins, flavonoids, steroids, and catechins have also been reported to contribute to the antimicrobial activity of the *Schinus* genus [38]. Interestingly, oleanolic acid a triterpenoid tentatively identified in this study (Table C.5) was shown in other reports to exhibit anti-*M. tb* and

reduce hepatotoxicity [39–42]. In addition, sesquiterpenes, terpenes, triterpenes, and triterpenoids that were also tentatively identified are well characterized for antimicrobial activity and may have contributed to the anti-*M. tb* activity.

The dichloromethane and methanol extracts *of Z. mucronata* showed strong activity against *M. aurum* A+, with MIC values of 0.04 mg/mL and 0.25 mg/mL, respectively. However, poor efficacy was observed for *Z. mucronate, S. brachypetala,* and *S. petersiana* against *M. tb* H37Rv. Results in our study agree with a report by Mativandlela et al., (2008) [43] which illustrated that bark extracts of *Z. mucronata* exhibit poor activity against *M. smegmatis*. On the contrary, other reports demonstrated that a combination of the *Z. mucronate* leaf, bark, and root extracts exhibit potent activity (MIC $\leq$ 1mg/mL) against *M. tb* [2, 44]. The poor activity observed against *M. tb* by crude extracts from *Z. mucronate*, *S. brachypetala,* and *S. petersiana*, does not completely reflect on lack of activity of the extracts *in vivo* because some of the compounds may be enzymatically activated or transformed *in vivo.* The transformed intermediates may then be potent against *M. tb*.

Results from the rigorous VSW returned only one hit compound (norajmaline) (Table C.6). In addition, norajmaline did not violate the *Rule of Five*. Lipinski's Rule of 5 defines the potential drug-likeness of a compound based on the relationship between physiochemical, and pharmacokinetics parameters active [45]. The ADME properties of norajmaline observed in this study were moderate but are within acceptable ranges ("Schrödinger Software Release 2015-2," 2015) tabulated in Table C.6. There is a high failure of drugs in the clinical phases because of poor pharmacokinetic properties [47, 48]. Thus, this investigation used the ADME parameters as filters to avoid the identification of false hits (Table C.6). The molecular docking regimes with increasing precision in this study were performed as a further refining stage, with the objective of obtaining

a hit molecule(s) based on extra-precision scores and gaining molecular insight into the binding mechanism. The XP docking score of norajmaline was –7.47 kcal/mol (Table C.6). Computational screening provides an efficient approach to identifying, characterizing, and modifying potential drug leads [49, 50].

To expand the essential knowledge about the binding dynamic trends of the phytocompound against *M. tb* PanK, molecular dynamics simulations, and ΔG_{Bind} were computed. Norajmaline formed a relatively stable complex with PanK with RMSD below 3 Å (Figure C.2). A comparison of the unbound native PanK RMSD and that of PanK-norajmaline revealed that the interaction of norajmaline with the residues of the binding pocket of PanK resulted in a slight structural change of PanK (Figure C.2, and Figure C.3). Norajmaline exhibited high affinity (ΔG_{Bind} –58.73 kcal/mol) to the hydrophobic binding domain of PanK based on the MM-GBSA calculations (Table C.7). The control ligand used in this study **1f** is an engineered triazole competitive inhibitor of PanK [23]. A comparison of the MD simulations of the **1f** and norajmaline revealed that the **1f** had a higher affinity (ΔG_{Bind} –67.70 kcal/mol) to the PanK binding pocket than the norajmaline (ΔG_{Bind} –58.73 kcal/mol) as shown by the post-MM-GBSA (Table C.7). For a natural product, norajmaline has an interestingly high affinity. The pre-MM-GBSA ΔG_{Bind} of the norajmaline-PanK complex (–37.64 kcal/mol) is more than that of post-MM-GBSA ΔG_{Bind} –58.73 kcal/mol. Molecular knowledge derived from virtual screening of phytochemicals revealed that norajmaline may potentially competitively inhibit *M. tb* PanK. On the other hand, norajmaline can be further modified to enhance the affinity and ADME properties while at the same time enhancing activity.

The current study investigated the cytotoxicity properties evaluated by *S. brachypetala, R. caffra, S. molle, Z. macronata*, and *S. petersian* indigenous to Limpopo Province, South Africa against

MDA-MB 231 triple-negative breast cancer cell line. The hexane crude extracts of *S. brachypetala* (H1), *R. caffra* (H2, D2 and EA2), *S. molle* (H3 and EA3), *Z. macronata* (H4 and C4), and *S. petersian* (C5 and D5) induced >50% cell growth inhibition against the MDA-MB 231 triple-negative breast cancer line at 62.5, 125 and 250 µg/mL. All the plant extracts had an IC_{50} value less than 30 µg/mL, which meets the criteria set by the American National Cancer Institute for a potent extract [51, 52]. According to these guidelines, an IC_{50} value less than 30 µg/mL indicates that the extract has the potential to effectively halt the growth of the cancer cells [51, 52].

Toxicological assays on medicinal plants of this study indicate that crude extracts exhibited low toxicity towards Vero monkey kidney cells, which agreed with previous findings of Tlphapi et al. (2020) [11]: they found that the *R. caffra* crude extract, fractions and pure compounds did not display any cytotoxic effects at a concentration of 50 µg/mL against HeLa cells. This is also supported by the observation from Sigidi et al. 2016 [53] where *S. petersiana* had the lowest toxicity against Vero monkey kidney cells at a concentration lower than 50 µg/mL. It showed that these plant species could be further experimented with since they do not display toxicity against healthy cells. In comparison, the study by Sigidi et al. (2016) [53] found that *Z. mucronata* extracts had lower toxicity, with IC_{50} values ranging from 150 to 250 mg/mL against Vero monkey kidney and MeWo cells. Moreover, an aqueous extract of *Z. mucronata* was found to have lower toxicity against HepG2/C3A cells with an IC_{50} value greater than 100 µg/mL that was reported by Da Costa Mousinho et al. (2013) [54]. These findings reinforce the observed low toxicity of *Z. mucronata* on HepG2/C3A cells in this study (Figure C.6 and C.7).

However, the hexane extract of *S. molle* showed the highest cytotoxic inhibition activity on the HepG2/C3A cell as reported by Nagah et al. (2021) [55] which is consistent with the results

obtained in this study (Figure C.6). The findings of the current study were supported by the study of Dzoyem et al. (2016) [44], which found that *Z. mucronata* and other indigenous plants were relatively safe compared to the positive control when tested on Vero cells. Additionally, bark extracts from *S. brachypetala* and *Z. mucronata* were found to be nontoxic against brine shrimp [44], which concurs with the results observed in this study on Vero cells (Figure C.7). Contrarily, Ruffa et al. (2002) [56] reported that the methanolic extract of *S. molle* was highly lethal to the human hepatoma HepG2 cell line, in contrast to extracts from other medicinal plants. This observation was supported by Hailan et al. (2022) [57], where nanoparticles synthesized from *S. molle* extracts were found to be highly toxic against HepG2 cells [58]. In that study, they attributed the high cytotoxicity of *S. molle* to several terpenoid compounds identified in its essential oil (Hailan et al. 2022). *S. molle* extracts are commonly used as insecticides due to their toxic properties [59, 60].

The results of this study indicate that various solvent extracts from five different plants (*S. brachypeta*, *R. caffra*, *S. molle*, *Z. mucronata*, and *S. petersiana*) have varying levels of cytotoxic activity against cancer cells. The IC_{50} values of these extracts ranged from 1.525 ± 0.458 to 26.26 ± 2.325, with the dichloromethane extract of *S. petersiana* (D5) and the ethyl acetate extract of *S. molle* (EA3) showing the highest potency with IC_{50} values of 1.525 ± 0.458 and 2.426 ± 0.141, respectively. Other extracts showed moderate antiproliferative activity with IC_{50} values ranging from 6.829 ± 0.37 to 8.625 ± 0.37. The control drug, cisplatin, showed potent antiproliferative activity with an IC_{50} value of 2.017 ± 0.09. It is worth noticing that the dichloromethane extract of *S. petersiana* (D5) and the ethyl acetate extract of *S. molle* (EA3) (2.426 ± 0.141) showed similar IC_{50} as the control drug (Figure C.8) (Table C.8).

The results of this study (Figure C.8 and Table C.8) are consistent with previous findings in the literature, which have shown that various plant extracts have antiproliferative activity against cancer cells. Nguefack et al. (2017) found that extracts from the root bark of *Rauvolfia vomitoria*, a close relative of *R. caffra*, showed significant antiproliferative activity against human breast cancer cells (MCF-7) [61]. Another study by Agbo et al. (2015) found that the ethanol extract of *Ziziphus mauritiana*, a close relative of *Z. mucronata*, showed antiproliferative activity against human liver cancer cells (HeLa cells) and breast cancer cells (MCF-7 cells) [62]. The IC_{50} values of the plant extracts in the present study are also similar to those reported for other plant-based anti-cancer agents. A study by Kim et al. (2010) found that the IC_{50} value of an ethanol extract from *Scutellaria baicalensis*, a traditional medicinal plant, was 2.55 ± 0.35 against human breast cancer [63]. Another study by Kim et al. (2012) found that the IC_{50} value of a methanol extract from *Tripterygium wilfordii*, another traditional medicinal plant, was 4.66 ± 0.54 against human leukaemia cells (HL-60 cells) through the mitochondrial pathway [64]. The results of this study indicate that various solvent extracts from *S. brachypeta, R. caffra, S. molle, Z. mucronata*, and *S. petersiana* exhibit varying degrees of antiproliferative activity against the MDA-MB 231 triple-negative cancer cells. To the best of our knowledge, this study is the first to report the cytotoxicity activities of these five medicinal plants from Limpopo against MDA-MB 231 triple-negative breast cancer. This highlights the potential of these plant extracts as alternative sources of anti-cancer agents against triple-negative breast cancer, but further studies are needed to confirm their activity and determine their potential as therapeutic agents [28].

Apoptosis is a type of cell death linked to cancer [65, 66] Apoptosis signalling has been used as a cancer therapy [65–68] The current study evaluated the apoptotic effects of various plant species on cancer cells using Annexin V and propidium iodide staining procedure via flow cytometry. The

results were presented in contour plots for apoptotic and necrotic cell death (Figure C.9 to C.10). Cisplatin (16.15 % apoptosis and 1.12 % necrosis) was used as a positive control and 0.25% DMSO as an untreated control. The study found that *S. petersiana* extract (D5) exhibited an early apoptotic stage at 41.97 % (Figure C.9.K and C.10), which is in agreement with literature that revealed the anticancer potential of the extracts [69]. *S. brachypetala* (H1) was observed to have a high percentage of live cells (84%) with low apoptotic and necrotic activities (7.51% and 0.29%, respectively) (Figure C.9.C and C.10) compared to other extracts, although phenolic compounds from *S. brachypetala* have been reported to have therapeutic effects against Alzheimer's disease [70, 71] The essential oil from *R. caffra* was found to be effective against MCF-7 breast cancer cells, possibly due to its high antioxidant activity [72, 73]. *S. molle* extracts were reported to have anticancer effects on human leukaemia cells (HL-60 cells) [12] and showed toxicity to the U-937 cell line and anti-tumour activity against human leukaemia monocyte lymphoma [73]. Silver nanoparticles synthesized from *S. molle* extracts were also observed to exhibit potential anticancer activity against HepG2 cells, [57]. All these findings agree with the finding from this current study. It is worth noting that the chloroform extracts from *Z. mucronata* (C4) were observed to induce higher late apoptotic cell death (29.79%) than the hexane extracts (H4) (22.21%) (Figure C.9.J, C.9.G and C.10), suggesting that compounds with anticancer activity could have intermediate or opposite polarity [75]. It is worth noting that the findings from this current study revealed that four plant extracts induced higher levels of apoptosis compared to the control drug, cisplatin. The highest level of apoptosis was observed with the dichloromethane extract of *S. petersiana* (D5) at 41.97%, followed by the chloroform extract of *Z. mucronata* (C4) at 24.74%, the ethyl acetates crude plant extracts of *S. molle* (EA3) at 23.89%, and *R. caffra* (EA2) at 19.75%. The control drug,

cisplatin, induced apoptosis at 16.15%. These findings suggest that the crude plant extracts tested in this study could potentially be used as alternative treatments for triple-negative breast cancer.

C.6 Conclusion

Medicinal plants are an invaluable source of potent bioactive anti-mycobacterial phytocompounds. *R. caffra* and *S. molle* which are prescribed by traditional healers in Limpopo are potent against *M. tb*and assist the developing community in alleviating and treating Tuberculosis. The integration of targeted virtual screening can rapidly and effectively be employed to identify potential lead compounds. Rigorous virtual screening comprising many filtering parameters reduces the chances of obtaining false positives. Norajmaline showed exceptionally high affinity to PanK and may be further modified to enhance affinity and ADME properties. This study offered a glimpse into the cytotoxic activities and antioxidant attributes of five selected medicinal plants from Limpopo, South Africa, against MDA-MB triple-negative breast cancer. The findings indicated that apoptosis was the form of cell death against MDA-MB triple-negative breast cancer.

Recommendation

The crude extracts of the plants comprise a wide variety of compounds that can further be separated and experimentally evaluated for anti-mycobacterial activity. Although the findings indicate that the selected medicinal crude plant extracts have the potential to treat triple-negative breast cancer, more exploration is needed to delve into the plants' mechanisms of action and isolate the bioactive elements responsible for the plant's anticancer properties in the *in vitro* study.

Ethics Approval and Consent to Participate

Ethical approval for this study was approved by the Research Ethics Committee: Biological and Environmental Safety (REC: BES) of Stellenbosch University with a reference number **BEE-2022-3188**. Ethical approval or permit for plants study were not applicable in this study however, institutional, national and international guidelines and legislation were followed in this study

Consent for Publication

Not applicable.

Data Availability Statement:

The datasets used and analyses during the current study are available from the corresponding authors on reasonable request.

Declaration of Competing Interest:

The authors declare no conflict of interest.

Funding: This work was supported by the South African Medical Research Association (SAMRC), the Centre of Excellence for Biomedical Tuberculosis Research (CBTBR), and the National Research Fund (NRF) [NRF GRANT UID129364].

Author Contributions: Conceptualization, V.M., V.M., A.T., F.A., L.M., and K.N.; methodology, F.A. V.M., A.T., V.M., K.N., K.G., and M.N; formal analysis, F.A., K.N., V.M.; investigation, F.A., V.M., K.N.; resources, V.M., A.T., M.N., L.S., K.G., and I.SN.; writing original draft preparation, F.A. V.M., K.N.; writing review and editing, V.M., A.T., M.N., L.M., K.G., K.I.T., L.B., L.S., N.T., and L.J; supervision, V.M., M.N., A.T., K.G., L.S., N.T., and I.S.N; project administration, L.S.; funding acquisition, V.M., and L.S. Molecular docking and Molecular

dynamics simulations, K.N., K.I.T., K.G., and M.N. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: We acknowledge the Open Access Fund of Stellenbosch University and Prof Maryna van de Venter, Nelson Mandela University South Africa. We would also like to extend our appreciation to the Center for High-Performance Computing, South Africa, for providing us with access to the infrastructure used to conduct the simulations.

8.7 References

1. Chihota V, Apers L, Mungofa S, Kasongo W, Nyoni IM, Tembwe R, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in regions of Southern Africa. International Journal of Tuberculosis and Lung Disease. 2007;11:311–8.

2. Green E, Samie A, Obi CL, Bessong PO, Ndip RN. Inhibitory properties of selected South African medicinal plants against *Mycobacterium tuberculosis*. J Ethnopharmacol. 2010;130:151–7.

3. Mphahlele M, Syre H, Valvatne H, Stavrum R, Mannsåker T, Muthivhi T, et al. Pyrazinamide resistance among South African multidrug-resistant *Mycobacterium tuberculosis* isolates. J Clin Microbiol. 2008;46:3459–64.

4. Evans JC, Murugesan D, Post JM, Mendes V, Wang Z, Nahiyaan N, et al. Targeting *Mycobacterium tuberculosis* CoaBC through Chemical Inhibition of 4′-Phosphopantothenoyl-L-cysteine Synthetase (CoaB) Activity. https://doi.org/10.1021/acsinfecdis.0c00904.

5. Traoré L, Yaro VSO, Soudré A, Ouédraogo-Koné S, Ouédraogo D, Yougbaré B, et al. Indigenous knowledge of veterinary medicinal plant use in cattle treatment in southwestern Burkina Faso (West Africa). South African Journal of Botany. 2020;128:189–99.

6. Nitcheu Ngemakwe PH, Remize F, Thaoge ML, Sivakumar D. Phytochemical and nutritional properties of underutilised fruits in the southern African region. South African Journal of Botany. 2017;113:137–49.

7. Vanisree M, Lee CY, Lo SF, Nalawade SM, Lin CY, Tsay HS. Studies on the production of some important secondary metabolites from medicinal plants by plant tissue cultures. Botanical Bulletin of Academia Sinica. 2004;45:1–22.

8. Makhado RA, Potgieter MJ, Policy F, Resources N. Colophospermum mopane Wood Utilisation in the Northeast of the Limpopo Province , South Africa. 2009.

9. Ajao AA, Moteetee AN. Tithonia diversifolia (Hemsl) A. Gray. (Asteraceae: Heliantheae), an invasive plant of significant ethnopharmacological importance: A review. South African Journal of Botany. 2017;113:396–403.

10. Ebeh Messanga R, Dominique Serge NB, Abouem A. Zintchem A, Norbert MN, Esther Del Florence MN, Patrick Hervé BD, et al. Rauvolfianine, a new antimycobacterial glyceroglycolipid and other constituents from Rauvolfia caffra. Sond (Apocynaceae). https://doi.org/101080/14786419201713568 32. 2017;32:1971–6.

11. Tlhapi DB, Ramaite IDI, Anokwuru CP, van Ree T, Hoppe HC. In Vitro Studies on Antioxidant and Anti-Parasitic Activities of Compounds Isolated from Rauvolfia caffra Sond. Molecules 2020, Vol 25, Page 3781. 2020;25:3781.

12. Garzoli S, Masci VL, Ovidi E, Turchetti G, Zago D, Tiezzi A. Chemical Investigation of a Biologically Active Schinus molle L. Leaf Extract. J Anal Methods Chem. 2019;2019.

13. Bvenura C, Kambizi L. Composition of Phenolic Compounds in South African Schinus molle L. Berries. Foods 2022, Vol 11, Page 1376. 2022;11:1376.

14. Mongalo NI, Mashele SS, Makhafola TJ. Ziziphus mucronata Willd. (Rhamnaceae): it's botany, toxicity, phytochemistry and pharmacological activities. Heliyon. 2020;6.

15. Arévalo JMC, Amorim JC. Virtual screening, optimization and molecular dynamics analyses highlighting a pyrrolo[1,2-a]quinazoline derivative as a potential inhibitor of DNA gyrase B of *Mycobacterium tuberculosis*. Scientific Reports 2022 12:1. 2022;12:1–13.

16. Ejalonibu MA, Elrashedy AA, Lawal MM, Kumalo HM, Mhlongo NN. Probing the dual inhibitory mechanisms of novel thiophenecarboxamide derivatives against *Mycobacterium tuberculosis* PyrG and PanK: an insight from biomolecular modeling study. https://doi.org/101080/0739110220201844055. 2020;40:2978–90.

17. Chiarelli LR, Mori G, Orena BS, Esposito M, Lane T, de Jesus Lopes Ribeiro AL, et al. A multitarget approach to drug discovery inhibiting *Mycobacterium tuberculosis* PyrG and PanK. Sci Rep. 2018;8.

18. Vojtek M, Marques MPM, Ferreira IM, Mota-Filipe H, Diniz C. Anticancer activity of palladium-based complexes against triple-negative breast cancer. Drug Discov Today. 2019;24:1044–58.

19. Nan Y, Su H, Zhou B, Liu S. The function of natural compounds in important anticancer mechanisms. Front Oncol. 2023;12:7082.

20. Song J, Zhang B, Li M, Zhang J. The current scenario of naturally occurring indole alkaloids with anticancer potential. Fitoterapia. 2023;165.

21. Tapfuma KI, Nyambo K, Adu-Amankwaah F, Baatjies L, Smith L, Allie N, et al. Antimycobacterial activity and molecular docking of methanolic extracts and compounds of marine fungi from Saldanha and False Bays, South Africa. Heliyon. 2022;8:e12406.

22. Baptista R, Bhowmick S, Shen J, Mur LAJ. Molecular Docking Suggests the Targets of Anti-Mycobacterial Natural Products. Molecules. 2021;26:475.

23. Bjorkelid C, Bergfors T, Raichurkar AK v., Mukherjee K, Malolanarasimhan K, Bandodkar B, et al. Structural and biochemical characterization of compounds inhibiting *Mycobacterium tuberculosis* pantothenate kinase. Journal of Biological Chemistry. 2013;288:18260–70.

24. Schrödinger. LigPrep | Schrödinger. Schrödinger Release 2018-1. 2021.

25. Kumar BK, Faheem, Sekhar KVGC, Ojha R, Prajapati VK, Pai A, et al. Pharmacophore based virtual screening, molecular docking, molecular dynamics and MM-GBSA approach for identification of prospective SARS-CoV-2 inhibitor from natural product databases. J Biomol Struct Dyn. 2022;40:1363–86.

26. Zong K, Xu L, Hou Y, Zhang Q, Che J, Zhao L, et al. molecules Virtual Screening and Molecular Dynamics Simulation Study of Influenza Polymerase PB2 Inhibitors. 2021. https://doi.org/10.3390/molecules26226944.

27. Ejalonibu MA, Elrashedy AA, Lawal MM, Mhlongo NN, Kumalo HM. Pharmacophore mapping of the crucial mediators of dual inhibitor activity of PanK and PyrG in tuberculosis disease. Mol Simul. 2022;48.

28. Adu-Amankwaah F, Tapfuma KI, Hussan RH, Tshililo N, Baatjies L, Masiphephethu MV, et al. Cytotoxic activity of Cape Fynbos against triple-negative breast cancer cell line. South African Journal of Botany. 2022;150:702–10.

29. Sowemimo A, Venables L, Odedeji M, Koekemoer T, Van De Venter M, Hongbing L. Antiproliferative mechanism of the methanolic extract of Enterolobium cyclocarpum (Jacq.) Griseb. (Fabaceae). J Ethnopharmacol. 2015;159:257–61.

30. Olaru OT, Venables L, Van De Venter M, Nitulescu GM, Margina D, Spndidos DA, et al. Anticancer potential of selected Fallopia Adans species. Oncol Lett. 2015;10:1323–32.

31. Adu-Amankwaah F, Tapfuma KI, Hussan RH, Tshililo N, Baatjies L, Masiphephethu MV, et al. Cytotoxic activity of Cape Fynbos against triple-negative breast cancer cell line. South African Journal of Botany. 2022;150:702–10.

32. Baharuddin AA, Roosli RAJ, Zakaria ZA, Md. Tohid SF. *Dicranopteris linearis* extract inhibits the proliferation of human breast cancer cell line (MDA-MB-231) via induction of S-phase arrest and apoptosis. Pharm Biol. 2018;56.

33. Milugo TK, Omosa LK, Ochanda JO, Owuor BO, Wamunyokoli FA, Oyugi JO, et al. Antagonistic effect of alkaloids and saponins on bioactivity in the quinine tree (Rauvolfia caffra sond.): Further evidence to support biotechnology in traditional medicinal plants. BMC Complement Altern Med. 2013;13.

34. Tlhapi DB, Ramaite IDI, van Ree T, Anokwuru CP, Orazio TS, Hoppe HC. Isolation, chemical profile and antimalarial activities of bioactive compounds from rauvolfia caffra sond. Molecules. 2019;24.

35. Bitombo AN, Zintchem AAA, Atchadé A de T, Moni Ndedi EDF, Khan A, Ngono Bikobo DS, et al. Antimicrobial and cytotoxic activities of indole alkaloids and other constituents from the stem barks of Rauvolfia caffra Sond (Apocynaceae). Nat Prod Res. 2022;36.

36. Turchetti G, Garzoli S, Masci VL, Sabia C, Iseppi R, Giacomello P, et al. Antimicrobial testing of schinus molle (l.) leaf extracts and fractions followed by gc-ms investigation of biological active fractions. Molecules. 2020;25.

37. Bernardes NR, Heggdorne-Araújo M, Borges IFJC, Almeida FM, Amaral EP, Lasunskaia EB, et al. Nitric oxide production, inhibitory, antioxidant and antimycobacterial activities of the fruits extract and flavonoid content of Schinus terebinthifolius. Revista Brasileira de Farmacognosia. 2014;24.

38. El-Nashar HAS, Mostafa NM, Abd El-Ghffar EA, Eldahshan OA, Singab ANB. The genus Schinus (Anacardiaceae): a review on phytochemicals and biological aspects. Natural Product Research. 2022;36.

39. Lin YN, Chen CJ, Chang HY, Cheng WK, Lee YR, Chen JJ, et al. Oleanolic acid-mediated inhibition of pregnane X receptor and constitutive androstane receptor attenuates rifampin-isoniazid cytotoxicity. J Agric Food Chem. 2017;65.

40. Ge F, Zeng F, Liu S, Guo N, Ye H, Song Y, et al. In vitro synergistic interactions of oleanolic acid in combination with isoniazid, rifampicin or ethambutol against *Mycobacterium tuberculosis*. J Med Microbiol. 2010;59.

41. Khameneh B, Iranshahy M, Vahdati-Mashhadian N, Sahebkar A, Fazly Bazzaz BS. Non-antibiotic adjunctive therapy: A promising approach to fight tuberculosis. Pharmacological Research. 2019;146.

42. Fadipe VO, Opoku AR, Singh M, Pereira AR, Rijo P, Mongalo NI. Antimycobacterial, antiplasmodial studies and cytotoxicity of oleanolic acid and its derivative from Syzygium aromaticum Linn (Myrtaceae). Biomedical and Biopharmaceutical Research. 2020;17.

43. Mativandlela SPN, Meyer JJM, Hussein AA, Houghton PJ, Hamilton CJ, Lall N. Activity against Mycobacterium smegmatis and *M. tb*by extract of South African medicinal plants. Phytotherapy Research. 2008;22.

44. Dzoyem JP, Aro AO, McGaw LJ, Eloff JN. Antimycobacterial activity against different pathogens and selectivity index of fourteen medicinal plants used in southern Africa to treat tuberculosis and respiratory ailments. South African Journal of Botany. 2016;102.

45. Khairy A, Hammoda HM, Celik I, Zaatout HH, Ibrahim RS. Discovery of potential natural dihydroorotate dehydrogenase inhibitors and their synergism with brequinar via integrated molecular docking, dynamic simulations and in vitro approach. Scientific Reports |. 123AD;12:19037.

46. Schrödinger Press Virtual Screening Workflow Virtual Screening Workflow Schrödinger Software Release 2015-2. 2015.

47. Lu W, Zhang R, Jiang H, Zhang H, Luo C. Computer-aided drug design in epigenetics. Front Chem. 2018;6 MAR:1–23.

48. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. 2018.

49. Yu MJ. Natural Product-Like Virtual Libraries: Recursive Atom-Based Enumeration. J Chem Inf Model. 2011;51:541–57.

50. Zong K, Xu L, Hou Y, Zhang Q, Che J, Zhao L, et al. molecules Virtual Screening and Molecular Dynamics Simulation Study of Influenza Polymerase PB2 Inhibitors. 2021. https://doi.org/10.3390/molecules26226944.

51. Abdel Fattah Hussein A, Mohammed A, ElMwafy H. Chemical and Bioactivity Studies on Salvia Africana-Lutea: Cytotoxicity and Apoptosis Induction by Abietane Diterpenes Isolated from Salvia Africana-Lutea. Journal of Basic and Environmental Sciences. 2018;5:72–9.

52. Wang X yang, Liu L ping, Zhu R xin, Kang T guo, Tong L jiang, Xie H, et al. Cytotoxic activities of some selected medicinal plants of the genus euphorbia. J Med Plant Res. 2011;5:6766–9.

53. Sigidi MT, Anokwuru CP, Zininga T, Tshisikhawe MP, Shonhai A, Ramaite IDI, et al. Comparative in vitro cytotoxic, anti-inflammatory and anti-microbiological activities of two indigenous Venda medicinal plants. Transl Med Commun. 2016;1:9.

54. Da Costa Mousinho NMH, van Tonder JJ, Steenkamp V. *In Vitro* Anti-diabetic Activity of *Sclerocarya Birrea* and *Ziziphus Mucronata*. Nat Prod Commun. 2013;8:1934578X1300800.

55. Nagah N, Mostafa I, Osman A, Dora G, El-Sayed Z, Ateya A-M. Bioguided isolation and in-silico analysis of Hep-G2 cytotoxic constituents from Laurus nobilis Linn. cultivated in Egypt. Egypt J Chem. 2021;0:0–0.

56. Ruffa MJ, Ferraro G, Wagner ML, Calcagno ML, Campos RH, Cavallaro L. Cytotoxic effect of Argentine medicinal plant extracts on human hepatocellular carcinoma cell line. J Ethnopharmacol. 2002;79:335–9.

57. Hailan WA, Al-Anazi KM, Farah MA, Ali MA, Al-Kawmani AA, Abou-Tarboush FM. Reactive Oxygen Species-Mediated Cytotoxicity in Liver Carcinoma Cells Induced by Silver Nanoparticles Biosynthesized Using Schinus molle Extract. Nanomaterials. 2022;12:161.

58. Díaz C, Quesada S, Brenes O, Aguilar G, Cicció JF. Chemical composition of *Schinus molle* essential oil and its cytotoxic activity on tumour cell lines. Nat Prod Res. 2008;22:1521–34.

59. Abdel-Sattar E, Zaitoun AA, Farag MA, Gayed SH el, Harraz FMH. Chemical composition, insecticidal and insect repellent activity of *Schinus molle* L. leaf and fruit essential oils against *Trogoderma granarium* and *Tribolium castaneum*. Nat Prod Res. 2010;24:226–35.

60. Huerta A, Chiffelle I, Puga K, Azúa F, Araya JE. Toxicity and repellence of aqueous and ethanolic extracts from Schinus molle on elm leaf beetle Xanthogaleruca luteola. Crop Protection. 2010;29:1118–23.

61. Nguefack J, Kamtchouing P, Nkengfack AE. Chemical composition, antioxidant and antiproliferative activities of Rauvolfia vomitoria root bark extract. Journal of Medical Plants Research. 2017;11:144–9.

62. Agbo FN, Olaleye MT, Okigbo RN, Afolayan AJ. In vitro anti-proliferative and antioxidant activities of Ziziphus mauritiana Lam. (Rhamnaceae) root bark extract. J Ethnopharmacol. 2015;174:304–9.

63. Kim HJ, Lee JH, Kim HS, Kim YH. Anti-proliferative and apoptotic effects of Scutellaria baicalensis root on human cancer cells. Journal of Ethnopharmacolog. 2010;132:466–71.

64. Kim HJ, Lee JH, Kim HS, Kim YH. Tripterygium wilfordii Hook F. root extract induces apoptosis in human leukemia cells through the mitochondrial pathway. J Ethnopharmacol. 2012;143:529–35.

65. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144:646–74.

66. Hanahan D, Weinberg RA. The Hallmarks of Cancer Review evolve progressively from normalcy via a series of pre. 2000.

67. Los M, Burek CJ, Stroh C, Benedyk K, Hug H, Mackiewicz A. Anticancer drugs of tomorrow: Apoptotic pathways as targets for drug design. Drug Discov Today. 2003;8:67–77.

68. Wong RS. Apoptosis in cancer: from pathogenesis to treatment. Journal of Experimental & Clinical Cancer Research. 2011;30.

69. Alshehri MM, Quispe C, Herrera-Bravo J, Sharifi-Rad J, Tutuncu S, Aydar EF, et al. A Review of Recent Studies on the Antioxidant and Anti-Infectious Properties of Senna Plants. Oxid Med Cell Longev. 2022;2022.

70. Hassaan Y, Handoussa H, El-Khatib AH, Linscheid MW, el Sayed N, Ayoub N. Evaluation of Plant Phenolic Metabolites as a Source of Alzheimer's Drug Leads. Biomed Res Int. 2014;2014:1–10.

71. Thakur A, Chun YS, October N, Yang HO, Maharaj V. Potential of South African medicinal plants targeting the reduction of Aβ42 protein as a treatment of Alzheimer's disease. J Ethnopharmacol. 2019;231:363–73.

72. Bendaoud H, Romdhane M, Souchard JP, Cazaux S, Bouajila J. Chemical Composition and Anticancer and Antioxidant Activities of Schinus Molle L. and Schinus Terebinthifolius Raddi Berries Essential Oils. J Food Sci. 2010;75:C466–72.

73. Calzada F, Solares-Pascasio J, Valdes M, Garcia-Hernandez N, Velázquez C, Ordoñez-Razo R, et al. Antilymphoma potential of the ethanol extract and rutin obtained of the leaves from schinus molle linn. Pharmacognosy Res. 2018;10:119.

74. Garzoli S, Masci VL, Ovidi E, Turchetti G, Zago D, Tiezzi A. Chemical Investigation of a Biologically Active *Schinus molle* L. Leaf Extract. J Anal Methods Chem. 2019;2019:1–6.

75. Kandawa-Schulz M, El-Sayed HRL. Anticancer, Antioxidant And Antimicrobial Screening Of Extracts From Ziziphus Mucronata, Heliotropium Ciliatum And Gnidia Polycephala From The Oshikoto Region Of Namibia A Thesis Submitted In Partial Fulfillment Of The Requirement For The Degree Of Master Of Science (Chemistry) Of The University Of Namibia By Secilia Kaenda Ilonga (200507290). 2012.