Strategic application of in silico drug discovery approaches to discover novel TB drugs

Miss Kimona Kisten

2023

A thesis submitted to the College of Health Sciences, School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, in fulfilment for the degree of Doctor of Philosophy in Medical Science

This is the thesis in which the chapters are written as a set of discrete research publications, with an overall introduction and final summary. Typically, these chapters will have been published in internationally recognised, peer-reviewed journals.

This is to certify that the contents of this thesis are the original research work of Miss Kimona Kisten.

As the candidate' supervisor, I have approved this thesis for submission.

Supervisor: Dr N.N. Mhlongo Signed: [redacted] Date: 30-11-2023

Co-supervisor: Dr H.M. Kumalo

PREFACE

This thesis is divided into the following chapters:

Chapter 1:

This chapter serves as an introduction into the background and rationale of the study comprising of the aim and objectives.

Chapter 2:

This chapter is an extensive literature review compiled on *M. tuberculosis* and the reason behind the continuous research facilitating the design, discovery, and development of more effective potential inhibitors for drug targets of interest, which address challenges that are posed in the efficacy of current market-available anti-TB drugs. This chapter includes the design, discovery, pathogenesis, transmission, virality, clinical features, mechanistic and structural characteristics, drug targets and statistics associated in this field.

Chapter 3:

This chapter explains the method perspectives of computer aided drug design by including various molecular dynamic approaches and applications, the computational tools and software required to study inhibition capability of drugs as well as methods to analyse binding affinity. This chapter conceptualises drug design by looking at an in depth understanding of the principles behind the methods.

Chapter 4:

This chapter is titled: Drug repurposing Approach against *Mycobacterium tuberculosis* Enoyl-[acyl-carrier-protein] reductase: Insight from Molecular Dynamics Simulations. The study investigates the outcome of attempted drug repurposing of inhA via a molecular modelling approach. This article has been published in *Molecular simulations*.

Chapter 5:

This chapter is titled: Evaluating the impact of C171Q mutation on the potency of Thiolactomycin to *M. tuberculosis* KasA binding pocket: Insights from molecular dynamics simulations and tailored-pharmacophore studies. The work investigates the outcome obtained from an enhanced tailored pharmacophore approach applied across a mutation using the approach of virtual screening.

Chapter 6:

This chapter is titled: Is Gene ontology an effective approach in drug repurposing? A case study on *M. tuberculosis DHPS (folP1)*: Insight from a docking and MM/PBSA perspective in estimating ligand binding affinities of potential inhibitors. The study investigates the effectiveness of the RepTB gene ontology-based approach by conducting a drug repurposing study on folp1.

Chapter 7:

This is the final chapter that proposes concluding remarks and future work.

ABSTRACT

Tuberculosis is an evolving cause for concern due to it being one of the major causes of mortality worldwide because of the onset of bacterial infection. It exists as a continuous field of study as the emergence of drug resistant strains have become rife whereby first- and second-line drugs can no longer be a means and method of treatment due to increasing ineffectiveness. Variations associated with proteins in the bacteria, *Mycobacterium tuberculosis*, can be attributed largely to poor compliance of patients and the existence of co-morbidities within a person, inevitably leading to coinfections and a diminished immune response. This evolution of the bacteria therefore calls for extensive research to be carried out on enhanced drug design and development techniques. Re-evaluation of previously identified drug targets, determination of newly viable drug targets and the identification and design of small molecule inhibitors form a basis on the forefront of this research.

Computer aided drug design techniques provide a novel method exponentially gaining popularity. Molecular modelling utilizing in silico developed methods, paired with drug repurposing, pose a viable, cost effective and efficient solution for drug design strategies. The use of chemical interaction data associated with small molecules forming complexes with drug targets of interest help understand the behaviour and mechanism of a protein before graduating to wet method techniques. Using the fundamentals of docking, molecular dynamics and virtual screening, various small molecules with the potential to provide extensive inhibition capability can be identified. This study investigates three major drug targets of *Mycobacterium tuberculosis (Mtb)*: Enoyl-[acyl-carrier-protein] reductase (*inhA*), β-ketoacyl ACP synthase (*KasA*), and Dihydropteroate synthase (*DHPS*/*folP1*) involved in the mycolic acid and folate pathways. Various tools inclusive of gene ontology, network-based inference, virtual screening and tailored pharmacophore as a function of molecular modelling and drug repurposing were used to identify potential potent inhibitors and comprehend the understanding of the structural changes, conformations and interactions associated with the protein and the response to suggested drug hits with the potential to affect an overall protein structure in consideration with the formation of a complex.

The first study (Chapter 4) employs the method of drug repurposing on a well-known target of interest: inhA. Introducing this approach in a detailed manner, current suggested FDA approved drugs, were complexed to *M. tuberculosis*-inhA and investigated using molecular modelling. Four proposed drug hits with DrugBank IDs of DB04007, DB05291, DB07453 and

DB08607 were analysed based on the increased overall binding free energies and MD analysis suggested and reiterated the potential future use of DB04007.

The second study (Chapter 5) introduces the efficient technique of a tailored pharmacophore paired with virtual screening. A thermodynamic molecular modelling-based approach is applied to investigate the impact of the C171Q mutation on the KasA-TLM complex. Using the pharmacophore moieties established from the standard drug complex's intermolecular forces of residues, a scaffold is built and screened to obtain a small molecule encompassing improved receptor-ligand interactions. The drug hit with PubChem ID 44207286 exhibited higher binding free energies associated with a stable and rigid system in comparison to the control.

The third study (Chapter 6) focuses solely on the gene ontology network-based inference platform of study. Drug repurposing was compared to traditional design and the *RepTB* approach was analysed by conducting a case study on folP1. Three gene ontology based predicted drugs with DrugBank IDs of DB03705, DB04047 DB04196 were analysed using molecular dynamics and compared to the control, DB03592. MD analysis and binding free energies showed the most stable systems being attributed to DB04047 and DB04196 thereby proving further reliability of the approach.

This series of studies carried out provides an in-depth in silico perspective into drug design research aimed at combatting *M. tuberculosis*. It serves as a basis in the identification of potential potent novel inhibitors. This approach can potentially serve as a platform to the development and discovery of novel drugs against a wide range of drug targets.

DECLARATION I – PLAGIARISM

I, Kimona Kisten, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been re-written, but the general information attributed to them has been referenced.
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference sections.

A detailed contribution to publications that form part or/and include research presented in this thesis is stated (include publications submitted, accepted, in press and published).

Signed: K. Kisten

DECLARATION II – LIST OF PUBLICATIONS

1. Kisten, K., Kumalo, H. M., Machaba, K. E., Ndagi, U., & Mhlongo, N. N. (2021). Drug repurposing approach against *Mycobacterium tuberculosis* Enoyl-[acyl-carrier-protein] reductase: insight from molecular dynamics simulations. *Molecular Simulation*, *47*(16), 1313-1325. (*Published*)

Contribution:

Kimona Kisten: contributed to the project by performing all calculations and manuscript preparation and writing.

Dr, N.N. Mhlongo: Supervisor

Dr. H.M. Kumalo: Co-supervisor

2. Tailored-pharmacophore and drug repurposing approach in the evaluation of the C171Q mutation in drug-resistant *Mycobacterium tuberculosis* β ketoacyl - ACP synthetase (*Mtb* - KasA) from a molecular modelling perspective. (*Not published*)

Contribution:

Kimona Kisten: contributed to the project by performing all calculations and manuscript preparation and writing.

Dr. K.E. Machaba: Assisted in manuscript writing

Dr, N.N. Mhlongo: Supervisor

Dr. H.M. Kumalo: Co-supervisor

3. Is Gene ontology an effective approach in drug repurposing? A case study on *M. tuberculosis DHPS (folP1)*: Insight from a docking and MM/PBSA perspective in estimating ligand binding affinities of potential inhibitors. (*Not published*)

Contribution:

Kimona Kisten: contributed to the project by performing calculations and manuscript preparation and writing.

Dr. K.E. Machaba: Assisted in manuscript writing

Nolwazi Ngidi: Assisted in calculations

Dr, N.N. Mhlongo: Supervisor

Dr. H.M. Kumalo: Co-supervisor

RESEARCH OUTPUT

A. LIST OF PUBLICATIONS

1. Drug repurposing Approach against *Mycobacterium tuberculosis* Enoyl-[acyl-carrier-protein] reductase: Insight from Molecular Dynamics Simulations.

B. SUBMITTED

2. Tailored-pharmacophore and drug repurposing approach in the evaluation of the C171Q mutation in drug-resistant *Mycobacterium tuberculosis* β ketoacyl - ACP synthetase (*Mtb* - KasA) from a molecular modelling perspective.

3. Is Gene ontology an effective approach in drug repurposing? A case study on *M. tuberculosis DHPS (folP1)*: Insight from a docking and MM/PBSA perspective in estimating ligand binding affinities of potential inhibitors.

INSPIRATION

“Fantasy is escapist, and that is its glory. If a soldier is imprisoned by the enemy, don't we consider it his duty to escape? If we value the freedom of mind and soul, if we're partisans of liberty, then it's our plain duty to escape, and to take as many people with us as we can!”

— J.R.R. Tolkien

DEDICATION

To all the beautiful souls who have passed on but have been a part of this journey, encouraging me, motivating me, supporting me, believing in me, or just existing as a beacon in my life.

ACKNOWLEDGEMENTS

I would like to thank:

My Family: My mother, who has driven me to the ends of the earth when need be and for allowing me to still remain a student while many others were continuing with the forward plans of industry. My sister, who understood my struggles emotionally, mentally and physically through this demanding time as she was also dedicated to her MSc, without our mutual venting sessions of setbacks and failures, I would not have made it through this process. To Casey, Flick and T, who have been my happy place, bringing me love and joy through the darkest of times. To all my babies, from Merida, to Peter, to Mateo for just making everyday more bearable.

My Friends: To all my friends from across the board who have encouraged and motivated me day in and day out (there are so many to name - you know exactly who you are). My word, where do I even begin? Melissa, you have had so many venting sessions with me, I lost count, you have been my voice of reason when making stupid decisions, my support system when I felt like I couldn't do it and my cheerleader when I finally achieved something. Serina and Funi, we started this journey of this degree move together, tutored on the way, been to multiple chemistry events together and most of all, had so many moments, from birthdays to just meeting up for a catchup sesh. Deni, we started chemistry together in our first year, and we may have changed colleges, but we remain in each other's hearts always. To Denise Dee (mama bear), for encouraging me to take on this journey even when we were both at our lowest sometimes, and for the most epic travel memories that will remain in history till kingdom come. To my dancing friends, my word, the amount of memories we've made together is insane: from belly dancing (Ali, Ness, Erica, Irene and the girls), Tribal (Sam, Margi, Kendra and the girls), Line dancing (Courteney and Candice, Aunty Mell – my soul sisters and other momma, Dawn, Belinda, Hayley and the entire DM studio), Hiphop (the best stress reliever class – Tam you've listened to me bitch and moan, Chels - you've helped me find the worth in life, Cath you have been the most amazing person to dance with and talk to and Dan you have absolutely no idea how much your classes have done for me through all those times), Latin & Ballroom (Trist for listening to me complain, Alain and Mell for helping me build my confidence through

everything and the D2di studio as a whole) and every single genre, inclusive of all the dancing people I've met over the years for giving me moments that I will cherish forever. These have helped me more than you know. To all the people scattered across the globe that have been a part of my life.

My Supervisors and Mentors: To Dr. Ndumiso Mhlongo, for taking me on as a student, helping train me, guiding me through the process, inspiring me, motivating for funding and training retreats, and allowing me the opportunity to grow and learn in this space. To Dr. Hezekiel Kumalo for setting up the first meeting, helping me troubleshoot, being a motivational co-supervisor and a joy to be around. To Dr. Kgotatso Machaba for helping me through this journey, inclusive of helping me decide on topics, helping me troubleshoot, reviewing my work, helping me with corrections and journal rebuttals, motivating me every step of the way, and being a friend when I just needed encouragement to continue.

The Laboratory Medicine and Medical sciences department: Without your friendship and support, I would have not made it through this journey. Firstly to Ola, for being my partner in crime through this PhD, training with me from the start, motivating and encouraging me, helping me peer review all my research outputs. To my group, Sabrina, Ntombi, Khanyo, Minnie, Makhosi and the rest of the computational group for working as a team and providing support. To Jivi, Maka, Thabani, Dudu, SQ, Priscilla, Terry, Bestinee, Kamini, Angela, Nkululeko, Lloyd, Thobeka, Lu, our dearest departed Nomali and the rest of my reading room soul mates, you have absolutely no idea how much I needed you and how much you've done for me over these past 4 years. Jiv, Maka and T, you have become my family! We were each other's biggest cheerleaders through this trying time. The degree is hard enough with all the failures though the process, the topic changes, the publication pressure, then throw in Covid, looting, flooding, university admin issues and you get a recipe for a continuous streak of meltdowns. To Dr. Khan, for providing motivation when needed and being our department mother and academic leader. To Letty for being a supportive mentor and friend through this entire process.

The College of Health Sciences: To the CHS, for funding me and providing financial assistance through the course of my degree. To Dr. Brenda De Gama, our academic leader, for continuously checking up on us, for helping wherever she could, holding support sessions and progress meetings to keep us motivated and understanding our setbacks, not just practically, but also mentally. To Prof for helping us through every step of the way whenever you could.

The Centre of High - Performance Computing: To CHPC, for providing computational resources, updates and troubleshooting tips when needed.

LIST OF ABBREVIATIONS

α	Alpha
θ	Angle
Å	Angstrom
β	Beta
°	Degree
Δ	Delta
%	Percentage
3D	Three-dimensional
BFE/ ΔG	Binding free energy
CADD	Computer-aided drug design
DCCM	Dynamic cross correlation matrix
GAFF	General Amber force field
GPU	Graphics Processing Unit
K	Kelvin
kDa	Kilodalton
LBVS	Ligand-based virtual screening
MD	Molecular dynamics
MM	Molecular Mechanics
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area
MMV	Molecule molecular viewer
mol	Mole
NMR	Nuclear magnetic resonance
NPT	Isobaric–isothermal ensemble
ns	Nanoseconds
PBVS	Pharmacophore-based virtual screening
PDB	Protein data Bank
PMEMD	Particle Mesh Ewald Molecular Dynamic
ps	Picosecond
QM	Quantum Mechanics
RDD	Rational drug design
RESP	Restrained electrostatic potential
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RoG	Radius of gyration

RSCB	Research Collaboratory for Structural Bioinformatics
SASA	Solvent accessible surface analysis
SBVS	Structure-based virtual screening
VMD	Visual molecular dynamics
VS	Virtual screening
WHO	World health organisation

LIST OF AMINO ACIDS

Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic Acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

LIST OF TABLES

Chapter 4

Chapter 5

Supplementary

Chapter 6

TABLE OF CONTENTS

CHAPTER 1

Introduction

1.1 Background and rationale

Tuberculosis (TB), an extremely transmissible disease, is brought about by the presence of the pathogen known as *Mycobacterium tuberculosis (M. tuberculosis)* [1]. It exists as an infection that infiltrates via the alveolar passage through aerosol droplets, attacking the lungs and further spreading throughout the body [2]. Current studies indicate that 85% of TB infections exist as pulmonary TB and is seen to cause long term bleeding, necrosis and lung cavitation [3]. TB remains the root of an estimated 2 million deaths annually. The overall infection rate increased from 10 million to an astounding 10.4 [4] million and the impermanence number jumping from 1.5 million people to 1.66 million individuals from the year 2018 to 2020 following though with the same trend in years 2021 through 2023 [5]. A recent global TB report states that in the 10 million people who contracted the disease, approximately 1.5 million deaths were reported. Treatment of TB via antibiotic therapy is thus required to diminish the mortality rate attributed to infections.

Antibiotic treatment of TB is required for a period of approximately 6 to 9 months [6]. A common set of drugs used for the treatment of TB are first line drugs known as isoniazid, ethambutol, rifampin and pyrazinamide [6]. Due to varying efficacy of the treatment and the constant lack of patient compliance to systematic regimen, multiple drug resistant strains of *M. tuberculosis* have aggressively emerged [7]. The extent of the varying resistance can be measured by the severity of the strain and the inefficiency in the response to the suggested drugs whereby the drug resistant (DR-TB) strains account for the overall response to just one of the first line treatments [8] and the multi drug resistant (MDR-TB) strains constitute the response to more than just one of the indicated first line drugs [9]. In the case of the presence of one of these variants, usage of second line drugs, namely fluoroquinolones, kanamycin, amikacin and capreomycin, is undertaken. Unfortunately, the existence of further strains called extensively drug resistant (XDR) and emerging extensively drug resistant (eXDR) strains are also considered to be additionally resistant to the set of second line drugs [9]. Due to the increase in mutations, new drugs and targets need to be discovered in an act to confront this disease.

Two of the major pathways attributing to overall functioning of TB remains the mycolic acid and folate pathways [10,11]. The prevention of the elongation of long chain fatty acids in the fatty acid synthetase II system has proven to decrease the biosynthesis of mycolic acid production thereby inhibiting the cell growth in *M. tuberculosis*. The mycolic acid biosynthesis is governed by the existence of both the mammalian fatty acid synthesis pathway (FAS I) and the bacterial fatty acid synthesis pathway (FASII) [10]. While the acyl-CoA primers (C_{16-18}) are produced in the FAS I pathway, the FAS II system elongates the meromycolates to C_{50-56} [12] . The FAS II system involved as a precursor in the production of mycolic acids, acts as a basis for survival of the Mycobacterium by diminishing the susceptibility of bacteria to antibiotics [13]. Folate is essential for the growth and multiplication of the bacteria. It catalyzes the condensation reaction of para-aminobenzoate (pABA) and 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (H_2PtPP) to pyrophosphate and 7,8- dihydropteroate [14]. According to the above precursors, the three major targets investigated in this study are Enoyl-acyl carrier-protein reductase (inhA), β-ketoacyl ACP synthase (KasA) and Dihydropteroate synthase (DHPS/ folP1). KasA and inhA are precursors to the production of mycolic acids and DHPS is a major contributor in the functioning of the folate pathway. Inhibition of enzyme functionality would contribute largely to the inhibition of bacterial cell growth in *M. tuberculosis*.

A modern approach more frequently applied is that of in silico-based computer aided drug design (CADD) [15]. Computational techniques are used to further accelerate the process of drug discovery decreasing the time restraint and monetary pressure governed by traditional drug design approaches, as well as provide an in depth in silico understanding of complex reactivity and binding during the formation of the system [16]. Molecular docking methods predict the probability of a ligand to bind to a target of interest, while molecular dynamics simulations presents atomic scale molecular interactions between a ligand and a target [17].

The increasing burden associated with infections and diseases, and the compounded effect of the exorbitant cost of drug development has initiated the implementation of beneficial methods such as drug repurposing. Drug repurposing, also more commonly known as drug repositioning and therapeutic switching, is an alternative approach whereby new pharmacological uses are evaluated for FDA approved drugs already existing in the market or previously abandoned

investigational prototypes [18]. Traditional drug development methods are considered to be expensive, time-consuming and often associated with failure in method validation and positive responses to antibiotics [18]. Drug repurposing is viewed as a modern strategy alternative to some of the laborious traditional drug discovery and development processes. It allows for the use of either *in silico*-based methods *in vitro* based experimental methods or a combination of both to identify or design and evaluate the inhibitory potential of a molecule on the target of interest. Literature indicates the use of gene ontology and network pharmacology approaches to propose drugs suggested to have high potency and inhibition capability against drug targets of interest. This approach is paired with structural similarity index, virtual screening, tailored pharmacophore and network-based inference combinations to analyse the target's activity [19]. The potential drugs, identified *via* gene ontology and network pharmacology protocols, are already known to possess previous anti-arrhythmic, anaesthetic, antimicrobial and antibacterial properties, making them supreme candidates for drug repositioning [19].

In this study, a combination of virtual screening, *in silico* drug repurposing, tailored pharmacophore and molecular dynamics simulation methods are applied to identify potential inhibitors of *M. tuberculosis* drug targets inhA, KasA and folP1, and to characterize the interaction of identified potential inhibitors against the respective drug targets in pursuit of new and effective TB treatment.

1.2 Aims and Objectives

The aim of this study is to identify possible inhibitors of three major drug targets in *M. tuberculosis*, namely inhA, KasA and folP1 respectively, through the application of molecular modelling methods.

The outlined objectives were as follows:

- To obtain crystal structures using RSCB: PDB and provide a comprehensive understanding of previously identified drug targets of interest: inhA, KasA and folP1.

- To conduct molecular dynamics simulations on ligand-receptor complexes, using the Amber software, to understand the structural dynamics involved in complex formation and evolution.

- To calculate RMSD and RoG of the ligand-receptor complex to analyse the stability of the system during the simulation.

- To calculate binding free energy using MM/PBSA method to predict the binding affinity of a proposed potential inhibitor to a target protein, and to decompose this energy into per amino acid contribution to identify residues of importance in proposed inhibitor binding.

- To identify intermolecular forces contributing to ligand-receptor complex formation.

- To calculate RMSF of target protein residues to understand conformational changes of amino acid residues induced by ligand binding.

- To apply ADME tool to predict physicochemical and pharmacokinetic properties of the proposed inhibitors.

- To assess the impact of the C171Q mutation on the binding of KasA to TLM.

- To identify pharmacophore moieties of KasA active site residues in a TLM-KasA complex to construct a scaffold model for virtual screening.

References

1. Delogu, G., Sali, M. & Fadda, G. The biology of *mycobacterium tuberculosis* infection. *Mediterr. J. Hematol. Infect. Dis.* **5**, (2013).

2. Bloom, B. R. & Murray, C. J. L. Tuberculosis: commentary on a reemergent killer. *Science (80-.).* **257**, 1055–1064 (1992).

3. Gadkowski, L. B. & Stout, J. E. Cavitary pulmonary disease. *Clin. Microbiol. Rev.* **21**, 305–333 (2008).

4. Floyd, K., Glaziou, P., Zumla, A. & Raviglione, M. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. *Lancet Respir. Med.* **6**, 299–314 (2018).

5. Chakaya, J. *et al.* Global Tuberculosis Report 2020–Reflections on the Global TB burden, treatment and prevention efforts. *Int. J. Infect. Dis.* (2021).

6. Chideya, S. *et al.* Isoniazid, rifampin, ethambutol, and pyrazinamide pharmacokinetics and treatment outcomes among a predominantly HIV-infected cohort of adults with tuberculosis from Botswana. *Clin. Infect. Dis.* **48**, 1685–1694 (2009).

7. Penn-Nicholson, A. *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.* (2021).

8. Manjelievskaia, J., Erck, D., Piracha, S. & Schrager, L. Drug-resistant TB: deadly, costly and in need of a vaccine. *Trans. R. Soc. Trop. Med. Hyg.* **110**, 186–191 (2016).

9. Organization, W. H. & others. *Multidrug and extensively drug-resistant TB (M*. (2010).

10. Takayama, K., Wang, C. & Besra, G. S. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.* **18**, 81–101 (2005).

11. Minato, Y. *et al. Mycobacterium tuberculosis* folate metabolism and the mechanistic basis for para-aminosalicylic acid susceptibility and resistance. *Antimicrob. Agents Chemother.* **59**, 5097–5106 (2015).

12. Schiebel, J. *et al.* Structural basis for the recognition of mycolic acid precursors by

KasA, a condensing enzyme and drug target from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **288**, 34190–34204 (2013).

13. Sullivan, T. J. *et al.* High affinity InhA inhibitors with activity against drug-resistant strains of *Mycobacterium tuberculosis*. *ACS Chem. Biol.* **1**, 43–53 (2006).

14. Zheng, J. *et al.* Para-aminosalicylic acid is a prodrug targeting dihydrofolate reductase in *mycobacterium tuberculosis*. *J. Biol. Chem.* **288**, 23447–23456 (2013).

15. Marshall, G. R. Computer-aided drug design. *Annu. Rev. Pharmacol. Toxicol.* **27**, 193–213 (1987).

16. Niranjan, D., Gill, S., Shrivastava, D., Jain, A. & Agrawal, O. P. Quantitative structure activity relationships in computer aided drug design: a review. *J. Drug Deliv. Ther.* **9**, 645–649 (2019).

17. Naqvi, A. A. T., Mohammad, T., Hasan, G. M. & Hassan, M. Advancements in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships. *Curr. Top. Med. Chem.* **18**, 1755–1768 (2018).

18. Rudrapal, M., Khairnar, S. J. & Jadhav, A. G. Drug repurposing (DR): an emerging approach in drug discovery. *Drug Repurposing-Hypothesis, Mol. Asp. Ther. Appl.* (2020).

19. Passi, A., Rajput, N. K., Wild, D. J. & Bhardwaj, A. RepTB: a gene ontology based drug repurposing approach for tuberculosis. *J. Cheminform.* **10**, 1–12 (2018).

CHAPTER 2

Literature review

2.1. Introduction

2.1.1. Background

Tuberculosis (TB) is an airborne, infectious bacteria caused by the pathogen *Mycobacterium tuberculosis (M. tuberculosis)* [1]. TB acts by attacking the lungs first then moves throughout the body [2]. Being referred to as the 'envy of human pathogens' [3], approximately one third of the population being skin tested exhibits a positive result for the presence of the bacteria [3]. *M. tuberculosis* is seen to manifest in 3 different stages: exposure, latent and active TB Once active, spreading over the body leads to the establishment of coinfections such as bronchitis or pneumonia [4]. The emergence of Covid19 posed further complications with comorbidities acting as contributing factors attributing to the severity of the disease on the immune system, thereby enhancing the virulence on the host [5].

Due to the increase in the presence of variants of the bacteria, currently marketed drugs are no longer effective in the treatment of available strains [6]. The increased study of targets of interest and the development of effective modes of treatment is therefore a priority in future analysis of the disease. This chapter provides a comprehensive account of the early discovery of the disease, the transmission life cycle, the major pathways, the current and future drugs, the main targets of interest, and the drug design techniques studied. It outlines the need for further understanding of the mechanism of virulence and pathogenesis, and the increased demand for the identification of possible drug candidates.

2.1.2. Discovery and history of M. tuberculosis

According to literature, the initial form of *Mycobacterium* has been said to exist for more than 150 million years as *Mycobacterium ulcerans* [7]. During the Middle Ages, a disease, known as *scrofula*, 'the king's evil', attacking the cervical lymph nodes was said to exhibit the same behaviour as the clinical form of TB further referred to as the 'white plague' [8]. Approximately 20000 years ago, the emergence of the first modern strain of *M. tuberculosis* was said to infect

people in East Africa [9]. Dating back to 2400BC, Egyptian mummies were said to exhibit skeletal deformities related to the onset of TB, [10] while both India and China documented lesions characteristic of tuberculosis nearly 2300 years ago [11]. Tubercular lesions on the lungs, coughing, sweating and blood ridden sputum [12] were reported as *schachepheth* in Hebrew history [13] and *phtisis* in Greek history [14].

In the 17th century, the study of pulmonary TB exhibited the initial attack of the respiratory tract but further spread to the joints, lymph nodes, nervous system, bones, liver, gastrointestinal tracts and skin further causing severe symptoms inclusive of necrosis, fibrosis, inflammation, spinal cord paralysis and abscesses with the onset of extrapulmonary TB [15,16]. In 1882, with the use of the methylene blue stainer, Robert Koch was successful in the isolation of the tubercule bacillus in animal serum earning him a Nobel prize in honorary medicine in 1905 [17].

2.1.3. Epidemiology and statistics

TB is the 10th largest cause of mortality worldwide and the top cause of mortality due to a bacterial infection [18]. This bacterial infection is said to be the leading cause of approximately 2 million deaths across the world annually, with the overall infection rate increasing from 10 million to 10.4 million people [19]. The mortality rate increased from 1.5 million to 1.66 million people from the year 2018 to 2020 leading to an 18% increase [18].

It has been reported that 80% of the cases globally can be attributed to the 22 countries referred to as high burden regions assuming an unequal distribution worldwide [3]. Approximately 87% of this amount are seen to be from a list of 30 developing countries, posing high risk factors. Examples of the highest infected countries include Nigeria, Philippines, Pakistan, and South Africa [20]. Developing countries are seen to exhibit higher mortality rates due to the overall compounded effect of coinfections established with other diseases such as the human immunodeficiency virus (HIV) and more recent, Covid19 [21]. This corresponds directly with the research carried out in both Vietnam and Brazil [22,23]. **Figure 1** represents the overlap of drug resistant TB with HIV coinfection in high burden countries.

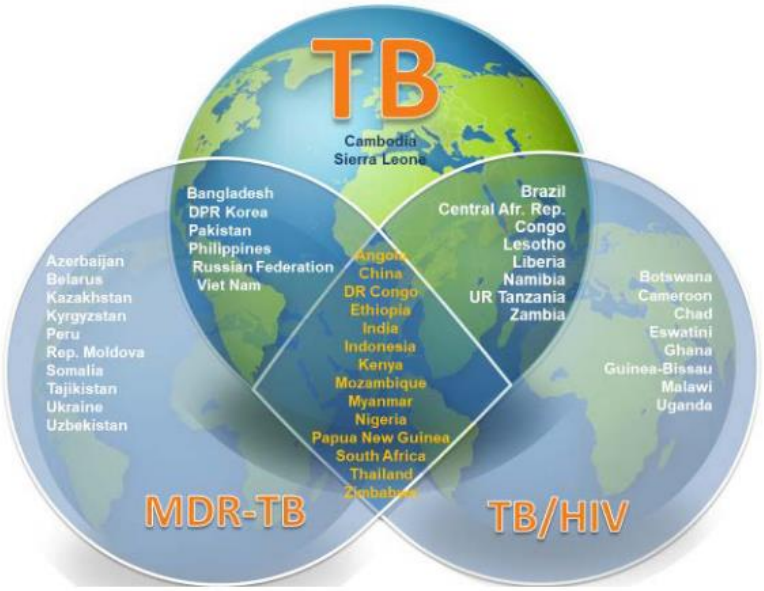


Figure 1: Overlap of TB and MDR-TB with HIV coinfection in high burden countries [24].

South Africa has exhibited an increase in infection of approximately 450000 people each year with approximately 270000 (60%) attributed to coinfection with HIV and 87000 (19%) resulting in overall death [25]. TB co-infection with the Covid-19 virus has adversely contributed to the TB infection rate, and complicated TB treatment protocols [26]. Current studies indicate that coinfection with HIV and Covid-19 enhances the virulence of TB in the host [5]. The exacerbated burden of TB demands the restructuring of current treatment protocols and development of multitarget inhibitors to potentially neutralize the impact of co-infection.

The occurrence of TB has thereby remained a rife problem in certain parts of the world more than others, indicating global areas exhibiting high risk factors. **Figure 2** is a graphical representation of the spread of TB indicating that the countries showing the highest infection rates belong to both the South-East Asia and Sub-Saharan African regions.

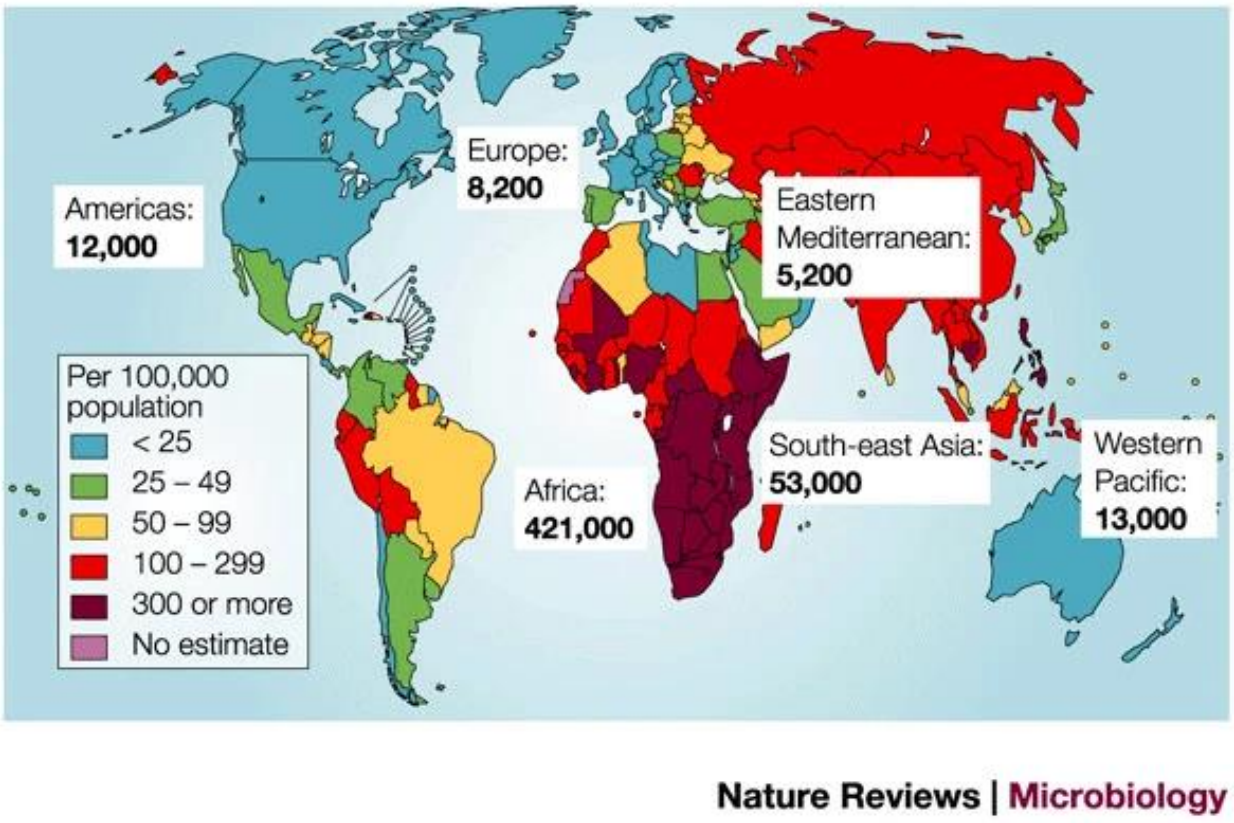


Figure 2: Showing the global regional spread of TB infections [27]**.**

Further statistical studies carried out on the South-East Asia region represented in **Figure 3** indicates the highest infection rates being attributed to both respiratory infections and TB across the board. With the knowledge of TB being a major contributor in the indirect cause of respiratory infections, a conclusion can be made that there is a definite overlap between the top two major causes.

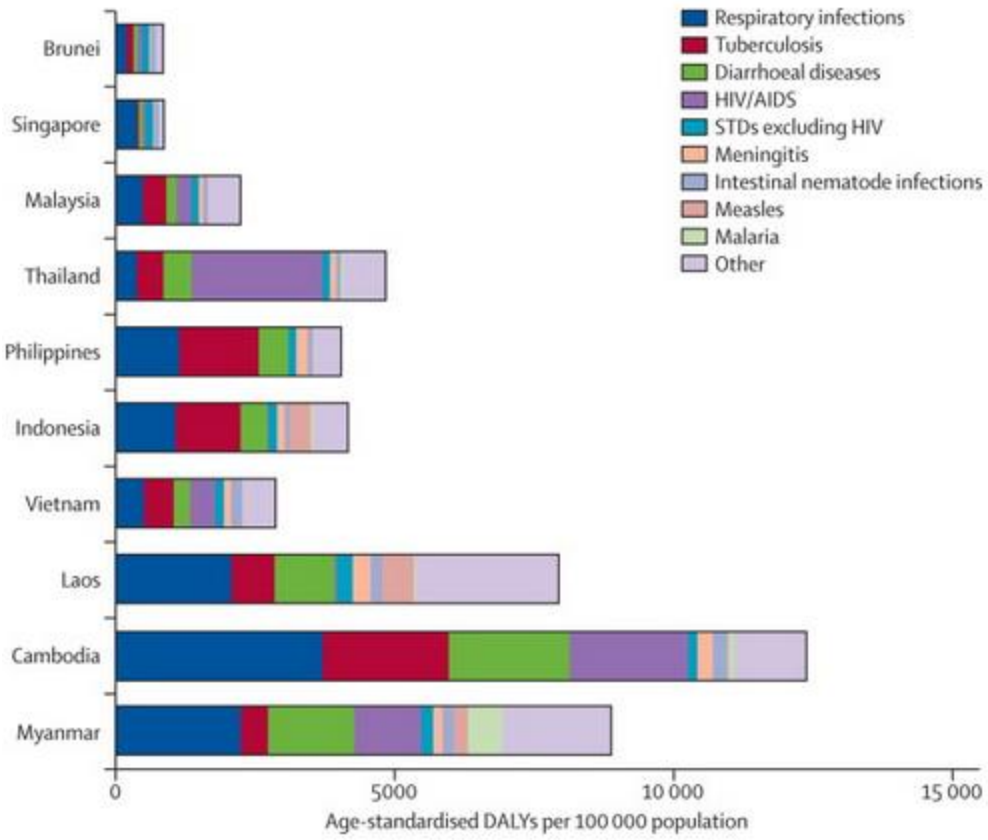


Figure 3: Major infectious diseases reported in South-East Asia [28].

2.1.4. Transmission, Life cycle and Prevalence

TB is a disease spread by means of infectious droplets spread by infectious individuals. The bacilli remain airborne for a couple of hours providing the opportunity for inhalation [3]. Phagocytosis, results in the inhalation of particles approximately >0.5 nm in size into the membrane of the plasma. This causes local inflammation thereby increasing the bacterial population count by making provision for fresh host cells [29]. The most prevalent form of the disease remains to be pulmonary TB giving rise to the formation of destructive granulomas. Latent and active TB refer to the stage in the process of granuloma formation [30]. The containment of microbial growth is successful in 90 to 95% of infected cases due to the establishment of early onset human response stemming in the confinement of the infectious foci within the granuloma resulting in latent TB [3]. The transmission of *M. tuberculosis* is presented in **Figure 4**.

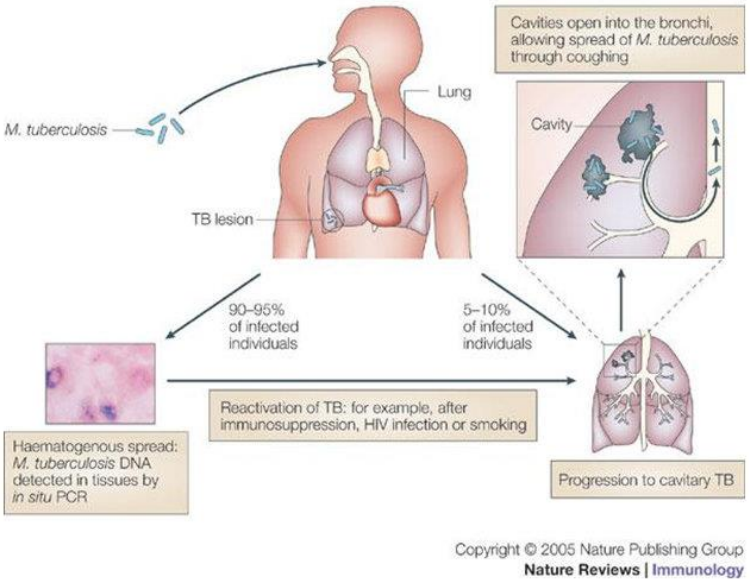


Figure 4: Transmission of TB *via* the alveolar passage [31].

Constant proliferation of dormant bacteria during primary infection leads to chronic progressive tissue damage further leading to necrosis then resulting in the formation of tubercular lesions [31]. Active granulomas eventually fracture, overturning a mass of infectious particles leading to the formation of a cough inevitably facilitating the aerosol spread of the bacilli [32]. The progression from immune response to active infection is represented in **Figure 5.**

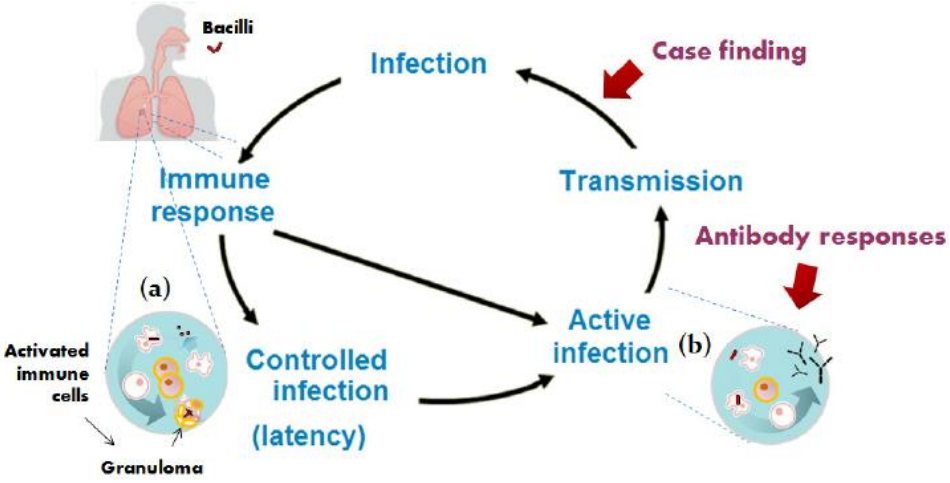


Figure 5: Progression of TB from latent to active infection [31].

Furthermore, *Th1* and *Th2* both form phenotype responses ultimately determining susceptibility and two types of immune response failures resulting in the generation of both type I in patients and type II TB in healthy infected individuals [33].

Inhibition of TB is based on inhibiting cell growth of the *M. tuberculosis*. **Figure 6** shows the general life cycle of *M. tuberculosis*.

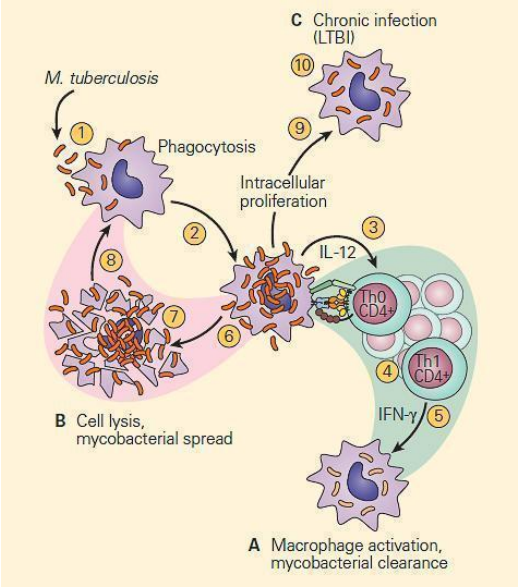


Figure 6: Life cycle of *M. tuberculosis.*
(https://www.immunopaedia.org.za/immunology/special-focus-area/3-immunity-to-tb/)

The process begins with phagocytosis resulting in the engulfing or ingestion of the bacteria further leading to either the viral degradation of the membrane causing bacterial spread, intracellular proliferation resulting in cell division ultimately leading to chronic illness or finally macrophage activation bringing about bacterial clearance [34,35].

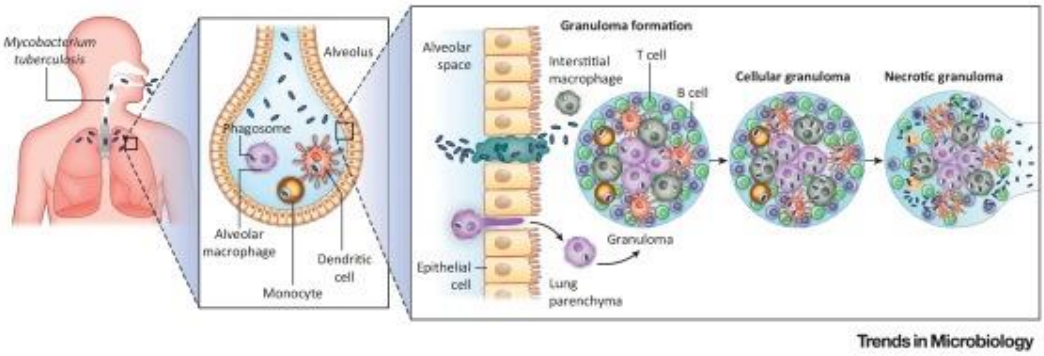


Figure 7: Life cycle of *M. tuberculosis* from alveolar entrance to granuloma formation [34].

Further investigation of the infection life cycle reveals the presence of a dendritic cell also known as an antigen presenting cell derived from the white blood cell monocyte, in the alveolar passage leading to possible boosting of adaptive immune response [36]. Seepage of the epithelial cell along with the presence of interstitial macrophages in the lung tissue leads to the formation of a granuloma [36]. The presence of B cells encourages the formation of antibodies thereby presenting it to the $CD4^+$ T cells tasked with early protection by controlling the bacterial growth encouraging apoptosis [37]. Clustering causing the formation of a granuloma emerges in response to inflammation caused by the infection then resulting in a necrotizing granuloma in response to the decaying or dead tissue caused by infection [38]. In the event of positive immune response, activated *Th1* cells in combination cytokine (IFN-γ) secreted by T cells activates the macrophage to kill the intracellular bacteria [39].

2.1.5. Drug resistance and Treatments

Long term treatment of tuberculosis requires the usage of antibiotics for a time period of approximately 6 to 9 months [40]. The original set of treatment known as first line drugs used were isoniazid, ethambutol, rifampin, and pyrazinamide [41]. These drugs were utilized to inhibit the specific drug targets of InhA [Enoyl-(acyl carrier-protein) reductase], bacterial RNA polymerase (RNAP), Arbinosyl Transferase and the S1 Component of 30S Ribosomal subunit [42]. Due to the varying efficacy of the treatment and the constant lack of patient compliance to existing courses, multiple drug resistant strains of *M. tuberculosis* have emerged [43].

The extent of the varying drug resistance can be measured by the severity of the strain and inefficiency of the response to the suggested drugs, [44] where the drug resistant (DR-TB) strains refer to the ineffective response to just one of the first line drugs and the multi-drug resistant (MDR-TB) strains constitute an ineffective response to more than one first-line drug. In the case of the initial drug resistant strains, the development of a range of second line drugs, inclusive of compounds such as fluoroquinolones, kanamycin, amikacin, and capreomycin was observed. Further strains now being referred to as extensively drug resistant (XDR) as well as emerging extensively drug resistant (eXDR) strains have surfaced and are unfortunately also resistant to the set of second-line treatments [40,45]. From the estimated 10 million TB infections being recorded annually, 2.9% of the DR-TB cases are seen to exist as MDR-TB and 6.2% of

the MDR-TB seen to fall within the category of XDR-TB [46,47]. Frequently emerging drug-resistant strains have been a long-standing hurdle in the treatment of TB [6]. The increase in the presence and emergence of drug resistance globally calls for the establishment and unequivocally warrants for the identification of new small molecules with novel mechanisms of action against future drug targets of interest. Thus being a forefront precursor in combating the disease [48].

2.1.6. Structure of M. tuberculosis cell wall and mechanistic pathways

The *M. tuberculosis* cell encompasses four distinct layers: the external capsule, peptidoglycan–arabinogalactan complex (AGP), plasma membrane and an asymmetrical mycomembrane [49]. Direct target of this envelope has been identified as the optimum interface for drug development. The capsule acts as a mode of bacterial survival imperative for a pathogen acting on an intracellular level. The organised layers allow for low permeability of the outer capsule [50]. **Figure 8** represents the structure of the cell wall of *M. tuberculosis*.

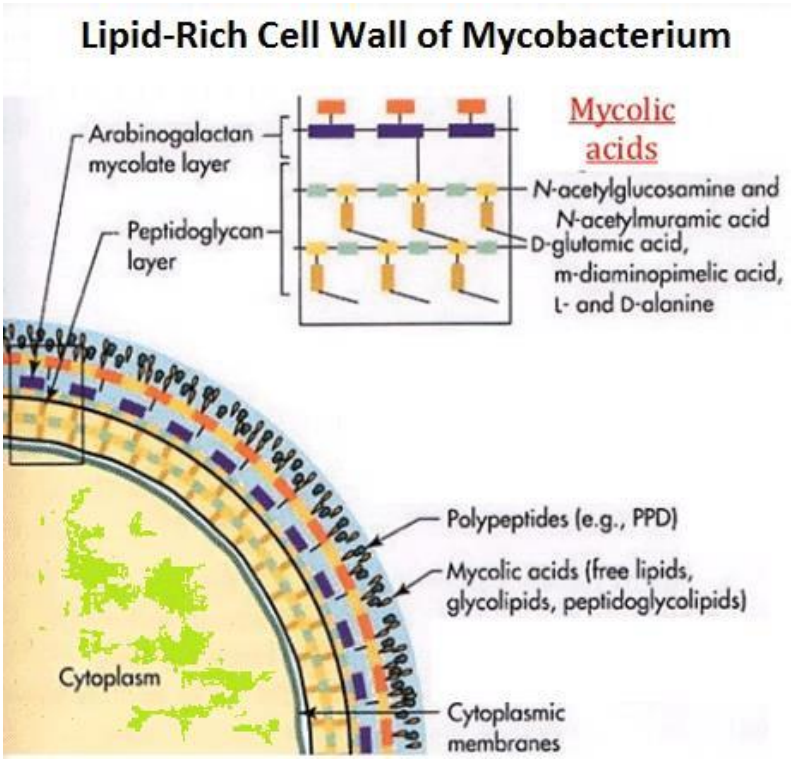


Figure 8: Structure of *M. tuberculosis cell wall* [50].

As is evident in **Figure 8**, the mycolic acid pathway forms a major part of the structure outside the cytoplasm. Inhibiting mycolic acid biosynthesis will further lead to the inhibition of cell growth in *M. tuberculosis.* Hence, disturbance of the mycolic acid pathway serves as a viable method of drug target investigation [51]. The mycolic acid biosynthesis is governed by the existence of both the mammalian fatty acid synthesis pathway (FASI) and the bacterial fatty acid synthesis pathway (FAS II) [52,53]. While the acyl-CoA primers (C_{16-18}) are produced in the FASI pathway, the FASII system elongates the meromycolates to C_{50-56} [53]. The FAS II system involved as a precursor in the production of mycolic acids, acts as a basis for survival of the Mycobacterium [54]. The survival of *M. tuberculosis* requires numerous essential metabolic pathways, another one of them being the folate biosynthesis pathway. This pathway offers proteins that have been identified as significant targets for antimycobacterial drug development. **Figure 9** describes the pathway attributed to and the proteins involved in mycolic acid production.

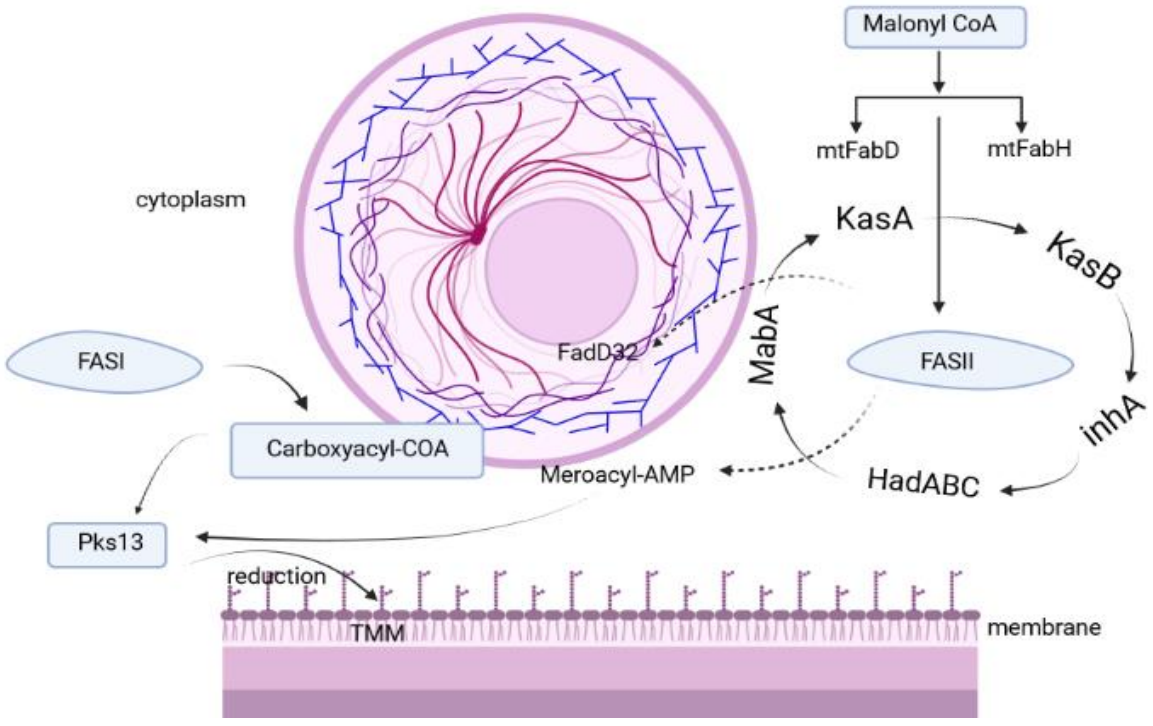


Figure 9: Mycolic acid biosynthesis pathway.

Figure 10 depicts the pathway and proteins involved in folate synthesis.

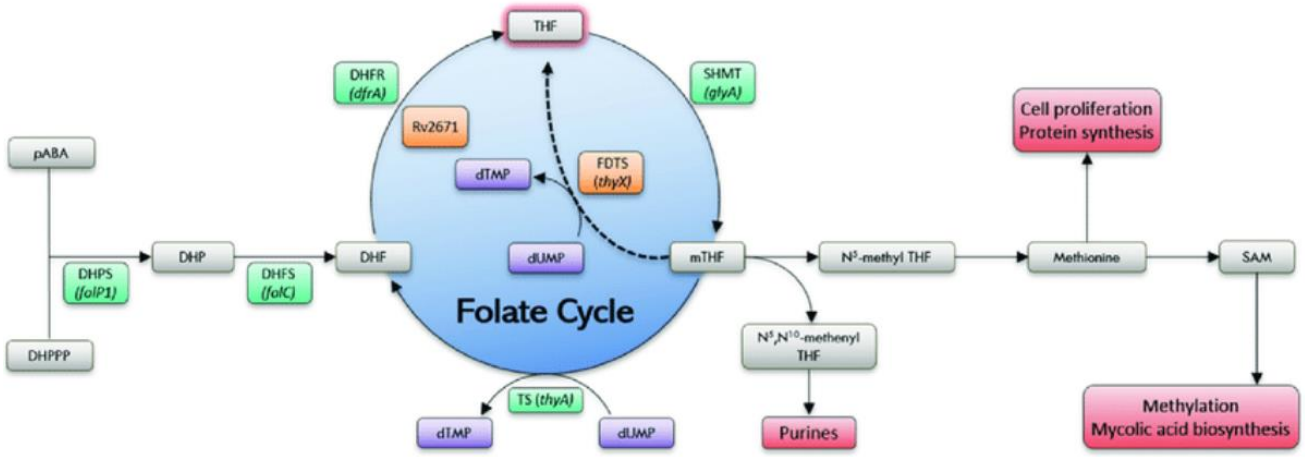


Figure 10: Folate pathway of *Mycobacterium tuberculosis* [55].

2.1.7. Targets

Evident from **Figure 11**, the common first line drugs (isoniazid, ethionamide, ethambutol and pyrazinamide) are used to inhibit the production of mycolic acids and FAS I halting the functionality of *M. tuberculosis* [54]. The second-line drugs (fluoroquinolone and capreomycin) are aimed at inhibiting DNA gyrase and ribosome production [56].

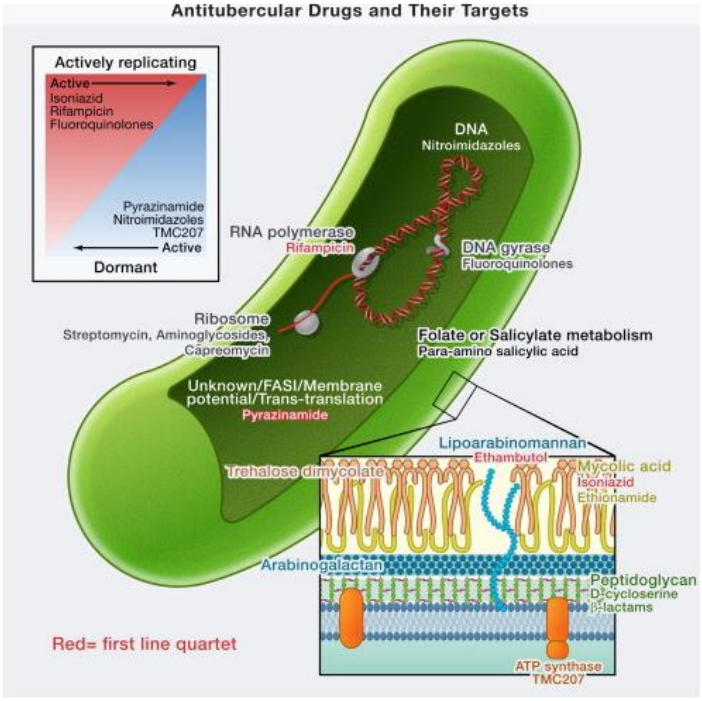


Figure 11: Targets and the known antitubercular drugs utilized [56].

Five major enzymes contribute to the production of mycolic acids: inhA, MabA, HadABC, KasA and KasB [53]. One of the major *M. tuberculosis* drug targets is InhA, initially inhibited by isoniazid [41]. InhA belonging to the fatty acid synthetase II system suggests that prevention of the production of these long chain fatty acids have proved to decrease mycolic acid biosynthesis inhibiting *Mycobacterium tuberculosis* [57]. InhA catalyses the β-nicotinamide adenine dinucleotide (NADH)-specific reduction of the 2-*trans*-enoyl-acyl carrier protein [58]. It acts as a vital step elongating C_{20} fatty acids to C_{60}-to-C_{90} in the FAS I (type 1 fatty acid synthase) [59]. **Figure 12** shows a graphical representation of the drug target inhA (contributing to mycolic acid biosynthesis) complexed to isoniazid.

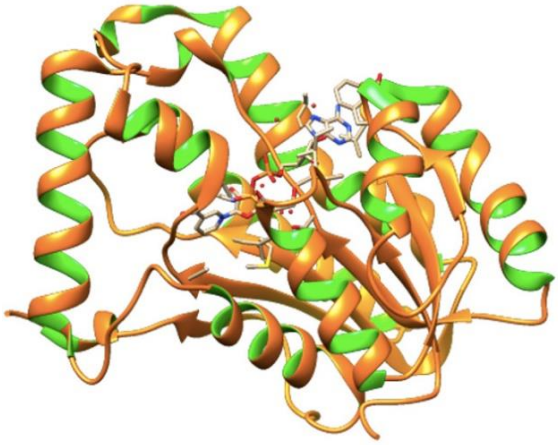


Figure 12: Crystal structure of the inhA-isoniazid complex [PDB ID: 4DRE] [60],

KasA and KasB exist as dimers that make up the β-ketoacyl-ACP fatty acid synthetase system [61]. The pathway exhibited in the cytoplasm directly links the functionality of KasA to the fatty acid synthetase system involved in the meromycolate backbone. The mycolic acid biosynthesis is governed by the existence of both the mammalian pathway (FASI) and the bacterial pathway (FASII) [52]. While the acyl-CoA primers ($C_{16\text{-}18}$) are produced in the FASI pathway, the FASII system elongates the meromycolates to $C_{50\text{-}56}$ [53]. The FASII system involved as a precursor in the production of mycolic acids, acts as a basis for survival of the Mycobacterium [54].This leads to the final condensation reaction catalysed by polyketide synthase (Pks13) thereby shielding the bacteria from the immune system rendering the drug ineffective [57]. A homodimer β-ketoacyl ACP synthase (KasA) acts as a catalyst in the Claisen condensation of the malonyl-AcpM and acyl-AcpM [62,61]. Thiolactomycin (TLM) is one of the known drugs that is reported to inhibit KasA [63].

Emergence of a mutation in KasA revealed the resistance of the protein to TLM. The initial structure of KasA is altered *via* a mutation of Cysteine to Glutamine at position 171 of the amino acid sequence. **Figure 13** depicts the structure of the C171Q KasA mutant complexed to TLM.

The original unbound apo and mutant structures of KasA were generated using Xray crystallography and resolutions of 1.8 – 2.2 Å were obtained [63]. The catalytic triad of His345, His311 and Cys171, all located in the core domain where the Cys171 newly mutated to Glu171 lies in the N-terminal domain and the other two residues located in the C-terminal domain of the α-helix [52]. Upon binding, the side chain of the newly established Glutamine generates a hydrogen bond from the oxygen to the nitrogen atom on Phe404 resulting in a 60 ˚ Phe shift and thereby 3.3 Å positional movement out of the active site as compared to the wild-type bound structure. The newly opened conformation generated by mimicking acylation insists that the Phe404 then has the ability to form favourable interactions with the TLM ring [64]. In the process, the Leu-116, initially present in the catalytic site, is expelled with the addition of the acylated chain while a third residue, Tyr126 acts as a solvent barrier in both wild-type and mutant bound and unbound structures [65].

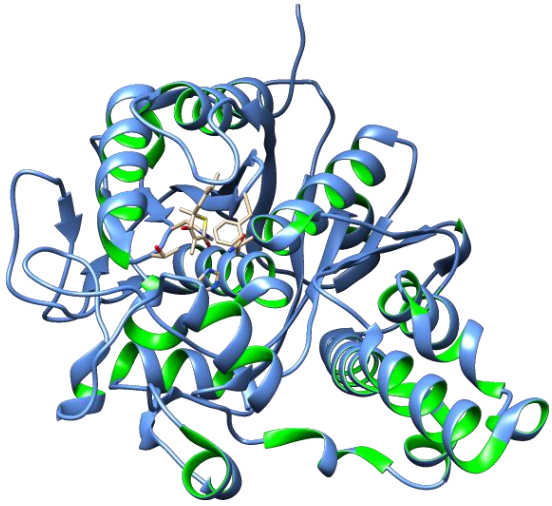


Figure 13: Crystal structure of C171Q KasA-TLM complex [PDB ID: 4C6X][60].

Dihydropteroate synthase (DHPS/ folP1) is required for the synthesis of folate, hence, essential for the growth and multiplication of the bacteria [66]. While folate is needed for cell synthesis and growth, DHPS catalyses the condensation reaction of para-aminobenzoate (pABA) and 6-

hydroxymethyl-7,8-dihydropterin pyrophosphate (H_2PtPP) to pyrophosphate and 7,8-dihydropteroate [67]. DHPS is a well-validated drug target [68]. It is studied as a selective drug target, due to it being present in microorganisms but absent in humans thereby rendering new methods and modes of action [69]. The *M. tuberculosis* DHPS protein structure is composed of 280 amino acid residues, with an eight β-strand central barrel encompassed by eight α-helices. The catalytic site amino acids are found in the central barrel (21, 86, 105, 177 and 213) [69]. The known drug for *M. tuberculosis* DHPS is Pterin-6-YL-Methyl-Monophosphate (PMM) [70]. **Figure 14** shows a graphical representation of the DHPS-PMM complex.

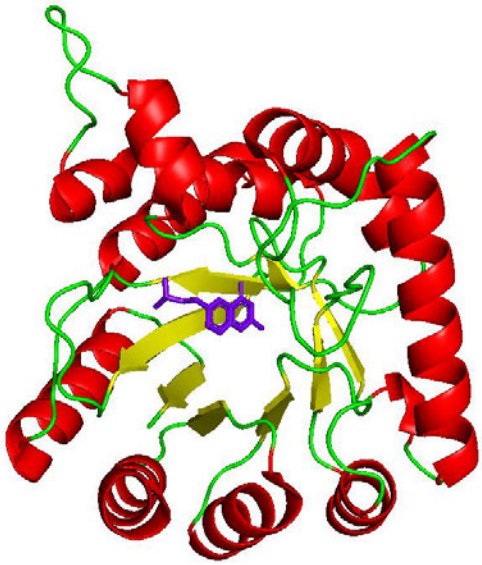


Figure 14: Crystal structure of DHPS-PMM complex [PDB ID: 1EYE][60].

2.1.8. Drug repurposing vs Traditional drug design

The overbearing weight of bacterial infections and diseases and the compounded effect of the price of overall drug discovery, design and development has called for the implementation of fast, efficient, and effective strategies such as drug repurposing with the utilisation of molecular modelling methods. Drug repositioning is utilized by complexing known, already clinically approved drugs to treat other ailments. One of the benefits of molecular modelling is to provide a platform for further investigation into the prediction of reactivity prior to biological studies.

Drug repurposing, also more commonly known as drug repositioning and therapeutic switching, is an alternative approach whereby new pharmacological uses are evaluated for FDA

approved drugs already existing in the market and abandoned investigational prototypes [71, 72]. The first case of drug repurposing was witnessed as an accidental result occurring in the 1920s [72]. It is a method targeted at investigating indications on specific and sometimes novel proteins that have not previously been tested against, using the small molecule hits [73]. It acts as an emerging, constantly evolving combat approach used more widely against homologies stagnant diseases, and mutations.

Traditional drug design is an expensive, and extremely time-consuming approach used to identify inhibitors and, failure in overall method validation and positive molecule response is a common occurrence [74]. Drug repurposing is said to decrease the chance of failure by 45% due to the omission of safety issues involved with toxicity experienced during traditional design rendering an average of a 5-to-7-year time save [75]. New, more efficient ways of drug development are sought after to bridge the gap existing in treatment creation. Drug repurposing allows for the use of either in-silico based computational studies, in-vitro based experimental studies or a combination of the two to identify, design and evaluate the effect of the small molecule on the protein of interest [76]. **Figure 15** shows the differences between in silico and activity-based drug repurposing.

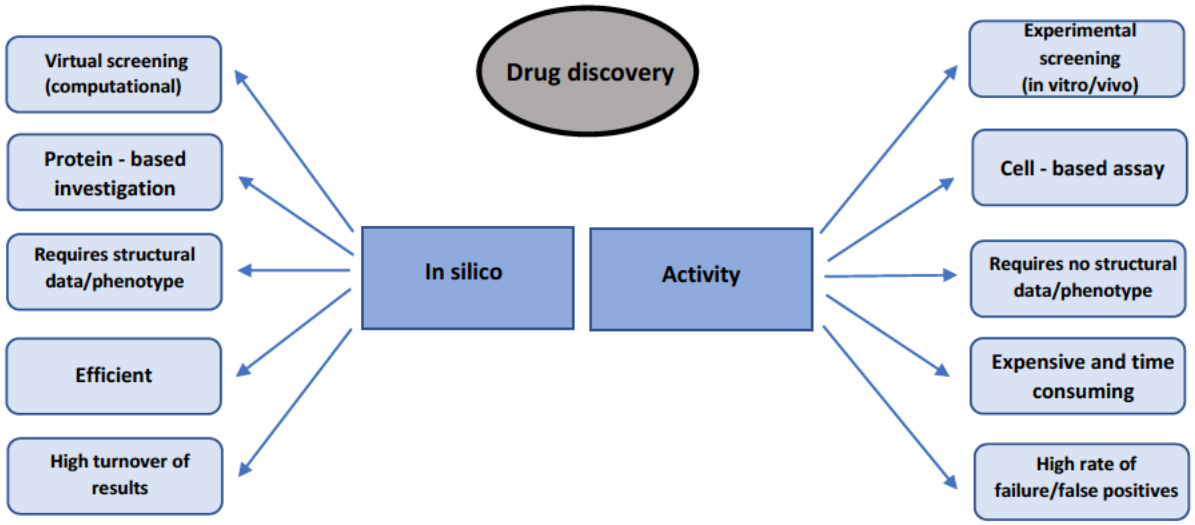


Figure 15: Differences of traditional and in-silico based approaches [72]

Observations have proven to show an annual 25% of pharmaceutical revenue being attributed to drugs that have been repurposed with 30% of the FDA approved drugs being derived from repositioned substances [77]. One well-known example of the approach is that of dimethyl

fumarate originally used to treat psoriasis and later repurposed to act as an inhibitor against multiple sclerosis [78].

Drug repurposing can be carried out using 3 different methods: target based, drug based or disease-based approaches. Structure based drug design aided by the method of molecular dynamics provides a pathway whereby the stages for ready to use drugs are less than that of traditional drug design [79–81]. Drug repurposing contains previous information on structural optimization, pharmacological, pharmacokinetic, toxicological data and safety information [82]. The stages of drug repurposing include in silico small molecule discovery, development and post market safety monitoring while traditional drug design requires the stepwise process of drug discovery, synthesis and expression, preclinical in vivo studies, clinical trial phases, FDA approval review and post market safety monitoring [74]. Research and development time and money attributed to drug repurposing works off an average of 3 to 12 years utilizing approximately 1.6 billion dollars in comparison to the 10 to 16 years utilizing approximately 12 billion dollars attributed to traditional drug design [77, 73].

2.1.9. Gene ontology: NBI DTI and tailored pharmacophore - virtual screening

Computational methods, have proven to be useful tools in the identification of potential inhibitors [83]. Drug repurposing encompassing gene ontology paired with a network-based inference approach assists in the study of inhibition potency. The known drugs of interest are seen to possess singly or a combination of biological properties such antibacterial, antimicrobial, anti-arrhythmic and local anaesthetic properties, making them supreme molecules for the drug repurposing journey [70]. Molecular modelling together with the preset of molecular docking methods aid in the prediction of the probability of a suggested drug hit binding to its drug target of interest, while the molecular dynamics simulations aspect presents the intermolecular interactions between a drug and its protein.

Gene ontology is an approach used in identifying potential inhibitors from a series of interactions analysed using network-based inference (NBI) data obtained from the evaluation of interactions between amino acids and their constituent side chains with the drug candidate

of interest [70]. *Passi et al* developed the method of *RepTB* which correlates interactions between drug target pairs. Evaluation of the drug target interactions act as a basis for further analysis of drug targets of interest and development of potential inhibitors. The aforementioned approach acts to serve as a predictive value in the identification of potential drug candidates and selected drug targets [70]. Atomic level analysis of interactions between a drug target of interest and potential small molecule inhibitor thereby proves to be a cornerstone in the starting point for drug design and development [84].

Virtual screening is a beneficial tool in the discovery of small molecules, posing as inhibitors of drug targets of interest. This technique is carried out using either a ligand- or receptor-based approach [85]. The ligand-based approach screens ligands based on beneficial structural properties of the small molecule, while the target-based approach screens ligands based on the structural property of the receptor [86,87]. While both approaches offer their respective benefits, a target-based approach is more advantaged as it relies on a possible complex rather than an isolated ligand. Machaba et al.[88] recently carried out an approach where ligand pharmacophores were tailored to counteract the effect of drug resistance-conferring mutations. This is a promising approach to identify novel inhibitors of drug targets that have developed drug resistance due to mutations, particularly in the treatment of TB. Tailored pharmacophore is a method whereby small molecules are built up from a base to acquire the most efficient set of protein-ligand interactions with the amino acids in the active site of the protein thereby leading to an increased binding energy [89]. The structure is dependent on specific functional groups that provide the most suitable electrostatic and hydrophobic interactions in a certain conformation within the binding site [88]. These vital moieties provide a starting point for the design of an improved inhibitor.

References

1. Delogu, G., Sali, M. & Fadda, G. The biology of *mycobacterium tuberculosis* infection. *Mediterr. J. Hematol. Infect. Dis.* **5**, (2013).

2. Bloom, B. R. & Murray, C. J. L. Tuberculosis: commentary on a reemergent killer. *Science (80-.).* **257**, 1055–1064 (1992).

3. Russell, D. G., Barry 3rd, C. E. & Flynn, J. L. Tuberculosis: what we don't know can, and does, hurt us. *Science (80-.).* **328**, 852–856 (2010).

4. Yamamoto, A. *et al.* Study on clinico-pathological features of active pulmonary tuberculosis found at autopsy in a general hospital. *Kekkaku:[Tuberculosis]* **84**, 71–78 (2009).

5. Daftary, A. HIV and tuberculosis: the construction and management of double stigma. *Soc. Sci. \& Med.* **74**, 1512–1519 (2012).

6. Manjelievskaia, J., Erck, D., Piracha, S. & Schrager, L. Drug-resistant TB: deadly, costly and in need of a vaccine. *Trans. R. Soc. Trop. Med. Hyg.* **110**, 186–191 (2016).

7. Hayman, J. Mycobacterium ulcerans: an infection from Jurassic time? *Lancet* **324**, 1015–1016 (1984).

8. Murray, J. F., Rieder, H. L. & Finley-Croswhite, A. The King's Evil and the Royal Touch: the medical history of scrofula. *Int. J. Tuberc. Lung Dis.* **20**, 713–716 (2016).

9. Kapur, V., Whittam, T. S. & Musser, J. M. Is *Mycobacterium tuberculosis* 15,000 years old? *J. Infect. Dis.* **170**, 1348–1349 (1994).

10. Morse, D., Brothwell, D. R. & Ucko, P. J. Tuberculosis in ancient Egypt. *Am. Rev. Respir. Dis.* **90**, 524–541 (1964).

11. Brown, L. The story of clinical pulmonary tuberculosis. (1941).

12. Pease, A. S. Some remarks on the diagnosis and treatment of tuberculosis in antiquity. *Isis* **31**, 380–393 (1940).

13. Daniel, V. S. & Daniel, T. M. Old Testament biblical references to tuberculosis. *Clin. Infect. Dis.* **29**, 1557–1558 (1999).

14. Adams, F. *The genuine works of Hippocrates*. vol. 1 (W. Wood, 1886).

15. Daniel, T. M. René Théophile Hyacinthe Laënnec and the founding of pulmonary medicine [Founders of Our Knowledge]. *Int. J. Tuberc. Lung Dis.* **8**, 517–518 (2004).

16. Duffin, J. *To see with a better eye: a life of RTH Laennec*. vol. 376 (Princeton University Press, 2014).

17. Barberis, I., Bragazzi, N. L., Galluzzo, L. & Martini, M. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *J. Prev. Med. Hyg.* **58**, E9 (2017).

18. Chakaya, J. *et al.* Global Tuberculosis Report 2020–Reflections on the Global TB burden, treatment and prevention efforts. *Int. J. Infect. Dis.* (2021).

19. Floyd, K., Glaziou, P., Zumla, A. & Raviglione, M. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. *Lancet Respir. Med.* **6**, 299–314 (2018).

20. Sharma, D., Sharma, J., Deo, N. & Bisht, D. Prevalence and risk factors of tuberculosis in developing countries through health care workers. *Microb. Pathog.* **124**, 279–283 (2018).

21. Trinh, Q. M. *et al.* Tuberculosis and HIV co-infection in Vietnam. *Int. J. Infect. Dis.* **46**, 56–60 (2016).

22. da Silva Escada, R. O. *et al.* Mortality in patients with HIV-1 and tuberculosis co-infection in Rio de Janeiro, Brazil-associated factors and causes of death. *BMC Infect. Dis.* **17**, 1–10 (2017).

23. Oni, T. *et al.* Patterns of HIV, TB, and non-communicable disease multi-morbidity in peri-urban South Africa-a cross sectional study. *BMC Infect. Dis.* **15**, 1–8 (2015).

24. Organization, W. H. WHO global lists of high burden countries for tuberculosis (TB), TB/HIV and multidrug/rifampicin-resistant TB (MDR/RR-TB), 2021–2025: background document. (2021).

25. Brust, J. C. M. *et al.* Improved survival and cure rates with concurrent treatment for multidrug-resistant tuberculosis–human immunodeficiency virus coinfection in South Africa. *Clin. Infect. Dis.* **66**, 1246–1253 (2018).

26. Saunders, M. J. & Evans, C. A. COVID-19, tuberculosis and poverty: preventing a

perfect storm. (2020).

27. Onyebujoh, P. & Rook, G. A. W. Focus: tuberculosis. *Nat. Rev. Microbiol.* **2**, 930–932 (2004).

28. Coker, R. J., Hunter, B. M., Rudge, J. W., Liverani, M. & Hanvoravongchai, P. Emerging infectious diseases in southeast Asia: regional challenges to control. *Lancet* **377**, 599–609 (2011).

29. Russell, D. G., Cardona, P.-J., Kim, M.-J., Allain, S. & Altare, F. Foamy macrophages and the progression of the human tuberculosis granuloma. *Nat. Immunol.* **10**, 943–948 (2009).

30. Via, L. E. *et al.* Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infect. Immun.* **76**, 2333–2340 (2008).

31. Wani, R. L. S. Tuberculosis 2: Pathophysiology and microbiology of pulmonary tuberculosis. *South Sudan Med. J.* **6**, 10–12 (2013).

32. Kaplan, G. *et al. Mycobacterium tuberculosis* Growth at theCavity Surface: a Microenvironment with FailedImmunity. *Infect. Immun.* **71**, 7099–7108 (2003).

33. Menon, J., Hoeppner, V. H., Judd, A., Power, C. A. & Bretscher, P. A. A hypothesis for the existence of two types of tuberculosis, reflecting two distinct types of immune failure to control the pathogen, based upon prevalence of mycobacterium-specific IgG subclasses. *Scand. J. Immunol.* **87**, e12665 (2018).

34. Koch, A. & Mizrahi, V. *Mycobacterium tuberculosis*. *Trends Microbiol.* **26**, 555–556 (2018).

35. Chai, Q., Zhang, Y. & Liu, C. H. *Mycobacterium tuberculosis*: an adaptable pathogen associated with multiple human diseases. *Front. Cell. Infect. Microbiol.* **8**, 158 (2018).

36. Pai, M. *et al.* Nature reviews disease primers. *Tuberculosis* **2**, 16076 (2016).

37. Skeiky, Y. A. W. & Sadoff, J. C. Advances in tuberculosis vaccine strategies. *Nat. Rev. Microbiol.* **4**, 469–476 (2006).

38. Chandra, P., Grigsby, S. J. & Philips, J. A. Immune evasion and provocation by *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* 1–17 (2022).

39. Mehta, P. K., Karls, R. K., White, E. H., Ades, E. W. & Quinn, F. D. Entry and

intracellular replication of *Mycobacterium tuberculosis* in cultured human microvascular endothelial cells. *Microb. Pathog.* **41**, 119–124 (2006).

40. Shah, N. S. *et al.* Worldwide emergence of extensively drug-resistant tuberculosis. *Emerg. Infect. Dis.* **13**, 380 (2007).

41. Chideya, S. *et al.* Isoniazid, rifampin, ethambutol, and pyrazinamide pharmacokinetics and treatment outcomes among a predominantly HIV-infected cohort of adults with tuberculosis from Botswana. *Clin. Infect. Dis.* **48**, 1685–1694 (2009).

42. Pedelacq, J., Nguyen, M. C., Terwilliger, T. C. & Mourey, L. A Comprehensive Review on *Mycobacterium tuberculosis* Targets and Drug Development from a Structural Perspective. *Struct. Biol. Drug Discov. Methods, Tech. Pract.* 545–566 (2020).

43. Santos, L. C. the molecular basis of resistance in Mycobaterium tuberculosis. *Open J. Med. Microbiol.* **2**, 24–36 (2012).

44. Organization, W. H. & others. *Multidrug and extensively drug-resistant TB (M*. (2010).

45. Shi, R., Itagaki, N. & Sugawara, I. Overview of anti-tuberculosis (TB) drugs and their resistance mechanisms. *Mini Rev. Med. Chem.* **7**, 1177–1185 (2007).

46. Saifullah, A. *et al.* Evaluation of risk factors associated with the development of MDR- and XDR-TB in a tertiary care hospital: a retrospective cohort study. *PeerJ* **9**, e10826 (2021).

47. Bonnet, M. *et al.* Treatment of tuberculosis in a region with high drug resistance: outcomes, drug resistance amplification and re-infection. *PLoS One* **6**, e23081 (2011).

48. Organization, W. H. *The global MDR-TB & XDR-TB response plan 2007-2008*. (2007).

49. Jankute, M., Cox, J. A. G., Harrison, J. & Besra, G. S. Assembly of the mycobacterial cell wall. *Annu. Rev. Microbiol.* **69**, 405–423 (2015).

50. Martynov, A. *et al.* Tuberculosis as an infectious pathology of immune system. *Ann. Mechnikov's Inst.* 8–14 (2016).

51. Jeffrey North, E., Jackson, M. & E Lee, R. New approaches to target the mycolic acid

biosynthesis pathway for the development of tuberculosis therapeutics. *Curr. Pharm. Des.* **20**, 4357–4378 (2014).

52. Schiebel, J. *et al.* Structural basis for the recognition of mycolic acid precursors by KasA, a condensing enzyme and drug target from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **288**, 34190–34204 (2013).

53. Takayama, K., Wang, C. & Besra, G. S. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.* **18**, 81–101 (2005).

54. Sullivan, T. J. *et al.* High affinity InhA inhibitors with activity against drug-resistant strains of *Mycobacterium tuberculosis*. *ACS Chem. Biol.* **1**, 43–53 (2006).

55. Hajian, B. *et al.* Drugging the Folate Pathway in *Mycobacterium tuberculosis*: The Role of Multi-targeting Agents. *Cell Chem. Biol.* **26**, 781-791.e6 (2019).

56. Kalscheuer, R. *et al.* The *Mycobacterium tuberculosis* capsule: a cell structure with key implications in pathogenesis. *Biochem. J.* **476**, 1995–2016 (2019).

57. Marrakchi, H., Lanéelle, G. & Quémard, A. InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II. *Microbiology* **146**, 289–296 (2000).

58. Dessen, A., Quémard, A., Blanchard, J. S., Jacobs Jr, W. R. & Sacchettini, J. C. Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*. *Science (80-.).* **267**, 1638–1641 (1995).

59. Gurvitz, A., Hiltunen, J. K. & Kastaniotis, A. J. Function of heterologous *Mycobacterium tuberculosis* InhA, a type 2 fatty acid synthase enzyme involved in extending C20 fatty acids to C60-to-C90 mycolic acids, during de novo lipoic acid synthesis in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **74**, 5078–5085 (2008).

60. Berman, H. M. *et al.* The protein data bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 899–907 (2002).

61. Wong, H. C., Liu, G., Zhang, Y.-M., Rock, C. O. & Zheng, J. The solution structure of acyl carrier protein from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **277**, 15874–15880 (2002).

62. Bhatt, A., Kremer, L., Dai, A. Z., Sacchettini, J. C. & Jacobs Jr, W. R. Conditional depletion of KasA, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis. *J. Bacteriol.* **187**, 7596–7606 (2005).

63. Luckner, S. R., Machutta, C. A., Tonge, P. J. & Kisker, C. Crystal structures of *Mycobacterium tuberculosis* KasA show mode of action within cell wall biosynthesis and its inhibition by thiolactomycin. *Structure* **17**, 1004–1013 (2009).

64. Rudraraju, R. S. *et al. Mycobacterium tuberculosis* KasA as a drug target: Structure-based inhibitor design. *Front. Cell. Infect. Microbiol.* 1352 (2022).

65. Schaefer, B., Kisker, C. & Sotriffer, C. A. Molecular dynamics of *Mycobacterium tuberculosis* KasA: implications for inhibitor and substrate binding and consequences for drug design. *J. Comput. Aided. Mol. Des.* **25**, 1053–1069 (2011).

66. de Castro Spadari, C., Vila, T., de Morais Barroso, V. & Ishida, K. New Targets for the Development of Antifungal Agents. *Encycl. Mycol.* 456–467 (2021) doi:10.1016/b978-0-12-809633-8.21026-1.

67. Zheng, J. *et al.* Para-aminosalicylic acid is a prodrug targeting dihydrofolate reductase in *mycobacterium tuberculosis*. *J. Biol. Chem.* **288**, 23447–23456 (2013).

68. Gengenbacher, M., Xu, T., Niyomrattanakit, P., Spraggon, G. & Dick, T. Biochemical and structural characterization of the putative dihydropteroate synthase ortholog Rv1207 of *Mycobacterium tuberculosis*. *FEMS Microbiol. Lett.* **287**, 128–135 (2008).

69. Baca, A. M., Sirawaraporn, R., Turley, S., Sirawaraporn, W. & Hol, W. G. J. Crystal structure of *Mycobacterium tuberculosis* 6-hydroxymethyl-7,8-dihydropteroate synthase in complex with pterin monophosphate: New insight into the enzymatic mechanism and sulfa-drug action. *J. Mol. Biol.* **302**, 1193–1212 (2000).

70. Passi, A., Rajput, N. K., Wild, D. J. & Bhardwaj, A. RepTB: a gene ontology based drug repurposing approach for tuberculosis. *J. Cheminform.* **10**, 1–12 (2018).

71. Dey, G. An overview of drug repurposing: Review article. *J. Med. Sci. Clin. Res.* **7**, 3–5 (2019).

72. Rudrapal, M., Khairnar, S. J. & Jadhav, A. G. Drug repurposing (DR): an emerging approach in drug discovery. *Drug Repurposing-Hypothesis, Mol. Asp. Ther. Appl.* (2020).

73. Aggarwal, S., Verma, S. S., Aggarwal, S. & Gupta, S. C. Drug repurposing for breast cancer therapy: Old weapon for new battle. in *Seminars in cancer biology* vol. 68 8–20 (Elsevier, 2021).

74. Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).

75. Xue, H., Li, J., Xie, H. & Wang, Y. Review of drug repositioning approaches and resources. *Int. J. Biol. Sci.* **14**, 1232 (2018).

76. Oprea, T. I. & Overington, J. P. Computational and practical aspects of drug repositioning. *Assay Drug Dev. Technol.* **13**, 299–306 (2015).

77. Cha, Y. *et al.* Drug repurposing from the perspective of pharmaceutical companies. *Br. J. Pharmacol.* **175**, 168–180 (2018).

78. Xu, Z. *et al.* Dimethyl fumarate for multiple sclerosis. *Cochrane Database Syst. Rev.* (2015).

79. Némethy, G., Leach, S. J. & Scheraga, H. A. The influence of amino acid side chains on the free energy of helix-coil transitions1. *J. Phys. Chem.* **70**, 998–1004 (1966).

80. Kokkinidis, M., Glykos, N. M. & Fadouloglou, V. E. Protein flexibility and enzymatic catalysis. *Adv. Protein Chem. Struct. Biol.* **87**, 181–218 (2012).

81. Chong, C. R., Chen, X., Shi, L., Liu, J. O. & Sullivan, D. J. A clinical drug library screen identifies astemizole as an antimalarial agent. *Nat. Chem. Biol.* **2**, 415–416 (2006).

82. Jin, G. & Wong, S. T. C. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov. Today* **19**, 637–644 (2014).

83. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Computational methods in drug discovery. *Pharmacol. Rev.* **66**, 334–395 (2014).

84. Palomino, J. C. & Martin, A. Is repositioning of drugs a viable alternative in the treatment of tuberculosis? *J. Antimicrob. Chemother.* **68**, 275–283 (2013).

85. Broccatelli, F. & Brown, N. Best of both worlds: on the complementarity of ligand-based and structure-based virtual screening. *J. Chem. Inf. Model.* **54**, 1634–1641 (2014).

86. Ripphausen, P., Nisius, B. & Bajorath, J. State-of-the-art in ligand-based virtual screening. *Drug Discov. Today* **16**, 372–376 (2011).

87. Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discov. Today* **7**, 1047–1055 (2002).

88. Machaba, K. E., Mhlongo, N. N., Dokurugu, Y. M. & Soliman, M. E. S. Tailored-pharmacophore model to enhance virtual screening and drug discovery: a case study on the identification of potential inhibitors against drug-resistant *Mycobacterium tuberculosis* (3R)-hydroxyacyl-ACP dehydratases. *Future Med. Chem.* **9**, 1055–1071 (2017).

89. von Korff, M. & Steger, M. GPCR-tailored pharmacophore pattern recognition of small molecular ligands. *J. Chem. Inf. Comput. Sci.* **44**, 1137–1147 (2004).

CHAPTER 3

Principles of computational methods

3.1. Introduction

Molecular modelling is a relatively new field in comparison to the traditional drug design approach. It is considered to be a highly economically viable process providing an efficient method of drug discovery making it a coveted strategy in future development [1]. By using computer aided drug design, a biological system comprising of a receptor and suggested ligand are complexed and the real time simulation mimicked using computational tools [2]. The interactions within the complex can thereby be analysed and the behaviour of the system varied to acquire maximum efficiency [3]. The complexed crystal structures are used as a base for the start of the study and the overall database for approved/well known drug hits aid in the process. While the field is still considered relatively new in comparison to other methods of drug design, the development and interest has provided substantial improvement in the computer programs required for further studies in a proportionately small-time frame. The combination of both computational and experimental data is required for successful drug design [4] but, the initial application of an in silico method helps greatly in the reduction of the unnecessarily high raw material and labour costs as well as the time frames needed for instant *in vitro* and *in vivo* biological studies [5].

Computational studies are a necessary tool in the efficient study of drug design and development, identification of functional groups exhibiting similar structural aspects and production of inhibitors. Pharmaceutical industries can consequently benefit from the enhanced tools of such a method in the process of identifying the novelty of drug targets and the appearance of potent inhibitors [6]. The heterogeneity of computer aided drug design approaches used lend host to a variety of principles they follow. Two major pieces of information are focused on in this field of study: the structure of the protein and small molecules and the chemical and physical properties associated with each [7]. Once complexed, the onus on the investigator to lead to the optimization and overall enhancement of these properties to obtain a positive result. The method is established by the foundation that post binding, the functionality of the characterised protein and the activity of the drug vary with the addition of interactions

reported in the catalytic site of the biological target of interest. Stability and flexibility of the complex lead to a directly proportional relationship with the inhibition capability of the small molecule suggestion [8]. The nature of all atoms in the binding pocket hence lead to the interaction data obtained from complexation.

The mechanisms of conformational analyses involved in the combined energetics of systems are investigated through a number of computational techniques including: molecular docking, molecular dynamic simulations, MD analysis, tailored pharmacophore and virtual screening [9]. This chapter provides a comprehensive understanding of the principles behind the invention and development of these techniques and tools.

There are two principles that govern the functioning of molecular modelling techniques via conformational changes granting changes in overall energy: molecular mechanics and quantum mechanics.

3.2. Molecular mechanics

Molecular based models are used to analyse and understand the qualitative and quantitative dynamics of the stability and reactivity associated with receptor-ligand complexes [10]. Molecular mechanics uses mathematical, empirical functions to approximate the intramolecular and intermolecular interactions between atoms and molecules within a system. The method was originally developed as a mechanical method for geometrical evaluation of macromolecule systems [11]. The aim was to decrease the cost associated with large molecules and the method depended largely on the polarization of force fields [12].

In 3-dimensional models, the change in geometry results in the change in position of individual atoms. Just as in the valence shell electron pair repulson (VSEPR) theory, the movement of the atoms result in an increase in the repulsion between lone pair and bonding pair electrons within the geometrical shape of the molecule [13]. The repulsion between atoms forces the bonds to rotate, bend and stretch altering the original bond angles [11]. The theory encompasses Newton's first law of motion [14] to describe the atoms of small organic hydrocarbons to large peptide molecules. The energy in this method is calculated without the use of the wave function [15].

These molecular mechanics simulations determine energetics of a model thus dictating the stability and compactness of a drug-drug target system thereby rendering a beneficial method in the discovery of potent inhibitors.

3.2.1. Potential energy

Potential energy is described as the stored energy within a body. This function initiates the method of molecular mechanics by understanding the information involved with the atomic bonding rather than solving the particle behaviour of the electrons [16]. The existence of different shapes prove that different bond lengths, angles and bond dissociation energies are established [17]. The assumption that similar atoms behave in a similar manner make up the backbone of this technique [18]. The composition of a molecule and its structural formula affects its reactivity. The technique thereby focuses specifically on the atom and bond types [19].

The potential energy function governed by the force field and the atomic type is defined by a set of equations following the first law of motion [20]. The morphologies created with the addition of intermolecular and intramolecular forces insists that the potential energy is generated using [21],[14]:

1. Bond stretching

$$E_{r} = K_r(r - r_0)^2$$

Figure 1: Graphical representation of the stretching of a bond [14].

Where K_r refers to the force constant for the bond, r_0 represents the equilibrium distance.

2. Angle bending

$$\boldsymbol{E_{\theta}} = \sum \boldsymbol{K_{\theta}(\theta - \theta_0)^2}$$

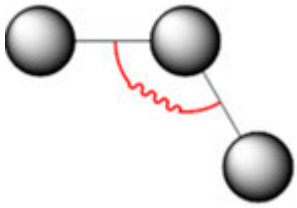


Figure 2: Graphical representation of the bending of a bond [14].

Where K_{Θ} refers to the force constant for the angle and Θ_0 represents the equilibrium angle.

3. Torsion

$$\boldsymbol{E_{\theta}} = \sum \boldsymbol{K_{\Phi}[1 + \cos(n_{\theta} - \Phi_0)]}$$

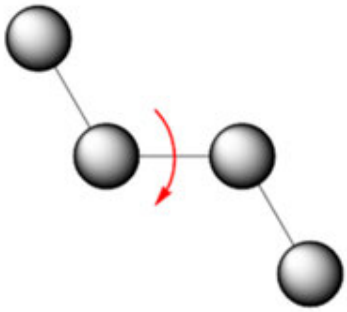


Figure 3: Graphical representation of bond torsion [14].

Where K_{Φ} refers to the dihedral angle and Φ_0 represents the equilibrium phase angle.

4. Electrostatic and van der Waals forces

$$E_{nb} = [\sum(\frac{A_y}{r_{ij}^{12}} - \frac{B_y}{r_{ij}^{6}})] + [\sum(\frac{q_i q_j}{D r_y})]$$

Figure 4: Graphical representation of intermolecular interactions [14].

Where r_{ij} refers to the distance, A_{ij} and B_{ij} refer to the parameters for vdW., D refers to the dielectric constant and q_i and q_j represent the charge points.

While the atoms are treated as spheres aiding the overlap resulting in the above equations, the final equation encompassing the addition of all the bonding types is as follows [22,19]:

$$\mathbf{Etotal = E_r + E_\Theta + E_\Phi + E_{nb}}$$

The force field chosen depends on the parameters chosen because of the type of energy data to be analysed. The most common force fields used are AMBER [23], CHARMM [24], GROMOS [25], ENCAD [26] and OPLS-AA [27]. The two major types of forcefields used are [28]:

- Type I force fields that need to contain cross terms to acquire accurate values for small systems
- Type II force fields that contain a simplified harmonic field for larger systems

3.3. Quantum mechanics

Quantum mechanics is a major component in the physics field acting as a baseline for further research. It studies the microscopic behaviour of physical systems. The initial idea of quantum mechanics, now referred to as 'the old quantum theory', belonged to the founding father Max Planck who discovered the packets of energy contained in blackbody radiation [29]. Further development of quantum mechanics was developed in 1923 by Louis de Broglie stipulating that electrons and all matter postulates a wave nature thus establishing the wave-particle duality theory [30]. Max Born, Werner Heisenberg, and Pascual Jordan continued the research to the development of matrix mechanics [31] and the breakthrough was further provided by Erwin Schrödinger who orchestrated the birth of wave mechanics [32].

The foundation of quantum mechanics is the understanding of the sub atoms that are contained in all matter: the core of an atom made up of protons, neutrons, electrons and photons [33]. Differing sciences are seen to all converge over central points where the quantum theory developed in physics, overlaps with the particles studied in chemistry thereby overlapping with the energetics of particles and bonds within biological systems [34]. The quantum theory acts by looking at a 3D model and mapping out the individual properties and using the electron density of specific systems [35]. Two major principles govern quantum mechanics: the Schrődinger wave function and the Born-Oppenheimer approximation theory [36].

3.3.1. Schrődinger wave function

In 1913, Niels Bohr developed the model of a singular atom stating that the nucleus, consisting of nucleons is surrounded by concentric circles of electrons with differing energy levels [37]. In 1926, the physicist Schrődinger developed a set of equations relating to the mathematical probability of the model [38]. While Bohr suggested that electrons existed as stationary entities whose position could be predicted, Schrődinger later made the observation that the path of an electron cannot be defined as it exists in constant motion [39,40].

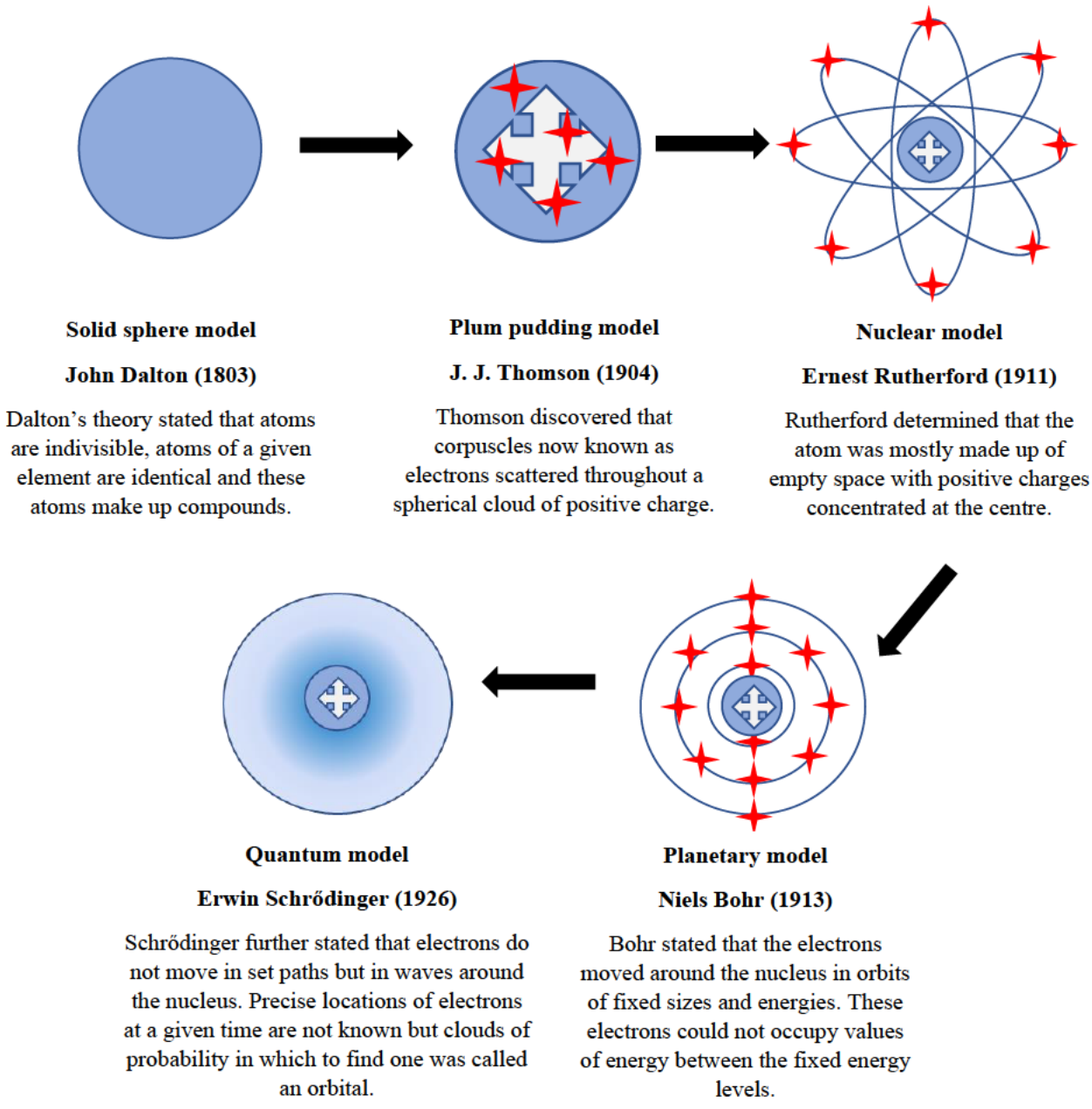


Figure 5: Schematic representation of the evolution of the atomic model from Dalton 1803 to Schrődinger 1926 (prepared by author).

Schrődinger proved in modern quantum mechanics that electrons do not occupy a fixed position in orbits but form a region of space known as an electron cloud generated by the constant motion of the sub atoms [41]. As evident in electronic configurations used to generate afbau diagrams, the periodic table follows quantum energy levels denoted by atomic subshells [42]. There is a directly proportional relationship between the quantum number and the energy

level harboured by that atomic subshell. As an electron gains energy, the quantum number will thereby jump up an energy level established by a new electron cloud in a subsequent subshell [43]. According to the particle theory, it can thus be established that particles contain a wide range of densities, and no momentum is in existence till observed [44]. The wave function in accordance with the potential set of equations governs this probability. The Schrődinger equation uses charge as a function of mass to predict the motion of sub atoms, specifically electrons within a system. The set of Schrődinger equation derivations are as follows [38,40]:

The Hamilton operator generates derivates dependant on space-based atoms by acquiring a summation of both the kinetic and potential energies of the atom:

$$\mathbf{H = T + V}$$

Where T refers to the Ek and V to the Ep

H can be further derived into a complex form of:

$$\boldsymbol{H} = \left[-\frac{\boldsymbol{h^2}}{\boldsymbol{8\Pi^2}} \sum \boldsymbol{i} \frac{\mathbf{1}}{\boldsymbol{m_j}} \left(\frac{\boldsymbol{\partial^2}}{\boldsymbol{\partial x^2}} + \frac{\boldsymbol{\partial^2}}{\boldsymbol{\partial y^2}} + \frac{\boldsymbol{\partial^2}}{\boldsymbol{\partial z^2}} \right) \right] + \sum \boldsymbol{i} \sum < \boldsymbol{j} \left(\frac{\boldsymbol{e_i e_j}}{\boldsymbol{r_{ij}}} \right)$$

The Schrődinger equation can therefore be viewed as:

$$\mathbf{H\Psi = E\Psi}$$

Where H refers to Hamiltonian operator, Ψ denotes the wave function and E refers to the eigenvalue energetics. One stipulation is that the wave function needs to be continuous, normalised and asymmetric and consisting of a single value.

3.3.2. Born-Oppenheimer approximation theory

When large molecular systems are used, the prediction of the potential of these thousands of atoms can be complex while just utilizing the Schrődinger equation. In such instances, the Born-Oppenheimer theory is used to work on a molecular scale rather than an atomic scale making it more manageable for predictions of large system [45]. Max Born and Robert

Oppenheimer developed the molecular quantum mechanics principle in 1927. The theory is built on the assumption that sub atoms contain different properties and masses thereby assuming that the functionality of the nucleus contributing the majority of the mass could be separated from the functionality of the electrons [46]. Interactions between the nucleus and electrons cement the nature of the heavier atoms remaining in a more fixed position while the smaller particles build up velocity. The attraction of the constantly moving electrons to the fixed nucleus forms the foundation of this theory [47]. The emergence of the effective nuclear charge aspect thus renders the nucleus arbitrary thereby making the assumption that the kinetic energy of the electrons makes up for the most part the kinetic energy of the molecule [48]. By subtracting the variable of the nucleus, the complexity of Hamiltons operator is therefore decreased [47].

The simplified wave equation can be written as follows:

$$\Psi(\boldsymbol{r_{elec}}) = \Psi(\boldsymbol{r_{elec}})(\Psi(\boldsymbol{r_{nucl}}))$$

The equation is then further denoted as:

$$(\boldsymbol{H_{elec+V_{NN}}})\Psi(\boldsymbol{r_{elec}}) = \boldsymbol{E_{EN}}\Psi(\boldsymbol{r_{nucl}})$$

Where H_{elec} refers to the adjustments made to stabilise positions and E_{EN} refers to the energy deduced from the electrons orbiting around the fixed nucleus.

This new equation is now seen as the electronic version of the Schrődinger equation [49]. Ground states are seen to exhibit a higher accuracy [50]. Once equilibrated and solved, the positions and potential energies can be established. The potential energy surface is an additional branch added to the increase in accuracy of the quantum mechanics method. It grounds off the potential energy associated with the probability of two atoms colliding [51]. The distribution and geometry of a molecule can then be based on high and low energy confirmations thus giving rise to the overall stability of a molecule further extending into the benefits associated with bioinformatics.

Although existing as two separate approaches, the combination of both principles and molecular mechanics can lead to the increase in accuracy of the data set acquired for large systems [52]. Molecular mechanics provides an in-depth analysis of intermolecular interactions such as permanent dipoles, van der Waals and hydrogen bonding while quantum mechanics provides information on the atomic and subatomic level interactions such as covalent, metallic

and ionic leading to electrostatic forces. In this instance the accuracy is guaranteed as the entire system is taken into account with the outer ions and solvent barriers [52]. This could also unveil any discrepancies along the mechanistic pathway thereby spotlighting significant energy variations and chemical reactivity.

3.4. Molecular dynamics simulations

3.4.1. Invention

The emergence of molecular dynamic simulations first reared its head with the introduction being established by the theoretical physicist, Berni Julian Alder [53]. After several advancements, Alder and collaborators found evidence of the first order kinetic transition of a fluid phase change in 1962 [54]. Statistical mechanics, utilizing socratic dialogue led to the first major breakthrough in simulations since the 1940s [53]. Computational methods thereby gained increasing popularity paired with theoretical evidence and experimental data to form an unstoppable method for future applications. Evolution of the technique from analysing small atoms and molecules to much larger macromolecule systems formed the forefront in the major field of study to come. Molecular dynamics provides the opportunity of understanding atomic level interactions established within medical biochemistry.

The initial study of simulations revealed two major types: Monte Carlo (MC) and Classic molecular dynamics (MD) simulations [55]. For ergodic systems, both methods should generate the same overall answer with the main difference being that MD generates data utilizing a time dependant average [55]. The superior method has been seen to be MD as a regard to the fact that not only information about the equilibrium structure is obtained but also the entire physical process while MC results in a loss of the physical process during sampling [56]. An extension of the difference lends to the usage of Newtons laws of motion in MD thereby rendering an increase in the efficiency of algorithms [57]. The conformational changes giving rise to intermolecular residual interactions help establish algorithmic energetics associated with the system.

The follow through of an MD simulation over a given time period generates a dynamic based trajectory capable of uncovering the in-depth system interactions of a complex. Structural

dynamics along with the energy contributions allow for the complex network of a system to be established.

3.4.2. Newton's laws of motion

Following Newton's principles, the first law governs the application of inertia where an object will only change its individual straight line of motion with the acting of an external force upon it [58]. Extending the initial discovery by Galileo, the movement of an object seemed to move farther on smoother surfaces thereby introducing the existence of friction [59]. The simplest form of the equation exists as:

$$\mathbf{F= d(mv)/dt}$$

Where F refers to the net force, m to the mass, v to the velocity and t to time.

The second law of force and acceleration states that the acceleration forms a directly proportional relationship with the net force but an inversely proportional relationship with the mass of the object [60]. The equation exists as:

$$\mathbf{F = ma} \qquad \text{further separated into} \quad \boldsymbol{F = \frac{d}{dt(mV)}}$$

Where F refers to force, m refers to mass and a refers to acceleration.

The third law states that for every singular force in nature, there is an equal but opposite force applied [61]. The equation is thus:

$$\mathbf{T = W = mg} \qquad \text{further separated into} \quad \boldsymbol{F_{ij} = G\frac{m_i m_j}{r^3 ij} r_{ij}}$$

Where F refers to force, m refers to the mass and g to the gravitational force.

Derivation of the final equation leads us to:

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

By utilizing steiner's rule for the moment of inertia [62], a condition must be set: that in order to obtain a stable system $\frac{d^2 I}{dt^2} < 0$.

3.4.3. Conditions for trajectory generation

There are certain conditions that need to be adhered to and a series of steps involved in the MD simulation process [63]:

- The choice of an interaction model
- The imposition of stipulated boundary conditions
- Determination of coordinates, properties, accelerations, and energetics of atoms
- Selection of initial positions and velocities of individual particles
- Selection of target temperatures and integration of pressure
- The application of a force field
- Equilibration of simulation to obtain thermodynamic values

After the generation of trajectories, the qualitative analysis of the system on a time dependant axis can be carried out.

3.4.4. Molecular docking

Molecular docking is the initial step in the process of drug discovery. It is used to determine the optimum configuration of a complex exhibiting the leading binding energy. Optimization of potential energy acts as a precursor in identifying the best configuration.

Receptor-ligand complexes act as the basis for the most common type of docking process [64]. Complexed and uncomplexed proteins are rigid structures undergoing the lock and key docking method [65].

The docking process investigates and predicts the optimum conformations and energetics in relation to the orientation of the ligand in the catalytic site of the protein [66]. By advancing the process of docking, further information on crystallographic structures and biological targets can be obtained. Initial docking strategies carried out are blind docking to find the most suitable active site within a protein. According to this information, there can be a range of inconsistencies that surface with the choice of incorrect binding sites including: unstable systems and unaccomplished small molecules of choice [66]. Docking results therefore need to be further validated by molecular dynamics to determine energy, stability, and flexibility of complexed systems.

3.4.5. MD analysis

MD analysis is run on the sequence of multidimensional positional vectors as a function of time characterised in a trajectory generated during a simulation. The calculations chosen are dependent on the direct objective of the study [67]. Graphical representations are generated from the assessment of said analyses. High resolution images and videos are then generated from large data files and further analysed.

A variety of analyses are carried out with a few being a necessity for complexed systems. The basic and major components of a computational study remain to be the stability of the biomolecular structure, the change in energetics during conformational fluctuations undergone during binding, and the atomic level features within the binding site of the complexed system.

3.4.5.1. System stability and convergence

The stability refers to a system existing in a thermodynamically stable state, thereby encompassing a lower potential energy. The existence of this neutrality decreases the chance of reactivity. The convergence of a system denotes the changes in bond angles of certain bonds during the repositioning of the protein during binding. The settling of the equilibration ligand-receptor complex is governed by the emergence of a levelled plateau rendering an energetically favourable conformation. The stability of the system can be established by using two types of analyses: root mean square deviation [68] and radius of gyration [69]. Convergence of a system is directly proportional to the stability of a receptor-ligand complex [70]. The change in the bond types and the evolution of the bond angles that occur during binding are measured throughout the system. Equilibrium of a system is reached, and stability is created with a plateau identifying a constant, non-changing bond angle. This accuracy then exhibits somewhat of a stable system [71]. The time frame associated with a trajectory is usually determined by the time stamp at which stability is seen to be obtained.

3.4.5.1.1. Root mean square deviation (RMSD)

The RMSD measures the spatial difference between α carbons of two structures during the course of a simulation [68]. The RMSD can be defined as follows:

$$\boldsymbol{RMSD} = \left(\frac{\sum \boldsymbol{N} \left(\boldsymbol{R_i} - \boldsymbol{R_i^0} \right)^2}{\boldsymbol{N}} \right)^{\boldsymbol{1/2}}$$

Where N refers to the number of atoms in the entire complex, R*i* refers to the position of the the *i* atom within the reference complex and R^0 refers to the alignment of the original conformation. The RMSD for a complex, ligand or receptor can hence be calculated by taking the average divided by the number of frames in the trajectory.

3.4.5.1.2. Radius of gyration (RoG)

The RoG is calculated by measuring the position of the atoms within a complex to their centre of gravity. The analysis of this method is usually referring to the over compactness of the system [69]. The RoG of a complex can be calculated using:

$$r^2g = \frac{\sum_{1-1}^{N} w_i(r_i - r^-)^2}{\sum_{1-1}^{N} w_i}$$

Where r refers to the centre mass of atom *i* and n refers to the position of that atom. Just like RMSD, the RoG is calculated as an average value across the frames of the trajectory.

3.4.5.2. Conformational changes and flexibility

3.4.5.2.1. Root mean square fluctuation (RMSF)

The RMSF measures the fluctuations of the Cα in the backbone of the protein in comparison to its original position throughout the trajectory. The RMSF allows for the investigation of the confirmational changes that occur during the binding process and measures the flexibility of the complex [72]. The RMSF can be measured by using the equation:

$$sRMSF_i \frac{(RMSF_i - RMSF)}{\sigma(RMSF)}$$

Where *i* refers to the i^{th} residue and the RMSF is subtracted from the RMSF of the i^{th} residue all divided by the standard deviation of the RMSF values.

3.4.5.3. Thermodynamics

Binding free energies are the major component of ligand-receptor interactions. These energetics determine whether the ligand will successfully bind to the receptor or not. The enthalpic and entropic contributions forming the basis of Gibbs free energy validates the stability of a ligand-receptor complex *via* the binding mechanisms associated with the complex. These energies are used in combination with binding free energies to quantitatively determine the energy values for each conformation [73]. The accurate and efficient techniques used with binding free energies are both the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics/GB Surface Area (MM/GBSA) [74] applications. Both methods estimate binding free energies with the use of continuum solvent models over the frames of the trajectory [75]. The equations below are utilised in calculating the total binding free energies and its components [76].

$$\Delta G_{binding} = \Delta G_{water}\,\textbf{(complex)} - [(\Delta G_{water}\,\textbf{(protein)} + (\Delta G_{water}\,\textbf{(peptide)}]$$

Refers to the procedure where the binding free energies could be estimated from the solvation free energies (G_{PB} + $G_{nonpolar}$), absolute energies in the gas phase (E_{gas}) and the translation, rotational and vibrational entropies. Each scheme is evaluated using the following:

$$\Delta G_{water} = E_{gas} + \Delta G_{solvation} - TS$$

$$G_{solvation} = G_{PB} + G_{nonpolar}$$

$$E_{gas} = E_{internal} + E_{electrostatic} + E_{vdW}$$

$$E_{internal} = E_{bond} + E_{angle} + E_{torsion}$$

where E_{bond}, E_{angle} and $E_{torsion}$ dictate the contributions to internal energy ($E_{internal}$) caused as a result of the overall strain given as a result of deviation from equilibrium positions, $E_{electrostatic}$ E_{vdW} denoting the contribution from electrostatic interactions and Van der Waals forces and -$T\Delta S$ representing the entropy derived in relation to temperature.

3.4.5.3.1. Per-residue energy decomposition

MM/PBSA was used to calculate the contribution of individual amino acid residues to the total binding energy [77]. The ligand-residue energetics are described by the following formula:

$$\Delta G_{\text{inhibitor- residue}} = \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{polar} + \Delta G_{non\text{-}polar}$$

Where the ligand residue interactions ($\Delta G_{\text{inhibitor-residue}}$) can be described by van der Waals contributions, electrostatic contributions, polar contributions and non-polar contributions.

3.4.5.4. Complimentary molecular dynamics analyses

3.4.5.4.1. Dynamic cross correlation matrix (DCCM)

DCC is a method used to quantitatively analyse the fluctuations reported in RMSF during the simulation. The presence of a mutation or the conformational changed experienced during ligand binding are both evident in the plots obtained by this analysis. The quantification states that +1 is a function of a correlated motion and -1 a function of an anti-correlated motion [78]. The equation used to calculate DCC is as follows:

$$C_{ij} \frac{\langle \Delta r_i \Delta r_j \rangle}{\left(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle \right)^{1/2}}$$

Where C*ij* refers to the cross-correlation coefficient, *i* refers to the *ith* residue, j referring to the *jth* residue and Δr refers to the vectors of displacement associated with each residue.

3.4.5.4.2. Principal component analysis (PCA)

Principal component analysis (PCA) is a method employed to simplify complex motions and flexibility within proteins to extract biological movements relevant from atomic motion. This method groups molecular conformational similarities together and determines the variability measured by directional eigenvalues (PC1 and PC2) [79]. It is established *via* the functionality

of covariance matrices where a positive value indicates a correlated motion, and a negative value indicates the presence of an anti-correlated relationship.

Vectors are calculated using eigenvectors and eigenvalues which stipulate that for every eigenvector, there is an eigenvalue corresponding to it [80]. These correspond to the multidimension of the molecule and clarify significance.

$$t_{k(i)} = x_{(i)} \cdot w_{(k)}$$

Where t refers to individual variables to obtain a maximum possible variance from x is the data matrix and w is the coefficient basis of matrix vectors.

3.5.1. Residue interaction network (RIN)

RIN is a method of analysing residue-residue interactions present within a system. The residue interaction network is a topology-based approach used in identifying key residue interactions present in proteins. Analysing these interactions could shed light on the differences in binding of the amino acids present in backbones. This is done by analysing atomic packing by means of using an approximately 0.25Å probe for each covalent and non-covalent interaction of every residue. Touching or overlap of 2 probes results in the generation of a contact dot [81].

3.5.2. Virtual screening – tailored pharmacophore

Virtual screening forms an integral part of the drug discovery process. It utilizes a variety of libraries containing small molecules which are continuously being updated. The millions of small compounds are screened against similar drug hits and drug targets of interest [82]. Virtual screening forms a method whereby criteria are put into place to decrease the number of compounds filtered through to identify alternative inhibitors. It plays on the fact that similar compounds, containing similar atoms will behave in the same manner if placed in an equivalent environment.

This technique is carried out using either a ligand-, receptor-, or tailored pharmacophore-based approach. The ligand-based approach screens ligands based on beneficial structural properties

of the small molecule and the consequent binding to targets of interest [83]. The target-based approach screens ligands based on the structural property of the receptor and beneficial aspects that will be governed by the catalytic site of the protein [84]. While both approaches offer their respective benefits, a target-based approach is more advantaged as it relies on a possible complex rather than an isolated ligand. Tailored pharmacophore is a promising approach to identify novel inhibitors of drug targets that have developed drug resistance [85]. Tailored pharmacophore is a method whereby small molecules are built up from a base dependent on interaction moieties to acquire the most efficient set of protein-ligand interactions with the amino acids in the active site of the protein thereby leading to an increased binding energy. It uses the interactions brought about by hydrogen binding, hydrophobic and hydrophilic interactions within a receptor-ligand complex [86]. The possible drug hits have proven to offer a feasible approach in many drug discovery instances.

References

1. Ekins, S., Mestres, J. & Testa, B. In silico pharmacology for drug discovery: applications to targets and beyond. *Br. J. Pharmacol.* **152**, 21–37 (2007).

2. Jambrina, P. G. & Aldegunde, J. Computational tools for the study of biomolecules. in *Computer Aided Chemical Engineering* vol. 39 583–648 (Elsevier, 2016).

3. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347**, 631–639 (1990).

4. Hoffer, L., Renaud, J.-P. & Horvath, D. Fragment-based drug design: computational and experimental state of the art. *Comb. Chem. High Throughput Screen.* **14**, 500–520 (2011).

5. Macalino, S. J. Y., Gosu, V., Hong, S. & Choi, S. Role of computer-aided drug design in modern drug discovery. *Arch. Pharm. Res.* **38**, 1686–1701 (2015).

6. Adelusi, T. I. *et al.* Molecular modeling in drug discovery. *Informatics Med. Unlocked* 100880 (2022).

7. Forster, M. J. Molecular modelling in structural biology. *Micron* **33**, 365–384 (2002).

8. Carlson, H. A. Protein flexibility is an important component of structure-based drug discovery. *Curr. Pharm. Des.* **8**, 1571–1578 (2002).

9. Barril, X. & Soliva, R. Molecular modelling. *Mol. Biosyst.* **2**, 660–681 (2006).

10. Feig, M. & Mirjalili, V. Protein structure refinement via molecular-dynamics simulations: what works and what does not? *Proteins Struct. Funct. Bioinforma.* **84**, 282–292 (2016).

11. Boyd, D. B. & Lipkowitz, K. B. Molecular mechanics: The method and its underlying philosophy. *J. Chem. Educ.* **59**, 269 (1982).

12. Rappe, A. K. & Casewit, C. J. *Molecular mechanics across chemistry*. (University Science Books, 1997).

13. Gillespie, R. J. The valence-shell electron-pair repulsion (VSEPR) theory of directed valency. *J. Chem. Educ.* **40**, 295 (1963).

14. Chang, C.-E. A., Huang, Y.-M. M., Mueller, L. J. & You, W. Investigation of

structural dynamics of enzymes and protonation states of substrates using computational tools. *Catal. (Basel, Switzerland)* **6**, 82 (2016).

15. Adcock, S. A. & McCammon, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* **106**, 1589–1615 (2006).

16. Tsai, P.-C. & Fang, T.-H. A molecular dynamics study of the nucleation, thermal stability and nanomechanics of carbon nanocones. *Nanotechnology* **18**, 105702 (2007).

17. Mannfors, B., Pietilä, L.-O., Palmö, K., Sundius, T. & Krimm, S. Molecular mechanics potential energy functions from vibrational spectroscopy and quantum chemistry. *J. Mol. Struct.* **408**, 63–69 (1997).

18. Ali, M. Atomic structure and binding of carbon atoms. (2022).

19. Vanommeslaeghe, K. & Guvench, O. Molecular mechanics. *Curr. Pharm. Des.* **20**, 3281–3292 (2014).

20. Botu, V. & Ramprasad, R. Learning scheme to predict atomic forces and accelerate materials simulations. *Phys. Rev. B* **92**, 94306 (2015).

21. Allinger, N. L., Zhou, X. & Bergsma, J. Molecular mechanics parameters. *J. Mol. Struct. THEOCHEM* **312**, 69–83 (1994).

22. Levitt, M., Hirshberg, M., Sharon, R. & Daggett, V. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comput. Phys. Commun.* **91**, 215–231 (1995).

23. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210 (2013).

24. MacKerell Jr, A. D., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolym. Orig. Res. Biomol.* **56**, 257–265 (2000).

25. Scott, W. R. P. *et al.* The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **103**, 3596–3607 (1999).

26. Wallner, B. & Elofsson, A. All are not equal: a benchmark of different homology modeling programs. *Protein Sci.* **14**, 1315–1327 (2005).

27. Dodda, L. S., Cabeza de Vaca, I., Tirado-Rives, J. & Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **45**, W331–W336 (2017).

28. Martín-García, F., Papaleo, E., Gomez-Puertas, P., Boomsma, W. & Lindorff-Larsen, K. Comparing molecular dynamics force fields in the essential subspace. *PLoS One* **10**, e0121114 (2015).

29. Klein, M. J. Max Planck and the beginnings of the quantum theory. *Arch. Hist. Exact Sci.* **1**, 459–479 (1961).

30. De Broglie, L. Waves and quanta. *Nature* **112**, 540 (1923).

31. Beller, M. Pascual Jordan's influence on the discovery of Heisenberg's indeterminacy principle. *Arch. Hist. Exact Sci.* **33**, 337–349 (1985).

32. Renn, J. Schrödinger and the genesis of wave mechanics. (2013).

33. Wilson, H. A. The ultimate particles: electrons, protons, photons, and neutrons. *Rice Inst. Pam. Univ. Stud.* **20**, (1933).

34. Tatulian, S. A. From the Wave Equation to Biomolecular Structure and Dynamics. *Trends Biochem. Sci.* **43**, 749–751 (2018).

35. Bohm, D. *Quantum theory*. (Courier Corporation, 2012).

36. Shankar, R. *Principles of quantum mechanics*. (Springer Science & Business Media, 2012).

37. Kragh, H. *Niels Bohr and the quantum atom: The Bohr model of atomic structure 1913-1925*. (OUP Oxford, 2012).

38. Schrödinger, E. Wave Mechanics. (1926).

39. McKagan, S. B., Perkins, K. K. & Wieman, C. E. Why we should teach the Bohr model and how to teach it effectively. *Phys. Rev. Spec. Top. Educ. Res.* **4**, 10103 (2008).

40. Schrödinger, E. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.* **28**, 1049 (1926).

41. Wahl, A. C. Chemistry by computer. *Sci. Am.* **222**, 54–71 (1970).

42. Anderson, J. B. A random-walk simulation of the Schrödinger equation: H+ 3. *J. Chem. Phys.* **63**, 1499–1503 (1975).

43. Dingle, A. & Carpi, A. Atomic Theory IV: Quantum Numbers and Orbitals.

44. Ericson, T. The statistical model and nuclear level densities. *Adv. Phys.* **9**, 425–511 (1960).

45. Combes, J.-M., Duclos, P. & Seiler, R. The born-oppenheimer approximation. in *Rigorous atomic and molecular physics* 185–213 (Springer, 1981).

46. Born, M. & Oppenheimer, J. R. On the quantum theory of molecules. *Сборник статей к мультимедийному электронному учебно-методическому комплексу по дисциплине «физика атома и атомных явлений»/отв. ред. Шундалов МБ; БГУ, Физический факультет* (1927).

47. Schlick, T. The 2013 Nobel Prize in chemistry celebrates computations in chemistry and biology. *SIAM News* **46**, 4 (2013).

48. Sutcliffe, B. T. & Woolley, R. G. On the quantum theory of molecules. *J. Chem. Phys.* **137**, 22A544 (2012).

49. Hermann, J., Schätzle, Z. & Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **12**, 891–897 (2020).

50. Kosloff, R. & Tal-Ezer, H. A direct relaxation method for calculating eigenfunctions and eigenvalues of the Schrödinger equation on a grid. *Chem. Phys. Lett.* **127**, 223–230 (1986).

51. Agostini, F., Abedi, A., Suzuki, Y. & Gross, E. K. U. Mixed quantum-classical dynamics on the exact time-dependent potential energy surface: a fresh look at non-adiabatic processes. *Mol. Phys.* **111**, 3625–3640 (2013).

52. Bakowies, D. & Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **100**, 10580–10594 (1996).

53. Ceperley, D. M. & Libby, S. B. Berni Julian Alder, theoretical physicist and inventor of molecular dynamics, 1925–2020. *Proc. Natl. Acad. Sci.* **118**, e2024252118 (2021).

54. Alder, B. J. & Wainwright, T. E. Studies in molecular dynamics. II. Behavior of a small number of elastic spheres. *J. Chem. Phys.* **33**, 1439–1451 (1960).

55. Jorgensen, W. L. & Tirado-Rives, J. Monte Carlo vs molecular dynamics for conformational sampling. *J. Phys. Chem.* **100**, 14508–14513 (1996).

56. Bartels, C. Analyzing biased Monte Carlo and molecular dynamics simulations. *Chem. Phys. Lett.* **331**, 446–454 (2000).

57. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **9**, 1–9 (2011).

58. Hecht, E. Origins of Newton's first law. *Phys. Teach.* **53**, 80–83 (2015).

59. Drake, S. Galileo and the Law of Inertia. *Am. J. Phys.* **32**, 601–608 (1964).

60. Pourciau, B. Newton's interpretation of Newton's second law. *Arch. Hist. Exact Sci.* **60**, 157–207 (2006).

61. Cornille, P. Review of the application of Newton's third law in physics. *Prog. energy Combust. Sci.* **25**, 161–210 (1999).

62. Dodig, M. Models and modelling of dynamic moments of inertia of human body. *Int. J. Sport. Sci.* 249–256 (2016).

63. Kofke, D. A. Getting the most from molecular simulation. *Mol. Phys.* **102**, 405–420 (2004).

64. Morris, G. M. & Lim-Wilby, M. Molecular docking. in *Molecular modeling of proteins* 365–382 (Springer, 2008).

65. Tripathi, A. & Bankaitis, V. A. Molecular docking: From lock and key to combination lock. *J. Mol. Med. Clin. Appl.* **2**, (2017).

66. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided. Drug Des.* **7**, 146–157 (2011).

67. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

68. Fukutani, T., Miyazawa, K., Iwata, S. & Satoh, H. G-RMSD: Root mean square deviation based method for three-dimensional molecular similarity determination. *Bull. Chem. Soc. Jpn.* **94**, 655–665 (2021).

69. Lobanov, M. Y., Bogatyreva, N. S. & Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* **42**, 623–628 (2008).

70. Sawle, L. & Ghosh, K. Convergence of molecular dynamics simulation of protein native states: Feasibility vs self-consistency dilemma. *J. Chem. Theory Comput.* **12**, 861–869 (2016).

71. Sargsyan, K., Grauffel, C. & Lim, C. How molecular size impacts RMSD applications in molecular dynamics simulations. *J. Chem. Theory Comput.* **13**, 1518–1524 (2017).

72. Bornot, A., Etchebest, C. & De Brevern, A. G. Predicting protein flexibility through the prediction of local structures. *Proteins Struct. Funct. Bioinforma.* **79**, 839–852 (2011).

73. Marelius, J., Hansson, T. & Åqvist, J. Calculation of ligand binding free energies from molecular dynamics simulations. *Int. J. Quantum Chem.* **69**, 77–88 (1998).

74. Wang, C., Greene, D., Xiao, L., Qi, R. & Luo, R. Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* **4**, 1–18 (2018).

75. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449–461 (2015).

76. Massova, I. & Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. drug Discov. Des.* **18**, 113–135 (2000).

77. Zoete, V. & Michielin, O. Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein–protein association using MM-GBSA: Application to the TCR-p-MHC complex. *Proteins Struct. Funct. Bioinforma.* **67**, 1026–1047 (2007).

78. Hünenberger, P. H., Mark, A. E. & Van Gunsteren, W. F. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* **252**, 492–503 (1995).

79. Altis, A., Nguyen, P. H., Hegger, R. & Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **126**, 244111 (2007).

80. Stein, S. A. M., Loccisano, A. E., Firestine, S. M. & Evanseck, J. D. Principal

components analysis: a review of its application on molecular dynamics data. *Annu. Rep. Comput. Chem.* **2**, 233–261 (2006).

81. Contreras-Riquelme, S., Garate, J.-A., Perez-Acle, T. & Martin, A. J. M. RIP-MD: a tool to study residue interaction networks in protein molecular dynamics. *PeerJ* **6**, e5998 (2018).

82. Broccatelli, F. & Brown, N. Best of both worlds: on the complementarity of ligand-based and structure-based virtual screening. *J. Chem. Inf. Model.* **54**, 1634–1641 (2014).

83. Ripphausen, P., Nisius, B. & Bajorath, J. State-of-the-art in ligand-based virtual screening. *Drug Discov. Today* **16**, 372–376 (2011).

84. Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discov. Today* **7**, 1047–1055 (2002).

85. von Korff, M. & Steger, M. GPCR-tailored pharmacophore pattern recognition of small molecular ligands. *J. Chem. Inf. Comput. Sci.* **44**, 1137–1147 (2004).

86. Machaba, K. E., Mhlongo, N. N., Dokurugu, Y. M. & Soliman, M. E. S. Tailored-pharmacophore model to enhance virtual screening and drug discovery: a case study on the identification of potential inhibitors against drug-resistant *Mycobacterium tuberculosis* (3R)-hydroxyacyl-ACP dehydratases. *Future Med. Chem.* **9**, 1055–1071 (2017).

CHAPTER 4

Drug repurposing Approach against *Mycobacterium tuberculosis* Enoyl-[acyl-carrier-protein] reductase: Insight from Molecular Dynamics Simulations

Kimona Kisten[1], Hezekiel M. Kumalo[1], Kgothatso E. Machaba[1], Umar Ndagi[2], and Ndumiso N. Mhlongo[1*]

[1]School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban 4001, South Africa

[2]Africa Centre of Excellence for Mycotoxin and Food Safety, Federal University of Technology, Minna

* Corresponding author: Ndumiso N. Mhlongo; Email: MhlongoN4@ukzn.ac.za

Telephone: +27 {0} 31 260 2428, Fax: +27 031 260 7792

Abstract

Enoyl-[acyl-carrier-protein] reductase (InhA), a major contributor in forming mycolic acids, is one of the major targets for the inhibition of *Mycobacterium tuberculosis* (*Mtb)*. The emergence and spread of drug resistant strains has resulted in a drug repurposing approach against *Mtb*-InhA. Herein, we introduce a drug repurposing approach from a molecular docking and molecular dynamic (MD) simulation perspective. Drug resistance and inappropriate use or failure to complete the full course of treatment is one of the most critical issues in health care. Hence, a drug repurposing approach was found hopeful to identify *in silico* hits with better binding affinities as compared to isoniazid, a known drug against *Mtb*. Four proposed drug hits (DB04007, DB05291, DB07453 and DB08607) all displayed higher total binding energies (ΔG_{tot}) in relation to isoniazid with DB04007 exhibiting a ΔG_{tot} of -41.11 as compared to the -7.89 kcal/mol observed from isoniazid. The RMSD of DB04007 stabilised with an average of 1.35Å while the control displayed an average of 1.99 Å. This pattern was reiterated by the lower RoG of DB04007 as compared to isoniazid. The RMSF of both DB04007 and the control exhibited a structural evolution denoted to common flexibility changes possibly related to the ligand induced characteristic conformational changes of the protein. A distinct increase in fluctuation was seen in the RMSF of DB04007 which could be attributed to Ser92 forming 2 hydrogen bonds with the inhibitor, the distal amino acids affecting stability and the evolution of residues 195-210 from a primary to a secondary structure. This prolonged achievement of stability resulted in a highly stable complex. DB04007, a drug of potential significance, exhibited a higher total binding free energy due to the residual VdW and hydrogen bonding interactions as compared to the control. The PCA analysis displayed negative loading on the PCA2 axis enhancing the previous results of decreased flexibility of the backbone also confirmed by the DCC correlations. This approach can potentially serve as a platform to the development and discovery of novel drugs against a wide range of *Mtb* targets.

Keywords

Mtb-InhA, Drug repurposing, Molecular dynamics simulation

Introduction

Tuberculosis is an infectious disease caused by the pathogenic bacteria *Mycobacterium tuberculosis (Mtb)* [1]. It is an airborne disease that enters through the alveolar passage, attacks the lungs then spreading to the rest of the body [2]. According to current studies, 85% of the total infected cases are pulmonary tuberculosis seen to cause lung cavitation, necrosis, and bleeding[3]. TB is the cause of approximately 2 million deaths worldwide every year, with the infection rate jumping from 10 million to 10.4 million and the mortality rate increasing from 1.5 million to 1.66 million people from 2018 to 2020 [4,5]. Approximately 87% of the total amount were from a list of 30 high risk, developing countries including examples like Nigeria, Philippines, Pakistan, and South Africa [6]. Developing countries exhibit higher mortality rates because of the compounded effect of coinfections with other diseases such as the human immunodeficiency virus (HIV). Although coinfection with HIV is seen to be approximately 3% worldwide, developing countries observe a much higher rate [7]. This is simultaneous with the studies carried out in Vietnam and Brazil [8,9]. South Africa has an infection increase of approximately 450000 people every year with 270000 (60%) attributed to coinfection with HIV and 87000 (19%) resulting in death [10].

Treatment of tuberculosis is required in patients exhibiting symptoms as well as patients carrying latent TB. Long term treatments require antibiotics for a period of 6 to 9 months. The most common drugs used are isoniazid, rifampin, ethambutol and pyrazinamide known as the set of first line drugs [11]. These drugs were used to inhibit the drugs targets InhA [Enoyl-(acyl carrier-protein) reductase], Bacterial RNA polymerase (RNAP), Arbinosyl Transferase and the S1 Component of 30S Ribosomal subunit [12]. Consequent to the variable efficacy of the drugs used in TB treatment and lack of patient compliance to current treatment regiment, drug resistant *M. tuberculosis* strains have emerged [13]. Extent of drug resistance can be determined by the ineffective response to one or more of the first line drugs where drug resistant (DR-TB) strains constitute the ineffective response to one of the first line drugs and multi drug resistant (MDR-TB) strains indicate ineffective response to more than one first line drug. In this case, use of the second line drugs, including fluoroquinolones, amikacin, kanamycin and capreomycin is observed. Unfortunately, further strains known as extensively drug resistant (XDR) and emerging extensively drug resistant (eXDR) strains have emerged and are also resistant to second line drugs [14,15]. Approximately 5% of the estimated 10 million TB infections each year are DR-TB, 2.9% of the DR-TB are MDR-TB and 6.2% of the MDR-TB are XDR-

TB [16,17]. Considering the exponential increase of drug resistance worldwide, new drugs and drug targets need to be identified to combat the disease [18].

One of the main *Mtb* drug targets is InhA, initially inhibited by isoniazid. InhA belongs to the fatty acid synthetase II system and prevention of the production of these long chain fatty acids have proved to decrease mycolic acid biosynthesis inhibiting *mycobacterium tuberculosis* [19].

A selected list of FDA-approved drugs predicted as potential InhA inhibitors through bioinformatics tools were repurposed against *M. tuberculosis*-InhA. The predicted drugs of interest are Bromo-WR99210, Lidocaine, 2-phenyl-4H-Benzo[H]Chromen-4-one and (5R)-5-(4-{[(2R)-6-Hydroxy-2,5,7,8-Tetramethyl-3,4-Dihydro-2H-Chromen-2-yl]Methoxy}Benzyl)-1,3-Thiazolidine-2,4-dione [20] with respective DrugBank IDs of DB04007, DB05291, DB07453, DB08607.

Analysis of literature suggests the potential inhibition capacity of these selected drugs against InhA using a network pharmacology and gene ontology approach. This was paired with network-based inference and structural similarity index studies in order to determine a starting point for the targets exhibiting oxidoreductase activity. The selected drugs are known to possess antibacterial, antimicrobial, anti-arrhythmic and local anaesthetic properties, making them ideal molecules for repurposing [20].

The increasing burden of diseases and the cost of drug discovery and development has called for the application of corner-cutting drug discovery strategies such as drug repurposing and molecular modelling methods. Drug repurposing strategies repositions known and clinically approved drugs to treat other diseases. This can be achieved by investigating the effects of such drugs on a drug target of another disease by applying molecular modelling methods prior to experimental work to predict the outcome. In this report, we assess the potential of selected drugs as inhibitors of the *M. tuberculosis* protein InhA through the application of molecular modelling methods mainly molecular docking and molecular dynamics simulations. Molecular docking methods predict the probability of a small molecule binding to its receptor, while molecular dynamics simulations present intermolecular dynamics between a small molecule and its receptor.

Although InhA has been studied extensively in the past, repurposing of these specific drug hits has not been previously investigated against this target using the molecular modelling method. In this report, we conduct molecular docking of the selected drugs into the protein structure of InhA and molecular dynamics simulations of the generated complexes to answer the following

questions: **{1}** Are the selected drugs capable of binding to the InhA protein structure? , **{2}** With what magnitude of binding affinity will they bind to InhA compared to the reference drug? , **{3}** Are the selected drugs capable of inducing an inhibitory effect on the protein *in silico*? , **{4}** What would be the mode of inhibition elicited by the selected drugs on InhA?

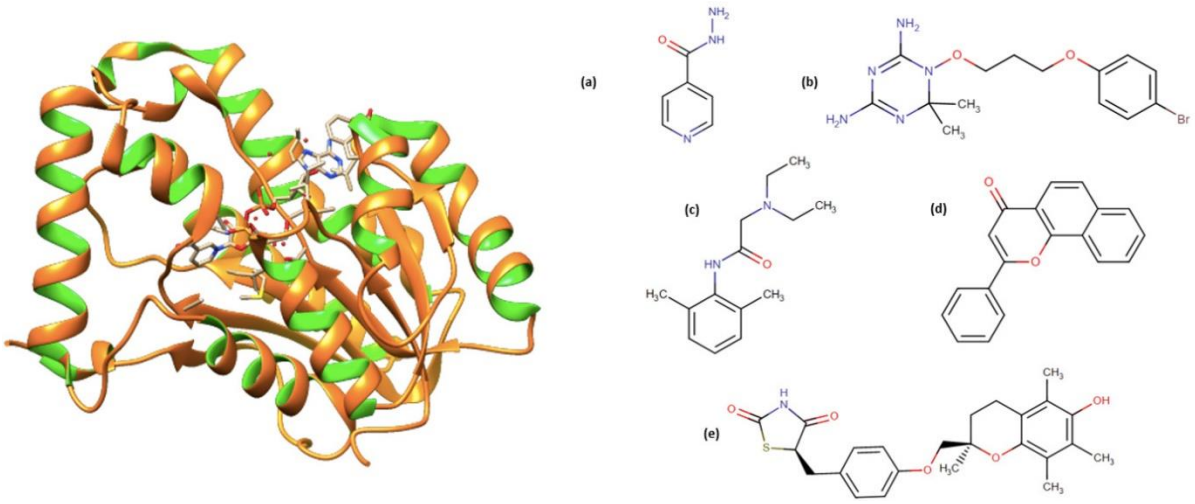


Figure 1: Structures of drug target *Mtb* inhA in complex with potential inhibitors and individual drugs of (a) Isoniazid (b) DB04007 (c) DB05291 (d) DB07453 and (e) DB08607

Materials and computational methods

Preparation for molecular docking

The x-ray crystallographic structure of *M. tuberculosis* InhA was obtained from the RSCB Protein Data Bank [PDB ID: 4DRE] [21]. Structures of drugs were obtained from DrugBank [22]. The protonation states of amino acids of InhA were fixed with the aid of the H++ server [23]. Hydrogen atoms were added to the protein using the UCSF Chimera [24] software and deleted from the ligands using the Avogadro [25] software.

Molecular docking calculations were performed using Autodock Vina version 4.2.6 supplied by MGL tools [26]. Throughout docking, Geister partial charges were distributed and profiled using the AutoDock Graphical manager interface provided by MGL tools [26]. Utilising the Lamarckian Genetic Algorithm [27], docked conformations were obtained. The binding site was identified using Autodock Vina with the following grid box parameters: centre coordinates of x = 2.083 Å, y = -31.306 Å and z = 14.944 Å and grid box dimensions of x = 48 Å, y = 46 Å and z = 46 Å. The resultant protein-ligand complexes were ranked according to their docking energy scores (kcal/mol) and saved for further molecular dynamics simulations.

Molecular dynamics simulation

Molecular dynamics simulations are used to determine physical behaviour of atoms in a system described by a particular force field [28]. Simulation of complexes was conducted using a GPU version of the Particle Mesh Ewald Molecular Dynamics (PMEMD) module of Amber18. The protein was described by the application of the Amber force field FF12SB [29]. Ligand parameters were set using Gasteiger charges in the Avogadro software. The Generalized Amber Force Field [29] was set using the Antechamber module. Hydrogen atoms were added to the protein structure and counter ions using the LEaP module [30] to neutralize the system. A TIP3P water box model with a 10 Å distance between the protein and water box boundary was applied. Long range electrostatic interactions with nonbonding cut-off distance of 12 Å were treated with the Particle mesh Ewald (PME) method [30]. The SANDER module was applied to conduct partial minimization of the system with a restraint potential of 500 kcal/mol $Å^{-2}$ for the solute for 1000 steps. Moderate heating of the systems from 0 to 300 K was performed using the

Langevin thermostat with harmonic restraints of 5 kcal/mol $Å^{-2}$ for the solute and a collision frequency of 1 ps. Systems were subsequently equilibrated without restrictions at 300 K and constant pressure of 1 bar for 2 ns. Bonds involving hydrogen atoms were treated with a SHAKE algorithm [29] and electrostatic interactions were treated with the PME method. Molecular dynamics simulations of 200 ns were performed in an isothermal-isobaric ensemble using a Berendsen barostat at pressure of 1 bar and 2 ps pressure-coupling constant. Analytical instruments namely root mean square deviation (RMSD), root mean square fluctuation (RMSF), hydrogen bond (Hbond) formation and principal component analysis (PCA) were calculated from the generated molecular dynamics trajectories using a CPPTRAJ module [31].

Post-molecular dynamics analyses

The resultant trajectories were analysed by the CPPTRAJ module of Amber 18. Graphical software programs were used to view and evaluate molecular structures. Origin version 8 software (Origin. 2003. Origin 8 OriginLab Corp., Northampton, MA) was used to generate graphs while ligand-protein interactions were analysed using LigPlus [32].

Binding free energy calculations

Ligand-protein binding affinity was calculated averaging snapshots over a 200ns trajectory using the Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) method [30] implemented in AMBER 18. An FF12SB force field utilising a radius of 1.4 Å was used [29]. This method uses the equations below to calculate the total binding free energy and its components.

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand}) \quad (1)$$

$$\Delta G_{binding} = E_{gas} + G_{sol} - T\Delta S \quad (2)$$

$$E_{gas} = E_{int} + E_{ele} + E_{vdw} \quad (3)$$

where $\Delta G_{binding}$ represents protein-ligand complex total binding energy, encompassing a sum of gas phase molecular mechanics energy (E_{gas}), solvation energy (G_{sol}) and entropy ($-T\Delta S$)

derived using the temperature. E_{gas} is a combination of internal energy (E_{int}), electrostatic (E_{ele}) and Van de Waals forces (E_{vdw}).

Per-residue energy decomposition

The MM/PBSA approach was used to calculate the contribution of individual amino acid residues to the total binding energy of the ligand. Ligand-residue energetics are described by the following formula:

$$\Delta G_{\text{inhibitor- residue}} = \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{polar} + \Delta G_{non\text{-}polar}$$

Where the ligand residue interactions ($\Delta G_{\text{inhibitor- residue}}$) can be described by van de Waals contributions, electrostatic contributions, polar contributions and non – polar contributions [29].

Results and Discussion

Molecular Docking

Molecular docking was applied to determine the most stable conformation of the complex, displaying the most suitable spatial orientation and binding affinity [33]. Figure 2 shows predicted drugs (DB04007, DB05291, DB07453 and DB08607) and isoniazid superimposed in the active site of *Mtb* InhA.

The docking results indicated values of -5.7, -6.4, -8.6, -7.3 and -10.2 kcal/mol for isoniazid, DB05291, DB07453, DB04007 and DB08607, respectively (Table 2). All drug hits were seen to occupy the binding site of the target and all displayed higher docking scores as compared to the control.

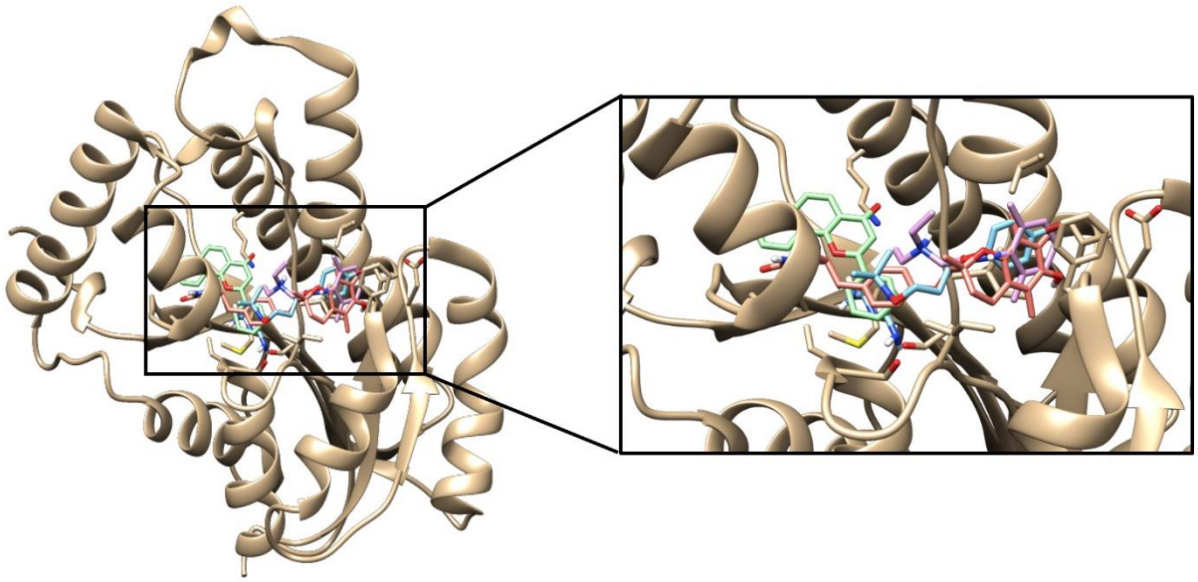


Figure 2: Receptor in complex with known inhibitor, isoniazid, DB04007, DB05291, DB07453 and DB08607 in the same active site pocket.

Complex stability and flexibility

RMSD

Root mean square deviation (RMSD) was calculated to determine the stability of all systems throughout the 200ns simulation. Figure 3 shows the assessment of the stability of each system with the respective RMSD values represented graphically [34].

Throughout the simulations, as evident in Table 1, the DB08607 system showed a higher average RMSD of 1.9 Å while the DB04007 system displayed a lower average value of 1.3 Å. The RMSD values of all conformations were within a reasonable fluctuation range of 1.3 - 1.9 Å. Hence, all systems conform to the standard average RMSD value of 2 Å [35]. As shown in Figure 3, all systems display a similar fluctuation pattern except for the DB04007 system. Thus, the presence of the DB04007 in the active site was found to impact the overall protein dynamic as compared to isoniazid, DB05291, DB07453 and DB08607.

We further analysed the DB04007 and isoniazid systems to get more insight into the stability during the simulation. Snapshots along the trajectories of 200ns MD simulations for the DB04007 and isoniazid conformations of InhA are given in Figure 4. Results, as shown in Figure 4, suggest that the binding of DB04007 to the InhA protein leads to an unstable conformation in residues 193 to 204. It has been reported that residues 195-210, containing an α6 helix, forms a disordered loop, eventually leading to the closure of the substrate binding pocket [36]. Hence, this can further support the fluctuations present in the RMSD pattern. The RMSD results presented herein provide solid information in an attempt to understand the dynamic behaviour of InhA bound to DB04007 and isoniazid. The current report highlights important features of an important InhA target, which could serve as an initial point in the process towards identifying more potential inhibitors.

There was an observed drift in the RMSD at approximately 180 ns for the DB04007 complex. Initially it was assumed that the size of the ligand in comparison to isoniazid was the reason for the distinctive change in dynamics. Further observation of the trajectories at timeframes in that regions of 155, 175 and 200 ns indicated a change in the ligand position and conformation within the active site thereby inducing changes in residues Ser92, Ile93 and Gly94 graduating to the structural evolution of the protein.

Table 1: RMSD averages of DB04007, DB05291, DB07453 and DB08407 against isoniazid (DB00951) in complex with *InhA*.

Drugs	Average RMSD
Isoniazid	1.86
DB04007	1.35
DB05291	1.63
DB07453	1.70
DB08607	1.99

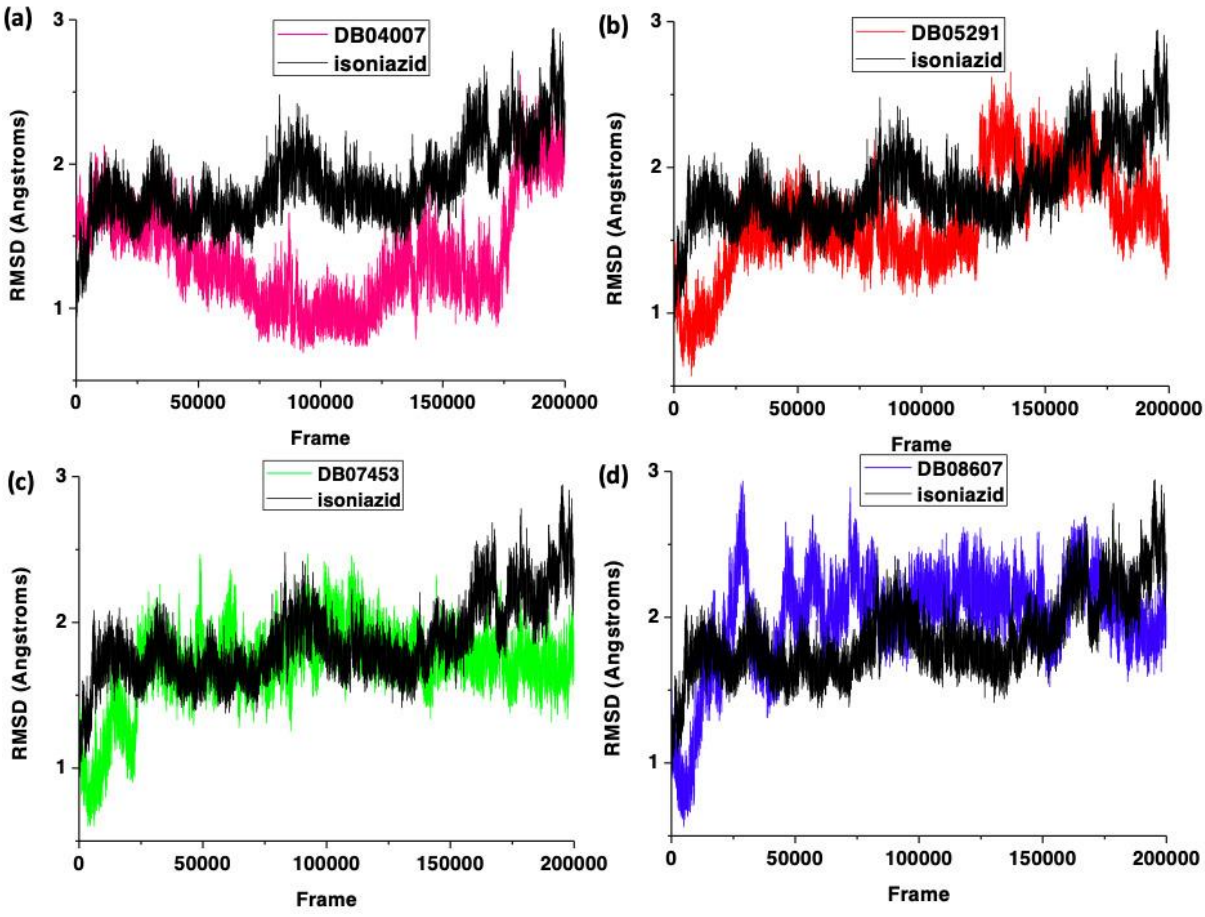


Figure 3: Graphs showing the RMSD of InhA in complex with (a) DB04007 (b) DB05291 (c) DB07453 and (d) DB08607 plotted against InhA-isoniazid complex (black).

RMSF

Direct interaction of a ligand and active site residues induces a conformational change of the protein structure [37]. The root mean square fluctuation (RMSF) is the measure of the flexibility of the protein structure in relation to the binding site amino acids as it compares the rotation of the Cα to its average position in the protein after binding [38]. RMSF was computed to investigate the flexibility and structural behaviour of the residues upon ligand binding. The RMSF was calculated from the 200 ns trajectories. Figure 6 indicates that all RMSF profiles follow a similar trend with constant fluctuations at 10-25, 35-60, 60-90, 90-125 and 140-160 indicating significant movement of the Cα atoms in relation to its original position due to ligand induced conformational changes. For all systems, restricted movement was seen at 25-35, 80-90, 125-140 and 140-180 denoting ligand-residue interaction and steric hindrance at the binding site. DB07453 restricts mobility to the same extent as isoniazid while DB05291 and DB08607 decrease the mobility of the amino acids at 10-25, 35-60 and 60-90. The DB04007 complex shows a higher magnitude of fluctuations compared to the corresponding monomers and showed increased flexibility at the 60-90 and 90-125 regions leading to decreased flexibility thereafter. This was confirmed by the investigation of snapshots annotating the high fluctuations to the ligand orientation and residues present in the interactions. This further confirms the longer stabilization period obtained for the RMSD of DB04007. Residues with higher fluctuations are located in the α-helix and β-sheet region, clearly indicating regions of higher mobility in the protein[39]. The fluctuation peaks present in the RMSF figures are denoted to individual amino acids present in the binding pocket. These amino acids in the catalytic site contribute significantly to the total binding energies of each InhA-inhibitor complex. A positive correlation is established between the conformational changes and binding energies attributed to electrostatic and vdW forces.

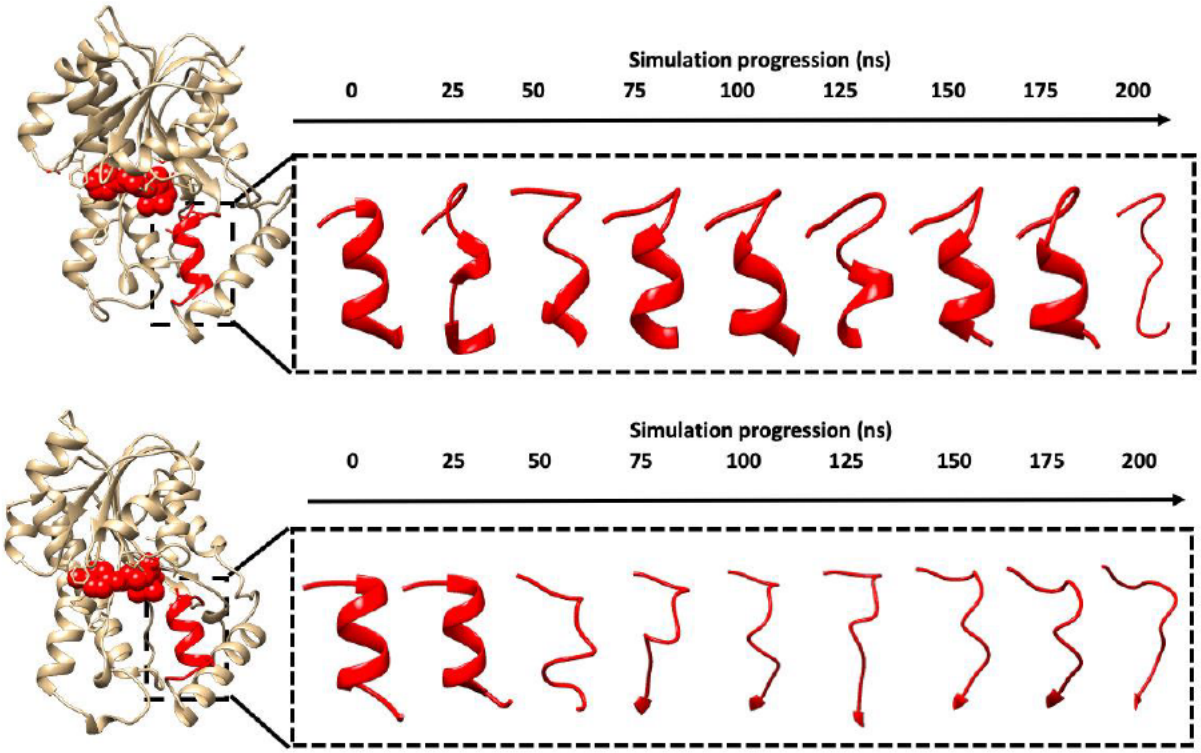


Figure 4: Graphical representation of the evolution of residues 193-204 of the inhA-DB04007 complex in comparison to the isoniazid-complex.

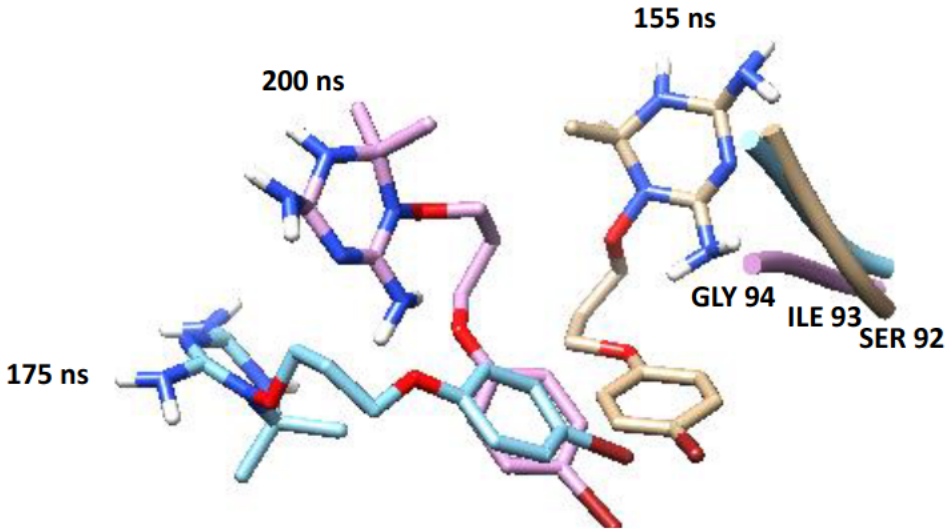


Figure 5: Positional changes and conformational evolution of DB04007 in the InhA active site. DB04007 colour-coded conformations were sampled at 155 ns, 175 ns and 299 ns of a 200 ns trajectory. Conformationally impacted nearby loop with residues 92 to 94 presented in corresponding colour code.

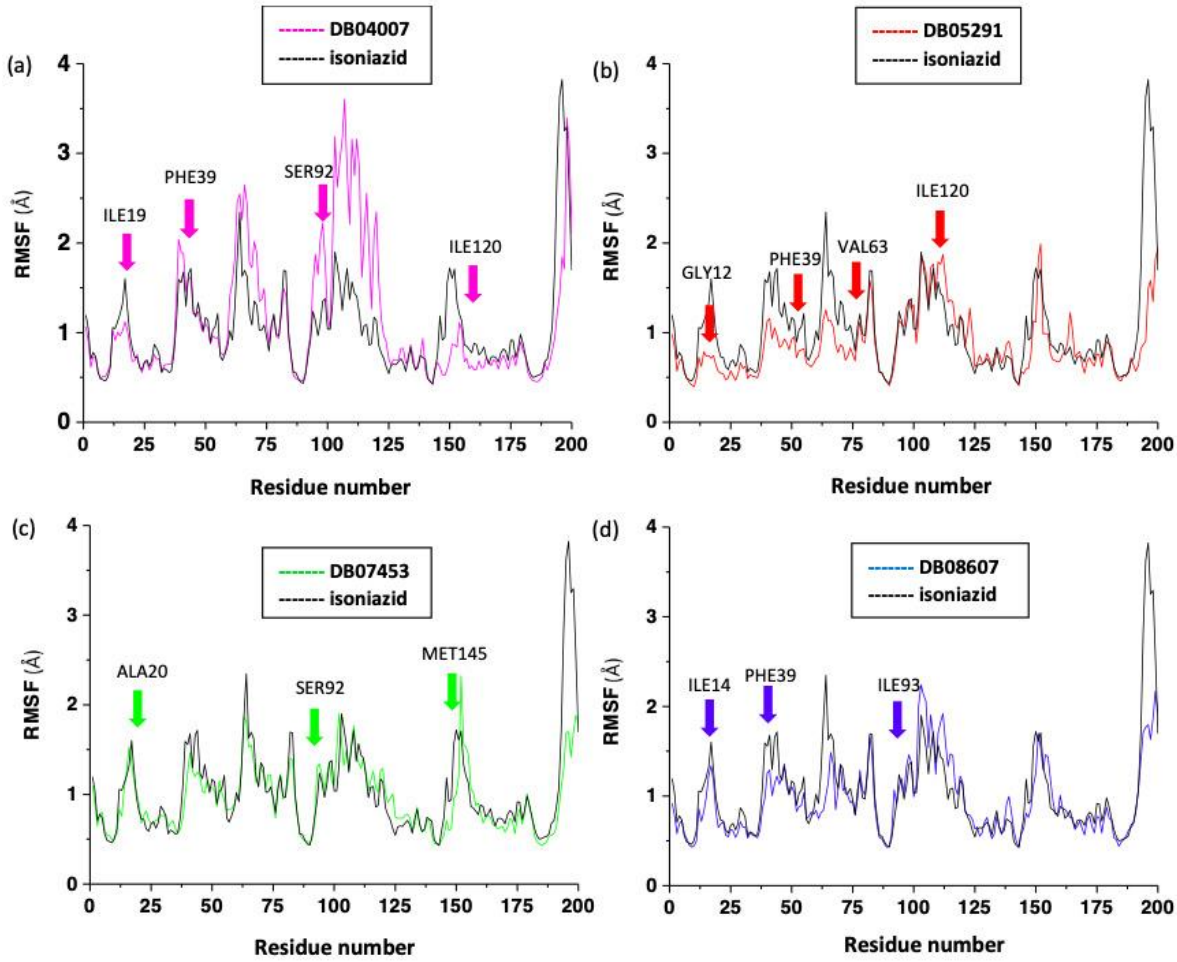


Figure 5: Graphs showing the RMSF of InhA in complex with (a) DB04007 (b) DB05291 (c) DB07453 (d) DB08607 and isoniazid.

RoG

The radius of gyration graphically presents insight into the equilibrium of the complexed system during the simulation [40]. All four potential inhibitors complexed with InhA were plotted against the control and presented in Figure 7. Analysis of RMSD indicated possible initial changes in conformational structure. This theory was confirmed by the initial fluctuations visible in the first 25ns of the RoG simulation analysis. Conformational change is due to the sudden introduction of inhibitors to binding site regions of the protein. The movement of the inhibitor in trying to find a suitable conformation in the binding site, introduces a range of plasticity to the static crystal structure resulting in expansion and delay in stability. DB05291,

DB07453 and DB08607 indicated a similar conformational range over the 200ns trajectory while DB04007 indicated a lower range of fluctuations over the simulation thereby reiterating the data obtained in RMSD and RMSF, which indicated a higher level of change and prolonged achievement of stability.

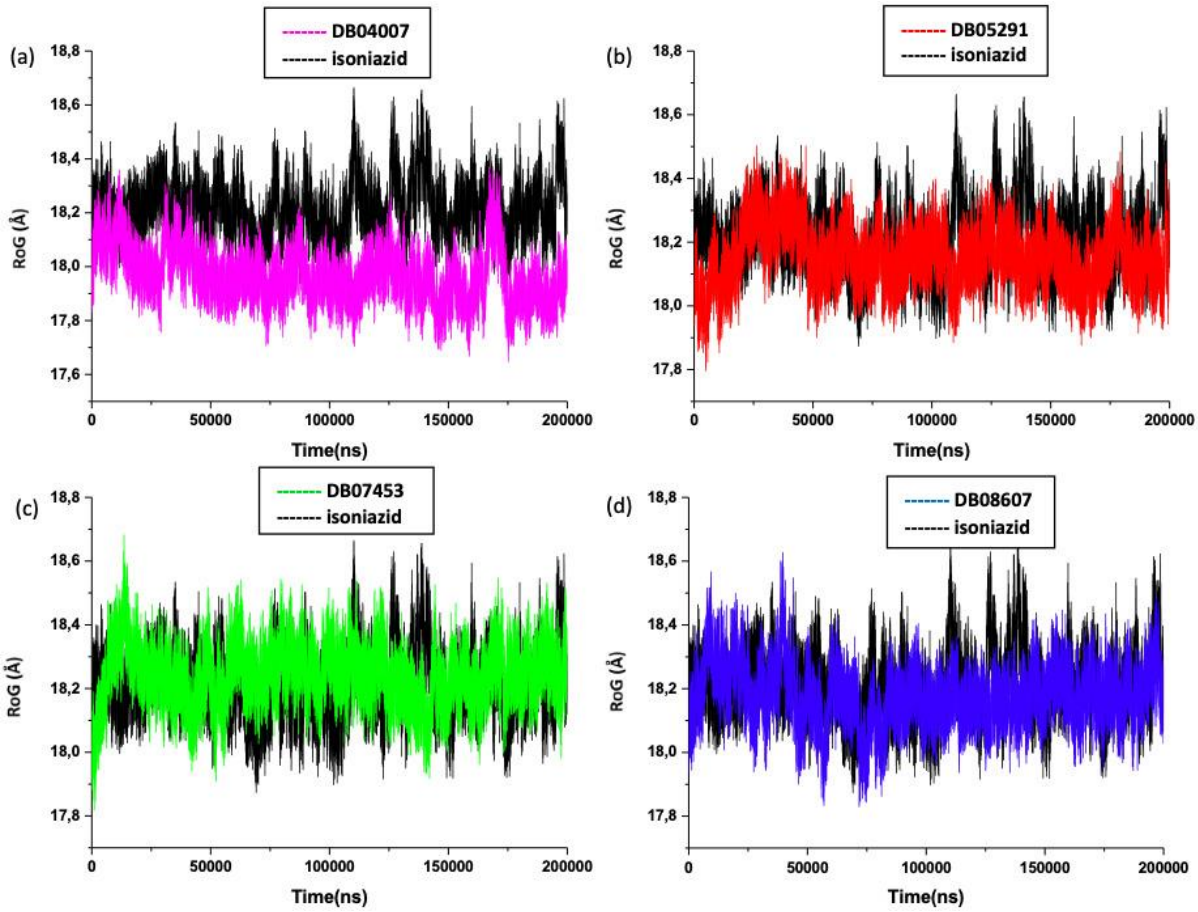


Figure 6: Graphs showing the RoG of InhA in complex with isoniazid against (a) DB04007 (b) DB05291 (c) DB07453, (d) DB08607-inhA complexes.

Binding free energies

Binding free energies indicate the strength of the forces between the inhibitor and corresponding amino acid residues. The efficiency of the inhibitor is dependent on the ability of the drug to bind into the binding site effectively [41].

Binding energy contributions to each individual inhibitor- InhA complex was calculated using the MMGBSA method from the 200 ns trajectory and tabulated in Table 2. All predicted drugs

show preferred total free binding energies as compared to isoniazid. Total energies of -26.6925 kcal/mol, -26.9317 kcal/mol and -39.5312 kcal/mol were observed for DB05291, DB07453 and DB08607, respectively, as compared with the -7.8633 kcal/mol from isoniazid. DB04007 denoted a highly stable system with a total binding energy of -41.1052 kcal/mol indicating favourable intermolecular forces between the drug target and predicted inhibitor.

Decomposition of energies was calculated to better understand the energetics of the system. This revealed most of the contribution stemmed from the Van de Waals and electrostatic interactions. For isoniazid, DB07453 and DB08607, the VdW contributions greatly outweigh the electrostatic interactions due to more hydrophobic arene rings being present in the binding pocket structure [42]. For DB04007, the electrostatic interactions form the base of the total binding free energy were due to the interactions occurring more from a longer heteroatom backbone chain [43]. This indicates a stronger set of intermolecular forces for DB07453, DB08607 and DB04007. This coincides with the data from RMSD and RMSF as DB04007 exhibited a highly stable complex coupled with strong interactions thereby exhibiting strong inhibitory potential.

Table 2: Molecular docking scores and Binding energies and contributions of DB04007, DB05291, DB07453 and DB08407 against isoniazid (DB00951) in complex with inhA (kcal/mol).

Drugs	**ΔG_{vdw}**	**ΔG_{eel}**	**ΔG_{egb}**	**ΔG_{esurf}**	**ΔG_{tot}**	**Docking score**
Isoniazid	-14.75 ± 2.32	-3.64 ± 5.87	12.77 ±5.20	-2.24 ± 0.27	-7.86 ± 2.10	-5.7
DB05291	-30.04 ± 3.26	-25.43 ± 6.94	32.72 ± 5.94	-3.94 ± 0.29	-26.69 ± 3.56	-6.4
DB07453	-36.00 ± 2.42	-8.03 ± 3.87	21.74 ± 3.02	-4.64 ± 0.21	-26.93 ± 2.06	-8.6
DB08607	-50.73 ± 3.38	-13.58 ± 6.17	30.84 ± 4.17	-6.06 ± 0.29	-39.53 ± 3.81	-10.2
DB04007	-41.67 ± 2.73	-71.76 ± 8.64	77.16 ± 8.20	-4.83 ± 0.22	-41.11 ± 3.13	-7.3

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{egb} (polar solvation energies) and ΔG_{esurf} (non-polar solation energies) to ΔG_{tot} (total)

Per residue energy decomposition

The binding energies were further analysed to show the contributions of individual active site residues to the total binding free energies in order to determine the structural impact on the interaction ability. Tables 3 to 7 show the per energy residue decomposition for all systems with the biggest contributors being Van de Waals and electrostatic interactions. The compounds showing the highest total binding free energies were DB04007 and DB08607 and the compounds showing the lowest were DB07453 and DB05291. For isoniazid, Ala126 (-0.48), Ala129 (-0.54) and Val173 (-0.65) contributed significantly to the total binding energy (VdW). DB04007 showed the major components as Gly12 (-4.96), **Ser92** (-9.28) and Ile 93 (-2.11) with the majority coming from electrostatic forces. DB05291 showed the highest interactions coming from **Gly12** (-2.33), Ile93 (-3.50) and Gly94 (-8.67) from electrostatic interactions. DB07453 had Ile19 (-1.34), **Met145** (-0.80) and Ile 192 (-1.08) showing significant contributions from VdW and DB08607 showed that **Ile14** (-1.27), **Phe39** (-3.34) and **Ile93** (-3.30) displayed the majority of the interactions from VdW.

Ser92 (DB04007), Gly12 (DB05291), Met145 (DB07453) and Ile14-Phe39-Ile93 (DB08607) all coincided with the data obtained in RMSF. The residues contributing vastly to the total binding energy of each system were also visible in the peaks created by structural change of the protein after binding. Each inhibitor displayed 1 catalytic site residue contributing significantly to the total binding energy. This is advantageous in better understanding a complexed system directly related to conformational change and per residue decomposition. Isoniazid, DB07453 and DB08607 contain a higher number of hydrophobic rings while DB04007 and DB05291 contain more hydrophilic chains, thereby constituting the contributions to binding. Both DB04007 and DB08607 displayed the highest energies with DB04007 denoting the highest contributions from electrostatic interactions and DB08607 showing significant contributions from VdW forces.

Table 3: Per residue energy (kcal/mol) decomposition of isoniazid-inhA complex.

Residue	ΔG_{vdw}	ΔG_{elec}	ΔG_{pol}	ΔG_{non-p}	ΔG_{total}
ALA 126	-0.48 ± 0.22	-0.29 ± 0.71	0.32 ± 0.37	-0.050 ± 0.032	-0.51 ± 0.46
ALA 129	-0.54 ± 0.22	0.016 ± 0.50	-0.011 ± 0.26	-0.013 ± 0.013	-0.54 ± 0.43
ALA 174	-0.037 ± 0.021	0.020 ± 0.058	0.0060 ± 0.054	0.00 ± 0.00	-0.011 ± 0.022
ARG 171	-0.011 ± 0.005	0.079 ± 0.26	-0.057 ± 0.26	0.00 ± 0.00	0.011 ± 0.0060
ASN 170	-0.024 ± 0.106	0.0060 ± 0.150	-0.00 ± 0.17	-0.00 ±- 0.0020	-0.018 ± 0.090
ILE 142	-0.041 ± 0.057	-0.012 ± 0.024	0.012 ± 0.022	-0.00 ± 0.00	-0.041 ± 0.060
PHE 172	-0.084 ± 0.040	-0.022 ± 0.17	0.080 ± 0.16	-0.0030 ± 0.0080	-0.029 ± 0.041
SER 184	-0.016 ±0.018	-0.011 ± 0.13	0.011 ± 0.068	-0.0010 ± 0.0050	-0.016 ± 0.083
TYR 125	-0.059 ± 0.027	0.013 ± 0.19	0.012 ± 0.17	-0.00 ± 0.0040	-0.035 ± 0.047
VAL 169	-0.046 ± 0.084	0.016 ± 0.13	-0.0050 ± 0.078	-0.0010 ± 0.0050	-0.035 ± 0.019
VAL 173	-0.65 ± 0.27	0.034 ± 0.26	0.033 ± 0.20	-0.069 ± 0.030	-0.65 ± 0.29

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total)

Table 4: Per residue energy (kcal/mol) decomposition of DB04007-inhA complex.

Residue	ΔG_{vdw}	ΔG_{elec}	ΔG_{pol}	ΔG_{non-p}	ΔG_{total}
GLY 12	-1.013 ± 0.557	-4.960 ± 1.120	3.970 ± 0.464	-0.124 ± 0.025	-2.127 ± 0.821
GLY 94	-0.867 ± 0.360	-1.754 ± 1.048	2.453 ± 1.051	-0.222 ± 0.053	-0.391 ± 0.227
ILE 14	-1.403 ± 0.368	0.999 ± 0.174	-0.923 ± 0.173	-0.252 ± 0.033	-1.579 ± 0.390
ILE 19	-1.415 ± 0.321	1.160 ± 0.501	-1.428 ± 0.195	-0.144 ± 0.024	-1.827 ± 0.428
ILE 93	-2.114 ± 0.343	0.478 ± 0.423	-0.178 ± 0.368	-0.197 ± 0.048	-2.011 ± 0.353
ILE 120	-0.432 ± 0.354	-0.202 ± 0.182	0.234 ± 0.184	-0.080 ± 0.060	-0.480 ± 0.400
MET 145	-0.098 ± 0.023	-1.194 ± 0.355	1.311 ± 0.356	-0.005 ± 0.006	0.014 ± 0.024
PHE 39	-1.685 ± 0.315	-0.015 ± 0.191	0.055 ± 0.124	-0.250 ± 0.045	-1.896 ± 0.357
SER 92	-0.674 ± 0.600	-9.282 ± 1.128	5.785 ± 0.938	-0.123 ± 0.031	-4.295 ± 0.769
VAL 63	-0.725 ± 0.287	0.270 ± 0.134	-0.163 ± 0.132	-0.076 ± 0.028	-0.694 ± 0.299

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total)

Table 5: Per residue energy (kcal/mol) decomposition of DB05291-inhA complex.

Residue	**ΔG_{vdw}**	**ΔG_{elec}**	**ΔG_{pol}**	**ΔG_{non-p}**	**ΔG_{total}**
GLY 12	-1.0020 ± 0.34	-2.33 ± 1.25	2.13 ± 1.09	-0.13 ± 0.032	-1.34 ± 0.56
GLY 94	-0.75 ± 0.64	-8.67 ± 1.30	5.78 ± 0.66	-0.26± 0.043	-3.90 ± 0.83
ILE 93	-1.95 ± 0.43	-3.50 ± 0.60	1.85 ± 0.26	-0.16 ± 0.044	-3.75 ± 0.67
ILE 120	-0.45 ± 0.24	0.50 ± 0.20	-0.50 ± 0.18	-0.065 ± 0.030	-0.52 ± 0.27
PHE 39	-1.06 ± 0.31	-0.24 ± 0.33	0.50 ± 0.28	-0.19 ± 0.033	-0.99 ± 0.30
PHE 95	-1.26 ± 0.27	1.38 ± 0.41	-0.92 ± 0.27	-0.20 ± 0.043	-1.01 ± 0.23
VAL 63	-0.73 ± 0.24	-0.89 ± 0.27	0.94 ± 0.31	-0.087 ± 0.032	-0.77 ±- 0.25

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total)

Table 6: Per residue energy (kcal/mol) decomposition of DB07453-inhA complex.

Residue	**ΔG_{vdw}**	**ΔG_{elec}**	**ΔG_{pol}**	**ΔG_{non-p}**	**ΔG_{total}**
ALA 20	-0.12 ± 0.035	0.082 ± 0.037	-0.025 ± 0.040	-0.0020 ± 0.003.	-0.063 ± 0.031
ALA 189	-0.28 ± 0.22	0.20 ± 0.072	-0.21 ± 0.088	-0.029 ± 0.019	-0.32 ± 0.25
ALA 258	-0.019 ± 0.011	-0.068 ± 0.027	0.086 ± 0.034	0.00 ± 0.00	-0.0010 ± 0.0050
GLY 190	-0.18 ± 0.077	-0.44 ± 0.13	0.56 ± 0.16	-0.0020 ± 0.0040	-0.063 ± 0.062
ILE 19	-1.34 ± 0.46	-0.096 ± 0.14	0.12 ± 0.11	-0.19 ± 0.046	-1.50 ± 0.50
ILE 23	-0.046 ± 0.021	0.049 ± 0.016	-0.043 ± 0.018	0.00 ± 0.00	-0.039 ± 0.023
ILE 192	-1.084 ± 0.27	0.14 ± 0.32	0.36 ± 0.25	-0.062 ± 0.019	-0.65 ± 0.30
MET 145	-0.80 ± 0.30	0.088 ± 0.22	0.31 ± 0.24	-0.042 ± 0.025	-0.45 ± 0.34
SER 92	-1.28 ± 0.42	0.056 ± 0.59	1.30 ± 0.39	-0.16 ± 0.035	-0.080 ± 0.40
VAL 236	-0.21 ± 0.14	0.066 ± 0.050	-0.067 ± 0.039	-0.0080 ± 0.011	-0.21 ± 0.15

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total)

Table 7: Per residue energy decomposition of DB08607-inhA complex (kcal/mol).

Residue	ΔG_{vdw}	ΔG_{elec}	ΔG_{pol}	ΔG_{non-p}	ΔG_{total}
GLY 94	-1.85 ± 0.32	-0.58 ± 0.96	1.78 ± 0.62	-0.26 ± 0.053	-0.91 ± 0.69
ILE 14	-1.27 ± 0.34	0.044 ± 0.12	0.060 ± 0.13	-0.31 ± 0.064	-1.47 ± 0.36
ILE 19	-0.99 ± 0.36	0.21 ± 0.20	-0.081 ± 0.19	-0.18 ± 0.044	-1.04 ± 0.33
ILE 93	-3.30 ± 0.53	-0.72 ± 0.55	0.85 ± 0.38	-0.25 ± 0.046	-3.42 ± 0.76
ILE 192	-0.49 ± 0.33	0.019 ± 0.57	0.079 ± 0.31	-0.079 ± 0.055	-0.47 ± 0.51
LEU 61	-0.30 ± 0.64	-3.43 ± 0.74	1.68 ± 0.18	-0.040 ± 0.013	-2.10 ± 0.47
PHE 39	-2.39 ± 0.41	-0.027 ± 0.17	0.47 ± 0.18	-0.33 ± 0.045	-2.28 ± 0.40
PRO 191	-0.057± 0.072	-0.15 ± 0.22	0.14 ± 0.14	-0.0010 ± 0.0040	-0.066 ± 0.17
SER 92	-0.66 ± 0.34	-0.52 ± 0.44	1.28 ± 0.39	-0.086 ± 0.038	0.0090 ± 0.41
THR 194	-0.81 ± 0.37	0.15 ± 0.55	0.15 ± 0.39	-0.24 ± 0.086	-0.74 ± 0.37
VAL 63	-0.80 ± 0.36	-1.45 ± 0.34	0.84 ± 0.13	-0.054 ± 0.016	-1.46 ± 0.33

Protein-ligand interactions

Figure 8 shows the active site intermolecular ligand-residue interactions for each drug. The contribution of Electrostatic and VdW contributions from individual amino acids coincides with the data from the per residue energy decomposition. Electrostatic and VdW forces are present throughout all the structures while hydrogen bonding occurs only in DB04007 and DB08607. Hydrogen bonding provides great insight into the extent of bonding as the presence of hydrogen bonding indicates an extensively bound system [44]. It is an integral part of the protein-ligand interaction as the number of hydrogen bonds coincides with how strongly the ligand binds to the active site via the amino acid residues [45]. The presence of hydrogen bonds greatly influences the stability of the complexed system. Further analysis of these bonds was carried out to determine the contribution (other than VdW and electrostatic interactions) to the binding of the ligand. Table 8 contains the true hydrogen bonds of the protein-ligand interaction network. Hydrogen bonds were only present in DB04007 and DB08607 dictating that DB04007 and DB08607 formed the most stable complexes with the drug target. Residue 92 present as a catalytic site residue in DB04007, showed interaction in the formation of two hydrogen bonds. This coincides with the peak present in the RMSF of DB04007. The

fluctuations present in the following residues could be due to the effect of distal amino acids o9n the protein. Bonding greatly affects the stability and energetics of a complex. The protein ligand interactions confirmed that the most suitable complex is indicated by DB04007 due to its overall stability, extensive bonding network and energetic contributions.

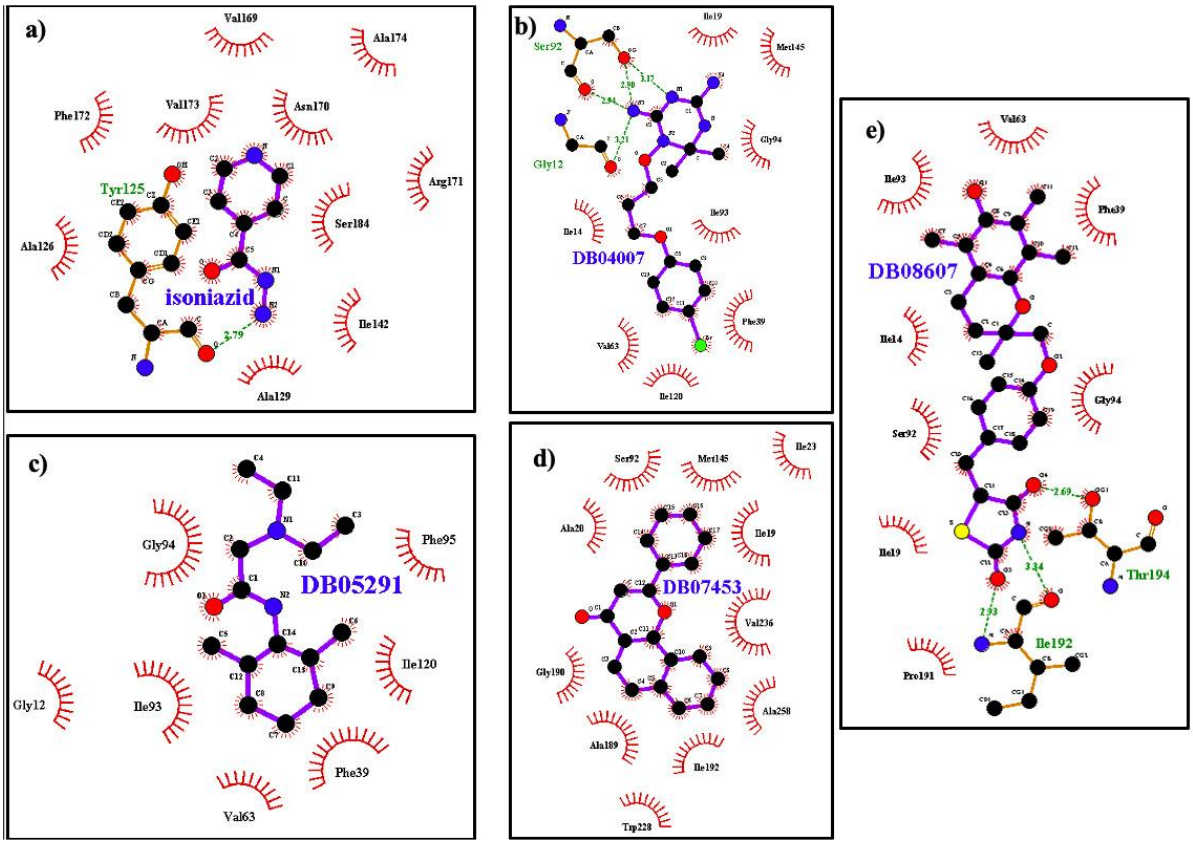


Figure 7: Intermolecular interactions of a) isoniazid-inhA b) DB04007-inhA c) DB05291-inhA d) DB07453-inhA and e) DB08607-inhA complexes.

Table 8: Table showing the H bonding data for the DB04007 and DB08607-inhA complexes.

	Acceptor	Donor H	Donor	Frame	Fraction (%)	Distance (Å)	Angle (˚)
DB04007	SER_92@OG	DB04007_268@H1	DB04007_268@N3	28083	14.04	2.8555	162.2711
	SER_92@O	DB04007_268@H1	DB04007_268@N3	8711	4.36	2.7999	153.2649
	GLY_12@O	DB04007_268@H	DB04007_268@N3	18652	9.35	2.8446	145.1558
DB08607	DB08607_268@O4	THR_194@HG1	THR_194@OG1	8032	4.02	2.7010	157.1268
	ILE_192@O	DB08607_268@H1	DB08607_268@N	2601	1.30	2.8635	146.3165
	DB08607_268@O3	ILE_192@H	ILE_192@N	93	0.5	2.9017	162.6294

Principal component analysis

Principal component analysis (PCA) is a method employed to simplify complex motions and flexibility within proteins to extract biological movements relevant from atomic motion [46]. This method groups molecular conformational similarities together and determines the variability measured by directional eigenvalues (PC1 and PC2) [47]. Figure 9 shows PCs of simulated complexes of assessed drugs versus the isoniazid complex plotted to compare patterns of flexibility. Evidently, all complexes display unique conformational behaviour. DB05291 and DB07453 show correlated motion and distinct compactness along the two components. This is indicative of minimal conformational impact induced on the protein upon ligand binding. DB08607 and DB04007 show a less correlated motion and increased early flexibility attributing to ligand induced conformational changes after binding. DB08607 just showed motion on one vector scale (PCA2) and DB04007 showed a large surface area and mobility on both vectors indicating maximum flexibility of the catalytic site. Movement on the PCA2 axis would indicate a negative loading indicating that DB04007 and DB08607 have a significant negative effect on the protein. This indicates that they would cause a reduced flexibility enhancing the effect of inhibition on the target enzyme rather than enhancing the functioning of it.

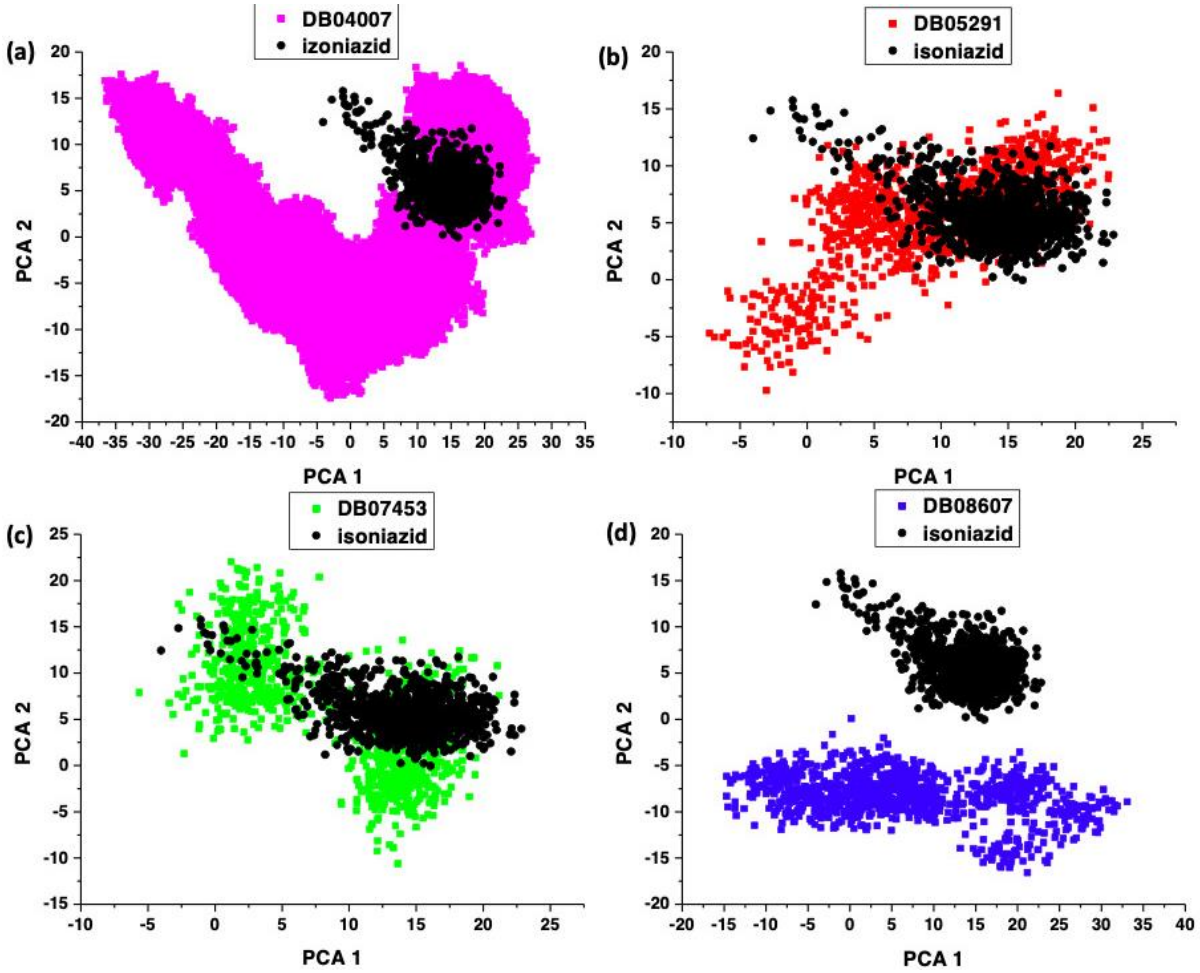


Figure 8: Graphs showing the principal component analysis of (a) isoniazid (b) DB04007 (c) DB05291 (d) DB07453 (e) DB08607 in complex with inhA.

Dynamic Cross Correlation

DCC analysis was performed to determine the positions of the Cα atoms during the conformational changes occurring upon simulation indicating the presence of correlated motions [33]. Figure 10 shows highly correlated motions represented as red to yellow and anti-correlated motions represented by the blue to black areas. All systems exhibited a similar overall correlated motion in comparison to the anti-correlated motions. All systems show the highest correlated area to fall approximately between residues 1 to 100 followed by a slightly correlated hydrophobic set of amino acids from 100 onwards. From the RMSF results, a higher fluctuation is seen in the amino acid range of 60-125 in the case of DB04007 displaying a slightly sooner peak of anti-correlated motions in that vicinity. This indicates that higher fluctuations in amino acids display a more anti-correlated movement. These results would

agree with the increased initial flexibility of the target complexed to DB04007 further confirmed by the results present in PCA.

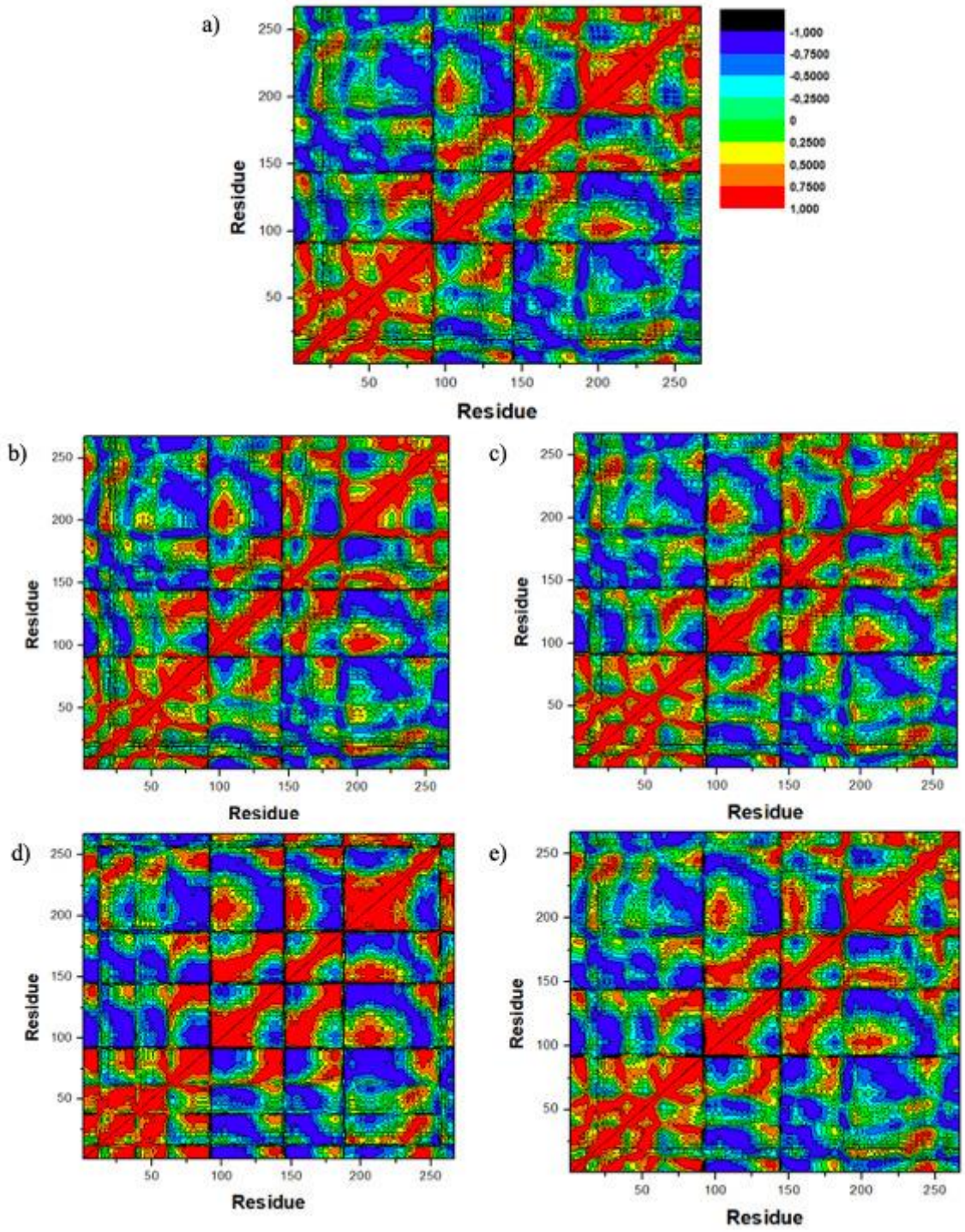


Figure 9: Dynamic Cross Correlation matrix representing the correlation of the residues of the covalently bound systems a) isoniazid b) DB04007 c) DB05291 d) DB07453 and e) DB08607 in complex with inhA.

Conclusion

Although the medicinal drug discovery and development of TB treatment regimens has been successful thus far, the drug resistance proves to be a challenging task. TB is still seen as one of the number one causes of death worldwide. Through literature, structural analyses, and molecular modelling, it can be observed that InhA highly contributes to the enzymatic and catalytic activity of *M. tuberculosis*, making it a drug target of extensive importance. Drug repurposing is an efficient and cost-effective way for modern medicine to move forward. In this course of study, four predicted drugs, as well as a control, were complexed to the drug target and evaluated for system stability. Further MD and post MD analyses were carried out to screen for inhibition capability. From the results obtained, DB04007 produced the most promising set of data. Binding affinity was seen to be -41.11 kcal/mol, while the docking score for DB04007 was -7.3 kcal/mol. DB04007 showed a stable system stabilising with an average RMSD value of 1.35 Å and increased structural rigidity visible in the RMSF plots. Protein-ligand interactions also added to the data by showing a good hydrogen bonding network through intermolecular interactions and major contributions from vdW forces due to the presence of hydrophilic chains. The PCA showed a slightly correlated motion indicating less flexibility also reiterated by the correlated motion present in the DCC plots. Overall, DB04007 showed better interaction data than any of the other predicted drug hits and the control. This gives an indication of the probability of DB04007 affecting the structural flexibility of InhA which could act to possibly inhibit the protein. Further synthetic analysis is therefore encouraged.

Conflicts of interest

The authors declare no intellectual or financial conflicts of interest.

Acknowledgements

The authors acknowledge the College of Health Sciences, School of Laboratory Medicine and Medical sciences and The Centre for High Performance Computing (www.chpc.ac.za) Cape Town, South Africa, for technical computational support, respectively.

References

1. Delogu, G., Sali, M. & Fadda, G. The biology of *mycobacterium tuberculosis* infection. *Mediterr. J. Hematol. Infect. Dis.* **5**, (2013).

2. Bloom, B. R. & Murray, C. J. L. Tuberculosis: commentary on a reemergent killer. *Science (80-.).* **257**, 1055–1064 (1992).

3. Gadkowski, L. B. & Stout, J. E. Cavitary pulmonary disease. *Clin. Microbiol. Rev.* **21**, 305–333 (2008).

4. Floyd, K., Glaziou, P., Zumla, A. & Raviglione, M. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. *Lancet Respir. Med.* **6**, 299–314 (2018).

5. Chakaya, J. *et al.* Global Tuberculosis Report 2020–Reflections on the Global TB burden, treatment and prevention efforts. *Int. J. Infect. Dis.* (2021).

6. Sharma, D., Sharma, J., Deo, N. & Bisht, D. Prevalence and risk factors of tuberculosis in developing countries through health care workers. *Microb. Pathog.* **124**, 279–283 (2018).

7. Trinh, Q. M. *et al.* Tuberculosis and HIV co-infection in Vietnam. *Int. J. Infect. Dis.* **46**, 56–60 (2016).

8. da Silva Escada, R. O. *et al.* Mortality in patients with HIV-1 and tuberculosis co-infection in Rio de Janeiro, Brazil-associated factors and causes of death. *BMC Infect. Dis.* **17**, 1–10 (2017).

9. Oni, T. *et al.* Patterns of HIV, TB, and non-communicable disease multi-morbidity in peri-urban South Africa-a cross sectional study. *BMC Infect. Dis.* **15**, 1–8 (2015).

10. Brust, J. C. M. *et al.* Improved survival and cure rates with concurrent treatment for multidrug-resistant tuberculosis–human immunodeficiency virus coinfection in South Africa. *Clin. Infect. Dis.* **66**, 1246–1253 (2018).

11. Chideya, S. *et al.* Isoniazid, rifampin, ethambutol, and pyrazinamide pharmacokinetics and treatment outcomes among a predominantly HIV-infected cohort of adults with tuberculosis from Botswana. *Clin. Infect. Dis.* **48**, 1685–1694 (2009).

12. Pedelacq, J., Nguyen, M. C., Terwilliger, T. C. & Mourey, L. A Comprehensive Review on *Mycobacterium tuberculosis* Targets and Drug Development from a Structural Perspective. *Struct. Biol. Drug Discov. Methods, Tech. Pract.* 545–566 (2020).

13. Santos, L. C. the molecular basis of resistance in Mycobaterium tuberculosis. *Open J. Med. Microbiol.* **2**, 24–36 (2012).

14. Shah, N. S. *et al.* Worldwide emergence of extensively drug-resistant tuberculosis. *Emerg. Infect. Dis.* **13**, 380 (2007).

15. Shi, R., Itagaki, N. & Sugawara, I. Overview of anti-tuberculosis (TB) drugs and their resistance mechanisms. *Mini Rev. Med. Chem.* **7**, 1177–1185 (2007).

16. Saifullah, A. *et al.* Evaluation of risk factors associated with the development of MDR- and XDR-TB in a tertiary care hospital: a retrospective cohort study. *PeerJ* **9**, e10826 (2021).

17. Bonnet, M. *et al.* Treatment of tuberculosis in a region with high drug resistance: outcomes, drug resistance amplification and re-infection. *PLoS One* **6**, e23081 (2011).

18. Organization, W. H. *The global MDR-TB & XDR-TB response plan 2007-2008*. (2007).

19. Marrakchi, H., Lanéelle, G. & Quémard, A. InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II. *Microbiology* **146**, 289–296 (2000).

20. Passi, A., Rajput, N. K., Wild, D. J. & Bhardwaj, A. RepTB: a gene ontology based drug repurposing approach for tuberculosis. *J. Cheminform.* **10**, 1–12 (2018).

21. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

22. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

23. Gordon, J. C. *et al.* H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, W368–W371 (2005).

24. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

25. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 1–17 (2012).

26. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

27. Guan, B., Zhang, C. & Ning, J. Genetic algorithm with a crossover elitist preservation mechanism for protein–ligand docking. *Amb Express* **7**, 1–11 (2017).

28. Perilla, J. R. *et al.* Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* **31**, 64–74 (2015).

29. Tian, C. *et al.* ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552 (2019).

30. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210 (2013).

31. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).

32. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **8**, 127–134 (1995).

33. Gschwend, D. A., Good, A. C. & Kuntz, I. D. Molecular docking towards drug discovery. *J. Mol. Recognit. An Interdiscip. J.* **9**, 175–186 (1996).

34. Duff, N. & Peters, B. Polymorph specific RMSD local order parameters for molecular crystals and nuclei: α-, β-, and γ-glycine. *J. Chem. Phys.* **135**, 134101 (2011).

35. Pontius, J., Richelle, J. & Wodak, S. J. <Pontius et al 1996 8950272.pdf>. 121–136 (1996).

36. Luckner, S. R., Liu, N., Am Ende, C. W., Tonge, P. J. & Kisker, C. A slow, tight binding inhibitor of InhA, the enoyl-acyl carrier protein reductase from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **285**, 14330–14337 (2010).

37. Kokkinidis, M., Glykos, N. M. & Fadouloglou, V. E. Protein flexibility and enzymatic catalysis. *Adv. Protein Chem. Struct. Biol.* **87**, 181–218 (2012).

38. Kuzmanic, A. & Zagrovic, B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys. J.* **98**, 861–871 (2010).

39. Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science (80-.).* **301**, 1196–1202 (2003).

40. Abseher, R. & Nilges, M. Are there non-trivial dynamic cross-correlations in proteins? *J. Mol. Biol.* **279**, 911–920 (1998).

41. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).

42. Meyer, E. A., Castellano, R. K. & Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chemie Int. Ed.* **42**, 1210–1250 (2003).

43. Micklefield, J. Backbone modification of nucleic acids: synthesis, structure and therapeutic applications. *Curr. Med. Chem.* **8**, 1157–1179 (2001).

44. Perrin, C. L. & Nielson, J. B. "Strong" hydrogen bonds in chemistry and biology. *Annu. Rev. Phys. Chem.* **48**, 511–544 (1997).

45. Patil, R. *et al.* Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PLoS One* **5**, e12029 (2010).

46. David, C. C. & Jacobs, D. J. Principal component analysis: a method for determining the essential dynamics of proteins. in *Protein dynamics* 193–226 (Springer, 2014).

47. Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L. & Maggio, R. M. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends Food Sci. Technol.* **72**, 83–90 (2018).

CHAPTER 5

Evaluating the Impact of C171Q Mutation on the Potency of Thiolactomycin to *M. tuberculosis* KasA Binding Pocket: Insights from Molecular Dynamics Simulations and Tailored-Pharmacophore Studies

Kimona Kisten [1], Kgothatso E. Machaba [1,2], Hezekiel M. Kumalo [1], Phiwayinkosi V. Dludla [3] and Ndumiso N. Mhlongo [1,*]

[1] School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban 4001, South Africa.

[2] Department of Biochemistry and Microbiology, University of Zululand, Private Bag X1001, KwaDlangezwa 3886, South Africa.

[3] Biomedical Research and Innovation Platform, South African Medical Research Council, Tygerberg 7505, South Africa

* Correspondence: MhlongoN4@ukzn.ac.za;

Tel: +27 (0) 31 260 2428, Fax: +27 031 260 7792.

Abstract

Molecular Dynamics (MD) simulations and Virtual Screening (VS) methods have proven to be useful tools in investigating protein mutations and identification of potential inhibitors. In the present work, we aimed at investigating the impact of C171Q mutation on the binding of Thiolactomycin (TLM) to *M. tuberculosis* KasA active site, and to further discover potential inhibitors that could possibly bind to KasA through the strategic application of MD and VS methods. Herein, we comparatively analysed the MD of TLM-KasA (wild-type and C171Q mutant) systems to understand the impact of C171Q mutation on the binding of TLM to KasA. A tailored-pharmacophore approach was applied to identify potential drug-like hits against both the wild-type and C171Q mutant KasA. The results of calculated binding free energies (ΔG_{tot}) of TLM to wild-type KasA are -33.12 kcal mol^{-1}, and -32.42 kcal mol^{-1} in the case of the C171Q mutant. The hit 44207286 identified through the tailored pharmacophore approach, displayed a higher ΔG_{tot} of -35.75 kcal mol^{-1} (wild-type) and -36.34 kcal mol^{-1} (C171Q mutation) compared to TLM. This study not only provides insight into the impact of the C171Q mutation on TLM binding to KasA but also identified a new potential drug-like compound against KasA. Molecular docking was performed using Autodock Vina version 4.2.6 supplied by MGL tools while molecular dynamic simulations were conducted using Amber18 GPU with the FF14SB Amber force field. The tailored pharmacophore was generated using Ligand scout and the RINanalyzer plugin was used to generate the residue interaction network.

Keywords: *Mtb*-KasA; C171Q mutation; thiolactomycin; MD simulation; post-MD analysis; tailored-pharmacophore

Introduction

Computational methods, have proven to be useful tools in the identification of potential inhibitors [1]. Virtual screening is a beneficial tool in the discovery of small molecules, posing as inhibitors of drug targets of interest. This technique is carried out using either a ligand- or receptor-based approach [2]. The ligand-based approach screens ligands based on beneficial structural properties of the small molecule, while the target-based approach screens ligands based on the structural property of the receptor [3,4]. While both approaches offer their respective benefits, a target-based approach is more advantaged as it relies on a possible complex rather than an isolated ligand. Machaba et al.[5] recently carried out an approach where ligand pharmacophores were tailored to counteract the effect of drug resistance-conferring mutations. This is a promising approach to identify novel inhibitors of drug targets that have developed drug resistance due to mutations, particularly in the treatment of Tuberculosis (TB). Tailored pharmacophore is a method whereby small molecules are built up from a base to acquire the most efficient set of protein-ligand interactions with the amino acids in the active site of the protein thereby leading to an increased binding energy [6]. The structure is dependent on specific functional groups that provide the most suitable electrostatic and hydrophobic interactions in a certain conformation within the binding site [5].

TB is caused by mycobacteria known as *Mycobacterium tuberculosis* (*Mtb*) [7], a 10th largest cause of mortality worldwide and the top cause of mortality due to bacterial infection. The year 2020 saw approximately 10.4 million TB infections and 1.66 million deaths worldwide, attributing an 18% overall increase in infections and deaths [8]. *Mtb* manifests in 3 different stages: exposure, latent and active tuberculosis [9]. Once active, the bacteria attack the lungs and can spread to various parts of the body, increasing the chance of co-infections such as bronchiectasis and pneumonia[9]. TB co-infection with Covid-19 virus have adversely contributed to TB infection rate, and complicated TB treatment protocols[10]. Previously, TB infections were monitored in proportion to viruses such as HIV to determine the extent of contribution from comorbidity factors to severity of the disease. Current studies indicate that coinfection with HIV and Covid-19 enhances the virulence of TB in the host [11]. The exacerbated burden of TB demands restructuring of current treatment protocols and development of multitarget inhibitors to potentially neutralize the impact of co-infection.

Original treatment of TB involved the usage of first-line drugs namely isoniazid, ethionamide and triclosan [12], while the more resistant strains were treated with second-line drugs, capreomycin and fluoroquinolone derivatives [13]. Frequently emerging drug-resistant strains have been a long-standing hurdle in the treatment of TB [14]. These strains are classified as multiple drug-resistant (MDR), extensively drug-resistant (XDR), and totally drug-resistant (TDR) depending on the severity of the strain and their response to first- and second-line drugs [15]. The emergence of these strains unequivocally warrants for the identification of new drug targets and development of inhibitors with novel mechanism of action as arsenals against TB. Mycolic acids are the major component of the *Mtb* cell wall. Hence, disturbance of the mycolic acid pathway (Figure 1) is therefore a viable method of drug target investigation [16]. The mycolic acid biosynthesis is governed by the existence of both the mammalian fatty acid synthesis pathway (FASI) and the bacterial fatty acid synthesis pathway (FASII) [17,18]. While the acyl-CoA primers (C_{16-18}) are produced in the FASI pathway, the FASII system elongates the meromycolates to C_{50-56} [18]. The FASII system involved as a precursor in the production of mycolic acids, acts as a basis for survival of the Mycobacterium [19].

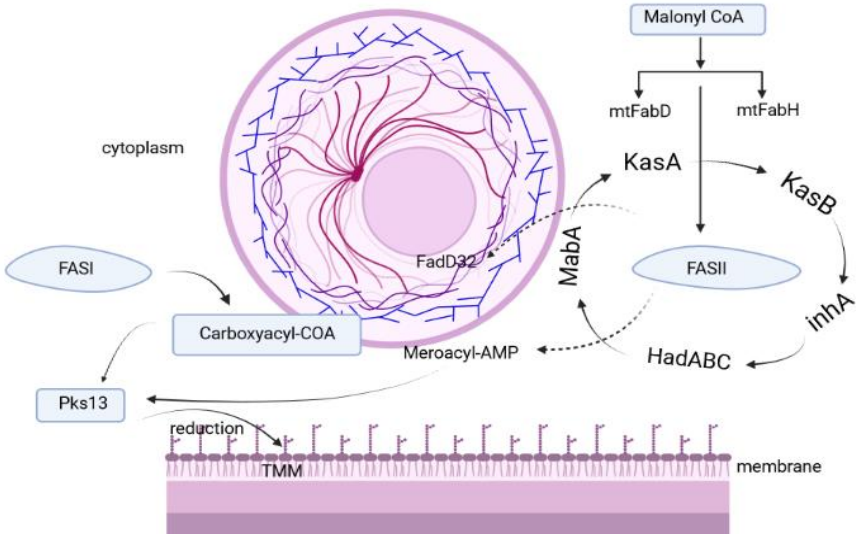


Figure 1. A mycolic acid biosynthesis pathway in *M. tuberculosis.* Five major enzymes contribute to the production of mycolic acids: inhA, MabA, HadABC, KasA and KasB. KasA and KasB exist as dimers that make up the β-ketoacyl-ACP fatty acid synthetase system. The pathway exhibited in the cytoplasm directly links the functionality of KasA to the fatty acid synthetase system involved in the meromycolate backbone. This leads to the final condensation reaction catalysed by polyketide synthase (Pks13) thereby shielding the bacteria from the immune system rendering the drug ineffective [20].

A homodimer β-ketoacyl ACP synthase (KasA) acts as a catalyst in the Claisen condensation of the malonyl-AcpM and acyl-AcpM [21,22]. Thiolactomycin (TLM) is one of the known drugs that is reported to inhibit KasA [23]**.** However, the impact of the C171Q mutation (located in the binding site) on the potency of TLM binding to KasA is not well established. Therefore, this study provides insight into TLM binding to KasA (wild-type and C171Q mutant) and to unravel molecular dynamics that occur during the course of a simulation, with an ultimate aim of identifying potent inhibitors to counteract the envisaged impact of mutation through a tailored-pharmacophore approach. The structure of KasA is altered via a mutation of Cysteine to Glutamine on position 171 of the sequence. The original unbound apo and mutant structures were generated using x ray crystallography and resolutions of 1.8 – 2.2 Å were obtained [23]. The catalytic triad of His345, His311 and Cys171, all located in the core domain where the Cys171 newly mutated to Glu171 lies in the N-terminal domain and the other two residues located in the C-terminal domain of the α-helix. Superposition did not reveal any major changes in the unbound structures prior to binding [17]. Upon binding, the side chain of the newly established Glutamine generates a hydrogen bond from the oxygen to the nitrogen atom on Phe404 resulting in a 60 ˚ Phe shift and thereby 3.3 Å positional movement out of the active site as compared to the wild-type bound structure. The newly opened conformation generated by mimicking acylation insists that the Phe404 then has the ability to form favourable interactions with the TLM ring [24]. In the process, the Leu-116, initially present in the catalytic site, is expelled with the addition of the acylated chain while a third residue, Tyr126 acts as a solvent barrier in both wild-type and mutant bound and unbound structures [25].

By applying a tailored-pharmacophore approach on this target, we aim to answer the following questions: 1) Does the mutation cause enhanced functionality or partial inhibition? 2) How does the change in amino acid fare in terms of the binding affinity to the inhibitor? 3) Can a new but similar structured drug be identified with improved binding affinities against both the wild-type and its variant? 4) How effective is the identified drug against both systems?
To the best of our knowledge, this is the first account of such comprehensive study on TLM binding to *Mtb*-KasA (wild-type and C171Q mutant): insights from molecular dynamics simulations and tailored-pharmacophore studies. In addition, we believe findings reported in this study would further enhance the understanding of TLM binding to *Mtb*-KasA (wild-type and C171Q mutant).

Computational methods

Preparation of protein and ligand structures for molecular docking

The crystal structure of *M. tuberculosis* KasA-Thiolactomycin complex (**Figure 2)** was acquired from the RSCB Protein Data Bank [PDB ID: 4C6X] [26]. One subunit of the dimer was used as a result of identical structures and heteroatoms and as well as crystallized molecules removed using Chimera. The protonation states of charged amino acids in KasA were fixed in H++ server [27]. Hydrogen atoms were deleted from the ligands and bond orders inspected and optimised using Avogadro [28].

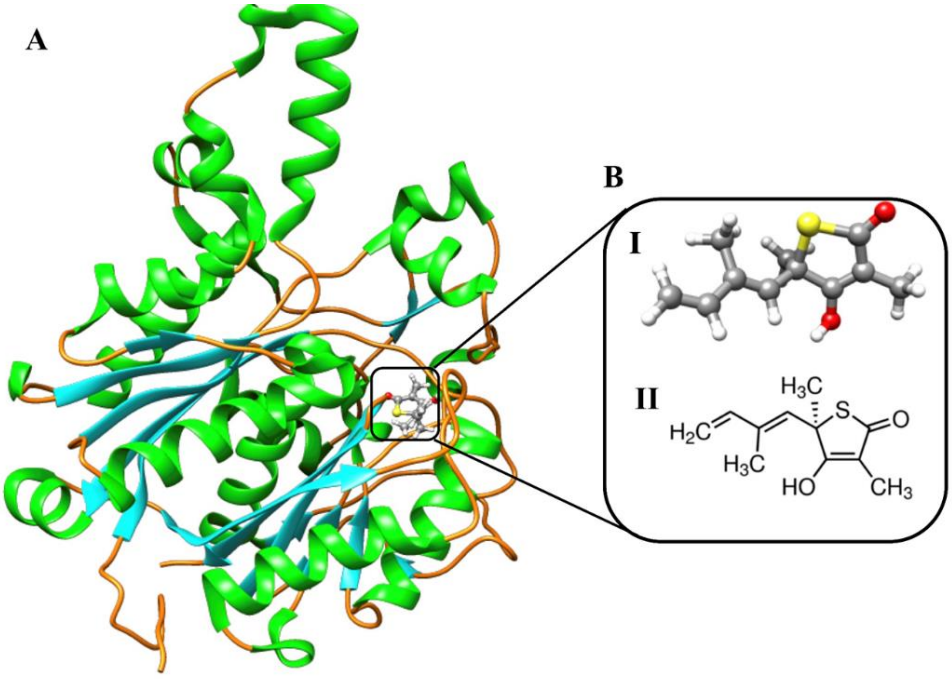


Figure 2. A) The 3D X-ray crystal structure of the KasA-TLM complex (where green = helix, light blue = ribbon, orange = loop) and B) I) 3D and II) 2D structures of TLM.

Molecular docking

Molecular docking was performed using Autodock Vina version 4.2.6 supplied by MGL tools [29]. Gasteiger charges were distributed using the Graphical manager interface provided by Autodock MGL tools. Docked conformations were obtained utilising the Lamarckian Genetic Algorithm [30]. The binding site was identified with the following grid box parameters: centre coordinates of x = -8.889 Å, y = -20.833 Å and z = 9.250 Å with grid box dimensions of x = 36 Å, y = 32 Å and z = 30 Å. The protein-ligand complexes were ranked according to their docking scores (kcal/mol) and saved for subsequent molecular dynamics simulations.

Molecular dynamics simulation

Molecular dynamic simulations were conducted using Amber18 GPU with the FF14SB Amber force field [31]. Ligand parameters were set using Avogadro [28]. The LEaP module was used to add hydrogen atoms to the protein structure and counter ions to neutralize the system. A TIP3P water box model with a 10 Å protein-water box boundary distance was applied. Energy minimization of systems was performed with SANDER, with a restraint potential of 500 kcal/mol $Å^{-2}$ for solute for 1000 steps. Long range electrostatic interactions with non-bonding cut-off distance of 12 Å were treated with the Particle mesh Ewald (PME) method [31]. Heating of the systems was performed from 0 to 300 K using the Langevin thermostat with harmonic restraints of 5 kcal/mol $Å^{-2}$ for the solute and a collision frequency of 1 ps. Systems were equilibrated without restrictions at 300 K and a constant pressure of 1 bar for 2 ns. Bonds involving electrostatic interactions were treated with the PME method and hydrogen atoms were treated with a SHAKE algorithm [32]. The 200 ns molecular dynamics simulations were performed in an isothermal-isobaric ensemble using a Berendsen barostat at a pressure of 1 bar and 2 ps pressure-coupling constant. Analytical methods namely root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration (RoG) were calculated from the generated trajectories using the CPPTRAJ module of Amber 18 [33]. Origin version 8 software [34] was used to plot graphs for MD analyses. Ligand-protein interactions were analysed using LigPlus[35], and BioRender (BioRender.com) was used to create the mechanistic pathway of *Mtb*.

Binding free energy calculations

Binding affinities of ligand-protein complexes were calculated from a 200 ns MD trajectory using the Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) method implemented in AMBER 18 [31]. The equations below are utilised in calculating total binding free energies and its components [36].

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand}) \quad (1)$$

$$\Delta G_{binding} = E_{gas} + G_{sol} - T\Delta S \quad (2)$$

$$E_{gas} = E_{int} + E_{ele} + E_{vdw} \quad (3)$$

where $\Delta G_{binding}$ represents the total binding energy of the protein-ligand complex, E_{gas} represents the sum of gas phase molecular mechanics energy, G_{sol} refers to the solvation energy and $-T\Delta S$ representing the entropy derived in relation to temperature. E_{gas} is comprised of the

addition of internal energy (E_{int}), electrostatic interactions (E_{eel}) and Van de Waals forces (E_{vdw}).

Per-residue energy decomposition

MM/PBSA was used to calculate the contribution of individual amino acid residues to the total binding energy. The ligand-residue energetics are described by the following formula [31]:

$\Delta G_{\text{inhibitor- residue}} = \Delta E_{vdW} + \Delta E_{eel} + \Delta G_{egb} + \Delta G_{esurf}$

Where the ligand residue interactions ($\Delta G_{\text{inhibitor- residue}}$) is described by the contribution of van de Waals contributions, electrostatic contributions, polar contributions and non-polar contributions.

Tailored pharmacophore and virtual screening

The total binding free energies and per residue energy decomposition were generated using the MM/PBSA approach [31] carried out on the 200 ns MD trajectories for both the wild-type and mutant complexes. Ligplot [35] was utilized to determine the hydrophobic interactions and hydrogen bond attributed to the amino acid components. The pharmacophore was then generated using Ligand scout [37] and modified by adding functional groups to generate a system complex of enhanced capability. The 3D tailored pharmacophore was determined using a structural formula and screened and fingerprinted against potential drug hits with the use of the Zinc database [38] and Pubchem [39] libraries. **Figure 3** shows the flow of the tailored pharmacophore approach carried out in the current study.

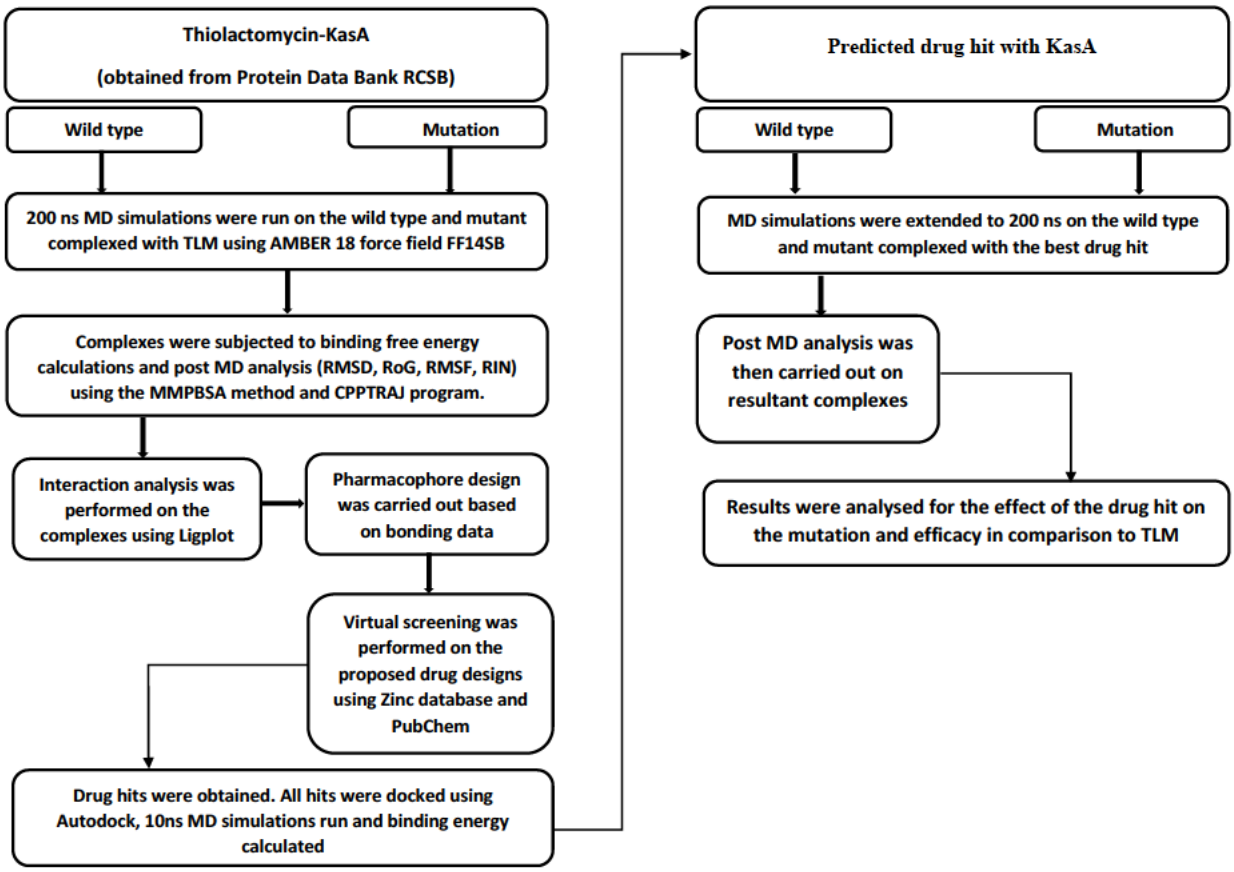


Figure 3. Flow diagram outlining the strategy adopted in the current study.

The pharmacophore structure-based skeleton containing moieties was generated using Ligandscout and screened on databases of Zinc database and PubChem multiconformational libraries [40,41]. The Zinc database, containing over 21 million compounds was screened for small molecules structurally similar to TLM via the tailored pharmacophore. The tool generated 5 decoys which were subsequently docked [42,43]. The PubChem database, containing over 69 million substances ranging from 27 million chemical structures, 449401 bioassays targeting over 7000 proteins. This tool generated 29 structures where 9 acted as substructure active compounds and the others as decoys [44,45]. Docking was performed on the protein for all small molecules. Criteria were then filtered using Lipinski's rule of five and only compounds fulfilling the pharmacophore properties were regarded as hits, compared to docking and further analysed using MD. While a diverse range of libraries exist, libraries created are based on many properties dependant on the user's requirements: target orientated, diversity orientated, chemically active orientated, natural products orientated, molecular property orientated etc [46].

The descriptors used in this study include properties such as the number of H-bonding acceptors, the number of H-bond donors, the logP, the number of chiral centres, the number of ring systems, the drug likeness and polar topological surface area in order to filter out potentials not just based solely on the molecular weight [47].

Residue interaction networks (RIN) analysis

RIN is a method of analysing residue-residue interactions present within a system [48]. Average structures were generated from the trajectories of the wild-type and mutant complexes. The 2D graphs were generated with the use of RING [49] utilizing PROBE [37] to define interactions. This is done by analysing atomic packing by means of using an approximately 0.25Å probe for each covalent and non-covalent interaction of every residue. Touching or overlap of 2 probes results in the generation of a contact dot. The RINanalyzer [49] plugin of Cytoscape [50] was used to visualise the overall residue interaction network.

Validation of docking approach

In order to validate the approach used in this study, the co-crystallized structure was superimposed with a docked *MtbKasA-TLM* complex presented in presented in **Figure 4.** This indicated a positive response to the methodic AutoDock parameters indicating an overall RMSD value lower than 2 Å.

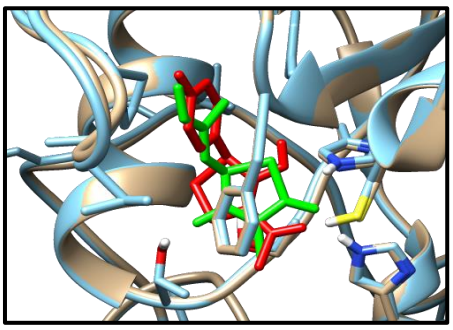


Figure 4. Superimposed structures of the co-crystallized and docked MtbKasA-TLM complexes.

Results and Discussion

C171Q mutation eliminates a TLM-C171 hydrogen bond in TLM-KasA complex

To provide insight into the effects of the mutation C171Q to TLM binding, thermodynamic calculations were carried out to gain insight into the binding free energy profiles. Utilizing the MM/PBSA approach, we performed thermodynamic calculations of the TLM-KasA (wild-type and mutant) complexes. In **Table 1**, the binding energy (ΔG_{tot})of TLM to wild-type and C171Q mutant is -33.12 kcal/mol and -32.41 kcal/mol, respectively.

Table 1. Binding energies (kcal/mol) of TLM and 44207286 to wild-type and C171Q mutant KasA.

Ligand	Protein	ΔG_{vdw}	ΔG_{eel}	ΔG_{egb}	ΔG_{esurf}	ΔG_{tot}
TLM	Wild-type	-37.10 ± 1.50	-14.39 ± 2.26	22.64 ± 1.96	-4.28 ± 0.07	-33.12 ± 1.18
	Mutant	-36.95 ± 1.89	-14.20 ± 2.58	23.03 ± 1.38	-4.28 ± 0.11	-32.41 ± 2.12
44207286	Wild-type	-38.01 ± 2.52	-27.74 ± 3.73	34.33 ± 1.96	-4.33 ± 0.08	-35.75 ± 2.12
	Mutant	-38.58 ± 1.80	-24.26 ± 2.25	30.93 ± 1.57	-4.42 ± 0.08	-36.34 ± 2.77

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{egb} (polar solvation energies) and ΔG_{esurf} (non-polar solation energies) to ΔG_{tot} (total).

It was observed that van der Waals forces contribute more towards the ΔG_{tot} of TLM to KasA. We can conclude that the introduction of a mutation resulted in a slight change in overall binding energy of TLM. Hence, our results provided here are in agreement with experimental findings that suggest that the presence of the C171Q mutation does not reveal any major changes that can be held accountable for changes in binding [23].

To understand of which residue and energy component has more impact on the total binding energy, total binding free energy for TLM was decomposed into individual amino acids contributions. The contributions of electrostatic interactions and van der Waals interactions to the relative binding free energy of TLM to wild-type and C171Q mutant was estimated. **Figure 5** shows per-residue energy decomposition analysis.

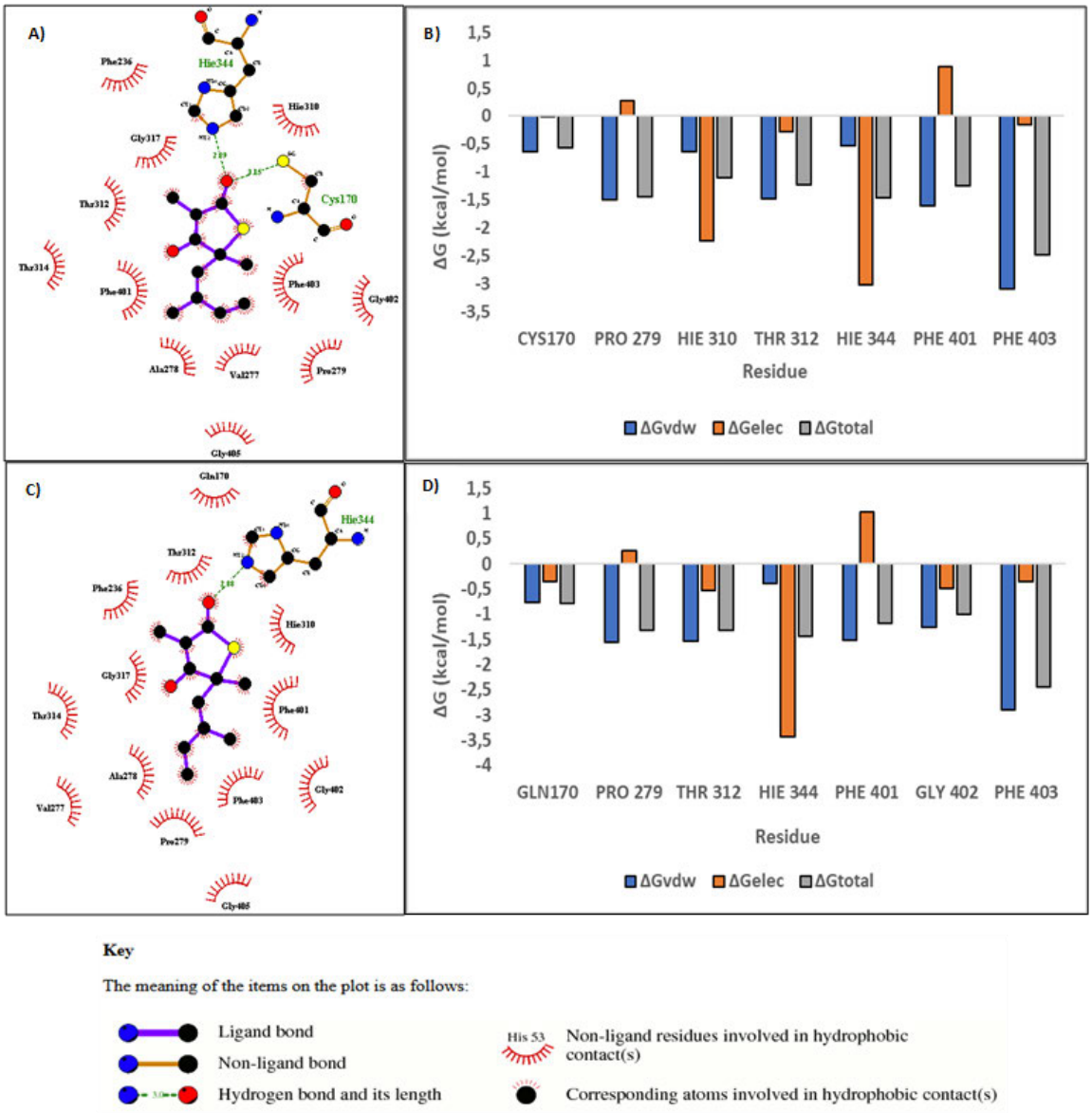


Figure 5. Interactions and per-residue energy decomposition (kcal/mol) of A) and B) wild-type and C) and D) mutant complexes with TLM.

As evident from per-residue energy contribution plots in **Figure 5**, PHE403 contributes more towards TLM binding in case of wild-type and mutant with ΔG_{tot} of -2.49 and -2.45 kcal/mol, respectively. Upon mutation, CYS171 displayed a ΔG_{tot} of -0.56 kcal/mol as compared to GLN171 with a ΔG_{tot} of -0.24 kcal/mol. In the wild-type complex, the analysis shows the presence of 2 distinct hydrogen bonds between the sulfur present on the CYS170 with the oxygen on TLM and the HIE344 nitrogen with the same TLM oxygen atom. In case of the

mutation as compared to the wild-type, HIE344 displayed a similar interaction with TLM. Hence, the presence of mutation lead to more hydrophobic interactions.

Based on the information presented in **Figures 5** and **Table 1** of both TLM complexes (wild-type and mutant), we apply a tailored-pharmacophore approach to screen for potential KasA inhibitors with improved binding affinity profiles on both systems. The tailored pharmacophore was designed based on improving the set of hydrophobic interactions of the ligand with the active site residues and reinstating the hydrogen bond that was lost due to the introduction of a mutation (**Figure 6)**.

The strategy of the model was based on two important factors:

1. Introducing a highly electronegative group containing a lone pair to act as a proton acceptor thereby creating a hydrogen bond [51].
2. Creating a structure containing more rings in order to increase the properties that enhance hydrophobicity on a molecular level [52].

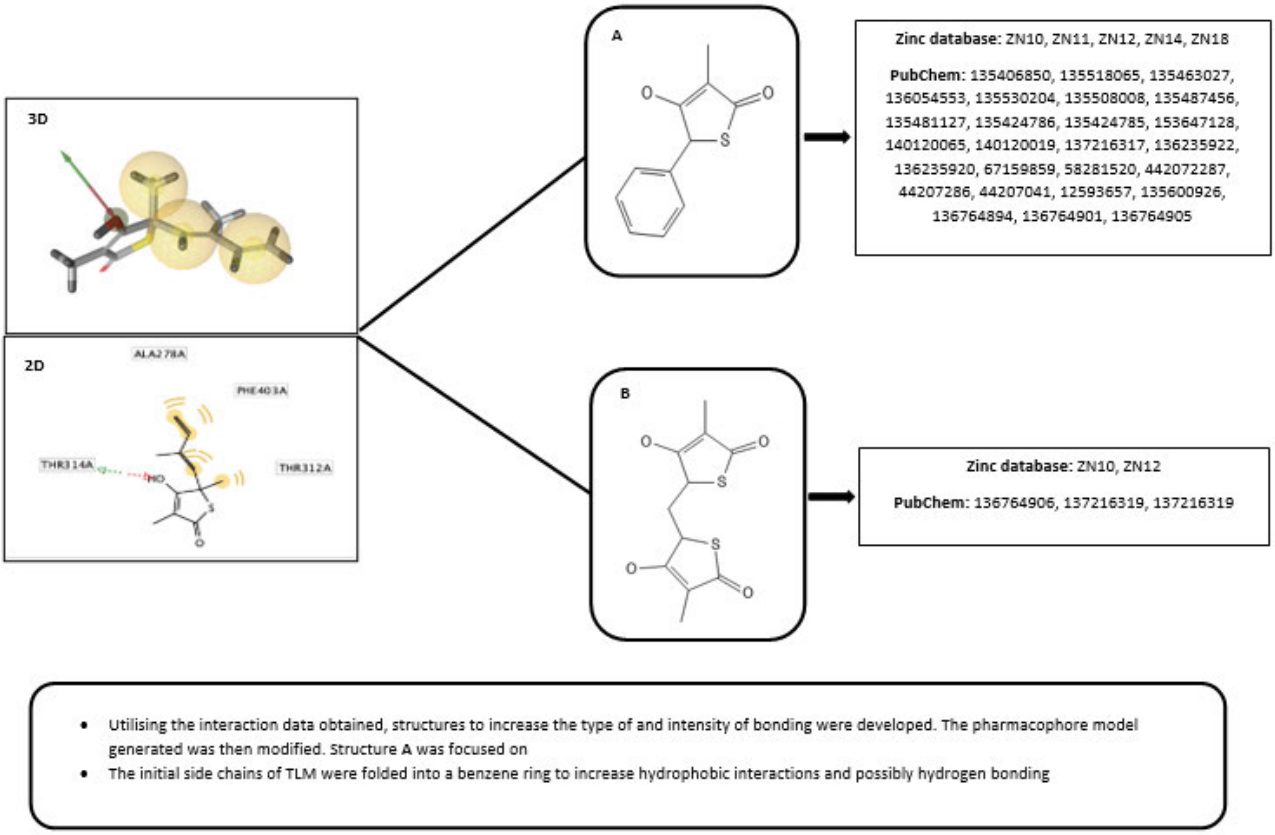


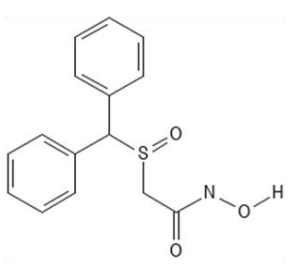
Figure 6. Tailored pharmacophore generation approach.

Two structures were selected as templates for virtual screening:

1. Proposed structure A was generated by encompassing the hydrophilic nature of the side alkyl chains and twisting them to form a ring thereby allowing the production of more hydrophobic interactions.

2. Proposed structure B was generated based on utilizing the strong hydrophobic and hydrogen bonding nature of the ring containing the multiple electronegative functional groups of sulfur and oxygen. The ring was doubled symmetrically to create a structure that had the ability to create hydrogen bonds of the same nature while minimizing the effect of lone pair repulsion.

The tailored pharmacophore models were then screened for hit compounds. Overall, 34 hits were obtained, 5 from Zinc database [38] and 29 from PubChem [39]. All 34 hits were optimized and docked into both the wild-type and mutant structures of KasA. Only hits with highest docking scores were further subjected to molecular dynamics simulations to determine their binding affinities. **Figure 6** shows binding profiles of selected hits in complex with wild-type and mutant structures of KasA.



Zinc ID: ZN12

IUPAC: 2-benzhydrylsulfinyl-*N*-hydroxyacetamide

Formula: $C_{15}H_{15}NO_3S$

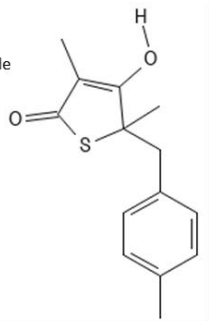
MW: 289.35 g/ mol

DS (wild): -8.5 kcal/ mol

DS (mutant): -8.6 kcal/ mol

ΔG (wild): -11.4509 kcal/ mol

ΔG (mutant): -31.4520 kcal/ mol



Pubchem ID: 13542478

IUPAC: 4-hydroxy-3,5-dimethyl-5-[(4-methylphenyl)methyl]thiophen-2-one

Formula: $C_{14}H_{16}O_2S$

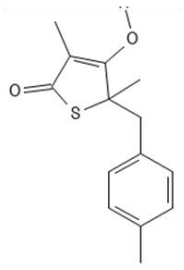
MW: 248.34 g/ mol

DS (wild): -8.6 kcal/ mol

DS (mutant): -8.7 kcal/ mol

ΔG (wild): -23.3463 kcal/ mol

ΔG (mutant): -13.6266 kcal/ mol



Pubchem ID: 136764894

IUPAC: (S)-3,5-Dimethyl-4-hydroxy-5alpha-(4-methylbenzyl)-2,5- dihydrothiophene-2-one

Formula: $C_{14}H_{16}O_2S$

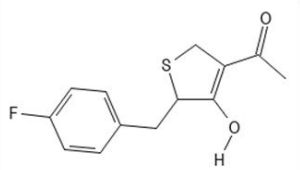
MW: 248.34 g/ mol

DS (wild): -8.7 kcal/ mol

DS (mutant): -8.6 kcal/ mol

ΔG (wild): -20.3921 kcal/ mol

ΔG (mutant): -21.1200 kcal/ mol



Pubchem ID: 44207286

IUPAC: 1-[5-(4-fluorophenyl)-4-hydroxy-2,5-dihydrothiophen-3-yl]ethanone

Formula: $C_{12}H_{11}FO_2S$

MW: 238.28g/ mol

DS (wild): -8.3 kcal/ mol

DS (mutant): -8.9 kcal/ mol

ΔG (wild): -35.7584 kcal/ mol

ΔG (mutant): -36.3433 kcal/ mol

Figure 7. Chemical properties, docking scores and binding energies (kcal/mol) of the wild-type and C171Q mutant in complex with 4 hits, ZN12, 135424785, 44207286, and 136764894.

From binding free energy calculations, compound **44207286** showed the highest binding energy to wild-type and mutant KasA, as compared to other hits. Compound **44207286** binding

was therefore further characterised to understand structural properties that favoured its binding to KasA, to compare these to TLM.

Hydrophobic properties potentiate preferential 44207286 binding to KasA active site

Compound **44207286** displayed the highest ΔG_{tot} of -35.75 and -36.34 kcal/mol in wild-type and mutant KasA, respectively, in comparison to TLM (**Table 2**). Favourable contributions were from vdW and electrostatic interactions for both wild-type and mutant KasA-**44207286** complexes. Per-residue energy decomposition was also calculated for both complexes to investigate the contributions of individual amino acids to the total binding energy. KasA-**44207286** interactions revealed the presence of mainly hydrophobic interactions (**Figure 8 A** and **B)**, contributing largely to the total binding free energy.

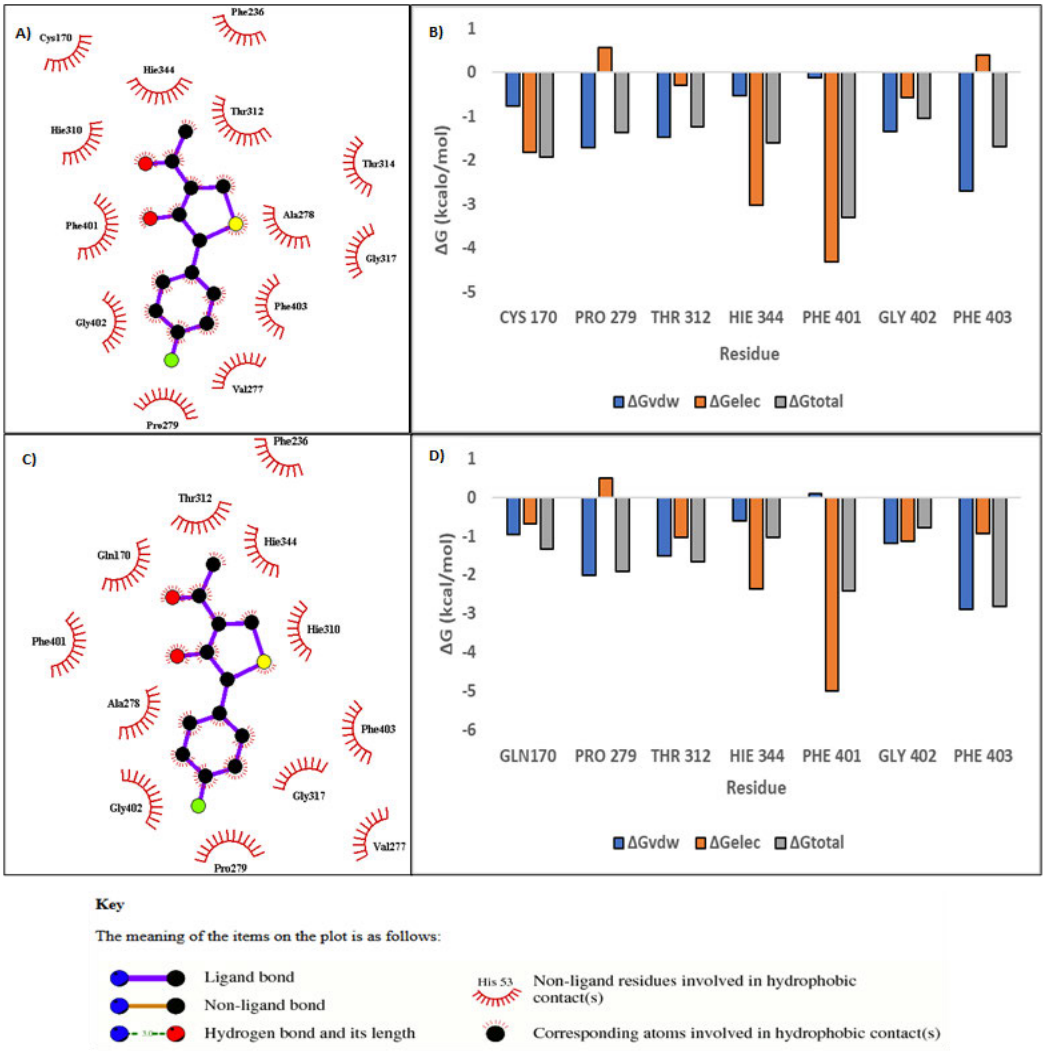


Figure 8. Interactions and per-residue energy decomposition of A) and B) wild-type and C) and D) mutant KasA-44207286 complexes.

Analyses suggest that the presence of electronegative groups in a ligand did not induce hydrogen bond formation but instead played a role in strengthening the hydrophobic interactions. From observed interactions, hydrophobic interactions contribute the majority to the total binding free energy. In KasA-**44207286** complexes, both CYS170 (-1.51 kcal/mol) and GLN170(-0.93 kcal/mol) displayed a highest ΔG_{tot} as compared to TLM complexes, where CYS170 displayed a ΔG_{tot} of -0.56 kcal/mol and -0.24 kcal/mol in GLN170. Results herein showed that the hit **44207286** exhibited an increase in ΔG_{tot} contribution as a result of a gain in hydrophobic interactions.

The aim of this study was to provide insight into TLM binding to *Mtb*-KasA (wild-type and C171Q mutant) and further apply a tailored-pharmacophore approach in search of novel potent drug hits. Hence, findings reported in this study enhance the understanding of TLM bound to *Mtb*-KasA (wild-type and mutant) and also introduce tailored-pharmacophore approach in discovering potential drug hit against *Mtb*.

Impact of compound 44207286 binding on KasA structure

To gain insight into the stability and flexibility of *Mtb*-KasA (wild-type and C171Q mutant) in complex with TLM and hit **44207286**, root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (RoG) and residue interaction network were studied.

Increase in stability and flexibility

Assessing stability and flexibility of a system allows for accurate sampling of a complex in a stable region[53]. The 200 ns trajectories of the wild-type and mutant complexes with TLM and **44207286** were analysed to determine the stability of the systems. The RMSD was calculated to determine the stability of both complexes and the results are presented in **Figure 9**.

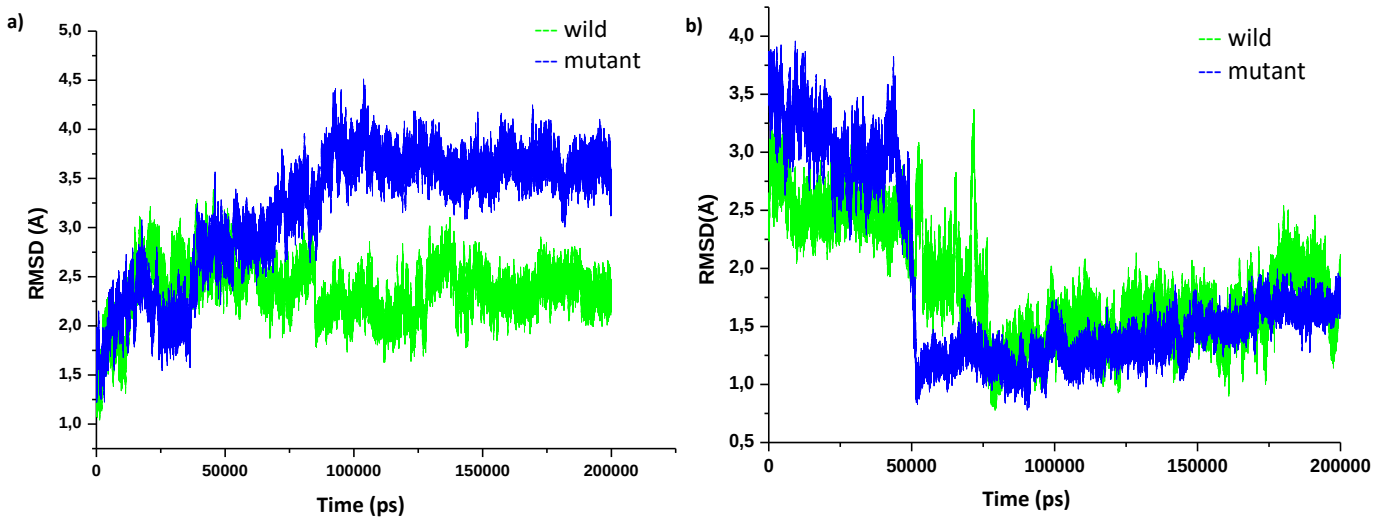


Figure 9. RMSD of the wild-type and C171Q mutant of KasA complexed with a) TLM and b) 44207286.

The estimated average RMSD values for the wild-type and mutant TLM complexes **(Figure 9a**) were 2.36 Å and 3.19 Å respectively. These values imply that the wild-type complex is more stable than the mutant complex. This indicates that the mutation potentially induced a decrease in stability of the complex [54]. The loss in stability could be due to the change in the conformation of the ligand in the presence of the mutation. A mutation of cysteine to glutamine could have introduced steric hindrance on TLM by a longer hydrocarbon chain in glutamine. With respect to wild-type and mutant 44207286 systems**,** the estimated average RMSD values were respectively 1.86 Å and 1.80 Å. These values denote similar magnitude of stability of both the complexes with a slight increase in stability of 0.06 Å in the wild-type.

During a 200 ns MD simulation, it can be observed that the TLM-KasA (mutant) shows high RMSD indicating large conformational changes compared to the wild-type. On the other hand, the 44207286 -KasA (wild-type and mutant) displays low RMSD during the 0-50 ns period. For further insight into the stability of the simulated complexes, we analysed the dynamics of the gating domain. This domain was observed to be dynamically unstable during the trajectory. Prominent structural changes in this domain were at a region of a 3_{10} helix, which gradually transformed into a loop and back into a partial helix (Figure 10 a and d). This transition may have compromised structural integrity and destabilised the overall protein structure.

A complete loss of a 3_{10} helix and constant formation of a full loop (Figure 10 b and c) highlights a degree of structural deformation and destabilization in the protein. These events explain the differences observed in RMSD of the respective complexes. Moreover, previous studies show that in the presence of a ligand, change in pH and other specific conditions, may lead to the disappearance or appearance of a 3_{10} helix due of its short length [55,56,57]. This may also account for the structural changes observed in this protein.

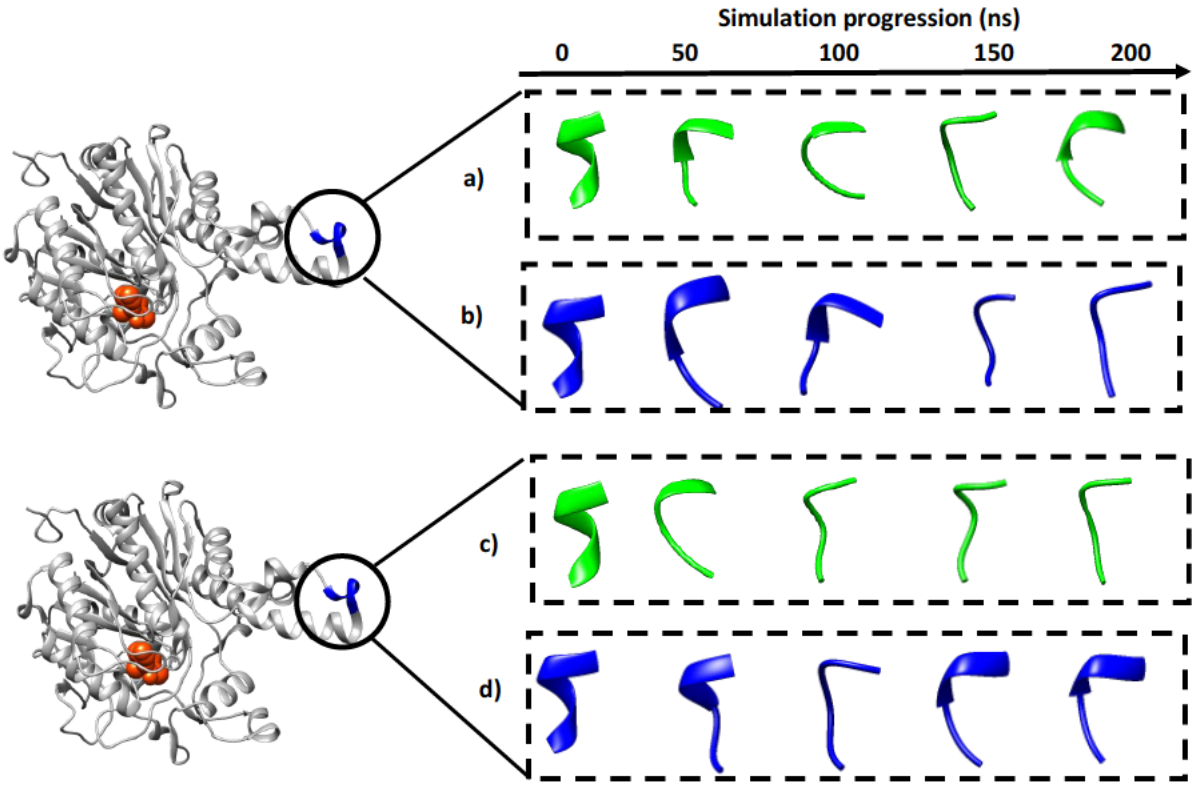


Figure 10. Conformational changes of a gating domain during a 200 ns MD trajectory. The evolution of a 3_{10} helix in a (a) TLM-KasA (wild-type); (b) TLM-KasA (mutant); (c) 44207286-KasA (wild-type); and (d) 44207286-KasA (mutant) conformations sampled at 50 ns intervals during a 200 ns MD trajectory. In a TLM-KasA (wild-type), the 3_{10} helix disappeared, with an increase in *loop length* at 50 and 100 ns and then start to decrease in *loop length* at 200 ns. However, the 3_{10} helix in case of TLM-KasA (mutant) disappeared, with an increase in *loop length* at 50 and 100 ns and then completely disappeared throughout the simulation. On the other hand, 3_{10} helix in case of 44207286 -KasA (wild-type) completely disappeared throughout the simulation after 50 ns while 3_{10} helixin case of the 44207286 -KasA (mutant).

Apart from the structural evolution of the protein discussed in Figure 10, Figures 11 and 12 indicate the observation of the trajectories from 40 to 90 ns. The drug hit 44207286 appeared to bring about significant changes in the structural evolution of the of the ligand in the active site of the wild and mutant structures in comparison to TLM. In conjunction with the interactions observed within the complexes, this could also be a result of the change in structure of the drug hit containing 2 hydrophobic rings while a straight chain paired with a ring is observed for TLM.

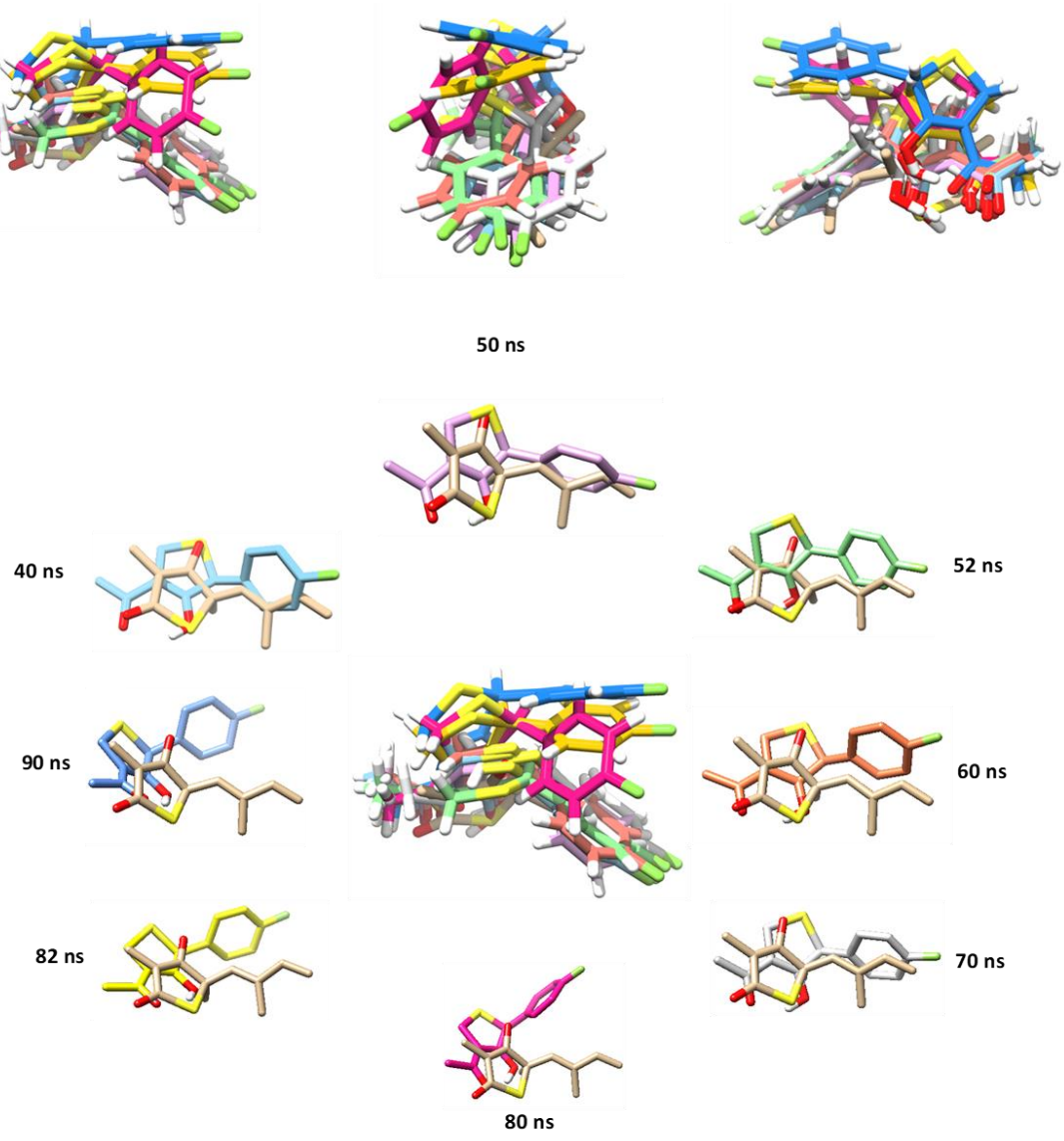


Figure 11: Conformational evolution of ligand 44207286 in wild type KasA active site. A cluster of ligand 44207286 colour-coded conformations sampled from 40 to 90 ns of a 200 ns trajectory versus a starting stricture conformation of TLM.

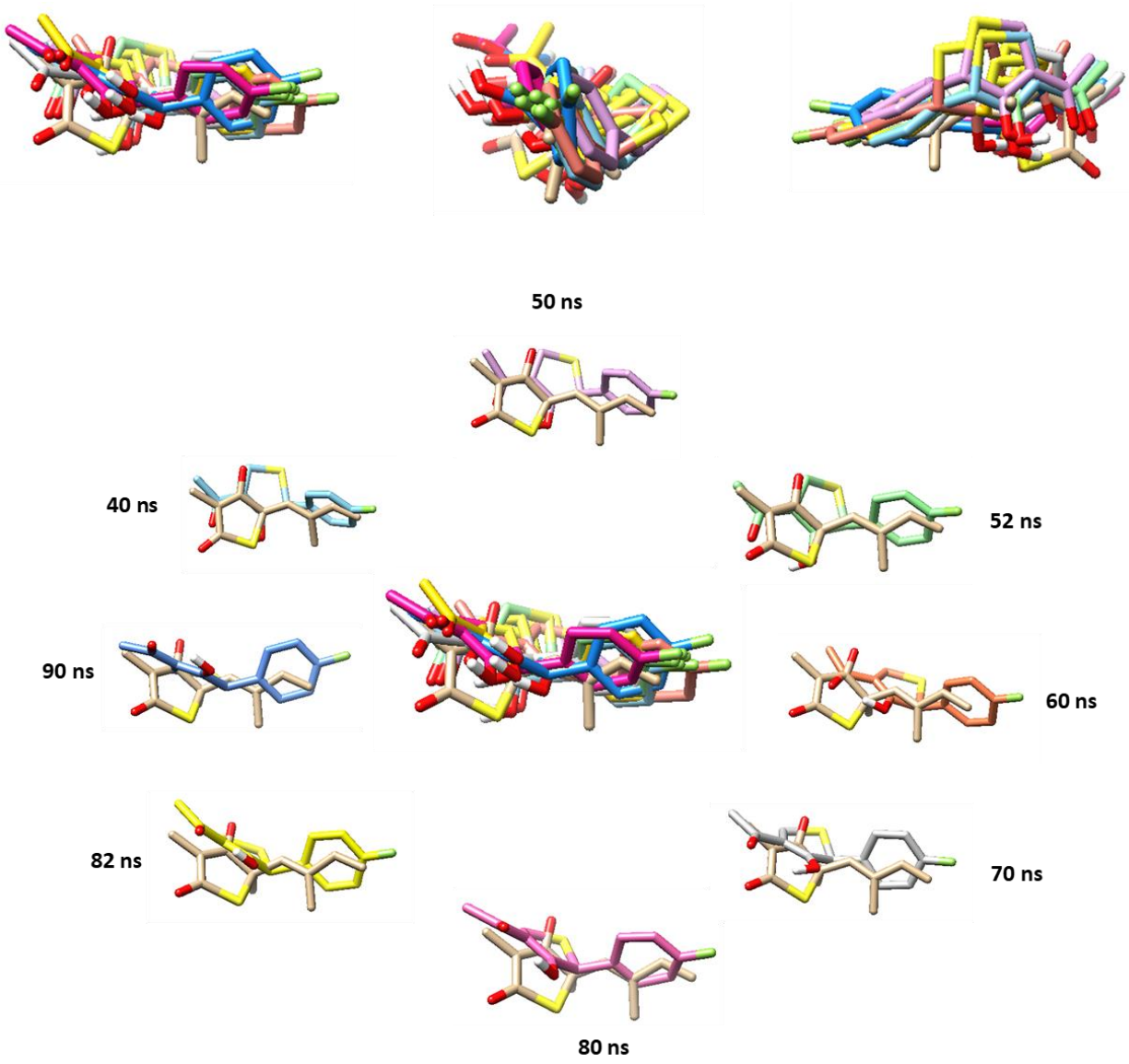


Figure 12: Conformational evolution of ligand 44207286 in mutant KasA active site. A cluster of ligand 44207286 colour-coded conformations sampled from 40 to 90 ns of a 200 ns trajectory versus a starting stricture conformation of TLM.

The radius of gyration (**Figure 11**) was analysed over the 200 ns simulation to determine the compactness of the system [58].

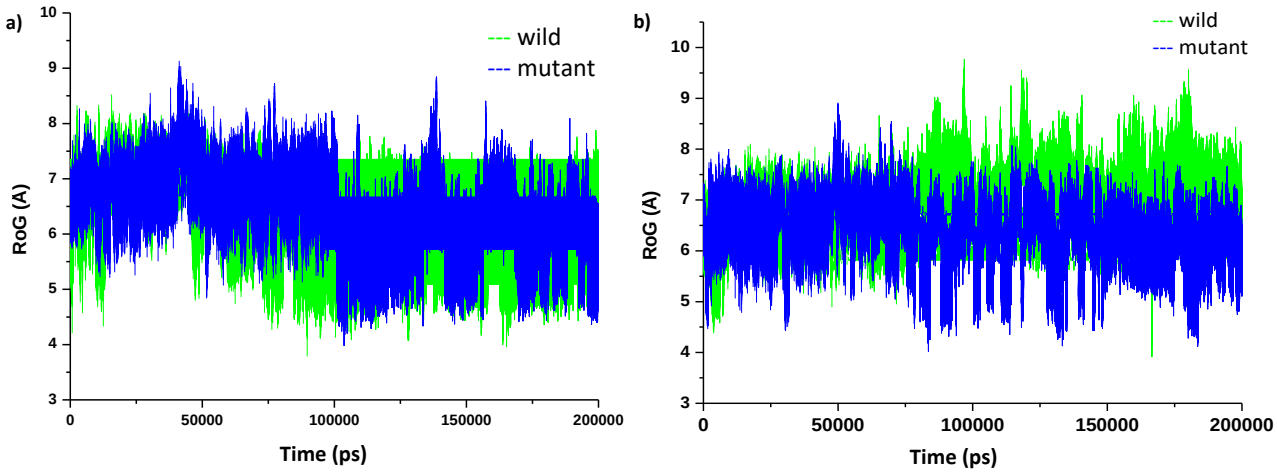


Figure 13. showing the RoG of the wild-type and C171Q mutant of KasA complexed with a) TLM and b) 44207286.

The RoG displayed a similar pattern for both the wild-type and mutant complexed with TLM with averages of 6.15 Å and 6.39 Å respectively. A higher RoG of approximately 6.97 Å was observed for the wild-type when complexed with **44207286**. The mutant complexed with **44207286** displayed a value of 6.32 Å possibly indicating a more compact therefore stable system [59]. This correlates to the results obtained in the RMSD analysis.

To determine the conformational changes and flexibility of the protein dependent on structural dynamics after binding [60], the RMSF was analysed using the 200 ns generated trajectories. **Figure 12** shows the RMSF of both TLM and **44207286** with the wild-type and mutant complexes.

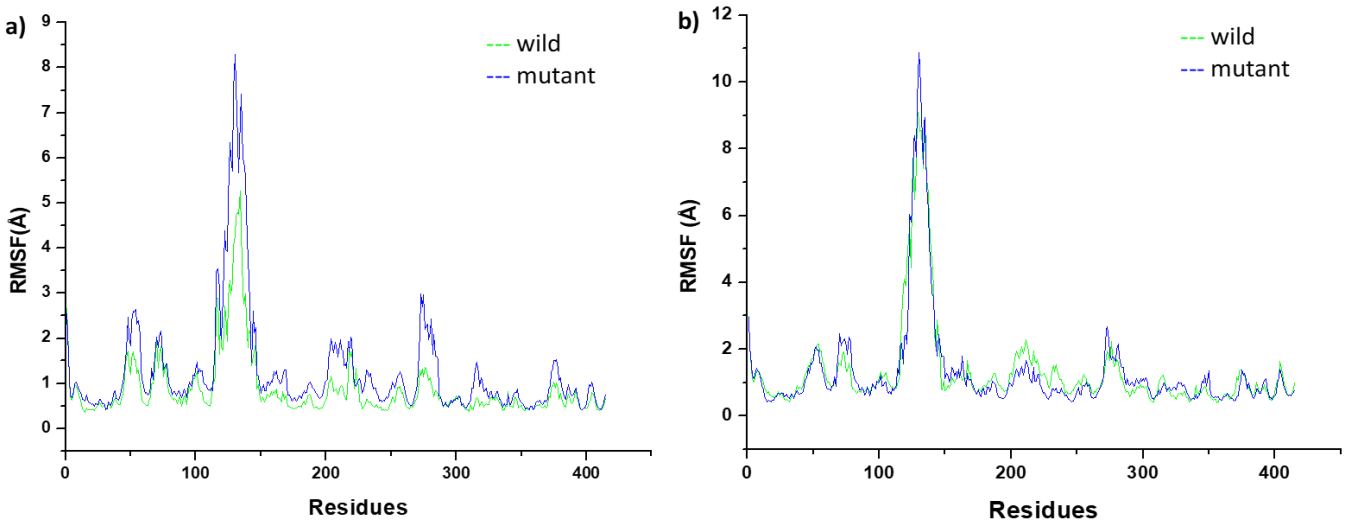


Figure 14. Root mean square fluctuation (RMSF) of the wild-type and C171Q mutation of KasA complexed with a) TLM and b) 44207286.

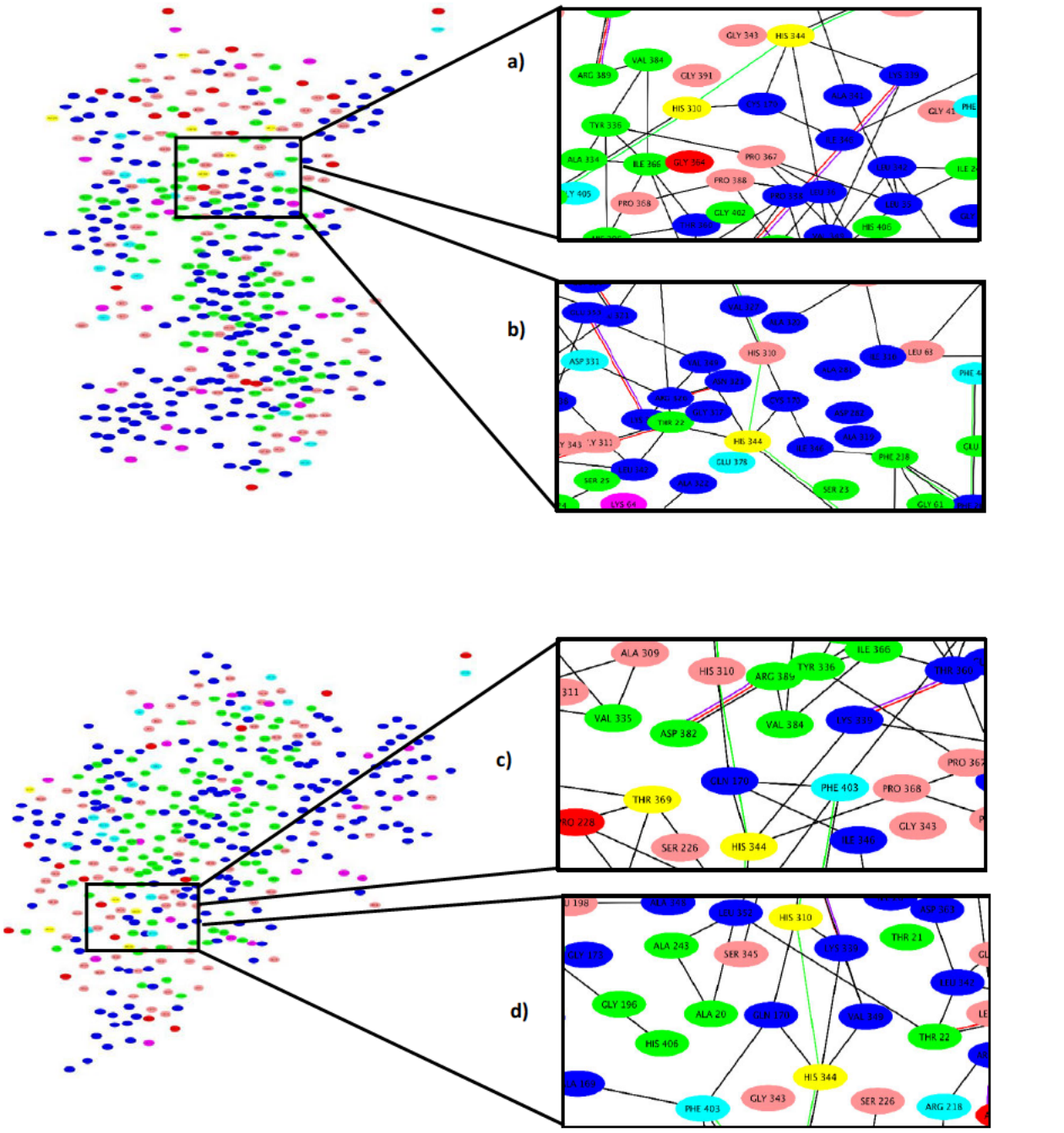
All 4 complexes showed similar patterns over the 200 ns simulation. General fluctuations were seen at residues Ser12, Lys52, Val82, Pro100, Lys135, Ile153, Ala225 and Ala280 which could be due to the initial backbone change or the influence of the distal amino acids [61]. For the TLM complexes, the presence of the mutation caused an increase in fluctuations as compared to the wild-type, with averages of 1.25 Å in comparison to 0.87 Å indicating an overall impact of fluctuations by the mutation. The peak at residue 135 is may be due to the evolution of a primary to a secondary structure during the simulation. After inspection of the **44207286** complexes, the mutation shows almost identical fluctuations as compared to the wild-type with both maintaining stability showing values of 1.29 Å and 1.33 Å respectively. This indicated that the mutation held no power over the drug hit. From these results, it can be concluded that with the presence of **44207286**, the effect of the mutation is weakened. These results correspond with the RMSD and RoG results and thereby attribute a successful tailored pharmacophore-virtual screening approach.

Proposed potential inhibitor increase indicates similar binding patterns

The residue network interaction is a topology based approach in identifying key residue interactions present in proteins [62]. Analysing these interactions could shed light on the

differences in binding of the amino acids present in backbones. RINs were generated using average MD structures generated from simulations. Graphical representation of the RIN networks are represented in **Figure 13**.

According to the viewing in Figure 11, there is a slight distortion in the network interaction due to the presence of the C171Q mutant. The presence of three Van der Waals forces are seen between CYS170 with HIS310, HIS344 and ILE346 in both the wild-type-TLM and wild-type-44207286 complexes while there is a single change in one of the interactions with the replacement of the CYS170 with GLN170 now interacting with PHE403 rather than ILE346 for both the mutant -TLM and -44207286 complexes. This correlated to the data associated with binding energy as the benzene ring present in phenylalanine differs from the energy associated with the isoleucine. The presence of the mutation does in fact affect the interaction network and thereby the conformation, altering the backbone and distorting the environment conducive to ligand binding [63].



DSSP code	Key	DSSP code	Key
B	Residue in isolated β-bridge (single pair beta sheet hydrogen bond formation).	E	Extended strand in parallel and/or anti-parallel beta sheet conformation. Minimum length 2 residues.
G	3-turn helix (3-10 helix). Minimum length 3 residues.	H	4-turn helix (alpha helix). Minimum length 4 residues.
S	Bend (the only non-hydrogen-based assignment).	T	Hydrogen bonded turn (3, 4 or 5 turn).
Interaction type			
Hydrogen bond		Ionic bond	Van der Waals

Figure 15. Showing the residue interaction network (RIN) of the a) wild-type and b) C171Q mutant of KasA complexed with TLM and a) wild-type and b) C171Q mutant of KasA complexed with 44207286.

In silico ADME predictions

ADME predictions allow for the estimated behaviour of the drugs within an organism dependant on its nature. Drug design is dependent on these specific physiochemical pharmacokinetics properties rendering approval. **Figure 14, Tables 2 and 3** both present the standard drug, TLM and drug hit, 44207286, evaluated using SwissADME [64]. This interface uses Lipinski's rule of five which states that any orally active drug can have no more than one violation of the following rules; number of hydrogen atom acceptors ≤10, number of hydrogen atom donors ≤5, molecular weight ≤500 g/mol, and the lipophilicity, Log p ≤ 4.15 [65]. Hence, high bioavailability scores and rule criteria indicates positive permeation and absorption as a function of Lipinski's rule. In case of the hit, ADME properties indicate a good chance of future TB treatment.

Table 2. Physiochemical parameters for the drugs.

Parameters	***Thiolactomycin***	***44207286***
Bioavailability score	0.85	0.85
Molecular weight (g/mol)	210.29	238.28
Hydrogen bond donors	1	1
Hydrogen bond acceptors	2	3
Lipophilicity (MLOGP)	1.37	1.97
Polarity : TPSA ($Å^2$)	62.60	62.60
Lipinski violations	0	0

Table 3. Pharmacokinetics parameters for the drugs.

Parameters	***Thiolactomycin***	***44207286***
Gastrointestinal absorption	High	High
BBB permeant	Yes	Yes
P-glycoprotein substrate	No	No
CYP1A2 inhibitor	No	Yes
CYP2C19 inhibitor	No	No
CYP2C9 inhibitor	Yes	No
CYP2D6 inhibitor	No	No
CYP3A4 inhibitor	No	No
Log k_p (skin permeant) (cm/s)	-5.59	-6.18

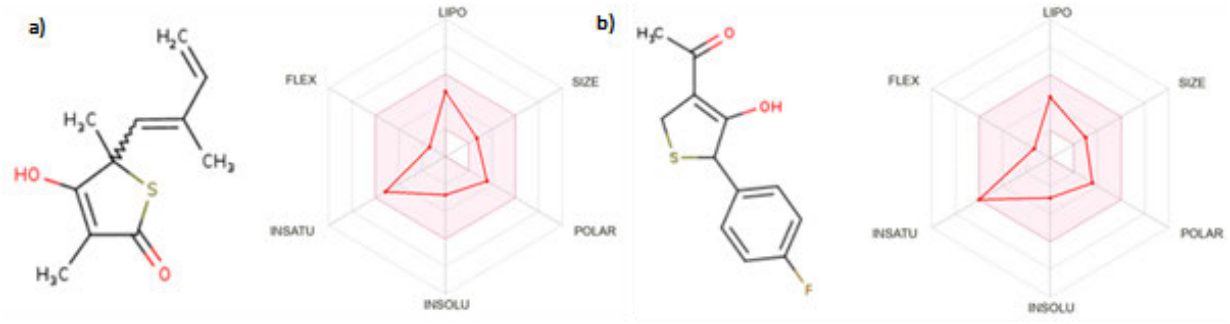


Figure 16. SwissADME bioavailability radars for a) TLM and b) 44207286. The pink area represents the optimal range for each property: saturation (INSATU), lipophilicity(LIPO) (XLOGP3), polarity (POLAR), solubility (INSOLU), molecular mass (SIZE), and flexibility(FLEX).

Conclusion

The evolution of *Mycobacterium tuberculosis* (*Mtb)* strains prompted novel approaches of drug design and discovery. In analysing the crystal structure of the β-ketoacyl ACP synthase (KasA) of *Mtb* in its wild-type and C171Q mutant form, systems were subjected to 200 ns MD simulations followed by post MD-analysis. Post-MD analyses including MM/PBSA, RMSD, RoG, RMSF and RIN assisted us to understand the effect of the C171Q mutation on the binding of thiolactomycin (TLM) to *Mtb*-KasA. Virtual screening 'tailored pharmacophore' approach was carried out to identify potential inhibitors of *Mtb*KasA. In the presence of TLM, post-MD analyses of the C171Q mutation system suggested a more flexible conformational nature of *Mtb*-KasA compared to the wild-type. Hence, the mutation results in higher residual flexibility, revealed by higher RMSF values, leading to a decrease in total calculated binding free energy ($\Delta G_{binding}$) of TLM to *Mtb*-KasA. In addition, RIN analysis suggested that the presence of the mutation affected the residue interaction network and ligand conformation. Eventually, a tailored pharmacophore approach identified a hit with improved binding affinities in both the C171Q mutant and wild-type systems when compared to TLM, a drug used in the clinical treatment of TB. Results further show that the hit binds to the protein, C171Q mutant and wild-type systems, and form a stable compact structure with a slight difference on the impact of the overall dynamics. Further experimental investigation is thereby encouraged in order to validate the findings of the new hit against *Mtb.* The findings of this study serve as a cornerstone for the discovery and design of drugs against a wide range of targets, mutant and/or wild-type.

Conflicts of Interest: The authors declare no intellectual or financial conflicts of interest.

Acknowledgments: The authors acknowledge the College of Health Sciences, School of Laboratory Medicine and Medical sciences and The Centre for High Performance Computing (www.chpc.ac.za) Cape Town, South Africa, for technical computational support, respectively. The work by Kgothatso E Machaba, reported herein was made possible through partial funding by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the Research Capacity Development Initiative (RCDI) Programme from funding received from the South African National Treasury. The content hereof is the sole responsibility of the authors and do not necessarily represent the official views of the SAMRC or the funders".

References

1. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Computational methods in drug discovery. *Pharmacol. Rev.* **66**, 334–395 (2014).
2. Broccatelli, F. & Brown, N. Best of both worlds: on the complementarity of ligand-based and structure-based virtual screening. *J. Chem. Inf. Model.* **54**, 1634–1641 (2014).
3. Ripphausen, P., Nisius, B. & Bajorath, J. State-of-the-art in ligand-based virtual screening. *Drug Discov. Today* **16**, 372–376 (2011).
4. Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discov. Today* **7**, 1047–1055 (2002).
5. Machaba, K. E., Mhlongo, N. N., Dokurugu, Y. M. & Soliman, M. E. S. Tailored-pharmacophore model to enhance virtual screening and drug discovery: a case study on the identification of potential inhibitors against drug-resistant *Mycobacterium tuberculosis* (3R)-hydroxyacyl-ACP dehydratases. *Future Med. Chem.* **9**, 1055–1071 (2017).
6. von Korff, M. & Steger, M. GPCR-tailored pharmacophore pattern recognition of small molecular ligands. *J. Chem. Inf. Comput. Sci.* **44**, 1137–1147 (2004).
7. Russell, D. G. *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nat. Rev. Mol. cell Biol.* **2**, 569–578 (2001).
8. Chakaya, J. *et al.* Global Tuberculosis Report 2020–Reflections on the Global TB burden, treatment and prevention efforts. *Int. J. Infect. Dis.* (2021).
9. Yamamoto, A. *et al.* Study on clinico-pathological features of active pulmonary tuberculosis found at autopsy in a general hospital. *Kekkaku:[Tuberculosis]* **84**, 71–78 (2009).
10. Saunders, M. J. & Evans, C. A. COVID-19, tuberculosis and poverty: preventing a perfect storm. (2020).
11. Daftary, A. HIV and tuberculosis: the construction and management of double stigma. *Soc. Sci. \& Med.* **74**, 1512–1519 (2012).
12. Vosátka, R., Krátk\`y, M. & Vinšová, J. Triclosan and its derivatives as antimycobacterial active agents. *Eur. J. Pharm. Sci.* **114**, 318–331 (2018).
13. Penn-Nicholson, A. *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.* (2021).
14. Manjelievskaia, J., Erck, D., Piracha, S. & Schrager, L. Drug-resistant TB: deadly,

costly and in need of a vaccine. *Trans. R. Soc. Trop. Med. Hyg.* **110**, 186–191 (2016).

15. Organization, W. H. & others. *Multidrug and extensively drug-resistant TB (M*. (2010).
16. Jeffrey North, E., Jackson, M. & E Lee, R. New approaches to target the mycolic acid biosynthesis pathway for the development of tuberculosis therapeutics. *Curr. Pharm. Des.* **20**, 4357–4378 (2014).
17. Schiebel, J. *et al.* Structural basis for the recognition of mycolic acid precursors by KasA, a condensing enzyme and drug target from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **288**, 34190–34204 (2013).
18. Takayama, K., Wang, C. & Besra, G. S. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.* **18**, 81–101 (2005).
19. Sullivan, T. J. *et al.* High affinity InhA inhibitors with activity against drug-resistant strains of *Mycobacterium tuberculosis*. *ACS Chem. Biol.* **1**, 43–53 (2006).
20. Marrakchi, H., Lanéelle, G. & Quémard, A. InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II. *Microbiology* **146**, 289–296 (2000).
21. Bhatt, A., Kremer, L., Dai, A. Z., Sacchettini, J. C. & Jacobs Jr, W. R. Conditional depletion of KasA, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis. *J. Bacteriol.* **187**, 7596–7606 (2005).
22. Wong, H. C., Liu, G., Zhang, Y.-M., Rock, C. O. & Zheng, J. The solution structure of acyl carrier protein from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **277**, 15874–15880 (2002).
23. Luckner, S. R., Machutta, C. A., Tonge, P. J. & Kisker, C. Crystal structures of *Mycobacterium tuberculosis* KasA show mode of action within cell wall biosynthesis and its inhibition by thiolactomycin. *Structure* **17**, 1004–1013 (2009).
24. Rudraraju, R. S. *et al. Mycobacterium tuberculosis* KasA as a drug target: Structure-based inhibitor design. *Front. Cell. Infect. Microbiol.* 1352 (2022).
25. Schaefer, B., Kisker, C. & Sotriffer, C. A. Molecular dynamics of *Mycobacterium tuberculosis* KasA: implications for inhibitor and substrate binding and consequences for drug design. *J. Comput. Aided. Mol. Des.* **25**, 1053–1069 (2011).
26. Berman, H. M. *et al.* The protein data bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 899–907 (2002).
27. Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.*

66, 27–85 (2003).

28. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 1–17 (2012).
29. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
30. Morris, G. M. *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662 (1998).
31. Tian, C. *et al.* ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552 (2019).
32. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210 (2013).
33. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).
34. Origin. 2003. Origin 8 OriginLab Corp., Northampton, M. No Title.
35. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **8**, 127–134 (1995).
36. Brandsdal, B. O. *et al.* Free energy calculations and ligand binding. *Adv. Protein Chem.* **66**, 123–158 (2003).
37. Word, J. M. *et al.* Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **285**, 1711–1733 (1999).
38. Sterling and Irwin. No Title. *J. Chem. Inf. Model* (2015).
39. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, D1388--D1395 (2021).
40. Kour, J. & Singh, K. Virtual screening using the ligand ZINC database for novel lipoxygenase-3 inhibitors. *Bioinformation* **9**, 583 (2013).
41. Xie, X.-Q. S. Exploiting PubChem for virtual screening. *Expert Opin. Drug Discov.* **5**, 1205–1220 (2010).
42. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **52**, 1757–1768 (2012).

43. Réau, M., Langenfeld, F., Zagury, J.-F., Lagarde, N. & Montes, M. Decoys selection in benchmarking datasets: overview and perspectives. *Front. Pharmacol.* **9**, 11 (2018).
44. Kim, S. Getting the most out of PubChem for virtual screening. *Expert Opin. Drug Discov.* **11**, 843–855 (2016).
45. 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.* **8**, 1–10 (2018).
46. Fukunishi, Y. & Lintuluoto, M. Development of chemical compound libraries for in silico drug screening. *Curr. Comput. Aided. Drug Des.* **6**, 90–102 (2010).
47. Lionta, E., Spyrou, G., K Vassilatis, D. & Cournia, Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top. Med. Chem.* **14**, 1923–1938 (2014).
48. Alfarano, C. *et al.* The biomolecular interaction network database and related tools 2005 update. *Nucleic Acids Res.* **33**, D418--D424 (2005).
49. Doncheva, N. T., Klein, K., Domingues, F. S. & Albrecht, M. Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* **36**, 179–182 (2011).
50. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
51. Aakeröy, C. B. & Seddon, K. R. The hydrogen bond and crystal engineering. *Chem. Soc. Rev.* **22**, 397–407 (1993).
52. Breslow, R. Hydrophobic effects on simple organic reactions in water. *Acc. Chem. Res.* **24**, 159–164 (1991).
53. Simmerling, C., Strockbine, B. & Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **124**, 11258–11259 (2002).
54. Luckner, S. R., Liu, N., Am Ende, C. W., Tonge, P. J. & Kisker, C. A slow, tight binding inhibitor of InhA, the enoyl-acyl carrier protein reductase from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **285**, 14330–14337 (2010).
55. Enkhbayar, P., Hikichi, K., Osaki, M., Kretsinger, R. H. & Matsushima, N. 310-helices in proteins are parahelices. *Proteins Struct. Funct. Bioinforma.* **64**, 691–699 (2006).
56. Vieira-Pires, R. S. & Morais-Cabral, J. H. 310 helices in channels and other membrane proteins. *J. Gen. Physiol.* **136**, 585–592 (2010).
57. Armen, R., Alonso, D. O. V & Daggett, V. The role of α-, 310-, and π-helix in helix→ coil transitions. *Protein Sci.* **12**, 1145–1157 (2003).

58. Lobanov, M. Y., Bogatyreva, N. S. & Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* **42**, 623–628 (2008).
59. Hoshen, J. Percolation and cluster structure parameters: The radius of gyration. *J. Phys. A. Math. Gen.* **30**, 8459 (1997).
60. Benson, N. C. & Daggett, V. A comparison of multiscale methods for the analysis of molecular dynamics simulations. *J. Phys. Chem. B* **116**, 8722–8731 (2012).
61. Némethy, G., Leach, S. J. & Scheraga, H. A. The influence of amino acid side chains on the free energy of helix-coil transitions1. *J. Phys. Chem.* **70**, 998–1004 (1966).
62. K Grewal, R. & Roy, S. Modeling proteins as residue interaction networks. *Protein Pept. Lett.* **22**, 923–933 (2015).
63. Kortemme, T. & Baker, D. A simple physical model for binding energy hot spots in protein--protein complexes. *Proc. Natl. Acad. Sci.* **99**, 14116–14121 (2002).
64. Gupta, N., Choudhary, S. K., Bhagat, N., Karthikeyan, M. & Chaturvedi, A. In silico prediction, molecular docking and dynamics studies of steroidal alkaloids of holarrhena pubescens wall. ex G. don to guanylyl cyclase C: Implications in designing of novel antidiarrheal therapeutic strategies. *Molecules* **26**, 4147 (2021).
65. Scow, D. T., Luttermoser, G. & Dickerson, K. Leukotriene inhibitors in the treatment of allergy and asthma. *Am. Fam. Physician* **75**, 65–70 (2007).

Supplementary material

Table 4: Docking scores and binding energies of the wild-type and C171Q mutant in complex with 4 drug hits (kcal/mol).

Ligand ID	Docking scores		Binding energies	
	Wild-type	Mutant	Wild-type	Mutant
ZN12	-8.5	-8.6	-11.4509	-31.4520
135424785	-8.6	-8.7	-23.3463	-13.6266
44207286	-8.3	-8.9	-35.7584	-36.3433
136764894	-8.7	-8.6	-20.3921	-21.1200

Table 5: Per residue energy (kcal/mol) decomposition of the KasA-TLM (wild-type) complex.

Residue	ΔG_{vdw}	ΔG_{elec}	ΔG_{pol}	ΔG_{non-p}	ΔG_{total}
CYS 170	-0.64 ± 0.13	-0.005 ± 0.22	0.09 ± 0.17	-0.004 ± 0.003	-0.56 ± 0.13
PRO 279	-1.50 ± 0.22	0.27 ± 0.06	-0.05 ± 0.06	-0.16 ± 0.025	-1.44 ± 0.20
HIE 310	-0.64 ± 0.44	-2.24 ± 0.73	1.81 ± 0.45	-0.03 ± 0.016	-1.10 ± 0.30
THR 312	-1.48 ± 0.26	-0.29 ± 0.19	0.68 ± 0.20	-0.14 ± 0.015	-1.23 ± 0.33
HIE 344	-0.54 ± 0.30	-3.03 ± 0.69	2.14 ± 0.30	-0.04 ± 0.015	-1.47 ± 0.26
PHE 401	-1.60 ± 0.29	0.89 ± 0.31	-0.43 ± 0.17	-0.11 ± 0.016	-1.25 ± 0.26
PHE 403	-3.09 ± 0.41	-0.16 ± 0.19	0.99 ± 0.28	-0.22 ± 0.023	-2.49 ± 0.43

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total).

Table 6: Per residue energy (kcal/mol) decomposition of the KasA-TLM (mutant) complex.

Residue	ΔG_{vdw}	ΔG_{elec}	ΔG_{pol}	ΔG_{non-p}	ΔG_{total}
GLN 170	-0.77 ± 0.23	-0.34 ± 0.20	0.33 ± 0.16	-0.01 ± 0.012	-0.79 ± 0.24
PRO 279	-1.56 ± 0.22	0.26 ± 0.12	-0.019 ± 0.11	-0.17 ± 0.03	-1.48 ± 0.21
THR 312	-1.53 ± 0.34	-0.52 ± 0.29	0.86 ± 0.30	0.13 ± 0.014	-1.32 ± 0.43
HIE 344	-0.38 ± 0.36	-3.44 ± 0.59	2.44 ± 0.20	-0.04 ± 0.015	-1.43 ± 0.23
PHE 401	-1.51 ± 0.22	1.03 ± 0.26	-0.59 ± 0.13	-0.09 ± 0.021	-1.17 ± 0.32
GLY 402	-1.26 ± 0.18	-0.49 ± 0.19	0.83 ± 0.16	-0.08 ± 0.020	-1.00 ± 0.17
PHE 403	-2.89 ± 0.40	-0.35 ± 0.23	1.01 ± 0.21	-0.22 ± 0.024	-2.45 ± 0.38

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total).

Table 7: Per residue energy (kcal/mol) decomposition of the KasA-24 (wild-type) complex.

Residue	ΔG_{vdw}	ΔG_{elec}	ΔG_{pol}	ΔG_{non-p}	ΔG_{total}
CYS 170	-0.76 ± 0.18	-1.83 ± 0.62	1.11 ± 0.27	-0.03 ± 0.011	-1.51 ± 0.47
ALA 278	-1.77 ± 0.15	-0.52 ± 0.14	0.46 ± 0.16	-0.09 ± 0.010	-1.92 ± 0.25
PRO 279	-1.71 ± 0.16	0.56 ± 0.10	-0.14 ± 0.08	-0.09 ± 0.013	-1.38 ± 0.17
THR 312	-1.48 ± 0.26	-0.29 ± 0.19	0.68 ± 0.20	-0.14 ± 0.015	-1.24 ± 0.33
HIE 344	-0.54 ± 0.32	-3.02 ± 0.69	1.99 ± 0.30	-0.04 ± 0.013	-1.61 ± 0.36
PHE 401	-0.13 ± 0.63	-4.32 ± 1.08	1.23 ± 0.29	-0.08 ± 0.012	-3.31 ± 0.89
GLY 402	-1.35 ± 0.19	-0.57 ± 0.29	0.92 ± 0.21	-0.04 ± 0.014	-1.04 ± 0.27
PHE 403	-2.70 ± 0.39	0.40 ± 0.24	0.78 ± 0.21	-0.18 ± 0.024	-1.69 ± 0.42

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total).

Table 8: Per residue energy (kcal/mol) decomposition of the KasA-24 (mutant) complex.

Residue	ΔG_{vdw}	ΔG_{elec}	ΔG_{pol}	ΔG_{non-p}	ΔG_{total}
GLN 170	-0.97 ± 0.22	-0.69 ± 0.26	0.75 ± 0.14	-0.02 ± 0.007	-0.93 ± 0.19
ALA 278	-1.29 ± 0.13	-0.44 ± 0.12	0.44 ± 0.13	-0.04 ± 0.013	-1.33 ± 0.25
PRO 279	-2.02 ± 0.19	0.49 ± 0.08	-0.22 ± 0.05	-0.17 ± 0.034	-1.92 ± 0.24
HIE 310	-1.00 ± 0.33	-0.83 ± 0.91	1.07 ± 0.52	-0.05 ± 0.009	-0.81 ± 0.51
THR 312	-1.52 ± 0.26	-1.03 ± 0.19	0.97 ± 0.18	-0.10 ± 0.026	-1.67 ± 0.29
HIE 344	-0.61 ± 0.16	-2.37 ± 0.72	1.97 ± 0.38	-0.027 ± 0.009	-1.04 ± 0.26
PHE 401	0.08 ± 0.56	-5.01 ± 0.58	2.58 ± 0.35	-0.095 ± 0.019	-2.43 ± 0.67
GLY 402	-1.18 ± 0.08	-1.14 ± 0.43	1.56 ± 0.41	-0.023 ± 0.008	-0.79 ± 0.19
PHE 403	-2.90 ± 0.45	-0.94 ± 0.15	1.24 ± 0.11	-0.24 ± 0.022	-2.81 ± 0.41

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total).

Table 9: Per residue energy (kcal/mol) decomposition of the KasA-24 (mutant) complex.

Residue	ΔG_{vdw}	ΔG_{elec}	ΔG_{pol}	ΔG_{non-p}	ΔG_{total}
GLN 170	-0.97 ± 0.22	-0.69 ± 0.26	0.75 ± 0.14	-0.02 ± 0.007	-0.93 ± 0.19
ALA 278	-1.29 ± 0.13	-0.44 ± 0.12	0.44 ± 0.13	-0.04 ± 0.013	-1.33 ± 0.25
PRO 279	-2.02 ± 0.19	0.49 ± 0.08	-0.22 ± 0.05	-0.17 ± 0.034	-1.92 ± 0.24
HIE 310	-1.00 ± 0.33	-0.83 ± 0.91	1.07 ± 0.52	-0.05 ± 0.009	-0.81 ± 0.51
THR 312	-1.52 ± 0.26	-1.03 ± 0.19	0.97 ± 0.18	-0.10 ± 0.026	-1.67 ± 0.29
HIE 344	-0.61 ± 0.16	-2.37 ± 0.72	1.97 ± 0.38	-0.027 ± 0.009	-1.04 ± 0.26
PHE 401	0.08 ± 0.56	-5.01 ± 0.58	2.58 ± 0.35	-0.095 ± 0.019	-2.43 ± 0.67
GLY 402	-1.18 ± 0.08	-1.14 ± 0.43	1.56 ± 0.41	-0.023 ± 0.008	-0.79 ± 0.19
PHE 403	-2.90 ± 0.45	-0.94 ± 0.15	1.24 ± 0.11	-0.24 ± 0.022	-2.81 ± 0.41

*ΔG = binding energy contribution from ΔG_{vdw} (van de Waals forces), ΔG_{eel} (electrostatic interactions) ΔG_{pol} (polar solvation energies) and ΔG_{non-p} (non-polar solation energies) to ΔG_{tot} (total)

CHAPTER 6

Is Gene ontology an effective approach in drug repurposing? A case study on *M. tuberculosis* Dihydropteroate synthase (DHPS/folP1*)*: Insight from a docking and MM/PBSA perspective in estimating ligand binding affinities of potential inhibitors

Kimona Kisten, Kgothatso E. Machaba, Nolwazi Ngidi and Ndumiso N. Mhlongo*

School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban 4001, South Africa

* Corresponding author: Ndumiso N. Mhlongo

Email: MhlongoN4@ukzn.ac.za

Telephone: +2731 260 2428, Fax: +2731 260 7792

Abstract:

Gene ontology is an widely used bioinformatics tool utilizing a range of methods to specialise in the understanding of cellular and molecular functionality within biological systems. One aspect is being employed as a starting point for predictive values in the process of drug discovery. Moreso, investigation of drug targets of interest and potential inhibitors has become more cost-effective and viable with the use of molecular dynamics. In this study, the probability of three drugs with DrugBank IDs of DB03705, DB04047 and DB04196 identified through a gene ontology approach, are repurposed as potential inhibitors of *Mycobacterium tuberculosis*-Dihydropteroate synthase (DHPS/folP1). Application of molecular docking, molecular dynamics simulations and binding free energy calculations were applied to evaluate the reliability of a gene ontology RepTB approach applied by *Passi et al* in identifying folP1 potential inhibitors (-5.6, -6.3 and -7.0 kcal/mol). Furthermore, the complexes were subjected to 200 ns MD simulations and MD analyses. Moreover, all complexes exhibited RMSD values lower than the benchmark value of 2 Å with averages of 2.04, 2.10, 2.23 and 1.67 Å respectively. Further analysis indicated no discernible increases in flexibility were seen for any of the complexes. Binding energies of DB03705 (19.99 kcal/mol) and DB04196 (23.87 kcal/mol) indicated better affinities than the control DB03592 (17.55 kcal/mol). Further interaction analysis indicated that the most significant contribution of residue energies was attributed to the presence of the rings, leading to increased hydrophobicity. The results therefore allude to the fact that the Gene ontology (RepTB) approach is a viable basis for the extensive method validation in drug design and discovery.

Keywords: *Mycobacterium tuberculosis*, *folP1*, drug repurposing, molecular dynamics, gene ontology, network-based inference.

Introduction

Gene ontology is described as a bioinformatics tool that can be seen as a primer in cataloguing gene functions within organisms [1]. It is sectioned into three categories: molecular function referring to biochemical activity linked to receptor-ligand binding, biological processes dictating the contribution of the gene in the mechanism and the cellular component denoting the position of the gene in the cell [2]. In current studies, gene ontology is an approach used in identifying potential inhibitors from a series of interactions analysed using network-based inference (NBI) data. The data obtained from the evaluation of interactions between a pool of drug targets and multiple drugs in the market are used for identification of potential drug candidates [3]. The aforementioned approach aids in the generation of a predictive value in the identification of potential drug candidates and selected drug targets. Gene ontology provides an extended in silico strategy applied to facilitate the drug discovery process.

Application of gene ontology is applied in the drug discovery of tuberculosis (TB) [4]. In this approach, targets which have no known variation in clinical isolates, no human homologs, are essential for the survival or virulence of *M. tuberculosis*, and whose predicted drugs may have a synergistic bactericidal effect, are prioritized. *Passi et al* developed the RepTB method which correlates interactions between drug-target pairs by analysing the predicted drug target interactions (DTIs) [3]. From the list of predicted DTIs, Dihydropteroate synthase (DHPS/folP1) was one of the selected targets of *M. tuberculosis*. According to Gengenbacher et al [5], DHPS (**Figure 1**), essential for the cell growth and survival of *M. tuberculosis*, is a well-validated drug target.

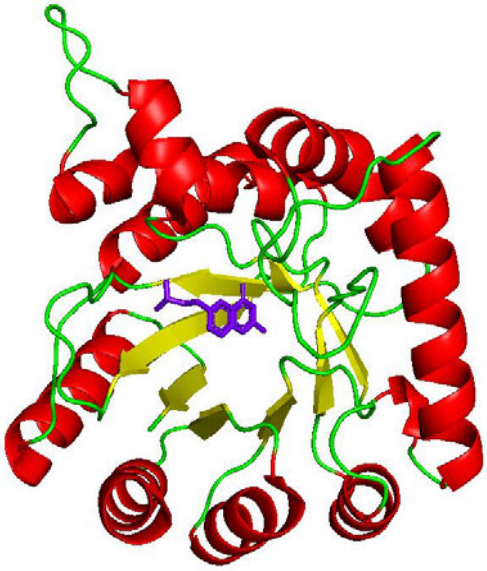


Figure 1: 3D x-ray crystallography structure of docked *M. tuberculosis*-DHPS [6]. DHPS/ folP1 is required for the synthesis of folate, hence, essential for the growth and multiplication of the bacteria [7]. While folate is needed for cell synthesis and growth, DHPS works by catalyzing the condensation reaction of para-aminobenzoate (pABA) and 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (H_2PtPP) to pyrophosphate and 7,8- dihydropteroate [8].

In pursuit of new anti-TB drugs, Passi *et al.* predicted three drugs (**Figure 2**) as potential inhibitors of *M. tuberculosis* DHPS. Further validation of the predicted DTIs via RepTB was thereby encouraged by the authors [3].

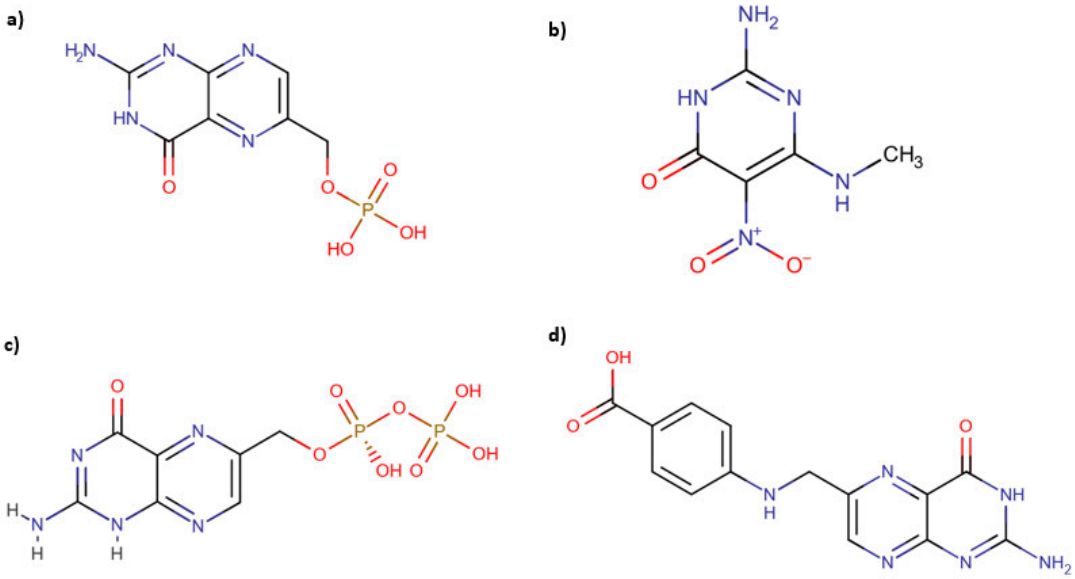


Figure 2: Chemical structures of a) DB03592 (PMM) b) PredDrug1 (6-methylamino-5-nitroisocytosine/DB03705) c) PredDrug2 (6-hydroxylmethylpterin-diphosphate/DB04047) and d) PredDrug3 (Pteroic acid/DB04196).

Therefore, this study provides molecular understanding of the binding mechanism and structural dynamics occurring during ligand-receptor interactions from a computational perspective. Application of molecular docking, molecular dynamics simulations and binding free energy calculations were applied to evaluate the reliability of a RepTB approach in identifying folP1 potential inhibitors. To achieve this, we apply molecular docking and MD simulations thereby cementing the importance of the application of in silico methods in the drug discovery process [9,10,11].

The aim of the study is to provide insight into the binding mechanism and energetics of the successive complex further emphasizing the soundness of the RepTB approach.

The study outcome was based on the following set of questions: 1. Do the predicted drugs show higher docking scores and binding affinities in comparison to the control? 2. Are there any similarities in the structures and thereby reactivity of the ligands? 3. Do any of the predicted drugs show enhanced binding capability? 4. What are the chemical interactions between the receptor and ligand? 5. Can further wet method analysis be continued after in-silico analyses? 6. Did the RepTB approach prove to be an effective method in probable inhibitor identification?

Materials and Methods

Preparation of structures for molecular docking and MD calculations

The *Mtb*DHPS X-ray crystal structure complex with PMM [6] was obtained from the Protein Data Bank with PDB ID of 1EYE [12] . The missing amino acids on the protein structure were fixed using DS Visualizer [13]. The ligands DB03705 & DB04196 were retrieved from DrugBank [14] and DB04047 was built using the Avogadro software [15]. The UCSF Chimera [16] and Avogadro software were utilized for structural preparations and optimization of the receptor and ligands.

Molecular docking

The AutoDock Vina [17], molecular docking tool was used for docking calculations. Docking was performed with default AutoDock Vina parameters. The gridbox housed the entire receptor *via* blind docking and generated the following parameters: exhaustiveness = 8, dimensions were x = 92, y = 98, and z = 92 and centers were x = -38.39, y = 2.08, and z = 31.61. Docked ligand conformations were generated using a Lamarckian genetic algorithm approach [18]. The complex exhibiting the best docking scores were visualized using UCSF Chimera [19] and considered for further molecular dynamic simulations.

Molecular dynamics simulations

MD is a computational technique used in the study of biological molecules to analyze the physical behavior of constituent atoms and molecules [20]. MD simulations give insight on the structural changes of the protein over a given time frame. The UCSF Chimera software was utilized for the preparation of the receptor and ligands for MD simulations. The MD simulations were carried out in the Amber 18 software, using the GPU version of PMEMD with, (GAFF) [21] a FF18SB Amber force field [22]. Amber modules, Antechamber and T Leap were utilized to generate partial charges for ligand and for addition of hydrogen atoms and counter ions, respectively. A TIP3P water box enclosed the system, and the distance was 10 Å between the system surface and box boundary. The system was minimized for 2500 steps and then gradually heated from 0 to 300 K with the collision frequency and harmonic restraints at 1 ps, 5 kcal mol^{-1} $Å^{-2}$, respectively using the Langevin thermostat [23]. The system was subjected to equilibration at 300 K,1 bar constant pressure and the SHAKE algorithm [24] restricted the system's bonds with hydrogen atoms. MD simulation was carried out for 200 ns using the

Berendsen barostat in an isothermal-isobaric ensemble, with pressure at 1 bar and pressure-coupling constant at 2 ps.

Binding free energy calculations

Molecular Mechanics Poisson Boltzmann Surface Area is a computational method used to enhance and validate analysis carried out on drug target pairs [21]. Gibbs free energy is calculated using entropic and enthalpic contributions on ligand-receptor complexes [25]. The best binding free energy is attributed to the complex with the most reliable and stable orientation [26].

The Amber18 implemented MM/PBSA method [27] was utilized for binding free energy calculations. Binding free energy calculations consist of a thermodynamic approach that gives the sum of all the intermolecular interactions between the protein and ligand [28].

Equations (1)-(4) describe the binding free energy calculations for all systems.

$$\Delta G_{binding} = G_{complex} - [G_{protein} + G_{ligand}] \quad (1)$$

$$\Delta G_{binding} = E_{MM} + G_{sol} - T\Delta S \quad (2)$$

$$\Delta E_{MM} = E_{ele} + E_{vdw} \quad (3)$$

$$G_{sol} = G_{polar} + G_{non\text{-}polar} \quad (4)$$

Where $\Delta G_{binding}$ is the protein-ligand complex's binding free energy, E_{MM} denotes the sum of molecular mechanics energy, G_{sol} is solvation free energy and the total entropy is $T\Delta S$. E_{ele} and E_{vdw} are electrostatic and van der Waals contributions, respectively and the sum of these two equals ΔE_{MM}. G_{sol} is the sum of polar (G_{polar}) and non-polar ($G_{non\text{-}polar}$) contributions. Per-residue energy decomposition was also estimated using the MM/PBSA approach and LigPlot [29] was used to visualize protein–ligand interactions.

MD analysis

MD trajectories obtained after the 200ns simulations were subjected to analysis calculations using CPPTRAJ. Amber18 modules were used to carry out the calculations. RMSD (root mean square deviation) evaluating the distance between groups of atoms, RMSF (root mean square fluctuation) measuring the displacement of atoms from a reference point and Rg (Radius of Gyration) estimating the distribution of atoms around the axis of the protein were calculated.

ADME predictions

To assess the drug-likeness of the investigated drugs, the SwissADME web tool [30] was utilized and the absorption, distribution, metabolism, and excretion (ADME) parameters reviewed. Drug design and approval require this data as it provides improved insight into essential features of the drugs namely: physicochemical properties lipophilicity, size, polarity, solubility, saturation, and flexibility. This study utilized default predictors as a means of assessment.

Results and Discussion

DB04047 and DB04196 exhibit promising structural integrity

The RMSD assesses the protein stability of a complex system by analysing the digression of the Cα atoms during the conformational changes experienced in binding [31]. **Figure 4A** represents the calculated results. The systems of DHPS-DB03592, DHPS-DB03705, DHPS-DB04047 and DHPS-DB04196 stabilized with RMSD systems lower than the benchmark with averages of 2.04, 2.10, 2.23 and 1.67 Å respectively [32] hence annotating a series of stable systems. DB04047 and DB04196 indicated the most stable complexes.

The RoG determines the distribution of the atoms around the central axis of the protein during binding. The calculations were performed and the results are presented in **Figure 4B**. It can be established that the results suggest that the presence of a ligand induces overall compactness of the system resulting in conformational stability. All complexes indicated similar convergence trends. The presence of a ligand therefore decreases the biomolecular flexibility of the protein which could potentially enhance the chance of inhibition. These results correlate with the RMSD observations.

The RMSF examines the positional fluctuations of the Cα atoms thereby providing insight into the behaviour of the individual amino acids during the course of the simulation [33]. These fluctuations act as an indicator of the overall flexibility and evolution of the amino acids of the protein as conformational changes are applied during the binding of the complex. Results are presented in **Figure 4C**. It can be observed that similar patterns are established for all the complexes with notable fluctuations occurring at amino acids Asp21, Glu51, Asn105, Hie141, Arg214 and Hie255. However, the apo indicated a higher fluctuation at Hie225 and DB04196 indicated a slightly higher fluctuation at Hie141 in comparison to the other ligand complexes. This could be due to the Hie141 sitting in the active site of the protein. Increased fluctuations are seen for the complex of DB03705 at Glu51, Ala115, Asp163 and Val200. This could be potentially due to the presence and influence of distal amino acids. From the aforementioned results, it can be established that the presence of the ligands in the catalytic site reduces the fluctuations of the protein thereby introducing rigidity into the protein structure leading to overall inhibition of the folate protein. DB04047 and DB04196 show the most promising observations indicating the least mobility. These results correlate with the results of RMSD and RoG indicating a stable, compact, and rigid system.

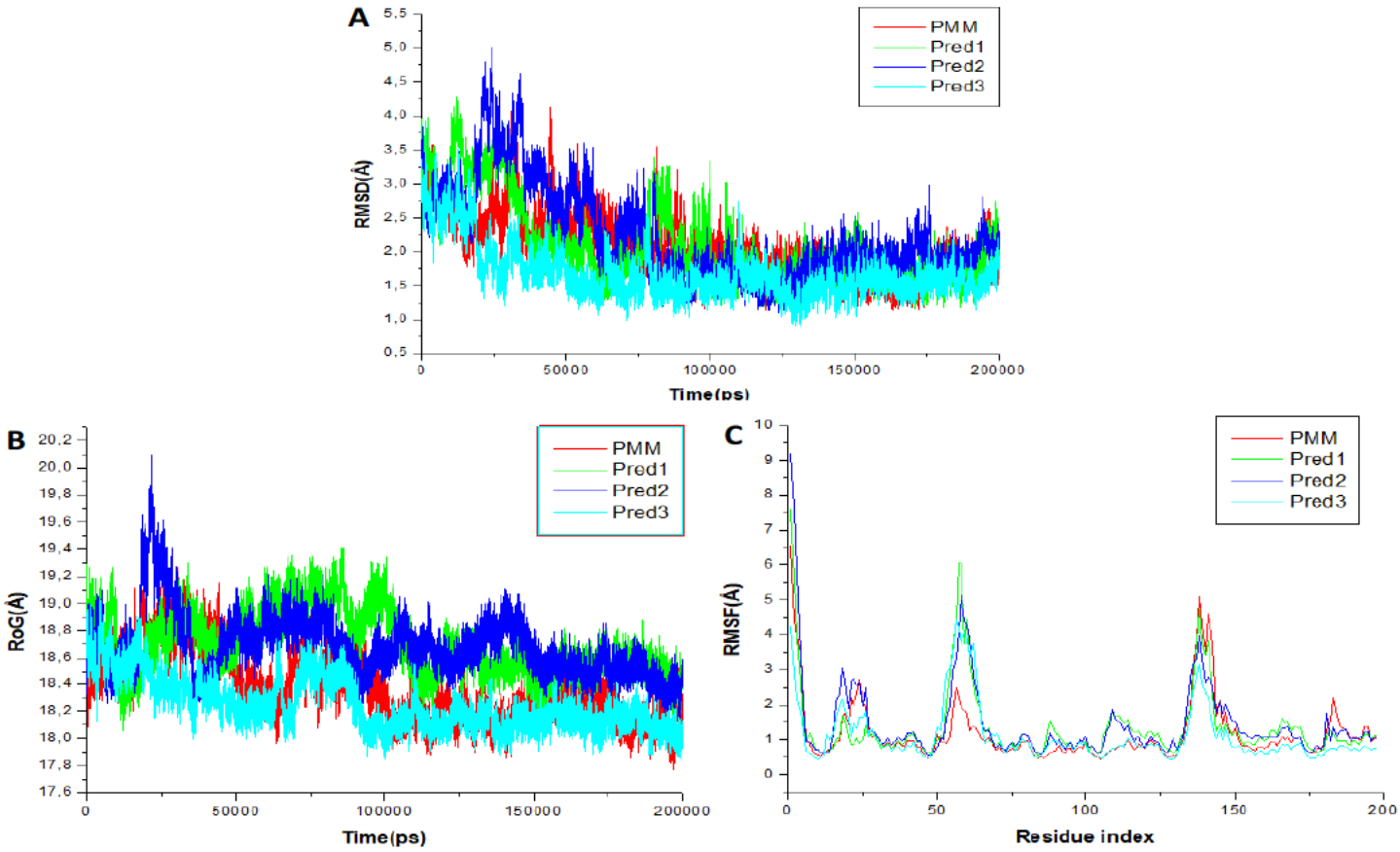


Figure 3: A) RMSD, B) RoG and C) RMSF plots of *M. tuberculosis* - DHPS complexed with DB03592 (PMM), DB03705 (PredDrug1), DB04047 (PredDrug2) and DB04196(PredDrug3).

Highest docking score coincides with favourable binding affinity

All suggested drugs were subjected to docking and further calculated using MM/PBSA to acquire the most stable conformation and favourable binding affinities. As presented in **Table 1**, the docking scores observed ranged from -5.6 to -7.0 kcal/mol. DB04196 was found to have the highest docking score (-7.0 kcal/mol) as compared to the other hits. Furthermore, the surface visualization of the docking of PMM, Pred1, Pred2 and Pred3 presented in **Figure 4** validated that all 4 drug hits occupied the same binding site area.

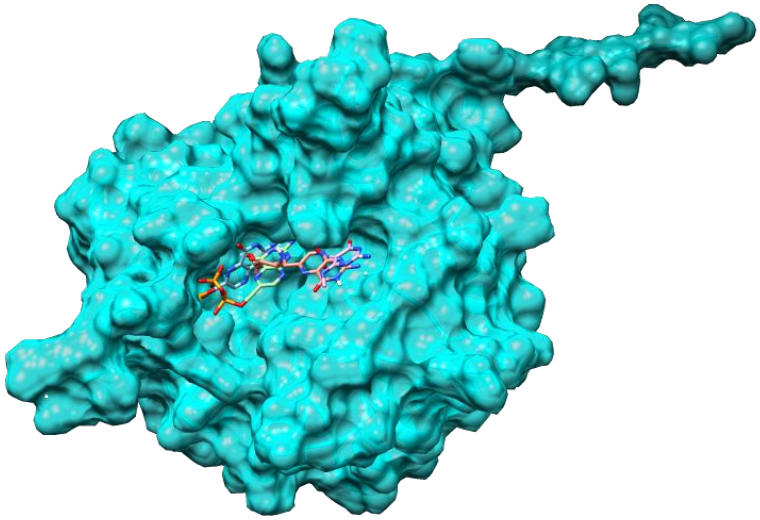


Figure 4: Surface visualization of DB03592 (PMM), DB03705 (PredDrug1), DB04047 (PredDrug2) and DB04196(PredDrug3) in the binding pocket.

Table 1: Binding free energy (kcal/mol) and the components of binding free energy of the different small molecules.

Complex	Docking scores	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
PMM	-6.9	-32.92 ±3.44	202.20±22.70	169.22±18.14	-186.83±18.18	-17.55±6.47
DB03705	-5.6	-23.92±2.84	-31.21±7.73	-54.88±8.28	34.99±5.28	-19.99±4.05
DB04047	-6.3	-37.32±4.07	359.40±29.86	322.08±28.95	-324.36±26.12	-2.28±7.88
DB04196	-7.0	-30.98±3.68	-41.39±14.00	-72.37±14.10	48.50±4.55	-23.87±4.55

***ΔE_{vdw} - Van der Waals; ΔE_{ele} - Electrostatic; ΔG_{gas} - gas-phase energy ΔG_{sol} - solvation energy; ΔG_{bind} – Binding energy**

The energy profiles represented by binding energies are shown in **Table 1**. The total binding free energies ranged from -2.28 to -23.87 kcal/mol with DB03705 (-19.99 kcal/mol) and DB04196 (-23.87 kcal/mol) both exhibiting higher overall binding energies than the suggested control, PMM (-17.55 kcal/mol). Both electrostatic and VDW interactions governed the majority contribution of the overall binding affinity for all complexes. According to the chemical structure of the small compounds, the presence of the benzene rings contributed

significantly to the VDW bonding via hydrophobicity [34]. The binding energy of DB04047 could be diminished by the presence of the large, highly electronegative phosphate chains thereby introducing the possibility of steric hindrance [35] and conformational change. **Figure 5** denotes the position of the major catalytic site residues in relation to the drug hit of interest.

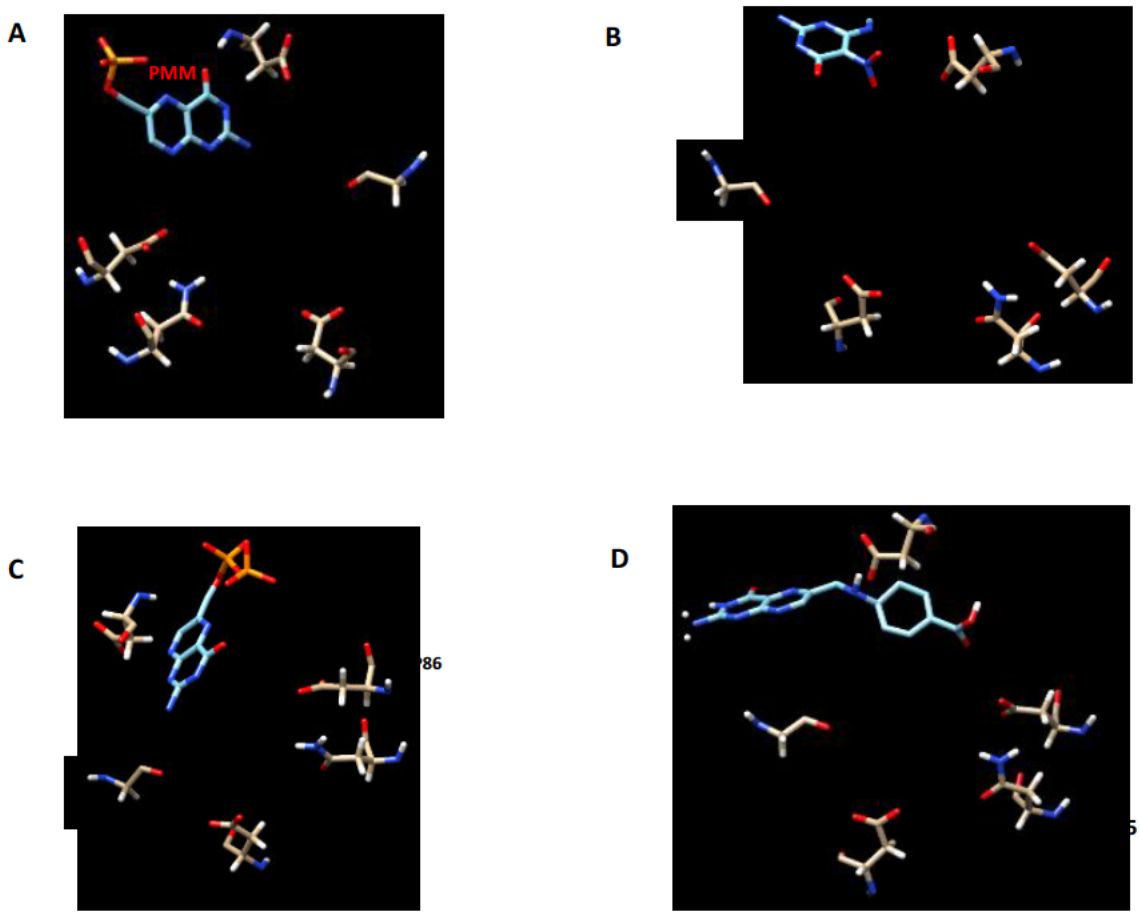


Figure 5: Binding site amino acids interacting with A) DB03592 (PMM), B) DB03705 (PredDrug1), C) DB04047 (PredDrug2) and D) DB04196(PredDrug3).

Electrostatic interactions form the bulk contribution to energetics

Per residue energy decomposition was carried to determine the individual contribution of amino acids to binding free energy. **Figure 6** shows the intermolecular ligand-residue interactions in the active site and the per residue energy decomposition for all drug complexes. The contributions of the van der Waals and electrostatic interactions of the individual residues

correlate with the data obtained in the per residue energy decomposition. All complexes indicated both hydrophobic and hydrogen bond interactions. Hydrogen bonding provides insight into the strength of a complex indicating a directly proportional relationship between the number of hydrogen bonds present and the strength and stability of the protein-ligand bound complex and hydrophobic interactions provide insight into the bonding of non-polar cyclic entities. **Table 2** shows further analysis into the hydrogen bond analysis of each complex. Valid hydrogen bonds fall within a benchmark value of 3 Å indicating the presence 3 in the DHPS-PMM complex, 2 in the DHPS-Pred1 complex, 3 in the DHPS-Pred2 complex and none present in the DHPS-Pred3 complex. All hydrogen bonds seemed to be established before the 50 ns mark of the trajectory.

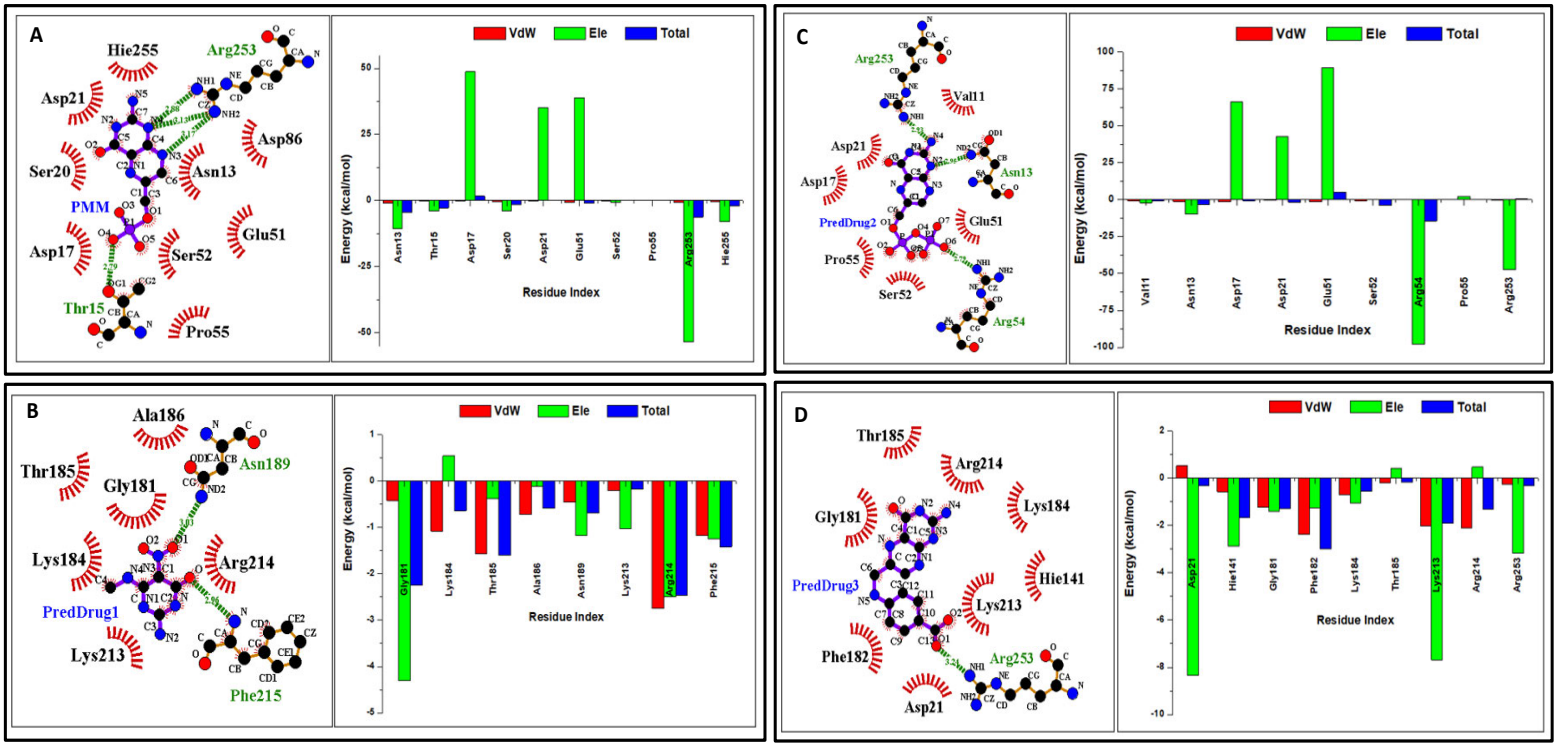


Figure 6: Interaction profile and per residue energy decomposition of A) DB03592 (PMM) B) DB03705 (PredDrug1) C) DB04047 (PredDrug2) D) DB04196 (PredDrug3).

Table 2: Hydrogen bond analysis of folP1 complexed to DB03592 (PMM), DB03705 (PredDrug1), DB04047 (PredDrug2) and DB04196 (PredDrug3).

Acceptor	Donor H	Donor	Frame	Fraction	Distance	Angle
PMM						
UNK_281@N4	ARG_253@HH22	ARG_253@NH2	8	0.0000	2.9260	146.1043
UNK_281@N4	ARG_253@HH11	ARG_253@NH1	112	0.0006	2.9405	144.2657
UNK_281@N3	ARG_253@HH11	ARG_253@NH2	6	0.0000	2.9490	151.2001
Pred1						
UNK_281@O1	ASN_189@HD22	ASN_189@ND2	2278	0.0114	2.8978	143.4333
UNK_281@O	PHE_215@H	PHE_215@N	47524	0.2376	2.8530	151.0553
Pred2						
UNL_281@N4	ARG_54@HH11	ARG_54@NH1	1	0.0000	2.9142	138.2943
ASN_13@ND2	UNL_281@HN2	UNL_281@N4	465	0.0023	2.9382	145.7162
UNL_281@O6	ARG_54@HH12	ARG_54@NH1	6061	0.0303	2.7895	153.7103

The interaction profiles of the standard drugs and potential inhibitors are graphically observed in **Figure 6.** Hydrogen bonds observed stipulated 4 bonds in the PMM complex, 2 in the DB03705 complex, 3 in the DB04047 complex and 1 in the DB04196 complex. The ascending order of the hydrogen bond analysis does not correspond to the ascending binding energy order hence, binding energies of the complexes are therefore largely attributed to the hydrophobic interactions present in the catalytic site. Further analysis of interactions and residue contributions showed the highest contribution occurring because of electrostatic interactions for the control, DB04047 and DB04196 while DB03705 showed an approximately similar contribution occurring from the establishment of van der Waals forces. The side chains of the predicted drugs cause the generation of the electrostatic interactions while the ring formation governs the hydrophobicity. Most of the electrostatic interactions stem from the residues Asp17, Asp21 and Glu51 for the PMM and DB04047 while Gly181 dominated DB03705 and Lys213 DB04196. The highest binding energy is seen to occur in DB04196. With only 1

hydrogen bond formed, it can be established that the interactions from per residue energy decomposition occur largely from the presence of van der Waals forces because of the presence of four 6-membered rings in the small molecule's chemical structure. Due to this selection of potent inhibitors, both the MM/PBSA and energy decomposition label DB04196 as a promising drug candidate.

Positive permeation and absorption

All small molecule drug hits were subjected to ADME predictions to estimate their potential bioavailability, the parameters observed were namely size, lipophilicity, polarity, solubility, saturation, and flexibility. These estimated predictions are essential in the approach of quality drug design and approval.

Table 3: Physicochemical parameters for the small molecules.

Parameters	***PMM***	***DB03705***	***DB04047***	***DB04196***
Bioavailability score	0.11	0.55	0.11	0.56
Molecular weight (g/mol)	273.15	184.14	353.12	312.28
Flexibility: #rotatable bonds	3	2	5	4
Hydrogen bond donors	4	3	5	4
Hydrogen bond acceptors	8	4	11	6
Lipophilicity (MLOGP)	-2.62	-1.53	-3.34	-1.39
Polarity : TPSA ($Å^2$)	174.12	129.62	230.46	146.88
Insolubility: Log S (ESOL)	0.12	-0.88	0.73	-2.05
Insaturation: Fraction Csp^3	0.14	0.20	0.14	0.07
Lipinski violations	0	0	0	0

ADME predictions estimate drug behaviour within an organism according to its nature. Approval of drug candidates is dependent on the physiochemical and pharmacokinetics. **Table 3** and **Figure 7** show the actual data and bioavailability radar of the evaluated parameters with the optimum physicochemical range for these parameters being; size: 150 - 500 g/mol, lipophilicity (XLOGP3): -0.7 - +5.0, polarity (TPSA): 20 - 130 $Å^2$, insolubility (Log S): 0 - 6, unsaturation (Fraction Csp^3): 0.25 – 1, number of hydrogen atom acceptors ≤10, number of hydrogen bond donors ≤5 and for flexibility: number of rotatable bonds should be less than 9

[30]. **Tables 2 and 3** both present PMM, DB03705, DB04047 and DB04196 results evaluated using SwissADME utilizing Lipinski's rule of five stating that any orally active drug cannot have more than one violation. Positive data indicates positive permeation and absorption for the suggested drugs.

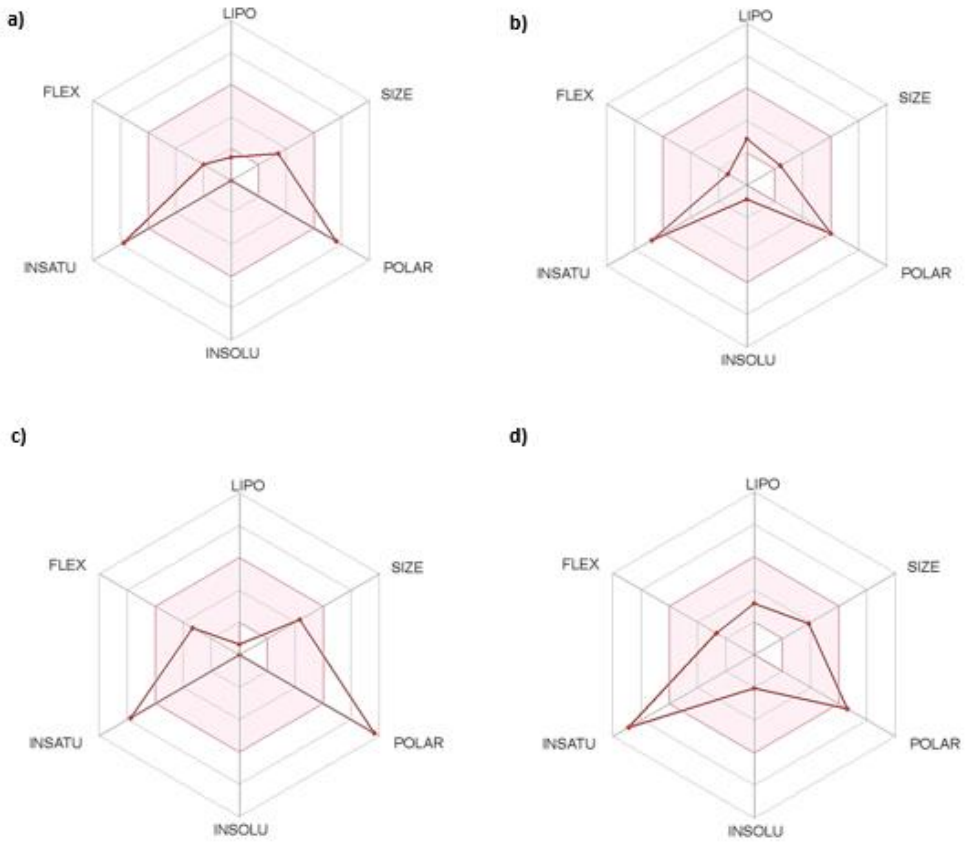


Figure 7: ADME predictions of small molecules a) PMM b) DB03705 c) DB04047 d) DB04196.

Table 4: Pharmacokinetic parameters for the small molecules.

Parameters	***PMM***	***DB03705***	***DB04047***	***DB04196***
Gastrointestinal absorption	Low	High	Low	Low
BBB permeant	No	No	No	No
P-glycoprotein substrate	No	No	No	No
CYP1A2 inhibitor	No	No	No	No
CYP2C19 inhibitor	No	No	No	No
CYP2C9 inhibitor	No	No	No	No
CYP2D6 inhibitor	No	No	No	No
CYP3A4 inhibitor	No	No	No	No
Log k_p (skin permeant) (cm/s)	-10.07	-7.79	-11.57	-8.18

The results presented in **Table 4** indicate that PMM and all the drug hits have relatively low human gastrointestinal absorption, were not substrates of P-glycoprotein, and didn't have blood-brain barrier permeability. The CYP enzyme drug metabolism exhibited negative inhibition.

Conclusion

The emergence and evolution of drug-resistant strains of *M. tuberculosis* pose a mediation in the world of novel drug design and discovery. A Gene ontology network-based inference approach (RepTB) developed by *Passi et al.* utilizing drug-target interactions is adopted in this study to evaluate the validity of the method. From previous literature, it has been established that *folP1* is the major contributor to the folate biosynthesis pathway resulting in the overall enzymatic functionality of *Mycobacterium tuberculosis*, making it an important drug target of interest. Drug repurposing offers an efficient way of in-silico method analysis for the inhibitory potential of small molecules. In this study, four drugs suggested via the RepTB approach, and control was docked into the drug target of interest. Complexes were then subjected to 200 ns molecular dynamic (MD) simulations and post-MD analysis to evaluate for system stability and inhibitory potential of each complex. MM/PBSA was used to calculate the binding energy of each complex. The highest binding energy was seen for complexes DHPS-DB03705 and DHPS-DB04196, both higher than the control, DHPS-PMM. The residue interaction data attributed the binding energy to electrostatic and hydrophobic interactions. The RMSD analysis indicated stable systems all round. The RMSF of the complexes indicated a decrease in structural flexibility for DHPS-DB04047 and DHPS-DB04196. Hence, DHPS-DB04196 showed the most promising set of results. Evaluation of the gene ontology RepTB method proved to be beneficial exhibiting positive responses to the hits and offering the best drug-target interactions. Further usage and development of this method is thereby recommended.

Conflicts of interests

The authors declare no intellectual or financial conflicts of interest.

Acknowledgements

The authors acknowledge the College of Health Sciences, School of Laboratory Medicine and Medical sciences and The Centre for High Performance Computing (www.chpc.ac.za) Cape Town, South Africa, for technical computational support, respectively.

References

1. Du Plessis, L., Škunca, N. & Dessimoz, C. The what, where, how and why of gene ontology—a primer for bioinformaticians. *Brief. Bioinform.* **12**, 723–735 (2011).

2. Botstein, D. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29 (2000).

3. Passi, A., Rajput, N. K., Wild, D. J. & Bhardwaj, A. RepTB: a gene ontology based drug repurposing approach for tuberculosis. *J. Cheminform.* **10**, 1–12 (2018).

4. Russell, D. G., Barry 3rd, C. E. & Flynn, J. L. Tuberculosis: what we don't know can, and does, hurt us. *Science (80-.).* **328**, 852–856 (2010).

5. Gengenbacher, M., Xu, T., Niyomrattanakit, P., Spraggon, G. & Dick, T. Biochemical and structural characterization of the putative dihydropteroate synthase ortholog Rv1207 of *Mycobacterium tuberculosis*. *FEMS Microbiol. Lett.* **287**, 128–135 (2008).

6. Baca, A. M., Sirawaraporn, R., Turley, S., Sirawaraporn, W. & Hol, W. G. J. Crystal structure of *Mycobacterium tuberculosis* 6-hydroxymethyl-7,8-dihydropteroate synthase in complex with pterin monophosphate: New insight into the enzymatic mechanism and sulfa-drug action. *J. Mol. Biol.* **302**, 1193–1212 (2000).

7. de Castro Spadari, C., Vila, T., de Morais Barroso, V. & Ishida, K. New Targets for the Development of Antifungal Agents. *Encycl. Mycol.* 456–467 (2021) doi:10.1016/b978-0-12-809633-8.21026-1.

8. Zheng, J. *et al.* Para-aminosalicylic acid is a prodrug targeting dihydrofolate reductase in *mycobacterium tuberculosis*. *J. Biol. Chem.* **288**, 23447–23456 (2013).

9. Brogi, S., Ramalho, T. C., Kuca, K., Medina-Franco, J. L. & Valko, M. In silico methods for drug design and discovery. *Frontiers in chemistry* vol. 8 612 (2020).

10. Hodos, R. A., Kidd, B. A., Shameer, K., Readhead, B. P. & Dudley, J. T. In silico methods for drug repurposing and pharmacology. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **8**, 186–210 (2016).

11. Ndagi, U., Mhlongo, N. N. & Soliman, M. E. The impact of Thr91 mutation on c-Src resistance to UM-164: molecular dynamics study revealed a new opportunity for drug design. *Mol. Biosyst.* **13**, 1157–1171 (2017).

12. Burley, S. K. *et al.* RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49**, D437–D451 (2021).

13. Kemmish, H., Fasnacht, M. & Yan, L. Fully automated antibody structure prediction using BIOVIA tools: Validation study. *PLoS One* **12**, e0177923 (2017).

14. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

15. Rayan, B. & Rayan, A. Avogadro Program for Chemistry Education: To What Extent can Molecular Visualization and Three-dimensional Simulations Enhance Meaningful Chemistry Learning? *World J. Chem. Educ.* **5**, 136–141 (2017).

16. Huang, C. C., Meng, E. C., Morris, J. H., Pettersen, E. F. & Ferrin, T. E. Enhancing UCSF Chimera through web services. *Nucleic Acids Res.* **42**, 478–484 (2014).

17. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

18. Allouche, A. Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares. *J. Comput. Chem.* **32**, 174–182 (2012).

19. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

20. Berendsen, H. J. C. Molecular Dynamics Simulations Methods and Applications. 1–23 (2004).

21. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. 20035_Ftp. *J. Comput. Chem.* **56531**, 1157–1174 (2004).

22. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).

23. Moustafa, S. G., Schultz, A. J. & Kofke, D. A. Effects of thermostatting in molecular dynamics on anharmonic properties of crystals: Application to fcc Al at high pressure and temperature. *J. Chem. Phys.* **149**, (2018).

24. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

25. Yang, Z., Wu, F., Yuan, X., Zhang, L. & Zhang, S. Novel binding patterns between ganoderic acids and neuraminidase: Insights from docking, molecular dynamics and MM/PBSA studies. *J. Mol. Graph. Model.* **65**, 27–34 (2016).

26. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B. & Huo, S. H. L Chong, M Lee, T Lee, Y Duan, W Wang, O Donini, P Cieplak, J Srinivasan, DA Case, TE Cheatham. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* **33**, 889–897 (2000).

27. Wang, C., Greene, D., Xiao, L., Qi, R. & Luo, R. Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* **4**, 1–18 (2018).

28. Wan, S., Bhati, A. P., Zasada, S. J. & Coveney, P. V. Rapid, accurate, precise and reproducible ligand-protein binding free energy prediction: Binding free energy prediction. *Interface Focus* **10**, (2020).

29. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **8**, 127–134 (1995).

30. Daina, A., Michielin, O. & Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 1–13 (2017).

31. Kirchmair, J., Markt, P., Distinto, S., Wolber, G. & Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes? *J. Comput. Aided. Mol. Des.* **22**, 213–228 (2008).

32. Fukutani, T., Miyazawa, K., Iwata, S. & Satoh, H. G-RMSD: Root mean square deviation based method for three-dimensional molecular similarity determination. *Bull. Chem. Soc. Jpn.* **94**, 655–665 (2021).

33. Sheffield, L. E. Molecular Dynamics Simulations Provide Insight into Stability of Hyperthermophilic Endoglucanases. (2021).

34. Graziano, G. Hydrophobicity of benzene. *Biophys. Chem.* **82**, 69–79 (1999).

35. Ingold, C. K. Quantitative study of steric hindrance. *Q. Rev. Chem. Soc.* **11**, 1–14 (1957).

CHAPTER 7

CONCLUSION

Tuberculosis (TB) remains a major contributor to worldwide deaths despite the avid research studies aimed towards combating it. The functionality and pathways of *M. tuberculosis* have been extensively studied over the years providing a basis for present and future research. The emergence of multiple drug resistant strains has threatened to offset the existing progress provided by previous treatments. The growing ineffectiveness can be attributed to the evolution of mutated residues occurring as a result of patient noncompliance. Computer aided drug design has proven to offer the chance to further the drug discovery process by alleviating the financial and time constraint pressure associated with traditional drug design. This need has given rise to the enhanced generation and improvement of in silico methods inclusive of molecular modelling. The technique is quickly gaining popularity as a cornerstone in the drug discovery industry.

This study was aimed at identifying specific drug targets of interest directly involved in two major *M. tuberculosis* pathways: mycolic acid production and folate synthesis. The three key enzymes: inhA, KasA and folP1 were investigated to determine the effectiveness of suggested drugs against protein functionality. Suggested drug hits were identified using a RepTB gene ontology-network based inference approach determined by *Passi et al.*, and a tailored pharmacophore-virtual screening approach. Structural integrity and inhibition potential were measured by means of correlations developed between stability, flexibility, structural dynamics, binding free energies, decomposition energies, conformations and intermolecular identities of receptor-ligand complexes. *Via* the in depth understanding of the proteins and small molecules, the identification of multiple potent inhibitors, bypassing the data provided by standard drugs, was seen to provide a successful outcome paving a future towards the hopeful extinction of drug resistance. This study thereby exhibits favourable results identifying not only the possibility of potentially potent small molecule drug hits, but also the atomic level understanding of enzymatic and complex functions and overall cementing the benefits of utilizing bioinformatics as a major drug discovery tool in the fight against multi drug resistant *Mycobacterium tuberculosis*.

Future perspectives

Extensive computational methods have proven to provide various insight into the overall workings of complexes. It exists as an ever-improving, cost effective and time efficient approach in obtaining potent inhibitors. Implementation of the techniques provided in this overall study can aid in the future identification and optimization of potential drug hits against many drug targets of interest.

The utilized in-silico studies have generated positive and successful results with stable complexes exhibiting highly favourable intermolecular interactions and could be further validated by future biological testing for drug validation.