Therapeutic Strategy to end Tuberculosis (TB) World: Structural and Functional Characterization of Potential Weak Hotspots of *Mycobacterium tuberculosis* Molecular Targets from Combinatorial *in silico* Perspective

Adeniyi Thompson, ADEWUMI

217079070

A thesis submitted to the College of Health Sciences, University of KwaZulu-Natal, Westville, in fulfilment of the requirements of the degree of Doctor of Philosophy

2021

Therapeutic Strategy to end Tuberculosis (TB) World: Structural and Functional Characterization of Potential Weak Hotspots of *Mycobacterium tuberculosis* Molecular Targets from Combinatorial *in silico* Perspective

Adeniyi Thompson, ADEWUMI
217079070
2021

A thesis submitted to the School of Health Sciences, University of Kwa-Zulu Natal, Westville, for the degree of Doctor of Philosophy

The chapters in this thesis are written as a set of discrete research publications with a general introduction and final summary. In most cases, these chapters were published in reputable, internationally recognized, peer-reviewed journals.

To certify that the contents of this thesis are the original research work of **Adeniyi Thompson, ADEWUMI**.

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

Supervisor:

Professor Mahmoud Soliman

Signed:

PREFACE

The thesis comprises of seven chapters, which includes:

Chapter 1

This chapter precisely highlights the background, rationale, and relevance of the study and further stated the intended aims and objectives. Lastly, the chapter was concluded with the general outline and structure of the thesis.

Chapter 2

The provision of a broad overview of the global impact of drug resistance tuberculosis (TB) morbidity and mortality and effectiveness and limitations of World Health Organization control plan (End TB strategy) are included in this chapter. Additional pieces of literature contain the cascade of *Mycobacterium tuberculosis* transmission cycle, prevention/vaccine, diagnosis, the therapeutic intervention of drug-resistance TB. Also discussed are the critical roles of the study drug targets of the Mtb. The targets include the demethylmenaquinone methyltransferase, N-acetylglucosamine-6-phosphate deacetylase, and secreted antigen proteins.

Chapter 3

This chapter dispenses a concise introduction into the computational chemistry, integrated molecular dynamics simulation protocols, and their applications in drug design and discovery process. The *in-silico* techniques theoretically discussed in this chapter were applied in this research work.

Chapter 4 (**Published work**): The chapter is presented in the required format of the Journal that published the full research article. Hence it is the final version of the accepted manuscript.

This chapter provides the dynamic structural and functional on the *NagA* due to its crucial biosynthesis role within the cell wall of *Mycobacterium tuberculosis*. The chapter also includes homology protocol amalgamated for the creation of the predicted structures of *NagA* enzyme homologs. Comparative sequence analysis was performed among the *NagA* of *smegmatis*, *marinum*, *leprae*, *ulcerans*, *africanum*, and *microti*. Mutation impact on the active site of then *NagA* was explored. The loop dynamics of *NagA* was investigated in this chapter to gain cognizance into the biological activity of the protein. This article has been published in The Protein Journal (IF = 1.317).

Chapter 5 (**Published work**): This chapter is presented in the required format of the Journal that published the full research article. Hence it is the final version of the accepted manuscript).

The chapter includes homology protocol integrated for the creation of the predicted structures of the methyltransferase family exhibited within the *Mycobacterium tuberculosis menG*. The study intimates the effectiveness of DG70 compound as a potential inhibitor of *menG* enzymes. Thermodynamics analysis incorporating per-residue energy decomposition were calculated to underline the critical amino acid residues of *menG* that are involved in binding and. Besides, the loop dynamics of *menG* was investigated to gain cognizance into the biological activity of the protein. This article has been published in the RSC Advances (IF = 3.070).

Chapter 6 (**Published work**): This chapter is presented in the required format of the Journal that published the full research article. Hence it is the final version of the accepted manuscript).

The chapter includes the creation of 3D Structures of the homologs of the *Mycobacterium tuberculosis* Ag85C enzyme using homology modelling protocol. Comparative multiple

sequence analysis was performed among the Ag85C protein of *tuberculosis*, *smegmatis*, *marinum*, *leprae*, *ulcerans*. Chapter 6 also unravels the structural dynamics and functional characterization of Ag85C protein. Thermodynamics analysis incorporating per-residue energy decomposition were calculated to underline the critical amino acid residues of Ag85C that are involved in the binding. Again, the loop dynamics of *menG* was investigated to gain cognizance into the biological activity of the protein. The loop dynamics of *NagA* was investigated in this chapter to gain cognizance into the biological activity of the protein. The article has been published in the RSC Advances (IF = 3.220)

Chapter 7

This chapter connotes the last chapter, where concluding remarks coupled with future works and perspectives proposed.

ABSTRACT

The world has witnessed several decades of Tuberculosis (TB) pandemic and numerous advanced scientific efforts to control the invasiveness of the newly evolving *Mycobacterium tuberculosis strains* (Mtb) resulting in drug resistance. TB disease has killed hundreds of millions of humans and left millions maimed that need to be rehabilitated; about 10.0 million infections and 1.5 million annually in the last decade. Drug-resistant TB has remained more challenging in the previous 20 years than drug-susceptible TB and is chromosomal mutations-associated in selected genes of the Mtb. Notable mutations identified by biomarkers are related to phenotypic drug resistance, and these include; an 81 bp region in *rpoB* gene with > 95 % mutations in rifampicin (RIF) clinical isolates and katG gene promoter of the *mabA*-inhA showed to be associated with INH-resistance. Different Strategies, including the recent WHO End TB approach, have been employed to alleviate or stop TB. The current identification of the critical roles of Mtb demethylmenaquinone methyltransferase (*menG*) target in the survival, pathogenesis, virulence, and drug resistance created an avenue for the development of efficacious therapeutics that can eradicate TB. *MenG* is a member of the methyltransferase superfamily. It catalyzes one of the last steps of the menaquinone biosynthesis pathway, requires for maintenance of the *Mtb* cell envelope. The other two studied targets investigated in this work are N-acetylglucosamine-6-phosphate deacetylase enzyme (*NagA*), which represents a critical enzymatic step in the production of essential amino sugar required by *Mtb* for the cell wall biosynthesis and the secreted antigen

85C enzyme (Ag85C) target. The latter target catalyzes the synthesis of trehalose derivatives and attachment of mycolic acids. These targets have gained considerable attention in drug discovery pipelines. However, there is little information about *menG*, as it lasks structural dynamics due to the lack of crystal structure, active site regions, and amino acids of it Mycobacteria homologs. Similarly, the dynamics of the *NagA* and Ag85C proteins structure are still unknown. Therefore, justifications led to the modelling of the 3D Structure of *menG* to understand the structural and functional features that could be investigated at the atomistic level. Homology models were also created for the five (5) mycobacterial homologs. Furthermore, the inevitable need for new drugs has led to the application of *in silico* techniques including molecular modelling and molecular dynamics simulations, which provide opportunities for the chemists to evaluate and assess numerous compounds that can lead to potential drugs against the mycobacterial disease. Furthermore, these computational techniques justify the present incorporation of several computational tools integrated into this study to provide insights into the conformational changes that illuminate potential inhibitory mechanism, identification of the binding site amino acids, and characterization. Here, we analyze the weak hotspots dynamics specific to each of the *Mtb* targets, most notably the loop and active residues around or within the ligand-binding sites to obtain useful findings for the design of higher efficacious potential antitubercular drugs. Molecular dynamics simulations were performed to gain molecular standpoints of the conformational binding of the experimental drugs, which were reported to be highly effective against each respective target. Structural dynamics and motions behaviour of *menG* upon the binding of inhibitor (DG70, biphenyl amide compound) were estimated. Additional *in silico* thermodynamic analyses were further employed to explore intuitions into the binding mode of each inhibitor mainly for the proposed binding site of *menG* to identify the residues for binding. Sequence analysis of the homologs of *Mycobacterium tuberculosis NagA* and Ag85C targets, including those of

smegmatis, *marinum*, *leprae*, *ulcerans*, were performed to obtain unique sequence similarities and differences and the structural and functional characterization upon the binding of the ligand. An experimental protocol let to the discovery of a selective covalent inhibitor, β-isomer monocyclic enolphosphorus Cycliphostin, of Ag85C SER-124. Moreover, chapter 4 also unravels the impact of the function of the non-synonymous single nucleotide polymorphisms of *NagA* target. The desired expectation is that the implementation of the information extricated from this study would provide the structural silhouette for pharmaceutical scientists and molecular biologists to abet in the identification and design of novel antimycobacterial drugs most especially for TB.

DECLARATION I: PLAGIARISM

I, Adeniyi Thompson ADEWUMI, declares that:

1. Research findings reported here in this thesis are originally mine, except where otherwise indicated.

2. We have not submitted this thesis for any degree or examination at any other university or institution of learning.

3. No other person's data, images, graphs, annotations or information is contained in this thesis unless explicitly acknowledged as being sourced from other persons.

4. This thesis contains no other person's writing unless expressly acknowledged as being sourced from other researchers. Moreover, in cases where we quoted other written sources, then:

a. their words have been re-constructed while the general information attributed to them appropriately referenced.

b. in places where their exact words used, their writings were italicized and quoted while also appropriately referenced.

No text, graphics or tables directly copied from the internet are pasted in this thesis unless explicitly acknowledged and sources appropriately indicated and referenced.

The detailed list of contributions to publications including those partly formed this thesis stated appropriately is presented.

Signed: ----------

DECLARATION II: LIST OF PUBLICATIONS

RESEARCH OUTPUTS

1 **Adeniyi Thompson Adewumi**, Pritika Ramharack, Opeyemi S. Soremekun, and Mahmoud E. S. Soliman (2020). Delving into the Characteristic Features of "Menace" *Mycobacterium tuberculosis* Homologs: A Structural Proteomics Perspective. *The Protein J.* **39**, 118–132.

Contribution:

Adeniyi Thompson Adewumi carried out technical and all the experimental methods coupled with the manuscript writing.

Pritika Ramharack: contributed to the technical and writing of the manuscript.

Opeyemi S. Soremekun: contributed by performing mutation analysis and writing.

Mahmoud E. S. Soliman: Supervisor

Appendix A: Pdf version of the publication

2 **Adeniyi Thompson Adewumi**, Opeyemi S. Soremekun, Mary B. Ajadi, and Mahmoud E. S. Soliman (2020). Thompson Loop: Opportunities for Antitubercular drug design by Targeting the Weak Spot in Demethylmenaquinone Methyltransferase protein. *RSC Advances*, **10**, 23466-23483.

Contribution:

Adeniyi Thompson Adewumi: contributed to the project by performing all the experimental work and manuscript preparation and writing.

Opeyemi S. Soremekun: contributed by reviewing the manuscript

Mary B. Ajadi: contributed by reviewing the manuscript

Mahmoud E. S. Soliman: Supervisor

Appendix B: Pdf version of the publication

3 **Adeniyi Thompson Adewumi**, Opeyemi S. Soremekun, Mary B. Ajadi, and Mahmoud E. S. Soliman (2020). Weak spots inhibition in the *Mycobacterium tuberculosis* antigen 85C target for Antitubercular drug design through selective irreversible covalent inhibitor-SER124. *Journal of Biomolecular Structure and Dynamics*,

Contribution:

Adeniyi Thompson Adewumi: contributed by performing technical and experimental work. Also, execute the manuscript preparation and writing.

Ahmed Elrashedy: contributed by setting up the covalent system and entropy calculation

Opeyemi S. Soremekun: contributed to the project by plotting the DCC matrix.

Mary B. Ajadi: contributed by performing technical work.

Mahmoud E. S. Soliman: Supervisor

Appendix C: Pdf version of the publication

4 Mary Ajadi, Opeyemi Soremekun, **Adeniyi T. Adewumi**, Hezekiel Kumalo and Mahmoud Soliman (**2020**), Leveraging on Active Site Similarities; Identification of Potential Inhibitors of Zinc-Finger and UFSP domain Protein (ZUFSP), *Current Pharmaceutical Biotechnology*

***Contribution**:*

Adeniyi T. Adewumi: contributed by performing technical work and reviewing of project

Mary Ajadi: contributed by performing technical and experimental work. Also prepared and wrote the manuscript.

Opeyemi S. Soremekun: contributed by performing technical work and reviewing of the study.

Hezekiel Kumalo: co-supervision

Mahmoud E. S. Soliman: Supervisor

Appendix B: Pdf version of the publication

5 Opeyemi Soremekun, Kehinde F. Omolabi, **Adeniyi T. Adewumi**, and Mahmoud Soliman (2020), Exploring the Effect of Ritonavir on SARS-CoV-2 and SARS-CoV Main Proteases – A Potential from Molecular Perspective, *Future Science OA*.

***Contribution**:*

Adeniyi T. Adewumi: contributed by performing part of the experimental work and writing the manuscript chapter one (Background, justification, aim, and objectives)

Opeyemi Soremekun: contributed by completing part of experimental work and technical

Kehinde F. Omolabi: contributed by performing technical and manuscript writing

Mahmoud E. S. Soliman: Supervisor

Appendix C: Pdf version of the publication

6 Mary Ajadi, Opeyemi Soremekun, **Adeniyi T. Adewumi**, Hezekiel Kumalo and Mahmoud Soliman (**2020**). Functional Analysis of Single Nucleotide Polymorphism in ZUFSP protein and Implication in Pathogenesis, *The Protein Journal* (manscript under peer-review)

Contribution:

Adeniyi T. Adewumi: contributed by performing technical and experimental work and review of the manuscript.

Mary Ajadi: contributed by performing technical and experimental work. Also prepared and wrote the manuscript.

Opeyemi Soremekun: contributed by performing technical work and reviewing of the study

Hezekiel Kumalo: Co-supervision

Mahmoud E.S Soliman: Supervisor

7 Ashish Kumar, **Adeniyi Thompson Adewumi**, K. P. Nandhini, Jonathan M. Collins, Fernando Albericio, and Beatriz G. de la Torre (2019). Troubleshooting When Using γ-Valerolactone (GVL) in Green Solid-Phase Peptide Synthesis. *Organic Process Research & Development*, **23**, 1096-1100.

Contribution:

Adeniyi Thompson Adewumi: contributed by performing technical and experimental work and writing part of the manuscript.

Ashish Kumar: contributed by performing part of experimental and manuscript writing

K. P. Nandhini: Technical contribution

M. Collins: Technical contribution

Fernando Albericio: Co-supervisor

Beatriz G. de la Torre: Supervisor

8 Mary Ajadi, Opeyemi Soremekun, **Adeniyi T. Adewumi**, Hezekiel Kumalo and Mahmoud Soliman (**2021**). Probing protein=protein interactions and druggable site identification: Mechanistic between Ubiquitin and Zinc finger with UFM1-specific peptidase domain protein (ZUFSP), *Chemistry & High Throughput Screening*- Accepted (**Impact Factor: 1.205)**

Contribution:

Adeniyi T. Adewumi: contributed by performing technical and review of the manuscript.

Mary Ajadi: contributed by performing technical and experimental work. Also prepared and wrote the manuscript.

Opeyemi Soremekun: contributed by performing technical work and reviewing of the study

Hezekiel Kumalo: Co-supervision

Mahmoud E.S Soliman: Supervisor

RESEARCH OUTPUT

A- List of Publications

1 Adeniyi Thompson Adewumi, Pritika Ramharack, Opeyemi S. Soremekun, and Mahmoud E. S. Soliman (2020). Delving into the Characteristic Features of "Menace" *Mycobacterium tuberculosis* Homologs: A Structural Proteomics Perspective. *The Protein J*. **39**, 118–132

2 Adeniyi Thompson Adewumi, Opeyemi S. Soremekun, Mary B. Ajadi, and Mahmoud E. S. Soliman (2020). Thompson Loop: Opportunities for Antitubercular drug design by Targeting the Weak Spot in Demethylmenaquinone Methyltransferase protein. *RSC Advances*, **10**, 23466-23483.

3 Adeniyi Thompson Adewumi, Opeyemi S. Soremekun, Mary B. Ajadi, and Mahmoud E. S. Soliman (2020). Weak spots inhibition in the *Mycobacterium tuberculosis* antigen 85C target for Antitubercular drug design through selective irreversible covalent inhibitor-SER124. *Journal of Biomolecular Structure and Dynamics* 1-21

B- Conferences

Poster Presentation of Delving into the Characteristic Features of "Menace" *Mycobacterium tuberculosis* Homologs: A Structural Proteomics Perspective- 13th National CHPC Conference, Sandton, Johannesburg, South Africa, 1-5 December 2019.

C- Contribution as a Co-Author

1 Mary Ajadi, Opeyemi Soremekun, **Adeniyi T. Adewumi**, Hezekiel Kumalo and Mahmoud Soliman (**2020**), Leveraging on Active Site Similarities; Identification of Potential Inhibitors of Zinc-Finger and UFSP domain Protein (ZUFSP), *Current Pharmaceutical Biotechnology*

2 Opeyemi Soremekun, Kehinde F. Omolabi, **Adeniyi T. Adewumi**, and Mahmoud Soliman (2020), Exploring the Effect of Ritonavir on SARS-CoV-2 and SARS-CoV Main Proteases – A Potential from Molecular Perspective, *Future Science OA*

3 Mary Ajadi, Opeyemi Soremekun, **Adeniyi T. Adewumi**, Hezekiel Kumalo and Mahmoud Soliman (**2020**). Functional Analysis of Single Nucleotide Polymorphism in ZUFSP protein and Implication in Pathogenesis, *The Protein Journal* (manscript under peer-review)

4 Ashish Kumar, **Adeniyi Thompson Adewumi**, K. P. Nandhini, Jonathan M. Collins, Fernando Albericio, and Beatriz G. de la Torre (2019). Troubleshooting When Using γ-Valerolactone (GVL) in Green Solid-Phase Peptide Synthesis. *Organic Process Research & Development*, **23**, 1096-1100.

5 Mary Ajadi, Opeyemi Soremekun, **Adeniyi T. Adewumi**, Hezekiel Kumalo and Mahmoud Soliman (**2021**). Probing protein=protein interactions and druggable site identification: Mechanistic between Ubiquitin and Zinc finger with UFM1-specific peptidase domain protein (ZUFSP), *Chemistry & High Throughput Screening*- Accepted (**Impact Factor: 1.205)**

6 Mary Ajadi, Opeyemi Soremekun, **Adeniyi T. Adewumi**, Hezekiel Kumalo and Mahmoud Soliman (**2021**). Probing protein=protein interactions and druggable site identification: Mechanistic between Ubiquitin and Zinc finger with UFM1-specific peptidase domain protein (ZUFSP), *Chemistry & High Throughput Screening*- Accepted (**Impact Factor: 1.205)**

ACKNOWLEDGEMENTS

First, I want to return ALL the glory to ALMIGHTY GOD the giver of life and everything thereof (John 3:27) because I could not have achieved this **milestone** if GOD did not release HIS enablement (*Phil. 4:13; "I can do ALL things through Christ that strengthens me"*). Thank you, Lord.

My Parents

I can not hesitate to acknowledge my lovely father, Late Pa Oluwatoyin Zacchaeus Adewumi and loving mother, Ma Bolanle Mabel Adewumi, for your good brought up in my life, love, prayers, and sacrifices. Pa, your wish was always for me to become a Doctor. Thank you so much for all your financial support for my education. I wish to see you again.

My Wife

Proverb 18:22 "He who finds a wife, finds a good thing and obtain mercy from the Lord."I thank God for the day He ordered your mind and steps to that Chemistry laboratory. Thank you so much for your love, affection, care, understanding, and sacrifices. You are the best gift I have ever get and I will always love you. Siyabonga" Standaswa mi". You are my ALL now.

Daughter

Thank you, my lovely daughter, Princess Araotun Nhlelozikathixo, for coming to the world during this memorable year. Thank you for your beautiful smiles that have continued to reassure me of more splendid in present and future.

My Siblings

Thank you, Big sister Sade, Kayode, Mopol Gbenga, and last born Muyiwa. I appreciate your belief in me and encouragement from you ALL. I, specially thank my immediate brother, Kayode Adewumi, I can not forget your sacrifice of love towards me during my first degree.

Extended Family members

Thank you for your love and prayers. Thanks to Uncle Adewumi Omotope Theophilus and his beautiful wife for always remembering me in their prayers and encouragement. I owe Uncle Ajulo Abiodun Lawrence a maximum thanks for his love, encouragement and prayers. He is more than what I call him. I love you, Sir. I must appreciate my South Africa mother, Late Manthwa Rebecca Frans, for your prayers and encouragement during the short time I knew you. She immediately adopted me as her son. You were fantastic. Adieu, Mama!!! Continue to rest at the bosom of God. I will see you again on the last day. My brother, Prince Oladapo Olagbegi and family, thank you.

Prof. M. E. S. Soliman

My uttermost appreciation goes to my supervisor, who stood by me genuinely at the crossroads of this study. Thank you for your fatherly role, company, laughter, friendship, and love.

Friends

Thank you ALL for your kind admonition and love. I appreciate it. The Redeemer's University of Nigeria, My Pastors; Dr Adelabu and family, Dr. Olumide Olukanni, Dr. Emmanuel Unuabonnah, Dr. Martins Omorogie, Alfred Moses, and other, I thank you, all. I

also want to thank my friends Mr. Sunday Abu, Prince Jide Aladesanmi (CANNADA), Prince Jide Jacobs (USA), FOLASADE RUKA (USA), and everyone that immensely contributed to this great achievement by sharing their time, love and encouragement towards me during the study period. I would fair if I forgot to appreciate my church in Durban, Redeemed Christian Chuech of God, Chapel of Praise, especially my father in the Lord and family, Pastor & Pastor (Mrs) Gabriel Adejimi. I am grateful to you all.

The University of KwaZulu-Natal, Durban and the College of Health Sciences.

Thank you for your financial supports

LIST OF ABBREVIATIONS

3D	Three-Dimensional
Å	Angstrom
α	Alpha
ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity
B	Beta
CADD	Computer-Aided Drug Design
DCCM	Dynamic cross correlation matrix
CHPC	Centre for High Performance Computing
GAFF	General Amber Force Field
HM	Homology Modelling
LD	Lethal dose
K	Kelvin
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area

MM/PBSA	Molecular Mechanics/Poisson-Boltzmann Surface Area
MW	molecular weight
MA	*Mycobacterium africanum*
Mtb	*Mycobacterium tuberculosis*
MM	*Mycobacterium marinum*
MS	*Mycobacterium smegmatis*
MU	*Mycobacterium ulcerans*
ns	Naanosecond
PCA	Principal Component Analysis
PDB	Protein Data Bank
ps	picosecond
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
RESP	Restrained Electrostatic Potential
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RoG	Radius of Gyration
RSCB	Research Callaboratory for Structural Bioinformatics
SASA	Solvent Accessible Surface Area
UCSF	University of California, San Francisco
vdW	van der Waals
WHO	World Health Organization
Wt	Wildtype

LIST OF AMINO ACIDS

One Letter code	Three Letter code	Amino acid name
A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartate
C	Cys	Cysteine
E	Glu	Glutamate
Q	Gln	Glutamine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine

K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosin
V	Val	Valine

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1 Introduction

1.1 Background and Rationale of the study

Tuberculosis (TB) has continued to remain the world most health-threatening pandemic. The disease is too complicated, and factors like latency, dormancy and drug resistance in the *Mycobacterium tuberculosis* (*Mtb*) make its treatment and cure challenging. TB infection is still the primary factor of mortality and socioeconomic disaster for millions of people globally[1] as its pathogen has persisted in the host man for more than 50 years. Despite several years of available TB control strategies encompassing effective therapy,[2] diagnostic tools; rehabilitation; preventive efforts; and a global awareness campaign to end its monster causing-organism, this common human enemy continues to succeed. TB has remained the leading cause of human death[3]; hence, it is regarded as the 'crown' cause of death globally, and ranked above HIV/AIDS infections (World Health Organization, WHO: 2019 reports). Specifically, the WHO estimated 10 million TB infection cases and about 1.2 million TB deaths among TB/HIV-negative co-infected persons in 2018; these numbers have remained unchanged in the last few years.[4] According to WHO, prevalence of TB cases is in the trend: South-East Asia (44 %), Africa (24 %), Western Pacific (18 %), Eastern Meditteranean (8 %), Americas (3 %), Europe (3 %). Notably, Nigeria (4 %) and South Africa (3 %) are two countries in Africa, accounting for two-thirds of worldwide's total TB incidences. The TB mortality rate in South Africa (SA) is relatively high in Africa continents because SA is the epicentre of the HIV epidemic globally, with 7.7 million people living with HIV in the country.[5] The association of TB with HIV is a global challenge since it increased patients' mortality rate with HIV.

More challenging about TB is the mutations-mediated drug resistances, which initially occurred as mono-drug resistance but now multidrug and extensively drug resistance TB (MDR/XRD-TB). The most common type of this mutation is the non-synonymous single-nucleotide polymorphisms (nsSNPs). Recently, a form of XRD-TB strains known as the total drug resistant-TB (TDR-TB), first reported in South Africa, was highly resistant.[6] There is a need for a combined radical approach of expanded preventive medication for persons with and without HIV and treatment with improved surveillance to improve global control of TB.

Drug resistance TB has caused several drawbacks to WHO EndTB strategy started in 2017, which compares the combination of bedaquiline and delamanid with current more extended treatment regimens for MDR-TB and XRD-TB. Almost 184 772 new cases of MDR/RR-TB persons were reported in 2018, which reflected 51 % confirmed bacteriological TB infected peoples.[4] TB drug resistances occur due to spontaneous mutation of genes to result in the resistance to antitubercular drugs. However, the most signalled reason for drug resistance is non-compliance with dosage regimens. The rifampicin/multidrug-resistant (MDR/RR-TB) patient treatment lasts for at least 18-24 months of administration of second-line drugs accompanied by counselling and monitoring.[7] Whereas, the standard treatment of TB only spans through 6-months regimens. The late emergence of drug-resistant strains and the continual treatment to eliminate the organism underlies the importance of compliance, and the risk involves early discontinuity of therapy.[8] Hence, non-adherence is very high with consequential poor outcomes, leading to the MDR strains dissemination. There are instances where drug resistance in the pathogens' genes are not traceable to mutations; thus, MDR/XRD TB is a complex phenomenon. Secondly, the success of TB pathogen is due to its complex mycobacterial cell wall comprising of three distinct macromolecules: peptidoglycan (PG), arabinogalactan (AG) and mycolic acids; which underly the slow growth rate of the pathogens and their characteristic thick, greasy, and impermeability nature of the cell

envelope. Also, the TB drugs undermine by adverse drug occurrences, lack of adherence to optimal therapy, and high costs of purchasing drugs also lead to drug resistance.[9] Also aggravate TB pandemic is the co-infection of some disease such as HIV and diabetes. The co-infection of HIV/AIDS or diabetes mellitus with RR/MDR- and XDR-TB also contributes to TB treatment's ineffectiveness, which further has a consequential impact on TB mortality.

Although the recent use of newer frontline drugs has ameliorated the therapeutic intervention of TB, several shortcomings, especially drug resistance, continue to plague their applications. Drug resistance causes arduous treatment and relapse rate to tuberculosis, leading to the disuse and discontinuation of some of these drugs. About one-half of the multidrug or extensive drug resistances have been affected by mutation-mediated or acquired resistance.[10, 11] Isoniazid, linezolid, and bedaquiline (BDQ) are few examples. TB treatment outcome using first-/second-line drugs and the newer drugs is discouraging, with success rates as low as 54 % and 30 % for MDR and XRD, respectively.[12] Drug design involves an interdisciplinary approach to enhance the discovery of more TB drugs because there is still a limited drug in the pipeline. New therapeutic strategies' goals include; ensuring TB cure without relapse; reducing TB mortality, contagions, and the formation of drug-resistant strains; search for low cost, little or no toxic model that interferes with biology *Mtb* protein functions.

Antitubercular therapy approach to developing a new chemical entity with a novel mechanism of action is a promising alternative, which shows significant activity against the *Mtb* strains and the expensive and time-consuming processes of drug discovery and development spans through 10-20 years of intensive laboratory research. Computer-aided drug design (CADD) methods have significantly improved drug discovery and development processes by overcoming some disadvantages with the subsequent possibility of having drug candidates with little or no toxicity. Applications of computational chemistry incorporated into biology provide clear-cut added advantages over the conventional drug discovery

process in a manner that reduces drug development steps by a fraction of the regular timescale. Computational techniques, including molecular modelling, hot spots pharmacophore identification, and Molecular dynamics (MD) simulations, furnish the biologists and medicinal chemists with information to design potent drugs to treat tuberculosis and other bacterial infections. The application of bioinformatics and molecular dynamics simulations methods have proven to alleviate the concept of "shooting in the dark" with experimental screening by reducing the protracted time scale of drug discovery. These methods have been substantially used in protein-ligand and protein-protein interactions. Taken together, identification and selection of critical *Mtb* receptors incorporated with the development of high inhibitory protocols that use low cost, non-toxic, high % potent compounds may contribute to the discovery of future lead compounds against TB.

The molecular and structural insights of three *Mycobacterium tuberculosis* targets; were unravelled in this study. They include the N-acetylglucosamine-6-phosphate deacetylase, demethylmenaquinone methyltransferase, and antigen 85C enzymes. The proteins drive the pathogen's virulence, survival, and drug resistance success against the frontline antitubercular drugs. The targeting of protein roles such as catalysis of sugar production steps, attachment of mycolic acids, and the synthesis of trehalose derivatives in the Mtb cell wall is an attractive approach towards the significant development of vaccines and drugs. Moreover, these findings' outcome is essential in the structure-based design of specific and selective small molecules that can inhibit the *Mtb* receptors, thereby discovering more potent, less toxic, and more specific antitubercular drugs.

1.2 Aims and objectives of the study

This thesis's primary aim unravels the structural dynamics and functional insights of selected crucial targets of *Mycobacterium tuberculosis* involved in the virulence, pathogenesis, and drug resistance using bioinformatics and computational techniques to identify and develop

probable potential inhibitors against *Mtb*. The *Mtb* targets investigated are N-acetylglucosamine-6-phosphate deacetylase (*NagA)*, demethylmenaquinone methyltransferase (*menG*) antigen 85C enzyme.

The specific goals include:

1. To investigate the structural and functional features of *Mtb NagA* protein, which aid the structure-based design of novel chemical entities with potential potent anti-TB activities. The following objectives were considered to achieve this goal:

1.1 prediction of 3D structures of the homologs of *MtbNagA*, which include *MSNagA*, *MANagA, MCNagA, MMNagA*, and *MUNagA* using the homology modelling protocol.

1.2 amino acids sequence analysis of *MtbNagA*, *MSNagA*, *MANagA, MCNagA, MMNagA*, and *MUNagA* using to obtain similarities and differences in their primary structures

1.3 computation of 350 ns MD simulations on *MtbNagA*, *MSNagA*, *MANagA, MCNagA, MMNagA*, and *MUNagA* and estimated post-MD analysis to validate and obtain the stability and flexibility.

1.4 investigation of the critical *Mtb NagA* mutants, including D267A, H56Q, H56N, and H56A on the amino acids using the online mutation resources (*Dynamut* and *Polyphen2*), to analyse the impact of functional of non-synonymous single-nucleotide polymorphisms.

1.5 comparative analysis of both static and dynamic structural conformational changes

2. To provide the structural and dynamic impacts of a biphenyl amide compound named DG70, we docked on demethylmenaquinone methyltransferase (*menG*), using integrated MD simulations. The following objectives were considered to achieve this goal.

2.1 prediction of 3D structures of the *MtbmenG* using the homology modelling protocol.

2.2 docking of DG70 and demethylmenaquinone, DMK9 (*menG* substrate) compounds following the identification of the poses with the most negative docking score and the active site identification-validation of the residues.

2.3 computation of 300 ns molecular dynamics simulations of unbound *menG* and DG70-bound *menG* and post-MD analysis.

2.4 investigation of the structural and functional binding landscape of DG70 against the *menG* protein.

2.5 obtaining the thermodynamic data of the target-inhibitor systems to calculate the binding energy between DG70 compound and *menG*

2.6 obtaining the energy contribution from each interacting residue in the binding analysis based on the MM/GBSA approach to identify the amino acid residues with highest binding free energy contributions.

3. Investigate structural characteristics and motions and dynamics of the *Mtb* target (Ag85C) upon the binding of the selective covalent β-isomer enolphosphorus Cycliphostin ($CyC_{8\beta}$)-bound and $CyC_{8\beta}$ open model conformations.

3.1 prediction of 3D structures of A85C proteins of the mycobacteria, including *smegmatis*, *marinum*, *leprae*, and *ulcerans* using homology modelling protocol

3.2 amino acids sequence analysis of the target Ag85C among the mycobacteria species including; *Mtbs*, *smegmatis*, *marinum*, *leprae*, and *ulcerans* to explore the levels of structural conservation and the unique similarities or differences, which could be vital in future for the identification and development of potential inhibitors against tuberculosis.

3.3 computation of 300 ns molecular dynamics simulations of unbound A85C and $CyC_{8\beta}$-bound A85C and the estimation of post-MD analysis including stability, flexibility, system and compactness.

3.4 investigation of the structural and functional behaviours of A85C protein upon the binding of $CyC_{8\beta}$-compound

3.5 obtaining the thermodynamic terms to determine the binding free energy

3.6 implementation of a PRED of each system based on the MM/GBSA approach.

1.3 Novelty and significance of this study

Although the global average TB incidence rate indicated a decline by 1.6 % per year from 2000 to 2018 and 2.0 % between 2017-2018, new drug-resistant Mtb strains are evolving to develop mutations in every frontline drug. Therefore, TB is still the number one global death cause and the principal cause of an infectious agent. For these reasons, the considerations of critical *Mtb* targets combined with selective inhibitory strategies furnish potential opportunities for some weak hotspots that produce highly effective TB treatment with little or no toxic level. The search for therapeutic targets and their drug design use to effectively treat tuberculosis is a very significant route. The roles of proteins, including *menG*, *NagA*, and Ag85C, are crucial to *Mtb* survival, virulence, and drug resistance attributes. They are involved in the organism's cell envelope's respiration or maintenance, especially *menG*, which has no crystallised structure. The targets are considered a means to end the menace of TB. Moreover, several CADD approaches that enhance experimental methods speed up drug discovery and present the means to overcome conventional therapy's shortcomings.

This research work focuses on investigating unique features and changes in the structures and functions of the conformations of the whole protein and specific loops of the homologs of *Mycobacterium tuberculosis NagA*, *menG*, and Ag85C of *Mycobacterium tuberculosis*. Also,

the homology model of *Mtb menG* was constructed due to the unavailability of the crystal structure. Similarly, the models of the *NagA* and Ag85C proteins of the other mycobacteria, including *smegmatis*, *marinum*, *leprae*, *africanum*, *microti*, and *ulcerans,* were constructed for sequence analysis determination. The study further pinpoints each target's sequential and structural behaviour to their respective enzyme family, thus elucidating the importance of understanding the binding landscape with their inhibitors and providing insights that could help design specific inhibitors to ameliorate drug resistance. The molecular perspectives into the binding of respective experimental inhibitors to the *menG* and Ag85C were provided here, specifically, bindings of DG70 chemotype to *menG* and $CyC_{8\beta}$ compound to Ag85C. These were achieved by characterising the proteins' active sites, analysing the structural and functional conformations, and identifying the absolute amino acids necessary for the binding. The design of the enzymes' plausible binding features proffers an approach to design inhibitors that are selective and distinct with polypharmacological properties, which would enhance the development of effective, stable, and efficacious small molecules inhibitors.

Further to the above *NagA* molecular insights, this study predicted the critical residues' mutation effect, including HIS56 and D267 at the binding site. The *NagA* mutants; D267A, H56N, H56Q, and H56A, have been reported to reduce the NagA protein's catalytic activity. On these notes, this thesis's research contents underlay the fundamental principle towards drug design against tuberculosis.

References

1. Nasiruddin, M., Neyaz, M. K. & Das, S. Nanotechnology-Based Approach in Tuberculosis Treatment. *Tuberc. Res. Treat.* **2017**, 1–12 (2017).
2. Palucci, I. & Delogu, G. Host Directed Therapies for Tuberculosis: Futures Strategies for an Ancient Disease. *Chemotherapy* **63**, 172–180 (2018).
3. Dheda, K. *et al.* Recent controversies about MDR and XDR-TB: Global

implementation of the WHO shorter MDR-TB regimen and bedaquiline for all with MDR-TB? *Respirology* **23**, 36–45 (2018).

4. Kendall, E. A., Malhotra, S., Cook-Scalise, S., Denkinger, C. M. & Dowdy, D. W. Estimating the impact of a novel drug regimen for treatment of tuberculosis: A modeling analysis of projected patient outcomes and epidemiological considerations. *BMC Infect. Dis.* **19**, 1–12 (2019).
5. World Health Organization, (WHO). *GLOBAL TUBERCULOSIS REPORT 2019*. (2019). doi:10.16309/j.cnki.issn.1007-1776.2003.03.004
6. Masquillier, C. *et al.* Sinako, a study on HIV competent households in South Africa: A cluster-randomized controlled trial protocol. *Biomed. Cent.* **21**, 1–14 (2020).
7. Laniado-Laborín, R. Clinical challenges in the era of multiple and extensively drugresistant tuberculosis. *Rev. Panam. Salud Pública* **41**, 1–6 (2017).
8. Palomino, J. C. & Martin, A. Drug Resistance Mechanisms in *Mycobacterium tuberculosis*. **3**, 317–340 (2014).
9. Andries, K. *et al.* Acquired resistance of *Mycobacterium tuberculosis* to bedaquiline. *PLoS One* **9**, 1–11 (2014).
10. Joshi, J. M. Tuberculosis chemotherapy in the 21st century: Back to the basics. *Lung India* **28**, 193–200 (2011).
11. Olaru, I. D., Heyckendorf, J., Andres, S., Kalsdorf, B. & Lange, C. Bedaquiline-based treatment regimen for multidrug-resistant tuberculosis. *Eur. Respir. J.* **49**, 1–4 (2017).
12. Adewumi, Adeniyi T, Ajadi, Mary B, Soremekun, Opeyemi S. and Soliman, M. E. S. Thompson loop : opportunities for antitubercular demethylmenaquinone methyltransferase protein †. *RSC Adv.* **10**, 23466–23483 (2020).
13. Koch, A., Cox, H. & Mizrahi, V. Drug-resistant tuberculosis: challenges and opportunities for diagnosis and treatment. *Curr. Opin. Pharmacol.* **42**, 7–15 (2018).

CHAPTER 2

2 Introduction

Tuberculosis pathogen remains the single infectious agent that has killed more than a billion people globally over the last two centuries.[1] Unfortunately, it was not later than 1882 that the incidences and mortality were well underway before Robert Koch announced his discovery of *Mtb*.[2, 3] Despite working alone, Koch completed outstanding medical-scientific achievements in less than just a year in human history. He named the TB pathogen *Mycobacterium tuberculosis* (*Mtb*) as globally called to date.[4] The genus *Mtb* belongs to the Mycobacteriaceae family and is an obligate pathogen thought to have co-existed with humans (its host) over several millions of years.[5,] *Mtb*'s ancient Origin is over 70 000 years, and it infected nearly 2 billion people recently.[3] *Mtb* is highly infectious and contagious, possesses complex immunological response, and chronic progression.[3,6, 7]

An estimation of ≥ 90 % of people who develop TB are adults (>15 years) in the ratio of male 2:1 female. Perhaps, the lack of available effective vaccines in adults contributes to the higher number of TB infections. The 30 high-TB burden countries are responsible for more than 90 % of the global cases annually.[8] Presently, more than 20 drugs are used for TB treatment, which was developed over 40 years ago.[9] Unlike drug-resistant TB, drug-susceptible TB can be effectively treated with anti-TB first-line medicines. Mutations in many of the *Mtb* genes are associated with drug resistance.[9] The Mtb's ability to acquire resistance by mutations was discovered when the first clinical trial with streptomycin was carried out about seven decades ago.[10] Figure 2.1 illustrates the proportion of infected people relative to men, women, children, and TB/HIV co-infection.

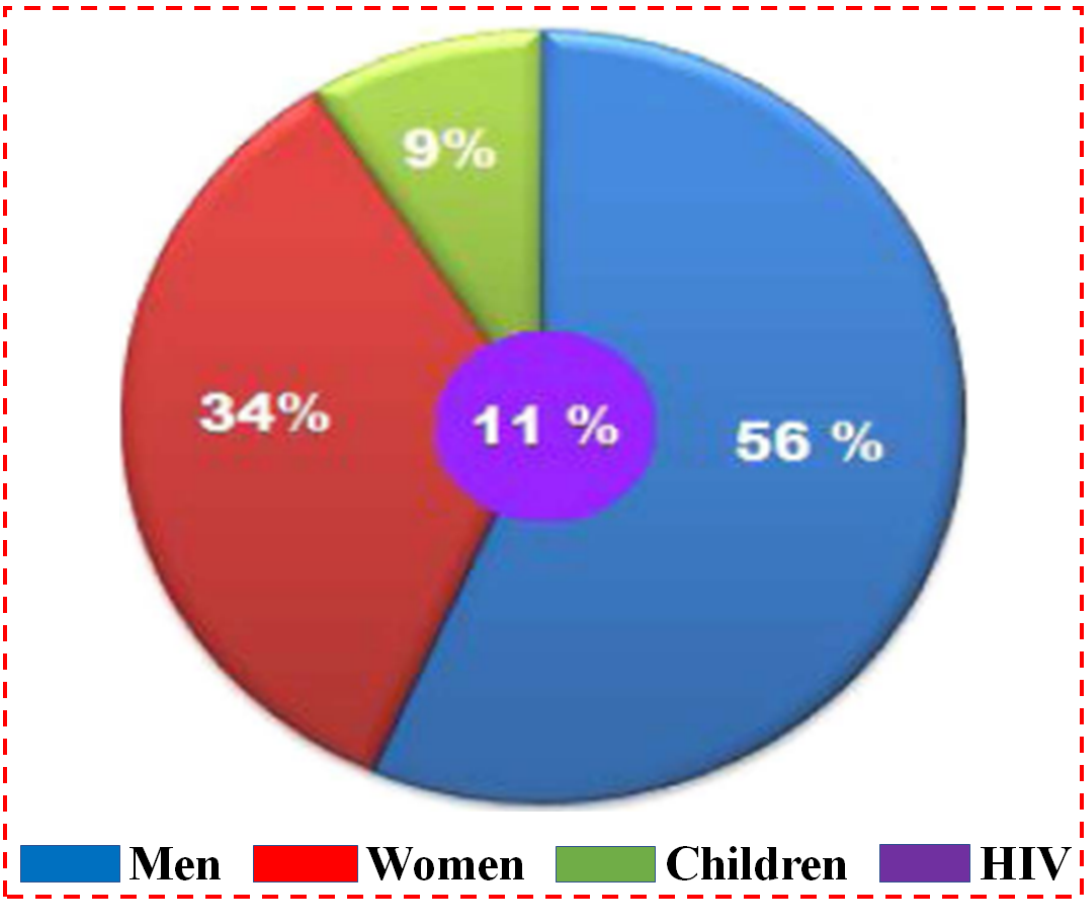


Figure 2.1. Epidemiology of Tuberculosis.[11]

WHO launched the End TB strategy in 2017 and unanimously adopted the milestones (2020 and 2025) and targets (2030 and 2035) to end the global TB pandemic. However, TB infection numbers must be falling by 10 % yearly (by 2025), and the mortality proportion falls by 6.5 % (same year) to achieve the milestones and goals of the WHO EndTB strategy[12]. The global incidence rate of TB (new cases per 100 000 population per year) falls by 2 %, with the average decreasing rate at 1.6 % per year in the period 2000-2018 and 2.0 % between 2017 and 2018,[12], indicating a low declining rate. Figure 2.2 illustrates graphically incidence and mortality curves projection needed to reach the EndTB targets and milestones between 2015-2035[13]. Achieving the WHO milestones and targets with the time range requires sustaining TB diagnosis, treatment, prevention services, and multisectoral action to address socioeconomic factors that drive TB epidemics.

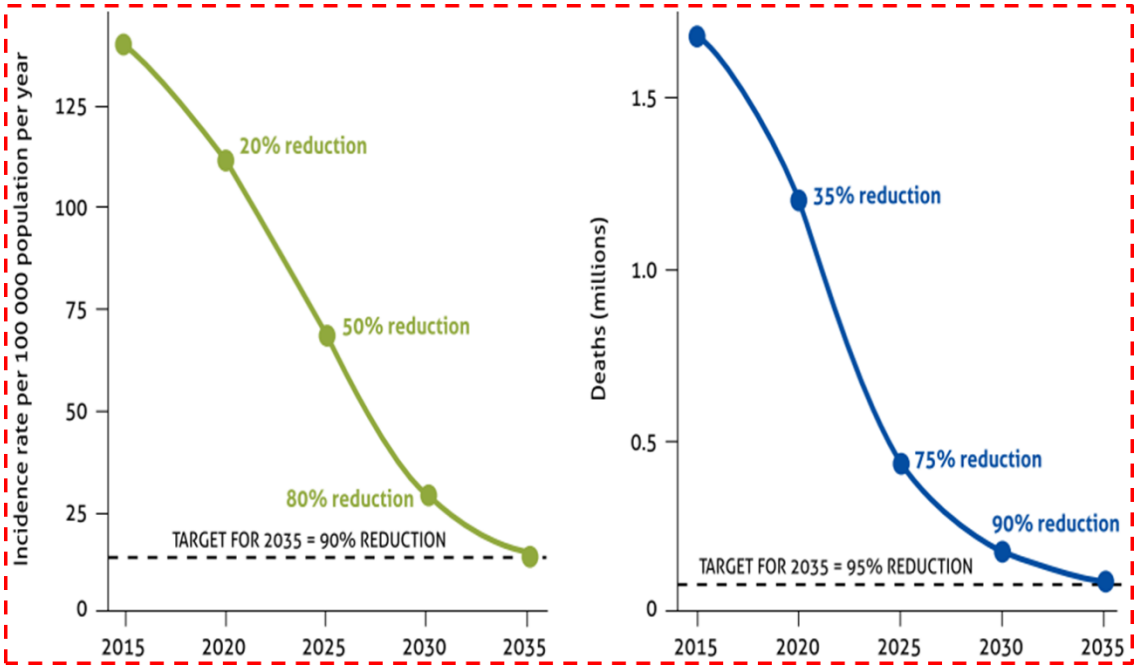


Figure 2.2. Predicted curves of incidence and mortality to attain the targets and milestones of the EndTB, 2016-2035 (WHO, 2017).

Mtb spreads when TB-infected people expel the bacteria into the air through coughing, sneezing, or laughing.[14, 15] Better still, the infection is transmitted from an infected person to an airborne particle susceptible person.[16] The transmission occurs when the inhaled droplet nuclei containing the pathogen travels via the mouth or nasal passages into the upper respiratory tract and further proceed to the bronchi until it enters the lung and the alveoli. Thus, ending the TB pandemic requires altering the germ transmission processes.[17] Figure 2.3 illustrates the Cascade of the *Mycobacterium tuberculosis* transmission cycle.

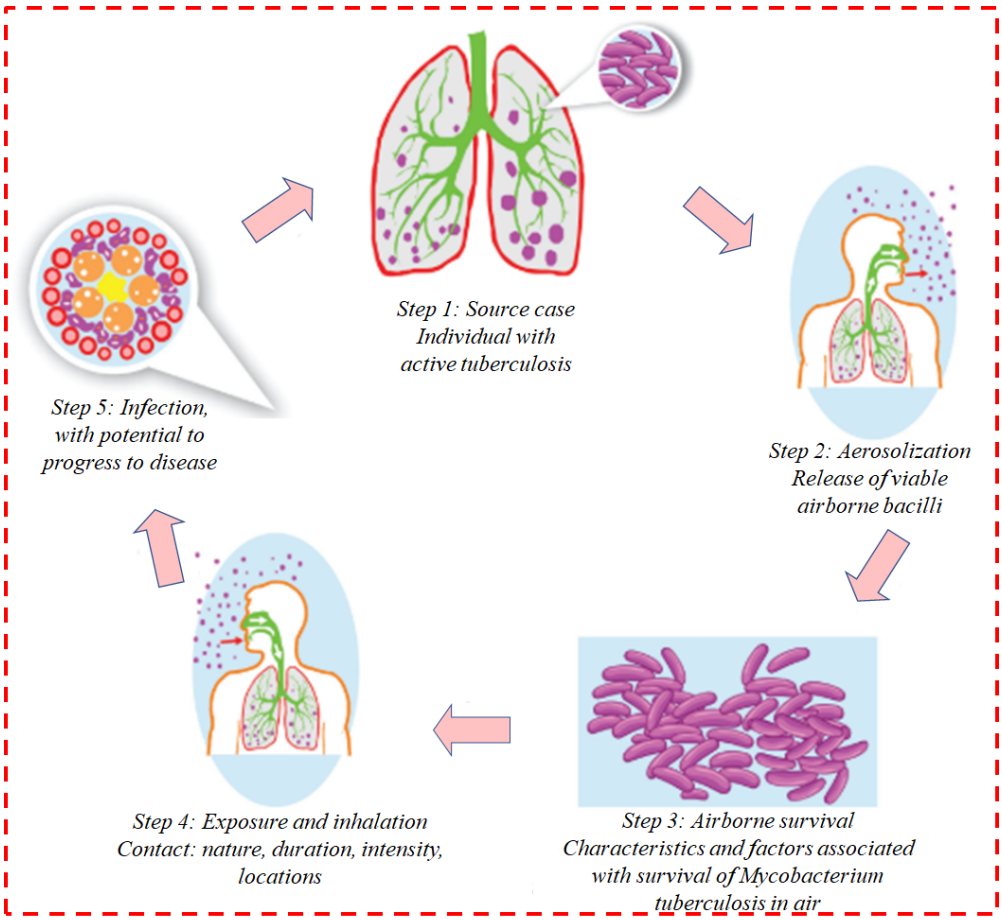


Figure 2.3. Cascade of *Mycobacterium tuberculosis* transmission cycle. Source: Japan Anti-Tuberculosis Association: Common sense of Tuberculosis, 2007

The TB infection can be pulmonary (lung) or extrapulmonary (hematogenous and lymphatic TB). Moreover, TB infection can be classified as latent and active. Latent TB occurs when a person harbours a germ, but the immune system holds them back from spreading, while active TB causes the pathogen's spread. A patient with latent TB shows no symptoms and is not contagious, but an active TB multiplies and causes sickness.[18] Latent TB is still alive and can become active anytime.

2.1 Multiple and Extensively Drug-Resistant Tuberculosis: Challenges & Treatment

MDR-TB and XRD-TB) have remained pervasive in the last couple of years.[19,20] Both the TB resistance types pose formidable challenges in diagnosis and treatment.[21, 22] Multidrug-

resistant TB occurs when the first-line drugs RIP and INH, that are the first two most effective, becomes resistant to by the pathogen. MDR-TB treatment requires less effective second-line drugs, more expensive, and associated with too many side effects and toxicity.[23] Drug resistance diagnosis is quite challenging because it may take up to 6-16 weeks and requires advanced laboratory tools. MDR-TB has a higher overall mortality rate than drug-susceptible TB.[19] It can be cured, but it takes at least three times longer to be cured than 18-24 months.[19, 24] In contrast, XRD-TB is the MDR-TB, which fails to respond to the multiple second-line drugs or resistance to not less than one drug in the two most important classes of medicines in an MDR-TB regimen.[8] XRD-TB treatments demand even more expensive and toxic third-line drugs such that a course of treatment is tailored to individual TB samples. The first remedy success against TB was introduced as the sanatorium cure in 1854.[3] The difficulty involved in diagnosing XRD-TB could be why many patients with XRD-TB do not survive tuberculosis infection. A more challenging situation is that HIV-positive patients are at greater risk of acquiring MDR-TB strains.

Initially, combined therapy comprises isoniazid (INH) plus para-aminosalicylic acid has been practised to overcome drug resistance development. Subsequently, the combination of pyrazinamide (PZA) and rifampicin (RIF) drugs are still being used. However, penurious drug prescription, impoverish patient adherence, drug supply/quality, and other factors lead to acquired drug resistance. MDR-TB can be primary, which occurs when a person is infected by someone having a resistant strain or contracted when regular TB is inadequately treated, allowing bacteria to develop resistance. MDR/XRD-TB's primary cause is the patients' failure to complete their total TB doses and regimen, which results in the disease relapse and resistance that is very challenging to cure. The treatments of MDR/XRD-TB are often longer than the standard therapies. The disease condition requires expensive second-line drugs that show no assurance of effectiveness and severe toxicity, resulting in an undesirable outcome.

Several newer medications, including bedaquiline, delamanid, SQ109, pretomanid, and linezolid, have effectively treated DR-TB. Unfortunately, they have "black-boxed" as a result of their highly toxic effect.

Moreover, new mechanisms of resistance have been identified. The resistance to antitubercular drugs dates back to the beginning of chemotherapy discovery, and there have been newly reported findings into the mechanisms of the *Mtb* strains resistance.

2.2 Diagnosis and Detection of TB infection

The ameliorative control towards ending TB pandemic tuberculosis includes prevention, diagnosis, and treatment plans and continuously posing more problems. Diagnosis and treatment must be specifically achieved to control drug-resistant TB. There is an urgent need for better testing, rapid diagnostics. Doing better testing of drug-susceptibility is not sufficient for maintaining drug resistance. The lack of adequate and straightforward point-of-care TB tools to accurately diagnose all kinds of TB, especially MDR and XRD, is challenging. Conventional TB diagnosis methods include the clinical diagnosis of active TB, Mantoux test, Chest X-ray radiographs, and classical laboratory methods. The sensitivity to the tuberculum in TB patient can be obtained by carrying out a test called Mantoux. Advanced stage of pulmonary tuberculosis is confirmed if lung caverns are shown by chest X-ray test.

The classical laboratory method is the most used worldwide due to its reliability. Therefore, continuing efforts towards improving the coverage and diagnosis quality, drug-resistant TB infected people, plus treatments are necessities. *Mtb* drug resistance can be determined using conventional methods, Agar proportion method, Fluorometric method (BACTEC 960 system), Microplate Alamar Blue Assay (MABA), molecular diagnosis tools, and the primary genetic markers to identify *Mtb*. Currently, the WHO recommends rapid molecular tests (Xpert, *MTB*/RIF assay (Cepheid, USA) to be the initial diagnostic test for TB disease[13, 25].

The *MTB*/RIF test (Xpert) is highly efficient in sputum samples to detect susceptible TB and RIP-resistance due to mutation within 2 hours. Besides, molecular data obtained from epidemiologic study can reveal the outbreak all forms of TB including drug-resistant TB hot spots. However, there are still significant and persistent gaps in detection.[15] The mortality rate of TB disease would continue to increase without timely diagnosis and essential operational research investment in diagnostic tools to enhance the cure of most TB patients. However, excellent knowledge of drug resistance mechanisms helps develop new tools for the rapid diagnosis of drug-resistant TB.

2.3 Prevention of TB infection

According to Desiderius Erasmus, a great Dutch philosopher (1500), "Prevention is better than cure." The menace of drug-resistant TB and its evolution must be pevented, especially in regions with prevalent HIV/AIDS cases. The reason being that co-infection of HIV and TB is more devastating that when single TB infection. The prevention of MDR/XRD-TB and TDR-TB require designed infection control strategies. The strategies include to involve campaigning to patients and the community, planning of infection control, collecting safe sputum samples, etiquette and hygiene surrounding coughing, rapid TB diagnosis and treatment, improve room air ventilation, etc. The discovery of effective TB vaccines could be a landmark for the reduction of TB infection. The bacilli Calmette-Guerin (BCG) vaccine, developed 100 years ago, is effective against TB in children and is recommended by WHO. Contrastingly, there is still no vaccine for TB in adults currently. Vaccine serves as a preventive treatment that elicited immune responses against TB. Although 14 vaccine candidates are in clinical trial Phase IIb, which could transform global TB preventive efforts, there could still be limitations in the effective control of the new and re-emerging strains of *Mycobacterium tuberculosis*.

2.4 Post-treatment Patient Rehabilitation

Tuberculosis may impose certain physical limitations and injuries that trigger reaction and chest wall retraction, compromising expansion.[26] Hence it is necessary to highlight the need for post-treatment rehabilitation of TB patients, including social, psychological, physiotherapy (exercise), and economic aspects.[27] Although most TB control programmes admit work completion after a patient is successfully cured, patients often experience prolonged sufferings with post-treatment or pulmonary rehabilitation (PR) or sequelae. Evidence supports the importance of a thorough and adequate evaluation of patients after TB cure to identify pulmonary rehabilitation beneficiaries. PR is effective in patients with tuberculosis[28], and the rehabilitation process involves a multidisciplinary approach.[27] Pulmonary rehabilitation makes the evolution and the increase in the chances of healing and social reintegration possible. Considering the length of TB treatment, in which the minimum is six months, in some cases, treatment may last for over five years, little or no attention is given to the social, psychological, and economic impact extended treatment can have on the patients' lives.[29] To a large extent, TB patients, especially in many high-burden countries, continue to face high levels of stigma and obstacles, even pre-diagnosis, which persist at every stage of the treatment process.

Research reports have shown that most individuals face high out-of-pocket expenses, including additional medication, nutritional supplements, and transport, despite the free diagnoses and treatment they received.[29] Nutritional supports and physiological challenges have not been adequately maintained.[30] Malnutrition increases the risk of TB mortality in patients with a low body mass index than those with a high body mass index. Psychological and emotional support should be provided to TB patients before diagnosis, treatment, and post TB sequelae. Therefore, it could be said that affected individuals do not often prepare for the financial, time, physiological, and psychological effects. Sometimes, the side effects of TB medication may be extreme to require extensive therapy and rehabilitation. Nutritional

supports and physiological challenges have not been adequately maintained. The friends, families, and even health workers of TB people often labelled them with various constant stigmas.

2.5 *Mycobacterium tuberculosis* drug resistance, targets and therapeutic interventions

The complete genome sequence of the *Mtb* provides opportunities to inhibit the organism target enzymes' activity by implementing screening campaigns that focus on identifying low molecular mass chemical compounds.[24] Currently, robust Whole-genome sequencing (*WGS*) technology is used to uncover the drug resistance mechanism in *Mtb*, and it rapidly sequences the complete mycobacterial genome.[31] Moreover, WGS can be used to identify signatures of convergent evolution and positive selection of newly associated drug resistance genes in *Mycobacterium tuberculosis*.[32] The essentiality of the *Mtb* protein target is very critical towards drug design development processes. The current consideration of targeted therapy, specifically the molecular targets towards TB treatment, has been the researchers' focal point. The needs for new drugs to improve the current TB therapy, especially with the worsening cases of drug resistance, are inevitable. Therefore, it is imperative to explore novel drug targets while developing new agents to enhance drug efficacy and circumvent resistance. Several *Mtb* targets have been identified over the years, but few are well-known drug targets, including InhA, FabD, FabC, KasA, etc.[33]

Many recognized promising pathways pertinent to drug targets include cell wall metabolism,[34] cellular respiration,[35] protein synthesis, etc. Table 2.1 highlighted some promising drug targets, functions, and inhibitors.[16] Several anti-TB drugs were discovered based on the screening compounds against the bacterial cells. Many drugs have more than one protein targets. However, some drug compounds may not enter the mycobacterial cell

wall because of their architectural structure; thick and greasy characteristics.[36,37] Yet, the target/activator's crystal structures can be resolved, improving the current drugs.

Table 2.1 Drug targets in *Mycobacterium tuberculosis*.

Target protein	Function	Inhibitor
Secreted antigen 85 phosphonate complex	TDM/TMM synthesis, and mycolic acid cell wall attachment	Cyclophostin ($CyC_{8\beta}$)
Demethylmenaquinone methyltransferase, *menG* (rv0558)	Menaquinone synthesis for cell wall maintenance	DG70 compound
N-acetylglucosamine-6-phosphate	Essential amino sugal (GlcN6P) for Cell wall biosynthesis & glycolysis betabolic pathways	M-methylhydroxyphos-phenyl-Dglucosamine-6-phosphate
Polyketide synthase(*pks13*)	Mycolic acid synthesis	Cerulenin, Thiopene
Acyl-AMP ligase (*fadD32*)	Mycolic acid synthesis	4, 6-diaryl-5,7-dimethyl coumarin
Mycobacterial transmembrane reporter protein (mmpL3)	Mycolic acid synthesis	Adamantyl ureas,
ATP synthase/cytochrome bc_1 complex (*qcrB*)	ATP synthase	Imidazopyridine amide (IPA), imidazo [1,2-a} pyridine, IPA compound Q203

2.6 Selected targets and their essentiality in the *Mycobacterium tuberculosis* Success

2.6.1 *Mycobacterium tuberculosis* N-acetylglucosamine-6-phosphate deacetylase, *NagA*

Mycobacteria species enciphers enzyme GlcNAc-6-phosphate deacetylase[34] which play critical functions in the production of acetylglucosamine-6-phosphate (GlcNAc6P) sugar and

its flow within the cell wall to build cellular envelopes.[34] *NagA* enzymes occur both in prokaryotic and eukaryotic, including the non- and mycobacteria.[34, 39] Essentially, they are members of the amidohydrolase superfamily. [40,41] The enzyme possesses a characteristic active site consisting of triose-phosphate isomerase (TIM)-like barrel fold and metal ions that are very critical to the protein functions [34, 40] Moreover, *NagA* enzymes has similar structure domains I and II. The D*omain I* contains eight alpha/beta $(\beta/\alpha)_8$-barrel-like,[34]. and *Domain II* includes a small β-barrel with unknown biological function. However, a study by Ahangar *et al.* 2018, suggested that *NagA* enzyme *Domain II* confers stability.

The whole cell of mycobacteria comprises of four thick layers called an envelope, which suggest why the pathogens are resistant to acid/alkaline media [42]. The cell wall of Mycobacterial is crucial for their survival and virulence [34] Embedded within the envelope are unique and diffused macromolecules. These include peptidoglycan, lipids, arabinogalactan, proteins, capsules of polysaccharides, and mycolic acids.[43] The glucosamine-6-phosphate (GlcN6P) is isomerised to form glucosamine-1-phosphate, processed to produce peptidoglycan. Cell wall's alteration is a potential therapeutic opportunity to terminate *Mycobacterium tuberculosis* or its homologs.

2.6.2 *Mycobacterium tuberculosis* demethylmenaquinone methyltransferase (menG)

MenG enzyme is a potential target that involves in the biosynthesis of menaquinone (MK) prokaryotes (such as *Mtb*) require for maintaining cell wall.[44,35,45] *Mycobacterium tuberculosis menG* amino acids are conserved and common among its homologs.[46 53] Demethylmenaquinone methyltransferase is a member of SAM-dependent methyltransferase (MTase) enzyme family, which have membrane domain and α/β/α fold pattern structure.[47,48] The domain is a potential molecular target towards designing an effective drug against TB. *Mtb menG* encodes Rv0558, which catalyses menaquinone (or known as [MK9(H_2)] production for respiration via the transfer of a C-methyl transfer from SAM structure.[49] Thereafter the catalytic transfer, SAM

is reduced to S-Adenosyl homocysteine (SAH).[48] The *MenG* protein is a Class I enzyme, which has at least two or more GXG motifs. One motif is found in the protein first β-sheet while the other motif is bound to a different β-sheet.[50,51] *Mycobacterium smegmatis menG* comprises of about 234 amino acids and a GXG motif.[52] Various classes of the methyltransferases are; Class I (Rossmann-like α/β), Class II (TIM α/β-α/β), etc.[53] Class transferase enzymes have remarkable compatible primary structures of 10 % similarity.[51]

2.6.3 *Mycobacterium tuberculosis* Secreted Antigen protein (Ag85C)

The mycobacterial cell envelope contains different proteins responsible for the success of the microorganisms in pathogenesis, survival, and resistance to many drugs.[54,34] *Mtb* antigen complex (Ag85A, Ag85B, and Ag85C) bind to the human fibronectin via heparin and cellular-binding domains or collagen-binding domain when it escapes the host immune system. Ag85 proteins are potential targets for treating drug-susceptible and drug-resistant TB because they catalyse mycolic acids' attachment in the cell wall and biosynthesis pathway of TMM and TDM. *Mycobacterium tuberculosis* cell envelope comprises the mycolyl-arabinogalactan-peptidoglycan. In other words, the macromolecules include mycolic acids, arabinogalactan (AG), and peptidoglycan (PG) (mAGP complex). PG is found outside the plasma membrane covalently linked to the AG. The mycolic acids are long-chain α-alkyl-β-hydroxyl fatty acids (C70-90) that occupy about 60 % of the whole cell wall and are attached to the AG. Long fatty acids undergo an esterification process to produce trehalose derivatives. These include monomycolate (TMM) and dimycolate (TDM), which serve as donors in mycolic acid metabolism. Moreover, the mycolic acids are interspersed with non-covalent glycolipids, proteins, and outer polysaccharide capsules.[34]

Mtb antigen 85 enzymes are crucial in the attachment of mycolic acids to the cell wall and the synthesis of trehalose moieties. Besides, this protein involves the efflux mechanism, which may result in phenotypic resistance. The enzymes share common structural and functional

similarities, such as mycolic acid donor TMM and 68-79 % conserved catalytic site sequence identity. Additionally, they have serine mycolyl esterase activity, carboxylesterase consensus sequence (GXSXG), among others.[55] Therefore, a single removal of one of the enzymes affects the complex containing the three proteins. Precisely, the alteration of the *fbpC2* gene result in a 40 % reduction in the AG-attached mycolic acids, and the reductive impact on either *fbpA* or *fbpB* genes leads to TDM reduction, which indicates low functional activity [56, 57]

References

1. Dheda, K. *et al.* Recent controversies about MDR and XDR-TB: Global implementation of the WHO shorter MDR-TB regimen and bedaquiline for all with MDR-TB? *Respirology* **23**, 36–45 (2018).

2. Murray, J. F., Schraufnagel, D. E. & Hopewell, P. C. Treatment of tuberculosis: A historical perspective. *Annals of the American Thoracic Society* **12**, 1749–1759 (2015).

3. 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. *Journal of Preventive Medicine and Hygiene* **58**, E9–E12 (2017).

4. Murray, J. F., Schraufnagel, D. E. & Hopewell, P. C. Treatment of tuberculosis: A historical perspective. *Annals of the American Thoracic Society* **12**, 1749–1759 (2015).

5. Gagneux, S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology* 1–12 (2018). doi:10.1038/nrmicro.2018.8

6. Korb, V. C., Chuturgoon, A. A. & Moodley, D. *Mycobacterium tuberculosis*: Manipulator of protective immunity. *International Journal of Molecular Sciences* **17**, (2016).

7. Maurizio de Martino, Lorenzo Lodi, Luisa Galli, E. C. Immune Response to *Mycobacterium tuberculosis* : A Narrative Review. 1–17 (2020). doi:https://doi.org/10.3389/fped.2019.00350

8. World Health Organization, (WHO). *Global tuberculosis report 2018. World Health Organization. http://www.who.int/iris/handle/10665/274453.* (2018).

9. Islam, M. M. *et al.* Drug resistance mechanisms and novel drug targets for tuberculosis therapy. *Journal of Genetics and Genomics* **44**, 21–37 (2017).

10. Laniado-Laborín, R. Clinical challenges in the era of multiple and extensively drugresistant tuberculosis. *Revista Panamericana de Salud Pública* **41**, 1–6 (2017).

11. Patil, K., Bagade, S., Bonde, S., Sharma, S. & Saraogi, G. Biomedicine & Pharmacotherapy Recent therapeutic approaches for the management of tuberculosis : Challenges and opportunities. *Biomedicine & Pharmacotherapy* **99**, 735–745 (2018).

12. (WHO), W. H. O. *Global Tuberculosis Report (Executive Summary)*. *Report* (2018).

13. WHO. *GLOBAL TUBERCULOSIS REPORT 2017: ENDTB*. (2017).

14. Sabbatani, S. Historical insights into tuberculosis. Girolamo Fracastoro's intuition on the transmission of tuberculosis and his opponents. History of an idea. *The infection in the history of medicine* **4**, 284–291 (2004).

15. World Health Organization, (WHO). *GLOBAL TUBERCULOSIS REPORT 2019*. (2019). doi:10.16309/j.cnki.issn.1007-1776.2003.03.004

16. Hameed, H. M. A. *et al.* Molecular targets related drug resistance mechanisms in MDR-, XDR-, and TDR-*Mycobacterium tuberculosis* strains. *Frontiers in Cellular and Infection Microbiology* **8**, (2018).

17. Gonzalo-asensio, J., Aguilo, N., Marinova, D. & Martin, C. Breaking Transmission with Vaccines : The Case of Tuberculosis. *Microbiol Spectrum* **5**, 1–11 (2017).

18. Drugs, A. *et al. Mycobacterium tuberculosis* Shikimate Pathway Enzymes as Targets for the Rational Design of Anti-Tuberculosis Drugs. *molecules* **25**, 1–36 (2020).

19. Günther, G. Multidrug-resistant and extensively drug-resistant tuberculosis: A review of current concepts and future challenges. *Clinical Medicine, Journal of the Royal College of Physicians of London* **14**, 279–285 (2014).

20. Pontali, E., D'Ambrosio, L., Centis, R., Sotgiu, G. & Migliori, G. B. Multidrug-resistant tuberculosis and beyond: An updated analysis of the current evidence on bedaquiline. *European Respiratory Journal* **49**, (2017).

21. Koch, A., Cox, H. & Mizrahi, V. Drug-resistant tuberculosis: challenges and opportunities for diagnosis and treatment. *Current Opinion in Pharmacology* **42**, 7–15 (2018).

22. Kurz, S. G., Furin, J. J. & Bark, C. M. Drug-Resistant Tuberculosis: Challenges and Progress. *Infectious Disease Clinics of North America* **30**, 509–522 (2016).

23. Reuter, A., Hughes, J. & Furin, J. Challenges and controversies in childhood tuberculosis. *The Lancet* **394**, 967–978 (2019).

24. Migliori, G. B. *et al.* MDR/XDR-TB management of patients and contacts: Challenges facing the new decade. The 2020 clinical update by the Global Tuberculosis Network. *International Journal of Infectious Diseases* **92**, S15–S25 (2020).

25. Khan, U. *et al.* The endTB observational study protocol: Treatment of MDR-TB with bedaquiline or delamanid containing regimens. *BMC Infectious Diseases* **19**, 1–9 (2019).

26. Wilches, E. C., Rivera, J. A., Mosquera, R., Loaiza, L. & Obando, L. Pulmonary rehabilitation in multi-drug resistant tuberculosis (TB MDR): A case report. *Colombia Médica* **40**, 436–441 (2009).

27. Mahler, B. & Croitoru, A. Pulmonary rehabilitation and tuberculosis: A new approach for an old disease. *Pneumologia* **68**, 107–113 (2019).

28. Visca, D. *et al.* Pulmonary rehabilitation is effective in patients with tuberculosis pulmonary sequelae. *The European respiratory journal* **53**, (2019).

29. Kumar, B. A. Rehabilitation of treated TB patients: Social, psychological and economic aspects. *International Journal of Mycobacteriology* **5**, S129–S130 (2016).

30. Visca, D. *et al.* Post-tuberculosis (TB) treatment: The role of surgery and rehabilitation. *Applied Sciences (Switzerland)* **10**, 1–18 (2020).

31. Cohen, K. A., Manson, A. L., Desjardins, C. A., Abeel, T. & Earl, A. M. Deciphering drug resistance in *Mycobacterium tuberculosis* using whole-genome sequencing: Progress, promise, and challenges. *Genome Medicine* **11**, 1–18 (2019).

32. Studer, R. A., Dessailly, B. H. & Orengo, C. A. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *Biochemical Journal* **449**, 581–594 (2013).

33. Jena, L. & Harinath, B. Anti-tuberculosis therapy: Urgency for new drugs and integrative approach. *Biomedical and Biotechnology Research Journal (BBRJ)* **2**, 16–19 (2018).

34. Ahangar Mohd Syed, Furze Christopher M., Guy, Collette S., Fullam, E., Cameron, A. D., Cooper, C., Graham, B. & Maskew, K. S. Structural and functional determination of homologs of the *Mycobacterium tuberculosis* N -acetylglucosamine-6-phosphate

deacetylase (NagA). *Journal of Biological Chemistry* **293**, 9770–9783 (2018).

35. Sukheja, P. *et al.* A Novel Small-Molecule Inhibitor of the *Mycobacterium tuberculosis* Demethylmenaquinone Methyltransferase MenG Is Bactericidal to Both Growing and Nutritionally Deprived Persister Cells. *American Society for Microbiology* **8**, 1–15 (2017).

36. Alderwick, L. J., Harrison, J., Lloyd, G. S. & Birch, H. L. The Mycobacterial Cell Wall - Peptidoglycan and Arabinogalactan. *Cold Spring Harb Perspect Med* **5**, 1–15 (2015).

37. Batt, S. M., Minnikin, D. E. & Besra, G. S. The thick waxy coat of mycobacteria, a protective layer against antibiotics and the host's immune system. *Biochemical Journal* **447**, 1983–2006 (2020).

38. Mouilleron, S. *et al.* The Mechanism of Acetyl Transfer Catalyzed by *Mycobacterium tuberculosis* GlmU . *Biochemistry* **57**, 3387–3401 (2018).

39. Popowska, M., Osińska, M. & Rzeczkowska, M. N-acetylglucosamine-6-phosphate deacetylase (NagA) of Listeria monocytogenes EGD, an essential enzyme for the metabolism and recycling of amino sugars. *Archives of Microbiology* **194**, 255–268 (2012).

40. Liu, A. & Huo, L. *Amidohydrolase Superfamily*. *eLS* (John Wiley & Sons, 2007). doi:10.1002/9780470015902.a0020546.pub2

41. Sugrue, E. *et al.* Evolutionary expansion of the amidohydrolase superfamily in bacteria in response to the synthetic compounds molinate and diuron. *Applied and Environmental Microbiology* **81**, 2612–2624 (2015).

42. Percival, S. L. & Williams, D. W. *Mycobacterium*. *Microbiology of Waterborne*

Diseases: Microbiological Aspects and Risks (Elsevier, 2014). doi:10.1016/B978-0-12-415846-7.00009-3

43. Abrahams, K. A. & Besra, G. S. Mycobacterial cell wall biosynthesis: A multifaceted antibiotic target. *Parasitology* **145**, 116–133 (2018).

44. Johnston, J. M., Arcus, V. L., Morton, C. J., Parker, M. W. & Baker, E. N. Crystal Structure of a Putative Methyltransferase from m *Mycobacterium tuberculosis*: Misannotation of a Genome Clarified by Protein Structural Analysis. *Society* **185**, 4057–4065 (2003).

45. Iqbal, I., Bajeli, S., Akela, A. & Kumar, A. Bioenergetics of Mycobacterium: An Emerging Landscape for Drug Discovery. *Pathogens* **7**, 1–30 (2018).

46. Lee, P. T., Hsu, A. Y., Ha, H. T. & Clarke, C. F. A C-methyltransferase involved in both ubiquinone and menaquinone biosynthesis: Isolation and identification of the Escherichia coli ubiE gene. *Journal of Bacteriology* **179**, 1748–1754 (1997).

47. Zou, X. W. *et al.* Structure and mechanism of a nonhaem-iron SAM-dependent C-methyltransferase and its engineering to a hydratase and an O-methyltransferase. *Acta Crystallographica Section D: Biological Crystallography* **70**, 1549–1560 (2014).

48. Ramharack, P. & Soliman, M. E. S. Zika virus NS5 protein potential inhibitors: an enhanced in silico approach in drug discovery. *Journal of Biomolecular Structure and Dynamics* **36**, 1118–1133 (2018).

49. Kwon, O. & Meganathan, R. Biosynthesis of Menaquinone (Vitamin K2) and Ubiquinone (Coenzyme Q). *EcoSal Plus* **3**, (2009).

50. Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransfer: A chronicle of convergence. *Trends in Biochemical Sciences* **28**, 329–335 (2003).

51. Struck, A. W., Thompson, M. L., Wong, L. S. & Micklefield, J. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **13**, 2642–2655 (2012).

52. Puffal, J., Mayfield, J. A., Moody, D. B. & Morita, Y. S. Demethylmenaquinone Methyl Transferase Is a Membrane Domain-Associated Protein Essential for Menaquinone Homeostasis in Mycobacterium smegmatis. *Frontiers in Microbiology* **9**, 1–12 (2018).

53. Wlodarski, T. *et al.* Comprehensive structural and substrate specificity classification of the saccharomyces cerevisiae methyltransferome. *PLoS ONE* **6**, e23168 (2011).

54. Nguyen, P. C. *et al.* Cyclipostins and Cyclophostin analogs as promising compounds in the fight against tuberculosis. *Scientific Reports* **7**, 1–15 (2017).

55. Ronning, D. M., Vissa, V., Besra, G. S., Belisle, J. T. & Sacchettini, J. C. *Mycobacterium tuberculosis* antigen 85A and 85C structures confirm binding orientation and conserved substrate specificity. *Journal of Biological Chemistry* **279**, 36771–36777 (2004).

56. Jackson, M. *et al.* Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the *Mycobacterium tuberculosis* cell envelope. *Molecular Microbiology* **31**, 1573–1587 (1999).

57. Nguyen, L., Chinnapapagari, S. & Thompson, C. J. FbpA-dependent biosynthesis of trehalose dimycolate is required for the intrinsic multidrug resistance, cell wall structure, and colonial morphology of Mycobacterium smegmatis. *Journal of Bacteriology* **187**, 6603–6611 (2005).

CHAPTER 3

3 Analysis of Biomolecular Structures and Therapeutic Targeting: Applications of Computational drug design and Molecular Modelling Analysis

3.1 Computational drug design and Molecular modelling

Computational chemistry (CC) is a field of Theoretical chemistry that has been very useful in molecular drug design,[65] and it focuses on finding solutions to chemical-related issues using computers. [66, 67, 68, 69] CC involves obtaining relevance of results to chemical problems and not directly developing theoretical methods. Simply put, CC allows the dynamics and motions, structure characterisation, and energetics of protein-ligand interactions, thus provide crucial application in drug design research. The incorporation of computational chemistry can unfold new and challenging issues to be solved during the search for lasting solutions.[70, 71] The standard CC protocols used in CC are Quantum mechanics and molecular dynamics.[72] The former method is used to study molecules' chemistry at the electronic level, while the latter used physics' classical laws without considering explicit electrons. The drug discovery process is quite complex, requiring interdisciplinary efforts to design effective and marketable plausible drugs.[67] The drug designing provides an avenue to search for chemical entities that can bind to a protein target cavity or active site geometrically and chemically. Cellular receptors serve as the mediators of several functions of the organism since they involve in the mechanism of physiology and disease.[69] A specific target disease can be modulated with an outcome lead compound that affects a protein's function.[69] The chemical

structures of leads can be modified to improve efficacy, reduce side effect, and improve selectivity based on the findings from receptor-ligand interaction.

Computational drug discovery has proved to be an effective strategy to accelerate and economise drug design/development processes.[73] Computer-aided techniques for drug design (CADD) is a modern rational design technology that facilitate researching the discovery of new drugs based on the knowledge of target structures'.[74] Conventional drug design methods include a random screening of chemical compounds found in nature or laboratory-based synthesised. The CADD techniques' ability to overcome traditional methods' shortcomings has accounted for its tremendous progress in the last years.[75, 73] The CADD application can make drug discovery more cost-effective by comparing the predicted and actual drug activity findings using the experimental data.[67], [69] Iteratively, this result provides clues to improve the properties of a chemical compound. Moreover, CADD techniques complements and enhances the efficiency of the drug discovery process and the technique.[68]

CADD techniques incorporate both structure-based accurate model, where protein structures are required, and ligand-based exact model, which ligand activities to design compounds interacting with the protein structure.[73] Structure-based drug design approaches include docking, pharmacophore modelling, *de novo* design and fragment-based drug design. Moreover, CADD methods combine computational chemistry applications with molecular modelling, which have been developed into techniques like molecular modelling (MD) simulations, free binding energy, which were applied in this study to achieve some of the objectives and the overall goal. Figure 3.1 highlights various computational approaches in drug design. CADD strategy can be classified into conformational modelling, property modelling and molecular drug design. Conformational modelling is defined as the study of the complexes of macromolecules and small molecules. The property modelling molecular drug design involves studying and optimising the biological, chemical, and physical

properties. This chapter discusses the applicable scope of computational and theoretical tools in this study.

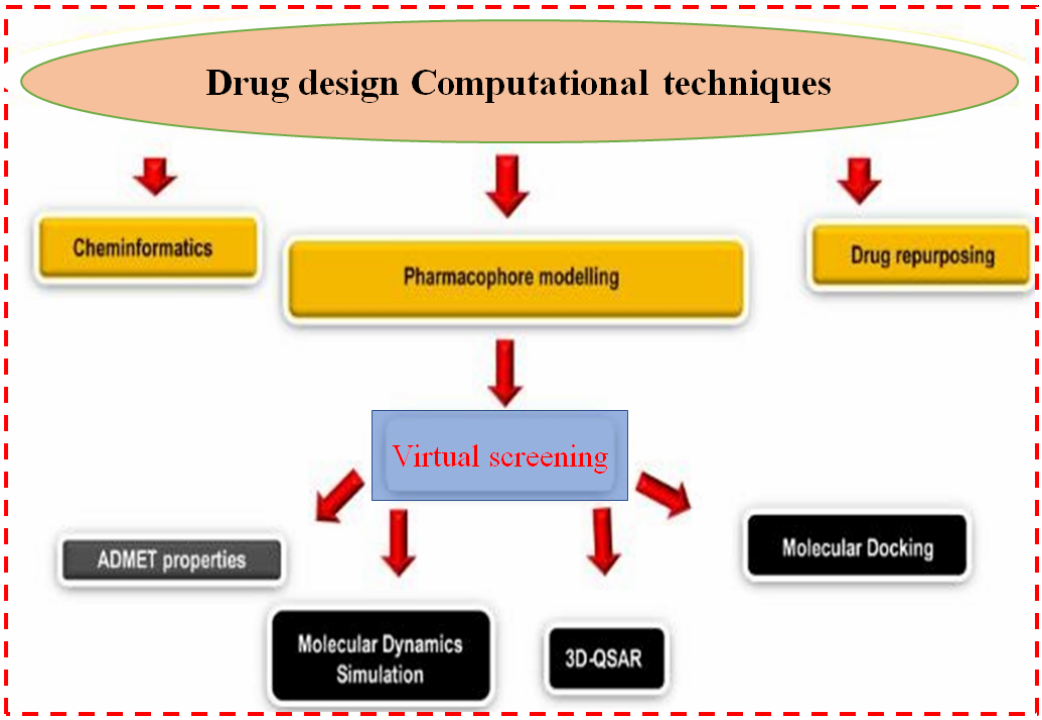


Figure 3.1. Computational approaches in drug design

Specifically, in drug design, molecular modelling allows scientists to use computers to visualise molecules, representing numerical molecular structures and describes all theoretical methods, including computational methods to simulate the biosystems and attributes of biochemical and chemical molecules. Computational drug design models can also be used for repositioning or repurposing and combination therapy design.[76], Figure 3.2. illustrates the basic molecular modelling strategies. Therefore, the computational drug design approaches are specially prioritised and outstanding features and applications in *de novo* designing.

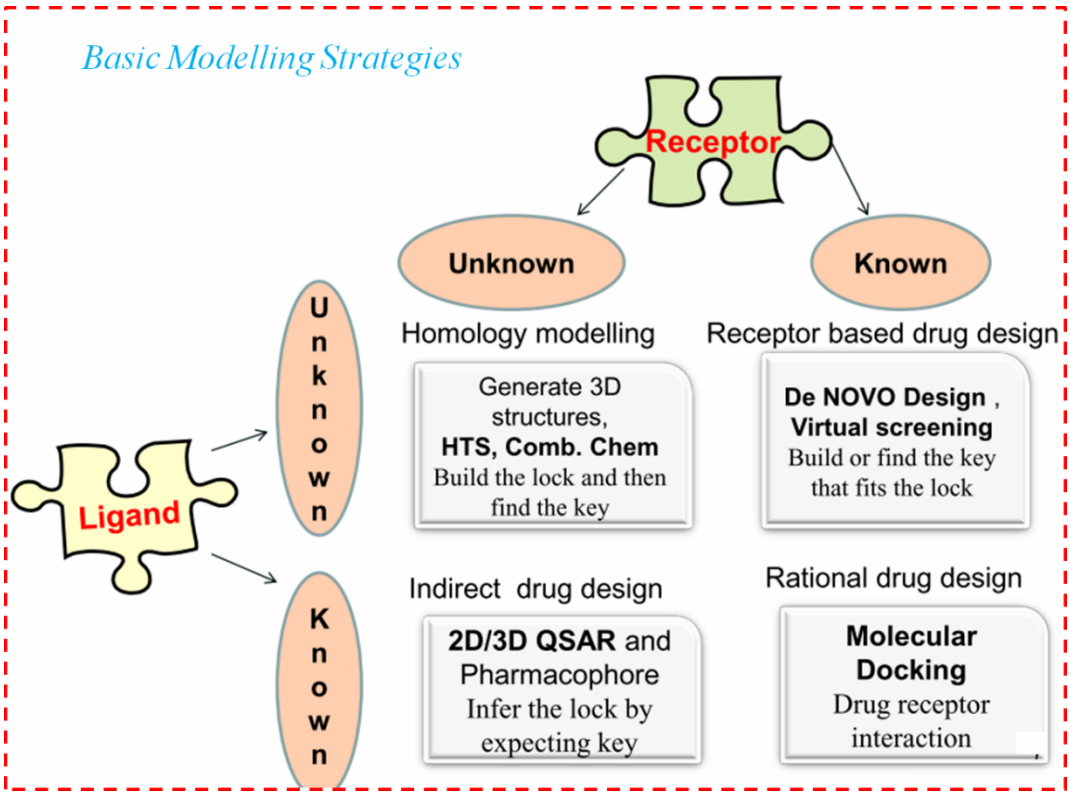


Figure 3.2. Basic molecular modelling strategies

3.2 Quantum mechanics and Schrodinger equation: Applications in drug design

Quantum mechanics (QM) involves calculating molecular orbital and offers the most detailed description of a molecule's chemical behaviours. The application of QM protein-drug interactions in empirical methods cannot be overplayed even though it provides limited applications. QM method considers molecules as collections of nuclei and electrons without referring to chemical bonds, detailing the understanding of system behaviour at the atomic level. The laws of quantum mechanics are applied in QM methods to solve the equations of Schrodinger and approximate wave function. The Schrodinger equation is fundamental to Physics and involves an analytical description of a quantum mechanical characteristic within a system.[77]

An Austrian physicist, Ervin Schrodinger, introduced the equation in 1926.[78] The Schrodinger equation's role is synonymous with Hamilton's laws of motion in mathematical

physics; non-relativistic of quantum and classical mechanics.[79] Schrodinger equation is either dependent or independent of time.[80] The time-independent Schrodinger equation is mostly used when solving computational chemistry problems. It expresses Hamiltonian operator as the potential energy that accumulate and kinetic energy.[80] Mathematical, simplest form of the equation is presented as:

$$H\psi = E\psi \quad \text{..... } \textit{Eqn.}\ 3.1$$

H is the Hamiltonian. H = kinetic energy (T) plus potential energy (V),

H = T + V,

wave function, ψ, advances the electron possibility in the nucleus within certain regions.

The E term is the system energy, y, given in the following H term relation

$$\mathbf{H} = -\frac{h^2}{8\pi^2}\left[\Sigma_i \frac{1}{mj}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + \Sigma_j \Sigma_{<j}\left(\frac{e_i e_i}{r_{ij}}\right)\right] \quad \text{..... } \textit{Eqn.}\ 3.2$$

1st term = the operative of the electron kinetic energy,
2nd term = the nuclei operator energy due to movement
Electron-nuclei attractions operator potential energy is the 3rd term,
Operator for electron-electron repulsions potential energy is estimated as the 4th term
Last term, operator for nuclei-nuclei repulsion potential energy, is equal to 5th term.

Schrodinger expression is very complex; hence it proves to be an in-executable solving molecular system, thus limited to solve the molecular structure of H_2.[81] Approximation of the equation by Born-Oppenheimer implementation is a compensation for molecular structures rather than atomic structures.

Furthermore, some few methods in QM are density functional theory (DFT), semi-empirical based calculations, and *ab initio* techniques. MP2 and CCSDT are electron correlation

methods that present better accurate QM calculations.[72] In contrast, the DFT calculations uses electron correlation approximation for analysis.[72] These methods may be used to obtain molecular system properties including but not limited to dipole moments, molecular structure equilibrium, reaction free energy, and vibrational frequencies, experimentally.

3.3 Born-Oppenheimer Approximations

Born-Oppenheimer approximation results in splitting the molecular Schrodinger equation into one part for the electronic wave function and the second part for the nuclear motions that resolve the vibration and rotation system.[72] The essential element of the approximation is as follows:

$$T_n\psi\,(R,r_1,r_2,...) = T_n[\phi(R)\,\Psi^{(R)}(r_1,r_2,...)] \cong \Psi^{(R)}(r_1,r_2,...)T_n\phi(R) \quad \text{..... } \textit{Eqn.}\ 3.3$$

The application of the nuclear kinetic energy operator in eq. 3.3 to the electronic wave function equals zero. The instantaneous positions of the nuclei are required rather than how they are moving. Practically, the approximation is used by clamping the atomic nuclei of a molecule of interest. The Born-Oppenheimer approximation is one of the best chemical physics approximations because it has shown to be valid in most situations.

Equation 3.4 presents the Schrodinger expression. It estimates the interactions of nuclei electrons in a fixed position that contained within the Software for *in silico* chemistry:

$$\mathbf{H^{elec}}\,\boldsymbol{\varphi^{elec}}\,\mathbf{(r,R) = E^{eff}\,(R)}\boldsymbol{\varphi^{elec}}\mathbf{(r,R)} \quad \text{..... } \textit{Eqn.}\ 3.4$$

3.4 Potential Energy Surface (PES)

The interactions between molecules and atoms, depending on the particles' forces, is a dynamic process. It is crucial to understand all the operating Forces within a chemical system; hence V (r) must be known. V (r) denotes the potential energy surface (PES) in a multi-dimensional complex system. Typically, the Potential energy surface (PES) is defined

within the Born Oppenheimer approximation. Electrons move faster than nuclei and adjust adiabatically to change the nuclear configuration because they are much lighter.[82] The phenomenon of Born-Oppenheimer stimulates variation in the electrons' position about the nuclei displacement, thus permitting the depiction of the potential energy surface. The potential energy of the surface exhibits electronic interaction regions unfavourable. Low energy areas serves as an nuclear reactions indicator to low energized molecular conformations and are accompanied by favourable electronic interactions.[83] Generally, reactants and products structure are stable and have relatively small energy due to position that most times correlated with a reaction coordinates' minima on the PES.

Examples of the potential energy surface from the QM methods include *ab initio* MP2 and DFTB. These energy surfaces show much lower barriers and more moderate landscapes. The electronic energy as a function of the 3N-6 internal nuclear degrees of freedom defines the potential energy surface (PES), which is very complicated and can have many minima and saddle points.

3.5 The Molecular Mechanics (MM)

Molecular mechanics (MM) is often used in large systems to calculate molecular structures and relative potential energies of molecular conformations or atomic arrangements.[84] MM is a computational method that computes the Potential energy surface for a particular sequence of atoms using Potential functions derived from classical physics. The system electrons are not explicitly considered, but every atom in the nucleus and the associated electrons is treated as a single particle. The Born-Oppenheimer approximation basis states that electronic and nuclear motions can be separated (uncouple)from each other is considered a justifiable reason for excluding electrons in MM. The difference in the energy between the system conformations during calculation is significant against the absolute potential energy values.

Molecular mechanics is an empirical method that relies on the force field parameters comprising a range of experimental data.[85]

Molecular mechanics is viewed as ball and spring atomic and model of molecule possessing classical Forces between the atoms and molecules. The Forces are obtained from the potential energy functions relating to the structures, including torsional angles, bond length, bond angles, etc. The energy contains designed parameters to reproduce exponential properties. Overall potential energy of a molecule, which expresses the MM, is described as the sum of bond-stretching energy (E_{str}), bond angle-bending energy (E_{bend}), torsion energy (E_{tor}) or twisting energy, and energy of interactions among the unbound atoms (E_{nb}). The energy contribution by the non-bonding particles is obtained from the van der Waals and electrostatic interactions and mathematically stated as follow:

$$E_{tot} = E_{str} + E_{bend} + E_{tor} + E_{vdW} + E_{elec} \quad \text{..... } Eqn.\ 3.5$$

$$E_{\text{tot}} = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{v_n}{2}[1 + cos\ (n\varnothing - r)] + \sum_{i<j} \left[\frac{A_{ij}}{r_{rj}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right] \quad \text{..... } Eqn.\ 3.6$$

E_{tot} is the potential energy, E_{vdW} is the energy contributions from van der Waal's unbound atoms. $E_{elec\ is}$ the energy given by the electrostatic interactions of the non-chemically bonded atoms.

3.6 Force fields

A force field can be defined to contain a set of functions and constants, known as parameters, which can be applied to correlate the system's energy following its particles.[86] Force field (FF) is the heart of any MD scheme and is used to describe the atomistic interactions analytically.[87] The molecular movement is governed by the atomic forces that can be divided into:

1. Those caused by interactions between atoms that are chemically

2. Those caused by interactions between atoms that are not bonded.

$$E_{total} = E_{bonded} + E_{nonbonded} \quad \text{..... } Eqn.\ 3.7.$$

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} \quad \text{..... } Eqn.3.8$$

$$E_{nonbonded} = E_{electrostastic} + E_{van\ der\ Waals} \quad \text{..... } Eqn.\ 3.9$$

FF presents a system's potential energy surface represented by a closed set of analytical potential energy functions. Moreover, the force fields' kinetic energy is also necessary; thus, the system moves across the potential energy surface's energy barriers. The implication is that the system experiences substantial changes such as conformational changes during the simulations. Provided the potential energy function can mimic the forces experienced by the 'real' atoms, the simulation results will be realistic. Equation 3.10 presents an example used to approximate the atomic forces that govern molecular movement.

$$E_{tot} = \overbrace{\sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - r\theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)]}^{\text{Bonded}} + \overbrace{\sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right]}^{\text{Non-bonded}} \quad \text{..... } Eqn.\ 3.10$$

Typically, chemical bonds and atomic angles are modelled in a typical, using simple springs (quadratic energy functions) that do not allow bond breaking. Dihedral angles are modelled using a sinusoidal function that approximate the energy differences between eclipsed and staggered conformations.

Examples of popular Force fields include GROMACS, GROMOS, AMBER, NAMD, LAMMPS, and CHARMM. They can be parametrised for atoms or biomolecules in simulation. In force fields, parameters are generated by *ab initio* method derivation or semi-empirical mechanical quantity calculations. The parameters could be experimental data sourced, e.g. X-rays, electron diffraction, NMR, and spectroscopy of neutrons..[86] There are

unique weaknesses and strengths associated with individual force field relative to the data, which allow a specific problem to be solved. Moderate low computational cost and accurate prediction protocols are incorporated in the application of force fields. Thus, force fields are attractive in molecular dynamics simulations and molecular mechanics calculations. Although there is a wide range of force fields applicable to any system, one must choose how cautiously every force field contains unique parameters designed for a specific molecule.

Therefore, choosing the correct force field is critical to achieving the study systems' theoretical structure-function's accuracy and applicability. Force fields cannot be used for bond breaking or formation calculations.[88] The AMBER force field was used in this study by applying the GENERAL AMBER FORCE FIELD (GAFF) parameters incorporated with the standard AMBER (14SB) force field[89] for the proteins.

3.7 Molecular dynamics

Molecular dynamics (MD) is a form of computer simulation in which atoms and molecules can interact for some time.[90] Simply, MD is the process of giving movement to proteins internally, which is produced by increasing the system's temperatures and cooling them rapidly in a concise time scale.[91] It is impossible to find the properties of complex systems analytically because molecular systems generally consist of a vast number of particles; MD simulation circumvents this problem by using numerical methods[92]. MD represents an interface between laboratory experiments and theory, which can be understood as a "virtual experiment".

Proteins are considered dynamic when in solutions,[93], and challenging to investigate their behaviour, structural flexibility, and motions in solution. The structure of small proteins can be solved and unravelled using the conventional techniques of x-ray crystallography techniques. The methods require strict periodic boundary conditions, which are very difficult

to obtain in non-crystalline structures.[93] Molecular dynamics simulations can predict a protein's state in solution and report the outputs as the trajectory states.[94] The MD simulation predicts the movement of large proteins in the solution, which is not possible in X-ray. MD method simulates the exact condition of a protein's existence and the output structures after MD is regarded as the best energy minimised and geometrically optimised. For this reason, the MD structure outputs are used in various experiments such as NMR, docking, and protein-ligand interactions.[95]

If the initial geometry of a system is derived from experimental data (X-ray or NMR structures), MD techniques can be used to sample the conformational space. The force and energy of all the particles within the system must be determined to set up an MD simulation protein system. MD incorporates the application of Newton's equation of motion for atoms on the energy surface. The goal of using MD is explicitly to integrate the use of Newton's equations to decipher and gain a fundamental understanding of the energies and structural deviation that may occur with a molecular network system. Nevertheless, initial states of the particles are needed. These include:

1 the position and velocity of each particle within the molecular network system,

2 a formidable force field to differentiate the forces between atoms, e.g. AMBER or CHARMM or GROMACS,

3 boundary conditions must be specified

4 The classical equation of motion can therefore be solved

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \quad \text{..... } \textit{Eqn.}\ 3.11$$

Terms denotation:

R_i (t) represents the particle position vector

T denoted the time-evolution,

M represents the mass of the particles

F_i implies the interacting force on the particles

The use of MD can be structured into four ongoing technical steps that are continuously processed to generate a trajectory. The four technical steps include:

1. the fundamental requirements (states) of the biomolecular system are defined:

a. the coordinates of each atom within the MD system

b. the bond characteristics present between each atom

c. the acceleration of each atom

2. The potential energy of each atom must be computed.

3. The equation of motion may be solved, integrating the potential energies extracted from step 2.

4. The cycle will proceed to step 1 once the new "state" of the system is saved and coordinated by each atom is changed.

The quantitative analysis will be performed on the system's time-evolution after the trajectory has been completed.

3.8 Post Molecular dynamics analysis

The outputs (trajectories) of the molecular dynamic simulation are obtained directly from the production run. The MD simulations trajectories could be outlined as successive snapshots defined by each atom's coordinate and velocity within the system simulated with time progression in phase space.[84] Essentially, it is necessary to consider the following when choosing a software to perform molecular dynamics analysis:

1. high qualitative visualisation properties; it enables accurate depiction of the trajectory's video clips and generation of high-quality snapshots/images

2. such software must be able to withhold and process large data volumes efficiently

3. contains a range of analytical tools that easily accessible within one program

The chosen post-MD analysis techniques must be directly dependent on the nature of the MD study. Detailed analysis remains critical to support any visualisation, even though the qualitative evaluation of data is essential. The following reasons detailed the need for the post-MD analysis of the trajectories:

1 determine the energetic and stability of conformation of the simulated system

2 calculate the thermodynamic energy fluctuations along the assembled system's trajectory

3 determine and characterise the binding landscape of small molecules

4 presentation of the dynamic conformational changes and variability that may occur within the bio-system throughout the MD trajectory. i

3.9 Stability of Simulated systems

3.9.1 System convergence (Stability)

Convergence of a system in molecular dynamics is commonly used to describe a protein structure's stability considering the bond types and vibration of bond angle during the system's structural motion.[96] A simulated system must represent a state of equilibrium to analyse a molecular dynamics trajectory accurately, which indicates energetic states and conformational plateau.[97] Thence, this plateau may be directly linked to a protein-ligand system's energetically stable conformations.

3.9.2 Root mean square deviations (RMSD)

The structural variation of systems such as protein-protein, protein-ligand can be obtained by the spatial differentiation between two static structures of the same trajectory.[98] This

(deviation measurement) is known as C-α atoms RSMD of an MD trajectory and can be calculated

$$\mathbf{RMSD} = \left(\frac{\sum_{N}(\mathbf{R_i} - \mathbf{R_1^0})^2}{\mathbf{N}}\right)^{\frac{1}{2}}$$ ……………………………………….*Eqn.* 3.12

Terms denotation equation:

N denotes the number of atoms in a protein-ligand complex

R_i indicates the vector position of the C-α atoms of particle ***i***

i is the conformation reference calculated after aligning the structure to an initial conformation (O) using the least square fitting protocol.

The average RMSD of a protein complex is the average structural deviation over each trajectory's number of frames. It can be calculated for simulated systems, including ligand, receptor, and complex.[11]

3.9.3 Radius of gyration (RoG)

The radius gyration of protein-ligand complex measures the Root Mean Square distance of the atoms from their common gravity centre.[99, 100] The C-α atom RoG estimates the compactness of the protein complex along a trajectory. The radius of gyration of the complex can be mathematically expressed from the following expression:

………………………………. *Eqn.* 3.13

Terms denotation in the expression are defined below:

r_i: = position of the i^{th} atom, W = mass/weight of each particle, and r = centre mass of atom i.

The mean value is calculated by estimating the C-α atoms RoG values over the frames number.

3.10 Conformational features of protein-drug systems

3.10.1 Root mean square fluctuations (RMSF)

The Root Mean Square fluctuation (RMSF) of a protein estimates the C-α atom fluctuations based on the average protein structure along the system's trajectory.[101] It further hypothesises the flexibility of a protein's region based on computed RMSF.[102] Standardised RMSF values are calculated using the following mathematical expression:

…… *Eqn.* 3.14

Terms denotations:

$RMSF_i$ denotes RMSF of the i^{th} residue from which the average RMSF is subtracted, divided by the RMSF's standard deviation [σ(RMSF)] to yield the resultant standardised RMSF $s(RMSF_i)$].

RMSF is a post-MD analysis different from the RMSD and RoG methods because it computes the total residue fluctuation along the trajectory instead of analysing every frame in the output trajectories.

3.11 Thermodynamics calculations

In computational chemistry, thermodynamics validates a profound understanding of chemical reactions, calculation of molecular properties and their derivatives, and enables the predictions of the chemical reactivities.[103] The essential usefulness of thermodynamics calculations is detailed because of its present contribution to quantum mechanics. The application of thermodynamics in drug design and development in both academic and commercial has continued to increase.[104] A critical aspect of drug design and development is optimising interactions of molecules between a drug agent and its binding target. The characterisation of the complex's binding affinity provides information about the balance of

energetic forces that drive the binding interactions.[104] This (thermodynamics) is critical for optimising molecular interactions.

Measurement of binding by the thermodynamic method is envisaged to provide useful insights regarding the nature and balance of drug-target interactive behaviour. Some binding interactions may be significantly more accessible in rational drug design, while some can be extremely difficult. The crucial parameter used to describe the molecular interactions between two bound partners such as protein-ligand and protein-protein systems is the free energy, ΔG, where the magnitude and the sign represent the possibility of biomolecular interactions.

3.11.1 Binding free energy calculations

Generally, binding free energy calculations play a significant role in correlating the structure and function of proteins. It provides information about the mechanism of binding between a ligand and an enzyme by integrating enthalpy and entropy terms.[105] The eventual aim of a structure-based drug design is a simple, robust process ranging from a high-resolution crystal structure to a validated biomolecular target. It reliably produces an easily synthesised, high-affinity small molecule with desired pharmacological properties. The calculation of binding free energy steers in developing different algorithms and approaches, including thermodynamic integration, linear interaction energy, free energy perturbation, and molecular docking calculations.[106, 107] There are large numbers of methods for computing binding free energies with atomistic molecular models. However, most of them are still under active study, and each of them has different trade-offs between accuracy and computational efficiency.

Two commonly used methods for calculating the binding energy between a small ligand and a biological macromolecule include the Molecular Mechanics/Poisson-Boltzmann Surface

Area (MM-PBSA) the Molecular Mechanics/Generalised Born Surface Area (MM-GBSA) approaches.[108, 109] The duo method depends on MD simulations of the complex systems to compute painstaking statistical-mechanical binding free energy within a specified force field.[110] MM-PBSA and MM-GBSA exhibit favourable use, while each of them displays keen accuracy and computational effort between empirical scoring and rigorous alchemical perturbation methods. The MM-PBSA is more stringent than the MM-GBSA and subsequently substitutes the latter approach in water.[111] But for the calculations involving protein-drug interaction, including nucleic acids and carbohydrates,[112] MM-GBSA method is the preferred.[113]

MM-PBSA and MM-GBSA methods

This (MM-PBSA) is the end-point free energy calculation method molecular mechanics with Poisson-Boltzmann and surface area (MM-PBSA).[105] It is a compromise between speed and accuracy for physics-based estimates of protein-ligand binding affinities. The method requires direct simulations of the bound and unbound states, being an end-point method. The simplification is based on the expectation of significantly larger intrinsic errors with MM-PBSA than other rigorous methods.[114] The application of binding free energy calculation may enhance virtual screening and molecular docking of therapeutic drugs.[115] The binding free energy between the receptor and ligand is given by:[116]

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{receptor}} - \Delta G_{\text{ligand}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Eqn.}\ 3.15$$

$$\Delta G_{\text{bind}} = E_{\text{gas}} + G_{\text{sol}} - T\Delta S \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Eqn.}\ 3.16$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Eqn.}\ 3.17$$

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{SA}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Eqn.}\ 3.18$$

$$G_{\text{SA}} = \gamma\text{SASA} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Eqn.}\ 3.19$$

Where ΔEgas denotes the molecular mechanical energy of the system in a vacuum,

ΔG_{GB} denotes the solvation free energy, $T\Delta S$ denotes the entropy term,

ΔE_{int} is the sum of bonded internal energy, ΔE_{vdW} represents the non-bonded van der Waals,

ΔE_{elec} is electrostatic energy.

Note that ΔG_{GB} consists of polar contributions that are accounted for by the generalised Born model and ΔG_{SA} denotes non-polar contribution.[108]

The dynamics analysis of binding affinity helps to determine the approximate inhibitory activity of each inhibitor.[114]

3.12 Principal Component Analysis (PCA)

The principal component analysis (PCA) is a technique that helps to understand the dynamic behaviour of biological systems.[117] PCA describes the atomic displacement and identifies conformation changes by removing various conformational modes of the protein complex during the MD simulation.[118] This (PCA) analysis defines the eigenvectors, the direction of motion and the eigenvalues, magnitude for protein systems and are carried out by constructing a covariance matrix of the C-α atom displacements.[11] Before the principal component analysis, the counterions (Na+) and solvent molecules from the MD trajectories are stripped.[119] After that, PCA will be performed on all the C-α backbone atoms, and average the data over several snapshots taken at a specified time (ps or ns) intervals using the CPPTRAJ module in AMBER 14 or any other simulation code to generate principal components. The first two principal components (PC1 and PC2) are often considered, which corresponds to the matrix's first two eigenvectors. The trajectories are presented as scatter plots generated using the Origin Software or other visual molecular dynamics Software. Here, Origin was used to create and analyse the plotted data.

3.13 Solvent Accessible Surface Area (SASA)

Solvent accessible surface area, often referred to as SASA, is an analysis used to measure a protein's exposure with solvent (water in most cases). In general, amino acid residues located on a protein's surface serve as active sites to interact with molecules and ligands. The accessibility of solvent plays a vital role in the structure and function of biological macromolecules. Essentially, SASA provides insight into the level of exposure of the protein residues buried in the active site to contribute to the hydrophobic stabilisation of a protein.[120] The hydrophobic effect is the driving for in protein folding wherein hydrophobic residues drift away from water into the solvent-shielded hydrophobic core. Protein folding occurred when a polypeptide changes from an unfolded state to its native state. When proteins transitioned from an unfolded state to their native state (folded), SASA is lost between the two states.[121] The calculation SASA for the unfolded protein is more complicated. Muller and colleagues (1987) developed the SASA calculation method to calculate SASA for polar, non-polar and total molecular surfaces. SASA can be used to calculate the solvent accessibility of molecules from their 3D structure and predict the accessibility of amino acid sequence using the computational procedure. SASA may also be used to identify the active site residues, binding sites in DNA binding proteins and functionally essential residues in membranes proteins. In this study, SASA was used with the computational procedure, AMBER14 CPPTRAJ module.[122] The trajectories are presented as line plots using the Origin Software or other visual molecular dynamics Software. Here, Origin was used to generate and analyse the plotted results.

3.14 Define Secondary Structure of Protein (DSSP)

Define Secondary Structure of Proteins is the standard tool for explicating secondary elements, including alpha helices, beta sheets, and coils from protein structures.[123] The secondary protein structures of result from the formation of hydrogen bonds formed between atoms of the polypeptide backbones. DSSP is based primarily on hydrogen bonding patterns

and some geometric constraints and assigns every residue to one of eight possible states. The corresponding forms of helical structures are α-, 3_{10}-, and π-helices. The typical secondary structure is the beta-pleated sheet. The β-bridges are short fragments that display β-sheets such as binding patterns. The β-ladder is formed from the multiple consultive β-bridges formed from the state. DSSP does not have a separate state for β-sheet residues, and the one for β-ladders can be used. The remaining three states are turn, bend, and others describe the loop structures. DSSP was used with the computational procedure, AMBER14 CPPTRAJ module.[122] The method allows a database of secondary structure assignments (DSSP) to be carried out to study the secondary structure elements. The trajectories are obtained at various time interval and presented as plots using the Origin Software as visual molecular dynamics Software. Here, Origin was used to generate and analyse the plotted results.

3.15 Dynamic cross-correlation matrices (DCCM)

Dynamic cross-correlation matrices (DCCMs) are one of the useful post MD analysis that aids in understanding correlated motions of residual-based fluctuations throughout a simulation. DCCM analysis is a 3D representation that graphically illustrates the time-correlated information among the protein systems' residues.[91, 124] Residue-based time-correlated data can be analysed using visual pattern recognition.[91, 124] This analysis (DCCM) is taken to depict the cross-correlated displacements of backbone C-α atoms in the trajectories using the following equation:

……. *Eqn.* 3.21

***Terms definitions**:*

C_y = correlation coefficient, $i = i^{th}$ residues, $j = j^{th}$ residue, where i^{th} and j^{th} are corresponding displacement vectors, respectively. Also, Δr_i and Δr_j denote the displacement of i^{th} and j^{th} atom from the mean, respectively. The cross-correlation coefficient, Cy, scales from -1 to +1.

The upper and lower limits correspond to a fully correlated (+1) and anti-correlated (-1) motion throughout the MD simulation. Like other post MD analyses, DCCM analysis can be executed using the CPPTRJ module of AMBER14 or AMBER18. Again, the Origin visual Software is used to generate and analyse the matrices.

3.16 Other Drug design techniques applied in the Study

3.16.1 Molecular docking

Molecular docking is one of the most employing computational tools nowadays. It involves determining the most optimal position of two molecules to each other. Molecular docking finds its application in the structured-based drug and has been reported as partly the main contributing factor to solving many global optimisation problems.[125] The molecular recognition of the lock and key mechanism existing between a ligand and a receptor determines their dynamical interaction level.[126] In bio-computational and drug design, docking is mainly defined as the positioning of a small molecule (e.g. an inhibitor or chemotype) called a ligand into the active site of biological molecules of a known structural conformation.[127] Successful prediction of a target protein binding site or cavity for small molecules is crucial in structure-based drug design because it allows the screening of virtual libraries[128] of "drug-like" molecules. The binding energy of the ligand-receptor complex is calculated[129] as follows:

$$E_{\text{binding}} = E_{\text{target}} + E_{\text{ligand}} + E_{\text{target-ligand}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad Eqn.\ 3.22$$

There are many molecular docking programs available for academic and commercial purposes. Examples of the Softwares include FlexX, AutoDock GOLD, GLIDE Ligand docking, and SurfflexV. Every docking program displays sufficient precision to the other, but the Autodock-Tool generates two-order of magnitude compared with the other Tools. Autodock-Tool is substantially accurate in its binding mode predictions.[130] The docking

method employed in this study is the advanced version of AutoDock Vina.[131] The principle is based on neglecting the presence of proton of the enzyme and inhibitors while generating the binding affinity. Thus, the scores are often said to be unreliable. However, the molecular dynamics do consider the protons and the solvent (water in most case). Therefore, the MD score for the binding affinity is more accurate than the docking methods.

3.16.2 Homology modelling

The prioritise aim of drug discovery is orchestrating bioactive molecules intended to target a disease condition with minimal side effects and significant efficacy.[132] The crystal structure of a pathogen's protein is fundamental in drug design/development for the productive outcome, especially amid newly evolving infection.[133] Some of the techniques used to make biological molecules' structures include electron spectroscopy, NMR spectroscopy, and X-ray crystallography. However, these (experiment-based) techniques have multiple shortcomings;[134], for instance, high-cost increments, considerable experimental time, among others, are associated with X-ray crystal structures of high-resolution 3D-structural properties.[134] Homology modelling (HM) is an emerging technique gaining prominent and preferential applications to construct 3D macromolecular structures.

In comparison to the other methods, homology modelling is easy and accurate. HM uses a protein sequence whose crystal structure is a reference template to construct the protein structure.[132] Homology model accuracy depends on the availability, detection, and quality of known template structures. Homology modelling does not consume time, cheap to use; many HM studies can be obtained within a short time. A 3D model of a target protein may enable complete characterisation and exhibition of the mechanism of interaction on a protein-ligand complex structural and molecular level, elucidating the mode of action (MOA) of drug molecules and thus facilitate drug design.

References

1. Mucs, D. & Bryce, R. A. The application of quantum mechanics in structure-based drug design. *Expert Opin. Drug Discov.* **8**, 263–276 (2013).

2. Ugi, I. *et al.* Computer Assisted Solution of Chemical Problems—The Historical Development and the Present State of the Art of a New Discipline of Chemistry. *Angew. Chemie Int. Ed. English* **32**, 201–227 (1993).

3. Baldi, A. Computational approaches for drug design and discovery: An overview. *Syst. Rev. Pharm.* **1**, 99–105 (2010).

4. Medina-Franco, J. L. Advances in computational approaches for drug discovery based on natural products. *Rev. Latinoam. Quim.* **41**, 95–110 (2013).

5. Hung, C. L. & Chen, C. C. Computational approaches for drug discovery. *Drug Dev. Res.* **75**, 412–418 (2014).

6. Ekins, S., Mestres, J. & Testa, B. In silico pharmacology for drug discovery: Methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* **152**, 9–20 (2007).

7. Noori, H. R. & Spanagel, R. In silico pharmacology: drug design and discovery's gate to the future. *Silico Pharmacol.* **1**, 1–2 (2013).

8. Arodola, O. A. & Soliman, M. E. S. Quantum mechanics implementation in drug-design workflows: Does it really help? *Drug Des. Devel. Ther.* **11**, 2551–2564 (2017).

9. Ou-Yang, S. S. *et al.* Computational drug discovery. *Acta Pharmacol. Sin.* **33**, 1131–1140 (2012).

10. Shanmugam, G. & Jeon, J. Computer-aided drug discovery in plant pathology. *Plant Pathol. J.* **33**, 529–542 (2017).

11. Baig, M. H., Ahmad, K., Rabbani, G., Danishuddin, M. & Choi, I. Computer Aided Drug Design and its Application to the Development of Potential Drugs for Neurodegenerative Disorders. *Curr. Neuropharmacol.* **16**, 740–748 (2018).

12. Kotelnikova, E., Yuryev, A., Mazo, I. & Daraselia, N. Computational approaches for drug repositioning and combination therapy design. *J. Bioinform. Comput. Biol.* **8**, 593–606 (2010).

13. Bialynicki-Birula, I. On the linearity of the Schrödinger equation. *Brazilian J. Phys.* **35**, 211–215 (2005).

14. Schrödinger, E. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.* **28**, 1049–1070 (1926).

15. Tomasi, J., Mennucci, B. & Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **105**, 2999–3093 (2005).

16. Barde, N. P., Patil, S. D., Kokne, P. M. & Bardapurkar, P. P. Deriving time dependent Schrödinger equation from Wave-Mechanics, Schrödinger time independent equation, Classical and Hamilton-Jacobi equations. 31–48 (2015).

17. Fisk, G. A. & Kirtman, B. Study of the born-oppenheimer approximation with application to H 2. *J. Chem. Phys.* **41**, 3516–3521 (1964).

18. Truhlar, D. G. *Potential Energy Surfaces*. *The enctclopedia of Physical Science and Technology* **13**, (Academic Press, 2001).

19. Long, J. *Use of potential energy surfaces (PES) in spectroscopy and reaction dynamics*. *Chemistry Department, University of Iceland Use* (2010).

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

21. Bornemann, F. A., Nettesheim, P. & Schütte, C. Quantum-classical molecular dynamics as an approximation to full quantum dynamics. *J. Chem. Phys.* **105**, 1074–1083 (1996).

22. González, M. A. Force fields and molecular dynamics simulations. *Collect. SFN* **12**, 169–200 (2011).

23. Harrison, J. A. *et al.* Review of force fields and intermolecular potentials used in atomistic computational materials research. *Appl. Phys. Rev.* **5**, (2018).

24. Honarparvar, B., Govender, T., Maguire, G. E. M., Soliman, M. E. S. & Kruger, H. G. Integrated approach to structure-based enzymatic drug design: Molecular modeling, spectroscopy, and experimental bioactivity. *Chem. Rev.* **114**, 493–537 (2014).

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

26. Bopp, P. A., Buhn, J. B., Maier, H. A. & Hampe, M. J. Scope and limits of molecular simulations. *Chem. Eng. Commun.* **195**, 1437–1456 (2008).

27. Arnold, G. E. & Ornstein, R. L. Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: Cytochrome P450BM-3. *Biophys. J.* **73**, 1147–1159 (1997).

28. De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **59**, 4035–4061 (2016).

29. Yang, L. Q. *et al.* Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms. *J. Biomol. Struct. Dyn.* **32**, 372–393 (2014).

30. Scott A. Hollingsworth, R. O. D. Molecular dynamics simulation for all. *Neuron* **99**,

139–148 (2018).

31. Ferreira, L. G., Dos Santos, R. N., Oliva, G. & Andricopulo, A. D. *Molecular docking and structure-based drug design strategies*. *Molecules* **20**, (2015).

32. 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).

33. Amadei, A., Ceruso, M. A. & Di Nola, A. On the convergence of the conformational coordinates basis set obtained by the Essential Dynamics analysis of proteins' molecular dynamics simulations. *Proteins Struct. Funct. Genet.* **36**, 419–424 (1999).

34. Kufareva, I. & Abagyan, R. Methods of protein structure comparison. *Methods Mol. Biol.* **857**, 231–257 (2012).

35. Adewumi, Adeniyi T, Ajadi, Mary B, Soremekun, Opeyemi S. and Soliman, M. E. S. Thompson loop : opportunities for antitubercular demethylmenaquinone methyltransferase protein †. *RSC Adv.* **10**, 23466–23483 (2020).

36. Lobanov, M. Y., Bogatyreva, N. S. & Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* **42**, 701–706 (2008).

37. Monajjemi, M. & Oliaey, A. Gyration Radius and Energy Study at Different Temperatures for Acetylcholine Receptor Protein in Gas Phase by Monte Carlo, Molecular and Langevin Dynamics Simulations. *Proteins* **5**, 195–201 (2009).

38. Martínez, L. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS One* **10**, 1–10 (2015).

39. 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).

40. Brien, R. O., Markova, N. & Holdgate, G. A. *Thermodynamics in Drug Discovery*. *Applied Biophysics for Drug Discovery* (John Wiley & Sons Ltd, 2017). doi:doi.org/10.1002/9781119099512.ch2

41. Chaires, N. C. G. and J. B. Thermodynamic Studies for Drug Design and Screening. *Expert Opin Drug Discov.* **7**, 299–314 (2012).

42. Agoni, C., Ramharack, P. & Soliman, M. E. S. Co-inhibition as a strategic therapeutic approach to overcome rifampin resistance in tuberculosis therapy: Atomistic insights. *Future Med. Chem.* **10**, 1665–1675 (2018).

43. Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **93**, 2395–2417 (1993).

44. Kamerlin, S. C. L., Haranczyk, M. & Warshel, A. Progress in Ab initio QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *J. Phys. Chem. B* **113**, 1253–1272 (2009).

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

46. Wang, J., Hou, T. & Xu, X. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput. Aided-Drug Des.* **2**, 287–306 (2006).

47. Hou, T., Wang, J., Li, Y. & Wang, W. Assessing the Performance of the MM_PBSA and MM_GBSA Methods. 1. The Accuracy.pdf. *J. Chem. Inf. Model* **51**, 69–82

(2011).

48. Tsui, V. & Case, D. A. Theory and applications of the Generalized Born solvation model in macromolecular simulations. *Biopolymers* **56**, 275–291 (2000).

49. Feig, M. & Brooks, C. L. Evaluating CASP4 predictions with physical energy functions. *Proteins Struct. Funct. Genet.* **49**, 232–245 (2002).

50. Sun, H., Li, Y., Tian, S., Xu, L. & Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* **16**, 16719–16729 (2014).

51. Wright, D. W., Hall, B. A., Kenway, O. A., Jha, S. & Coveney, P. V. Computing clinically relevant binding free energies of HIV-1 protease inhibitors. *J. Chem. Theory Comput.* **10**, 1228–1241 (2014).

52. Aalten, D. M. F. V., Findlay, J. B. C., Amadei, A. & Berendsen, H. J. C. Essential dynamics of the cellular retinol-binding protein evidence for ligand-induced conformational changes. *Protein Eng. Des. Sel.* **8**, 1129–1135 (1995).

53. Decherchi, S., Masetti, M., Vyalov, I. & Rocchia, W. Implicit solvent methods for free energy estimation. *Eur. J. Med. Chem.* **91**, 27–42 (2015).

54. Bhakat, S., Martin, A. J. M. & Soliman, M. E. S. An integrated molecular dynamics, principal component analysis and residue interaction network approach reveals the impact of M184V mutation on HIV reverse transcriptase resistance to lamivudine. *Mol. Biosyst.* **10**, 2215–2228 (2014).

55. Post, M., Wolf, S. & Stock, G. Principal component analysis of nonequilibrium molecular dynamics simulations. *J. Chem. Phys.* **150**, 1–11 (2019).

56. Soremekun, O. S., Olotu, F. A., Agoni, C. & Soliman, M. E. S. Recruiting monomer for dimer formation: resolving the antagonistic mechanisms of novel immune check point inhibitors against Programmed Death Ligand-1 in cancer immunotherapy. *Mol. Simul.* **45**, 777–789 (2019).

57. Durham, E., Dorr, B., Woetzel, N., Staritzbichler, R. & Meiler, J. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *J. Mol. Model.* **15**, 1093–1108 (2009).

58. Lu, S. & Wagaman, A. S. On methods for determining solvent accessible surface area for proteins in their unfolded state. *BMC Res. Notes* **7**, 1–7 (2014).

59. Wang, J. & Hou, T. Develop and test a solvent accessible surface area-based model in conformational entropy calculations. *J. Chem. Inf. Model.* **52**, 1199–1212 (2012).

60. Tanwar, Himani and Doss, C. G. P. An Integrated Computational Framework to Assess the Mutational Landscape of α-L-Idurondase IDUA Gene. *J. Cell. Biochem.* 555–565 (2018). doi:10.1002/jcb.26214

61. 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).

62. Eweas, A. F., Maghrabi, I. A. & Namarneh, A. I. Advances in molecular modeling and docking as a tool for modern drug discovery. *Der Pharma Chem.* **6**, 211–228 (2014).

63. Du, X. *et al.* Insights into protein–ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* **17**, 1–34 (2016).

64. Ramachandran, K. I., Deepa, G. & Namboori, K. *Computational chemistry and molecular modeling: Principles and applications*. *Computational Chemistry and*

Molecular Modeling: Principles and Applications (2008). doi:10.1007/978-3-540-77304-7

65. Talele, T., Khedkar, S. & Rigby, A. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem.* **10**, 127–141 (2010).

66. Nervall, M. *Binding Free Energy Calculations on Ligand-Receptor Complexes Applied to Malarial Protease Inhibitors*. (2007).

67. Vajda, S. & Kozakov, D. Convergence and combination of methods in protein-protein docking. *Curr. Opin. Struct. Biol.* **19**, 164–170 (2009).

68. 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).

69. Toomula, Nishant,, D Sathish Kumar, V. P. K. A. Computational Methods for Protein Structure Prediction and Its Application in Drug Design. *J. Proteomics Bioinfor atics* **4**, 289–293 (2011).

70. Ashburn, T. T. & Thor, K. B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–683 (2004).

71. Forrest, L. R., Tang, C. L. & Honig, B. On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys. J.* **91**, 508–517 (2006).

CHAPTER 4

Delving into the Characteristic Features of "Menace"a *Mycobacterium tuberculosis* Homologs: A Structural dynamics and Proteomics Perspectives

Adeniyi T. Adewumi[a], Pritika Ramharack[a], Opeyemi S. Soremekun[a], Mahmoud E. S. Soliman[a]*

[a]Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

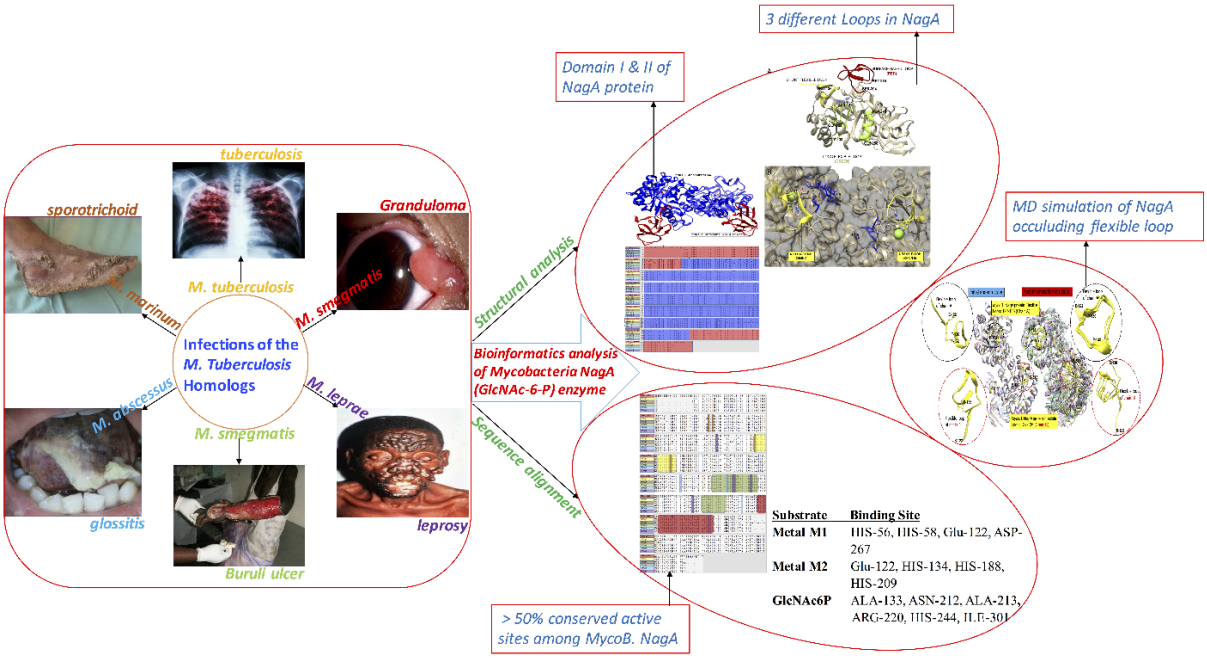
*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 8048

Fax: +27 (0) 31 260 7872

4.1 Graphical Abstract



Human Mycobacteria diseases; detrimental morbidity rate, WHO (https://www.who.int/tb/areas-of-work/zoonotic-tb/en/), Padayatchi *et al* (2017) *SAMJ, S. Afr. med. J; 107 (6)*

4.2 Abstract

The global increase in the morbidity/mortality rate of Mycobacterial infections, predominantly renascent tuberculosis, leprosy, and Buruli ulcers have become worrisome over the years. More challenging is the incidence of resistance mediated by mutant *Mycobacterium strains* against front-line antitubercular drugs. Homologous to all *Mycobacteria* species is the GlcNAc-6-phosphate deacetylase (NagA), which catalyses essential amino sugars synthesis required for cell wall architecture, hence, metamorphosing into an important pharmacological target for curtailing virulence and drug-resistance. This study used integrated bioinformatics methods, MD simulations, and *DynaMut* and *PolyPhen2* to; explore unique features, monitor dynamics, and analyze the functional impact of non-synonymous single-nucleotide polymorphisms of the six NagA of most ruinous Mycobacterium species; *tuberculosis* (*Mtb*), *smegmatis* (*MS*), *marinum* (*MM*), *ulcerans*, *africanum*, and *microti respectively*. This approach is essential for multi-targeting and could result in the identification of potential polypharmacological antitubercular compounds. Comparative sequential analyses revealed ≤ 50% of the overall structure, including the catalytic Asp267 and reactive Cys131, remained conserved.

Interestingly, *MS*-NagA and *MM*-NagA possess unique hydrophobic isoleucine (Ile) residues at their active sites in contrast to leucine (Leu) found in other variants. More so, unique to the active sites of the NagA is a 'subunit loop' that covers the protein active site; probably crucial in binding (entry and exit) mechanisms of targeted NagA inhibitors. Relatively, nsSNP mutations exerted a destabilizing effect on the native *NagA* conformation. Structural and dynamical insights provided, basically pin-pointed the "Achilles' heel" explorable for the rational drug design of target-specific 'NagA' inhibitors potent against a wide range of mycobacterial diseases.

Keywords: Mycobacterial infections, peptidoglycan, bioinformatics tools, Molecular dynamics (MD) simulations, GlcNAc-6-phosphate deacetylase protein, , mutation

4.3 Introduction

Mycobacteria species were discovered over a century ago and include; the non-pathogenic [1], pathogenic (*tuberculosis*), *and* opportunistic pathogens (*non-tuberculosis*) [2]. Examples of pathogenic *Mycobacterium species* include; *Mycobacterium tuberculosis* (*Mtb*) that causes tuberculosis (TB) (the most-deadly disease) [3] and *Mycobacterium ulcerans* (*MU*) that causes chronic-necrotizing Buruli ulcer. [4]. Similarly, Opportunistic Mycobacteria such as *Mycobacterium marinum (MM)* and *Mycobacterium avium complex (MAC)* cause cutaneous diseases involving plastic surgery [5]. *Mycobacterium smegmatis*, a non-pathogenic agent, causes granuloma. They destroy the human immune system and activate phagocytes-causing a respiratory tract or cutaneous disease [5]. About 20 *Mycobacteria species* cause different kinds of conditions in humans, but more than 140 of them are recognized [4,7]. Majority of them are environmental opportunistic pathogens that exist as saprophytes in soil and water [6]. Table 4.1 summarizes the human clinical mycobacterial infections and their respective causative agent [7].

Table 4.1 Mycobacterial clinical diseases in humans and their causative mycobacterial species

Disease	Agent
Lymphadenopathy	*M. avium complex, M scrofulaceum*
Skin lesion: Post-trauma abscess, Swimming pool granuloma, Buruli ulcer	*M. chelonae, M. fortuitum, M. ulcerans M. terrae, M.* Mar*inum*
Pulmonary disease	*M. kansasii, M. xenopi, M. avium complex, M. tuberculosis*
Disseminated disease: AIDS-related, Non-AIDS related	*M. avium complex, M.genelonae, M. chelonae*

Mycobacterial infections have been a burden to almost every country in the world [8]. Despite the decades of "war" combating these pathogens, a greater commitment is still required to eliminate the menace of the infections. Specifically in 2018, the World Health Organization (WHO) again reiterated the estimate of TB to be; 10.0 million people had the infection of which 5.8 million were men, 3.2 million were women, and children being the remaining people [9]. TB may co-infect with either HIV/AIDS or chronic pulmonary diseases in human, [10, 11] hence, results in a quicker death of the TB patients. The new cases of MDR and XDR tuberculosis complicate the impending status of mycobacterial infections.[12] Although the TB incidence is falling by 2% annually, about 558 000 new cases of TB (known as active TB, multidrug- or extensive drug-resistant TB (MDR/XDR) occur annually in countries including China, India, and South Africa [13]. Almost one-quarter of MDR-TB cases cause resistance to at least a first-line drug including Isoniazid, rifampicin, streptomycin, and ethambutol [14]. More challenging is the increasing MDR/XDR mediated by mutations to newer drugs like bedaquiline and delamanid, and these mutation-mediated resistances have subsequently remained incurable. Thus, TB remains a public health crisis and health security threat [15,16]. Therefore, the TB morbidity and mortality rate demands urgent medical treatment and preventative measures.

Mycobacteria species encode a unique enzyme GlcNAc-6-phosphate deacetylase (Ahangar *et al*., 2018) that controls the synthesis and intracellular flow of N-acetylglucosamine-6-phosphate (GlcNAc6P), an amino sugar used by bacteria to build cellular envelopes.[17] *NagA* enzyme is abundant in nature and are found in eukaryotic and prokaryotic including; mycobacteria and non-mycobacteria organisms[17], [18] and are a member of amidohydrolase superfamily [19, 20]. The enzyme possesses a characteristic triose-phosphate isomerase (TIM)-like barrel fold and catalytic metal ions in the active region. [17, 19] All the characterized *NagA* enzymes comprise of two similar architectural domains (I and II); *Domain I* contains $(\beta/\alpha)_8$-barrel structure, [17] while *Domain II* comprises of small β-barrel with unknown biological function. However, a study by Ahangar *et al.,* (2018) suggested that *Domain II* enhances the stability of *NagA* enzyme.

The mycobacterial cell wall is made up of four thick layers, thus explaining bacteria's resistance to acid or alkaline conditions altogether called envelope [4]. The cell wall of Mycobacterial is crucial for their survival and virulence [17.] The cell wall consists of distinct and interspersed constituents including; peptidoglycan, arabinogalactan, mycolic acids, lipids, polysaccharide capsules, and proteins [21]. The mechanism of biosynthesis involves a nucleophilic attack of GlcNAc6P substrate via conserved aspartic acid residues by activating a water molecule to protonate amine the group [22]. The glucosamine-6-phosphate (GlcN6P) is isomerized to form glucosamine-1-phosphate, processed to produce peptidoglycan (Figure. 4.1). The alteration of *Mycobacterial* cell wall provides a therapeutic platform to inhibit pathogenic mycobacteria.

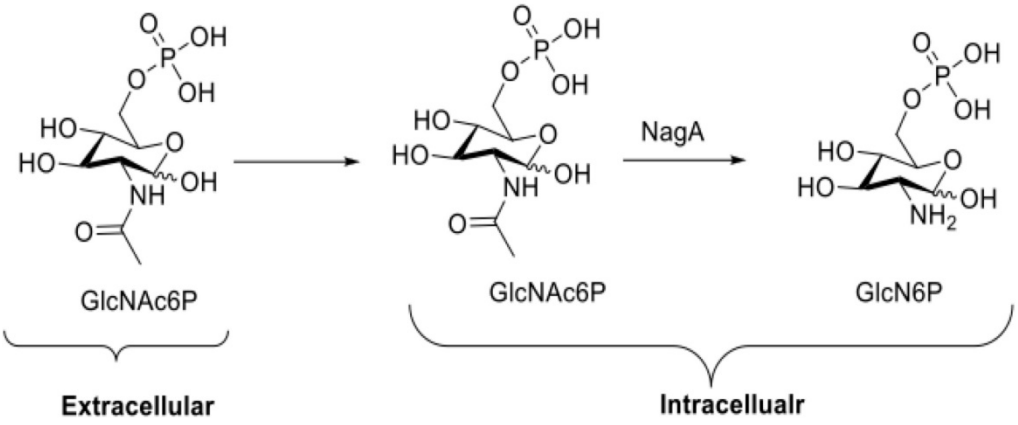


Figure 4.1. Metabolic synthesis pathway of GlcN6P by Mycobacteria NagA enzyme

Bioinformatics techniques have contributed significantly to the study of proteins by promoting the correlation between proteins sequence analysis and the characterization of intrinsic disorders [23,24]. These computational tools allow the investigation into the nature of protein disorders from large databases, including Ensemble and Swiss-Pro.[25, 26] Additionally, the use of bioinformatics tools to analyze, simulate and extrapolate chemical information allows a reduction in time-consuming and of laboratory-based research and financial expenditure [27,28]. We employed integrated bioinformatics methods to explore unique features pertinent to *Mycobacteria NagA* across six most detrimental *Mycobacterium* species; *tuberculosis* (*Mtb*), *smegmatis* (*MS*), *marinum* (*MM*), *ulcerans*, *africanum*, and *microti*. In addition, it provides a comparative analysis of the 3D homology models of six mycobacterium species *NagA* and the effect of the mutation on the protein, thus elucidating on the potential conserved drug binding sites. This information may be utilized in the identification of inhibitors that may target multiple pathogenic *mycobacteria.*

4.4 Methodology

4.4.1 Mycobacteria *NagA* amino acids sequence acquisition and homology modelling

Two X-ray crystal structures of *Mycobacterium smegmatis NagA* enzymes were retrieved from the RCSB PDB (accession ID: 6FV3 and 6FV4)[17]. The 6FV3 (Wild type) was used for

the MD simulations, and 6FV4 (mutant structure) was used for the mutation analysis. The 3D structures of the remaining five other *Mycobacterium species* NagA including *tuberculosis, smegmatis, marinum, ulcerans,* and *microti* were not available. Hence, the amino acid sequences of the enzymes were retrieved from the UniProt database ([29], http://www.uniprot.org/) (Table 4.2). We performed the homology modeling using both the SWISS-MODEL ([30], [31] online resource and the Modeller Software version 9.1 [32], [33], add-on in UCSF Chimera tools version 1.13.1 [34], in which the chain A of template (6FV3) was selected to build the model. Furthermore, multiple sequence alignment was done using the CLUSTALW server [35], [36] where the Chain A of *Mycobacterium smegmatis* (6FV3) was evidenced to have the best template with the amino acid sequences of the targets.

Table 4.2 UniProt information of homologs of the five mycobacteria NagA enzymes.

Mycobacterium species *NagA*	**Accession code**	**Amino acid residue**
MA-NagA	A0A120J2Z3	383
MC-NagA	A0A120IYI0	383
MM-NagA	B2HDT2	388
Mtb-NagA	O53382	383
MU-NagA	A0PNQ6	383

Table 4.3 showed the criteria including sequence identity, sequence similarity, and Global Model Quality Estimate (GMQE) used for choosing the template and the quality (QMEAN Z Scores) of the models. The sequence identity (%) *MANagA, MCNagA, MMNagA, MtbNAgA,* and *MUNagA* models are 56.0, 56.0, 51.61, 56.0, and 51.34 and the sequence similarity are 0.45, 0.45, 0.52, 0.45, and 0.43 respectively. GMQE value ranges between 0 and 1. The GMQE values of *Mycobacterium species NagA; MANagA, MCNagA, MMNagA, MtbNAgA,* and *MUNagA* models were 0.80, 0.80, 0.76, 0.80, and 0.46 respectively. The GMQE values

showed that the models are highly stable. To validate the models, the template and the models were uploaded on the Chimera Tool and superimposed.

Table 4.3 Summary of template criteria and Quality for choosing Homology models of Mycobacterium species NagA proteins

NagA model	Template	Seq. (%) Identity	Seq. (%) Similarity	QSQE	QMEAN	GMQE
MA-NagA	6FV3.1. A	56.00	0.45	0.44	-1.35	0.80
MC-NagA	6FV3.1. A	56.00	0.45	0.44	-1.35	0.80
MM-NagA	6FV3.1. A	51.61	0.52	0.44	-1.20	0.76
Mtb-NagA	6FV3.1. A	56.00	0.45	0.44	-1.35	0.80
MU-NagA	6FV3.1. A	51.34	0.43	0.51	-0.51	0.46

QSQE = Quartenary Structure Quality Estimate, QMEAN = scoring function (Z scores)

GMQE = Global Model Quality Estimate

4.4.2 Sequence and structural analysis

Sequence alignments and structural investigations among the active site regions of the six (6) mycobacterial *NagA* enzymes were undertaken using the alignment tools available through Chimera, with default settings applied [25]. The conserved regions across the 6 mycobacterial *NagA* were observed and classified as: fully conserved (identical), partially conserved and poorly conserved. The active sites were obtained from previous studies [17] and validated by identifying the residues interacting with the substrate (N-acetylglucosamine-6-phosphate-GlcNAc6). This was completed using the UCSF Chimera visualization software. The other vital features of Mycobacterial *NagA* enzymes include a flexible loop, conserved Asp267, binucleate metal-centre, Cys131 residue, and *NagA* mutants especially D267A.

4.4.3 Molecular dynamics (MD) simulation

MD simulation is one of the important tools for the theoretical study of molecular systems using time-dependent behaviour to provide detail information on the fluctuations, conformational flexibility, and thermodynamics of proteins or other biological systems. The

free structures of *Mycobacterium* species were prepared, by removal of the hydrogen atoms, using the UCSF Chimera software packages. The protein systems were set up for 350 ns MD simulations using the Graphic Processor Unit (GPU) version of Particle Mesh Ewald Molecular Dynamics (PMEMD) incorporated into the AMBER18 package [37]. The AMBER force field was used to parametrize the protein systems and LEAP module was used to add hydrogen atoms and counter (Na^+ and Cl^-) ions to neutralize the systems. All systems were solvated in an orthorhombic TIP3P water box of a size of 10 Å. Two minimization steps were carried out; first is a partial minimization of 2500 steps using a weak restraint potential of 10 kcal.mol-1 and second is a full minimization of 2500 steps without restraining the conjugate energy. The heating of systems occurred gradually from 0 to 300 K for 5 psec in a canonical ensemble (NVT), using harmonic restraint of 10 $kcal.mol^{-1}$ Å and a Langevin thermostat with a 1-ps random collision frequency. Equilibration was carried out at a temperature (T) of 300 K in NPT ensemble for 350 ns of a time step of 2 fs without any restraint using the Berendsen-Barostat was used to maintain the pressure of 1 bar.

For each system, 350 ns MD simulations were conducted in an isothermal-isobaric (NPT) ensemble using Berendsen-Barostat at a pressure of 1 bar and a pressure-coupling constant of 2 psec. The coordinates and trajectories of the apo structures were saved and analyzed using CPPTRAJ and PTRAJ modules of AMBER18 GPU. Subsequently, we determined the following post-analysis: root mean square deviation (RMSD) and root means square fluctuations (RMSF) for both the whole structure and the loop covering the active site. The visualization and graphics software such as Visual molecular dynamics was used to visualize trajectories; UCSF ChimeraTools and Maestro software were used to prepare the structures; and Origin software (OriginLab, Northampton, MA) were used for the plotting and generation of graphs.

4.5 Results and Discussion

4.5.1 Unique characteristics of structural features of Mycobacteria *NagA* proteins

The sequence analysis of the static structure of *Mycobacterium species NagA* protein revealed some similarities in the domain region, loop, mutational impact, and the reactive CYS131 residue. All the structures of *NagA* enzyme, including *non-mycobacteria* and *mycobacteria* have two overall common architectural domains [38,39]. In this study, we modelled five mycobacteria tuberculosis *NagA* homologs as the targets and subsequently aligned the structures to the crystallized structure of *MS-NagA* enzyme. Figure 4.2 represents the two structural domains present in the Mycobacterium *NagA* protein. Of importance, the amino acid sequence of *NagA* enzymes showed more than 50 % conserved residues.

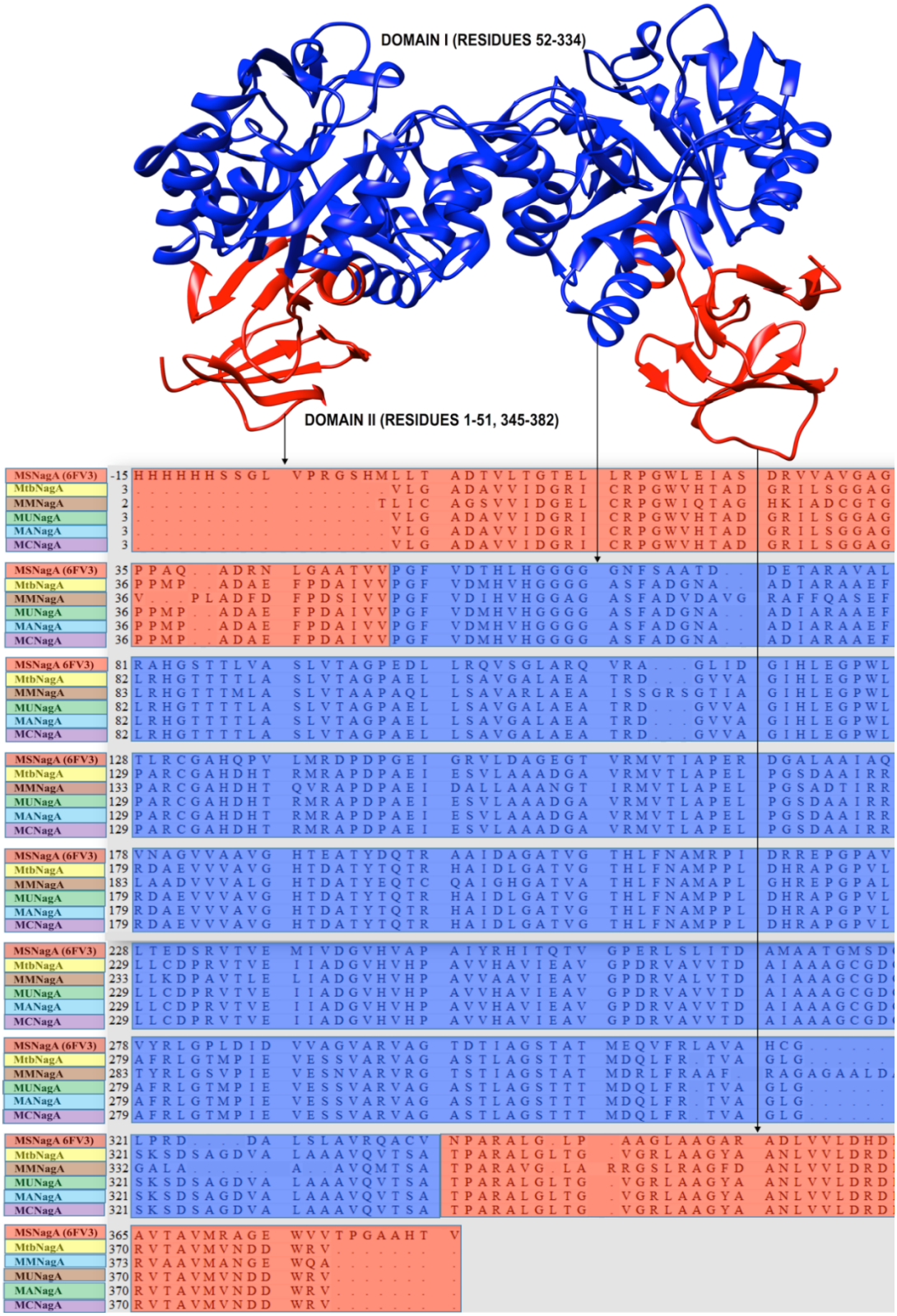


Figure 4.2. Overall structure and sequence analysis of the 6 Mycobacteria of; smegmatis, tuberculosis, marinum, ulcerans, africanum, & microti. Blue colour represents domain I (residues: 52 – 344), while the red colour represents domain II (1 – 51 & 345 – 382).

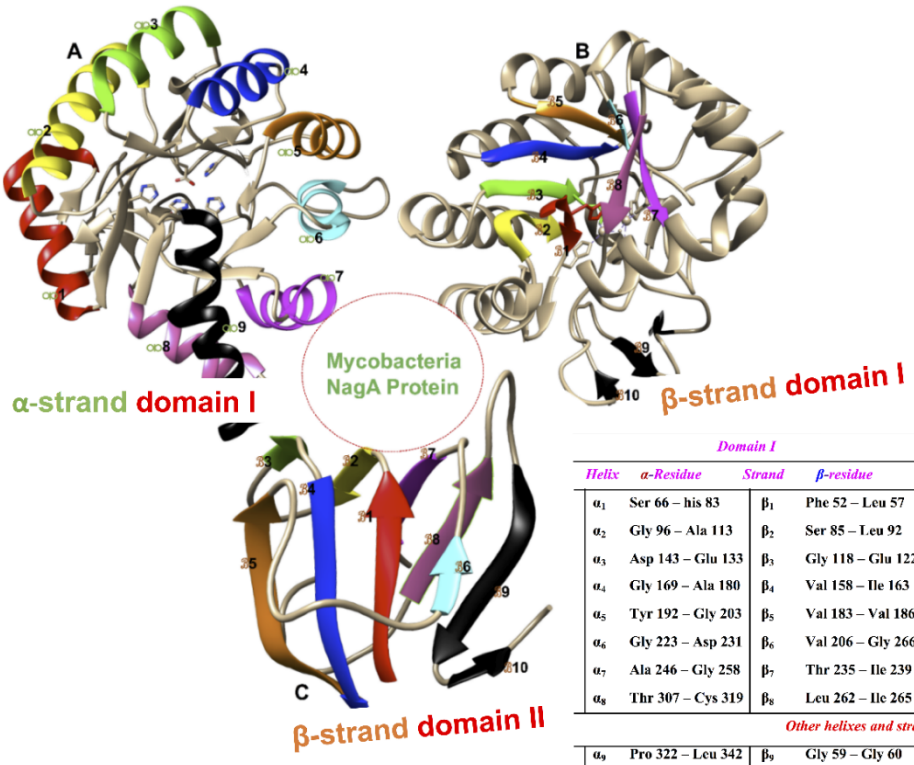
4.5.2 Domain I

The domain is a twisted $(\beta/\alpha)_8$ - triosephosphate isomerase (TIM) located within 52-344 residues which housed the ligand and the metal ion-binding sites. Figure 4.3 revealed 8 folded *alternating* β-strands and 8 α-helices[17]. Generally, *NagA* proteins undergo dimeric interactions to form a homodimer between two monomeric subunits, and each monomer contains the active sites as seen in *MS-NagA* enzyme. The H-bonds are formed between residues on α_8 and α_9 while salt bridges are also formed between Glu230 and Arg251 and both confer stability to the structure.

Roles of α-helix and β-strand barrel structures of NagA

The α/β TIM-barrel protein structures occur in about 14 enzymes including/g Mycobacteria *NagA* enzymes. The Mycobacteria *NagA* TIM-barrel enzymes assume a circular geometry and did not deviate from the paradigm-layered structure of the residues in its interior. The assembly of α-helices and β-strands secondary structures of a protein is determined by the available low-energy conformations and packing of the residues. The first six β-sheets (β_1 – β_6) in domain I are approximately parallel, while the other two are parallel to each other. However, the extra helices and stranded sheets involve structurally in the stability of the enzyme. For instance, α_9 forms an H-bond with α_8 of the neighbouring subunit monomer to sustain the stability of the homodimer. The same α_9 is also involved in the flexible loop formed from residues 203 – 223 and 248 – 258.

The β-strands are characterized by hydrophobic residues and account for why they are found in the interior of the protein structure. The domain contains more β-strands than domain II though they are not alternating. We represented α-helices and extra β-strands in black colour.



Domain I				Domain II
Helix	α-Residue	Strand	β-residue	β-residue
α_1	Ser 66 – his 83	β_1	Phe 52 – Leu 57	Met 1 – Ala 5
α_2	Gly 96 – Ala 113	β_2	Ser 85 – Leu 92	Thr 7 – Leu 9
α_3	Asp 143 – Glu 133	β_3	Gly 118 – Glu 122	Leu 14 – Leu 15
α_4	Gly 169 – Ala 180	β_4	Val 158 – Ile 163	Gly 18 – Ala 23
α_5	Tyr 192 – Gly 203	β_5	Val 183 – Val 186	Arg 26 – Ala 32
α_6	Gly 223 – Asp 231	β_6	Val 206 – Gly 266	Arg 41 – Asn 42
α_7	Ala 246 – Gly 258	β_7	Thr 235 – Ile 239	Thr 47 – Pro 50
α_8	Thr 307 – Cys 319	β_8	Leu 262 – Ile 265	Leu 357 – Leu 360
Other helixes and strands				
α_9	Pro 322 – Leu 342	β_9	Gly 59 – Gly 60	Val 366 – Arg 371
α_{10}	Pro 326 – Leu 344	β_{10}	Gly 63 – Asn 64	Glu 374 – Trp 375
		β_{11}	Gly 277 – Val 278	
		β_{12}	Ile 286 – Val 288	
		β_{13}	Ala 293 – Val 295	

Figure 4.3. Illustration of the Secondary structure of Mycobacterium smegmatis NagA enzyme

Long flexible loop: In the overall structure of Mycobacteria *NagA*, this is a formation of a long flexible loop from residues 203–223 of one subunit that extends out to α_9 of the neighbouring subunit residues 243–258. The loop provides additional stability for mycobacteria *NagA* protein and is loop is represented in green, as shown in Figure 4.4 (A).

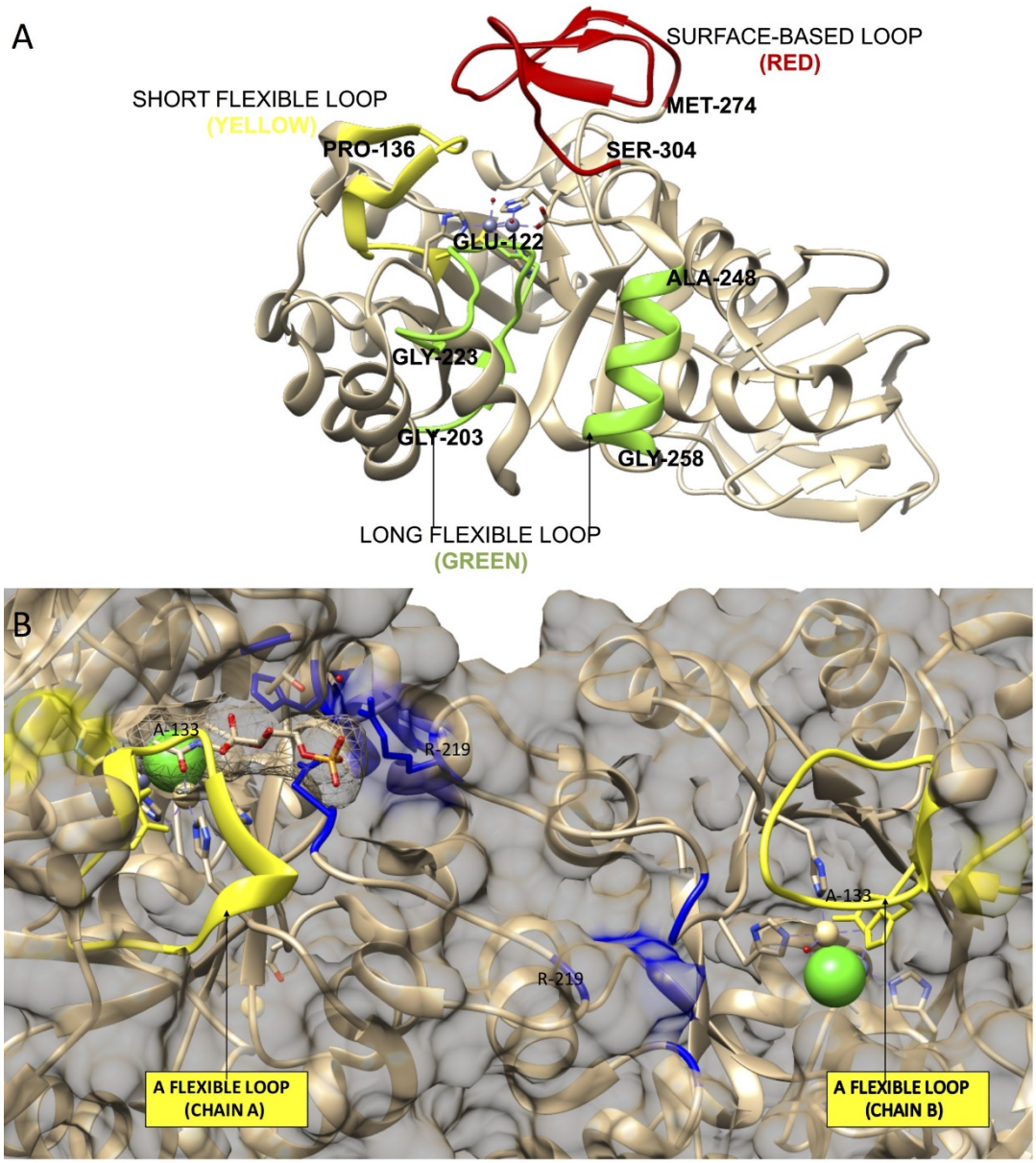


Figure 4.4. A is the Illustration of three Mycobacteria NagA loops; short flexible loop (yellow), Long flexible loop (green) and surface-based loop (red). **B** is the ligand-binding sites (blue) and a short flexible loop covering the ligand site

Surface-based loop: This second loop is formed at the surface of *NagA* and comprises of amino acids between Met274 – Ser304. It forms a dynamic "lid" that may enhance the binding of substrate and ligand through an opening/closing mechanism. This loop is represented in red in Figure 4.4 (A) structure of the Mycobacteria *NagA* enzyme.

Short flexible loop: This is another structural loop and is called a flexible loop that is formed Glu122 to Pro136 at the dimeric interface of mycobacteria *NagA* enzymes. The loop is

represented in yellow (Figure 4.4A). As observed in Figure 4.4B, the loop hinders the binding of GlcNAc6P by blocking the active site of the ligand in Chain B [17]. Hence, it showed the potential to cover any prospective inhibitor. Therefore, the occlusion provides deduction towards designing inhibitors that may prevent the synthesis of amino sugar and subsequently destroys mycobacterial cell wall synthesis. Figure 4.4B further reveals the binding of GlcNAc6P as shown in stick representation of Chain A. The ligand-binding site is shown in blue.

4.5.3 Domain II

This *domain* (II) region comprises of residues 1 – 51 and 345 – 382 containing small β-barrel strands contributed from the N- and C- termini of *NagA* enzymes. Over the years, the biological functions of *domain II* have not been understood. However, Ahangar *et al.* suggested that the β-barrel domain has a stabilizing role in the enzyme's structure. Structurally, Mycobacteria *NagA* enzyme β-barrel strands are conserved in mycobacteria NagA enzyme. Protein thermodynamics of folding is very important to protein study and is determined by protein stability and the hydrophobicity effect of the side-chain residues is the driving force [40].

However, the structural sequence analysis showed that while *MSNagaA* and *MMNagA* contain 381 and 385 residues, respectively, each of the other four *NagA* structures contain 382 residues. Hence, the superimposition of the six NagA revealed a closer similarity between *MSNagA* and *MMNagA* structures while *MANagA*, *MCNagA*, *MtbNagA*, and *MUNAgA* were similar. In addition, the first 16 residues of *MSNagA* did not have corresponding residues in any of the modeled structures, though they are not within the binding sites. Second, a hydrophobic leucine (L304) present in *MSNagA* and *MMNagA* is important for the binding of the substrate. However, we observed a variation in which the

leucine was replaced with isoleucine (I302) in *MANagA*, *MCNagA*, *MtbNagA*, and *MUNAgA* proteins as observed in Figure 4.5.

Furthermore, another variation observed in secondary structures of *MSNagA* when compared with other structures is the loops that were formed in *MANagA*, *MCNagA*, *MtbNagA*, and *MUNAgA* enzymes rather the helices found the template (*MSNagA*) structure. For instance, GLY278 and ARG281 residues constitute part of a long helix present in the *MSNagA* active site but were observed to form a loop in *MANagA*, *MCNagA*, *MtbNagA*, and *MUNAgA* enzymes.

4.5.4 Sequence analysis of Mycobacteria NagA proteins

Bioinformatics is a multidisciplinary research area [41] in which the computer is used to store, analyze, and interpret a large volume of biological information including; sequence analysis, protein structure prediction, comparative genomics, etc. [42, 43] Sequence alignment analysis has revealed great applications between the protein regulation and sequence properties. Figure 4.5 is the sequence multiple alignments of *MS-NagA* (template) and the homologs; *Mtb-NagA*, *MM-NagA*, *MU-NagA*, *MA-NagA*, and *MC-NagA* enzymes.

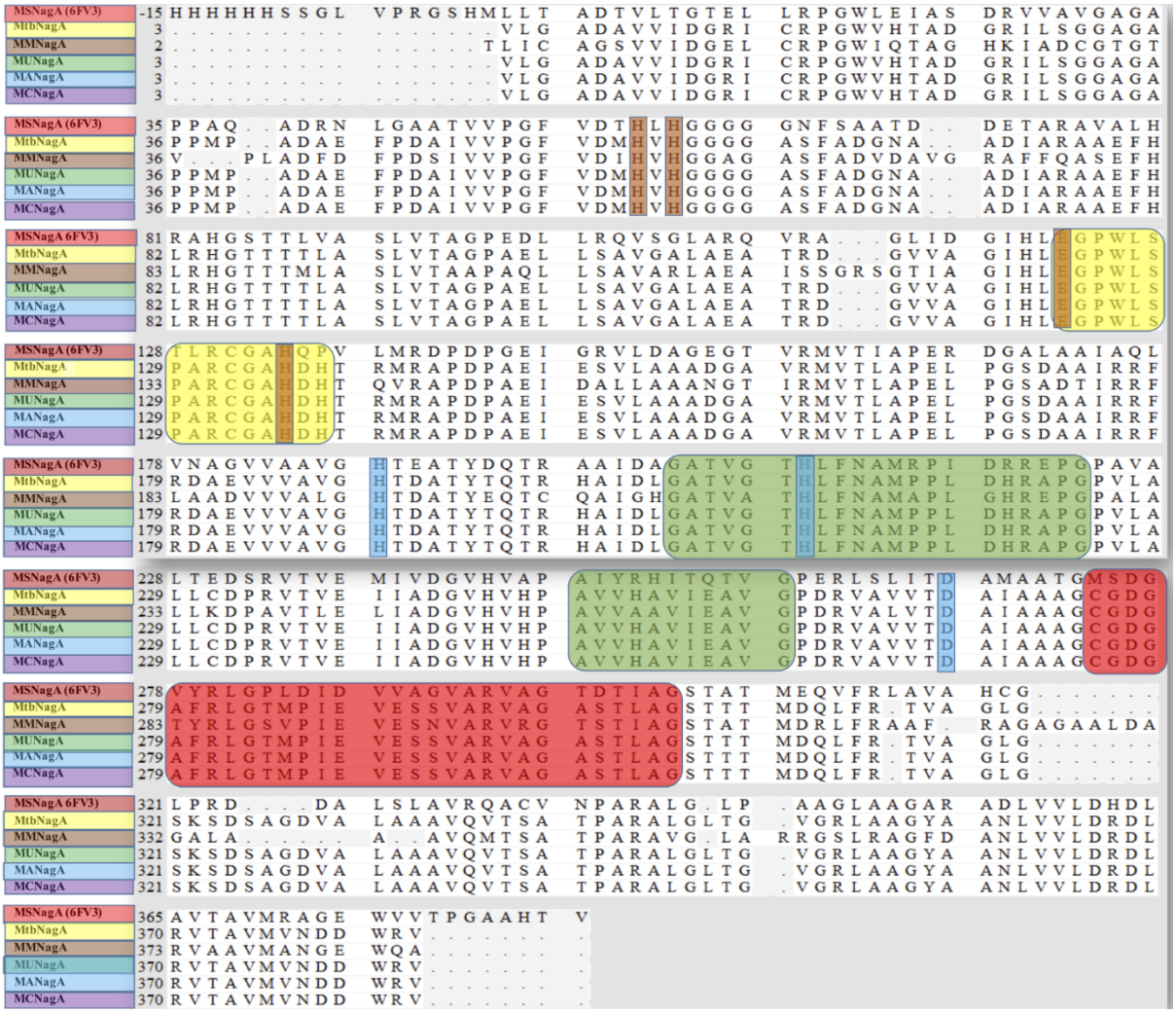


Figure 4.5. Illustration of the active site regions and sequence analysis of Mycobacterial NagA enzyme; metal- (M1 & M2) binding sites residues are highlighted in blue and brown colour, whilst the ligand-binding site residues are shown in pink colour. The HXH motif residues are seen in M1-binding site.

The region of maximum conserved residues is between Glu118 and Met214. There is a high similarity of *NagA* amino acid sequence of ≥ 45 % identical residues, 28.76 % less conserved, and 27.24 % poorly conserved protein residues among the six Mycobacteria *NagA* enzyme. The 3-D structures of *NagA* enzyme including those of Mycobacteria contain two active sites the metal-binding and the substrate-binding (GlcNAc6P) sites located within domain I [17] and provided in Table 4.4.

Table 4.4. Mycobacteria NagA active sites consisting of; M1 (Zn), M2 (Zn/Cd), and GlcNAc6P

Substrate	Binding Site Residue
Metal M1	**HIS**-56, **HIS**-58, **GLU**-122, **ASP**-267
Metal M2	**GLU**-122, **HIS**-134, **HIS**-188, **HIS**-209
GlcNAc6P	**ALA**-133,**ASN**-212,**ALA**-213,**ARG**-220, **HIS**-244, **ILE**-301

The conserved residues in the ligand- and metal-binding sites include histidine (H), asparagine, arginine (N) and alanine (A). aspartate (D) and glutamate (E). These active regions contain both hydrophobic and polar residues. Moreover, the regions also contain acidic (D and E) and basic (H and R) amino acid residues. Interestingly, the hydrophobic Isoleucine residue found in *MS-NagA* and *MM-NagA* is replaced by Leu302 in *Mtb-NagA*, *MU-NagA*, *MA-NagA*, and *MC-NagA*.

Another feature of *NagA* protein is the presence of divalent metal ions binding site that involved in its catalytic processes. Mycobacteria *NagA* enzyme possess metals ions (Zn and Cd) affects *NagA* functions including its stability and the design of new inhibitors. Moreover, they serve as co-factors which controls metabolism and regulate signal transport. This metal binding site is characterized with histidine residues that may impact the overall function of acid/base balance maintenance of the site by donating a proton which serving as cation [44, 45].

Moreover, the conserved aspartic acid controls the catalytic mechanism of conversion of GlcNAc6P substrate to GlcN6Pand acetate [34, 35]. Unlike the non-mycobacteria *NagA* protein, Mycobacteria *NagA* possesses a conserved cysteine (Cys-131)[17] within the flexible loop (122 – 136).

The cysteine-131 is unique to all Mycobacteria *NagA* enzymes[17]. It may be manipulated to bind with an inhibitor through which the catalytic activity of *NagA* may be hindered because cysteine residues are generally reactive. For instance, cysteine-targeted therapeutic approaches had been used to prepare irreversible peptidomimetic inhibitors in the hepatitis C virus protease containing a non-catalytic cysteine [48]. Therefore, Cys131 residue is a potential target in *NagA* protein for the inhibitors.

Mycobacteria NagA is a metalloenzyme, and a metal-binding site is a feature in *NagA* enzymes crucial to the catalytic activity of mycobacterial species [20,49,50]. However, there are interesting differences in the *NagA* enzyme metal-binding sites[17] across bacteria species. The *NagA* enzyme metal-site may be a mononuclear or binuclear center. Mycobacteria *NagA* enzymes, such as *T. marima* and *B. subtilis NagA*, have bi-nuclear metal-binding active sites [20,50] while Gram-negative non-mycobacterium *E. coli* has a single metal in the active site. M. *smegmatis* has a binuclear metal-binding site.[17] A defined electron density confirmed the two bound metal ions with one buried in the cavity of the $(\alpha/\beta)_8$ barrel for ligand-free and a larger metal ion in ligand-bound.

Another feature of mycobacteria *NagA* enzyme is the formation of an isologous association of the enzyme.[51] Importantly, the organization of dimers plays a crucial role in the biological processes. This gives rise to closed structures with an intrinsic symmetry, enhanced conformational stability, and thermal stability [52]. The protein subunit functions in the allosteric regulation and catalysis of protein functions [51]. the dimer-conformation remains conserved among all the analyzed *mycobacteria*. The active site of the Mycobacteria *NagA* enzymes is formed within the homodimer at the interface of the domain I of *NagA* protein. Based on the minimal amount of literature available, the mechanism of dimeric interactions may differ from the general enzyme oligomerization. However, the interactions of the *mycobacterial NagA* enzymes include H-bonding and salt bridges.

4.5.5 Effect of mutation on *NagA* proteins

Single nucleotide polymorphisms occurring in protein-coding regions may lead to structural defects in the 3D structure of a protein, thus, leading to pathological effects. Both the single and the two points mutations have a reductive effect on the catalytic activity of *NagA* proteins. Mycobacterial species *NagA* enzymes possess a conserved His56 – His58 (HXH) motif at the metal-binding sites as seen in the mycobacteria. The mutation of the single point amino acid of mycobacterial species causes the complete loss of catalytic activity. Mutants such as D267A, H56Q, H56N, and H56A have been reported in the structure of *NagA* [17].

We, therefore, used the *DynaMut* [53] and *Polyphen2* [54] resources to study the functional impact of these nsSNPS occurring in *NagA* proteins. *PolyPhen2* is used to determine if the amino acid substitution in the sequence occurs in a conserved region and predicts if the substitution has any deleterious impact on the structure of the protein [54]. *PolyPhen* employs the position-specific independent count (PSIC) score to ascertain the functional and structural impact of the substitution. A PSIC score of 1.0 is regarded as damaging. The SNPs are classified as benign, possibly damaging and probably damaging [54]. *DynaMut* is a web server that predicts the impact of the mutation on protein conformation, flexibility, and stability [53].

Table 4.5 Respective predicted Free energies and Scores from DynaMut and Polyphen resources for Mycobacteria NagA mutants.

	DynaMut		***Polyphen***		
Mutant	▲▲G	▲Vibration Energy	**Inference**		Score
D267A	-0.155	0.655	Destabilizing	Probably Damaging	1
H56Q	-0.834	0.480	Destabilizing	Probably Damaging	1
H56N	-0.834	0.480	Destabilizing	Probably Damaging	1
H56A	-0.663	1.033	Destabilizing	Probably Damaging	1

As predicted by DynaMut, change in folding free energy (ΔΔG) of the D267A, H56Q, H56N, and H56A variants were -0.155, -0.834, -0.834 and -0663 respectively. While the Δ vibrational entropy energy of the variants was 0.655, 0.480, 0.480 and 1.033 respectively (Table 4.5). PolyPhen also predicted the four variants to be probably damaging with a PCIS score of 1. Overall, these variants were predicted to be destabilizing with increased residual fluctuations (Figure 4.6) as corroborated by the Δ vibrational entropy energy.

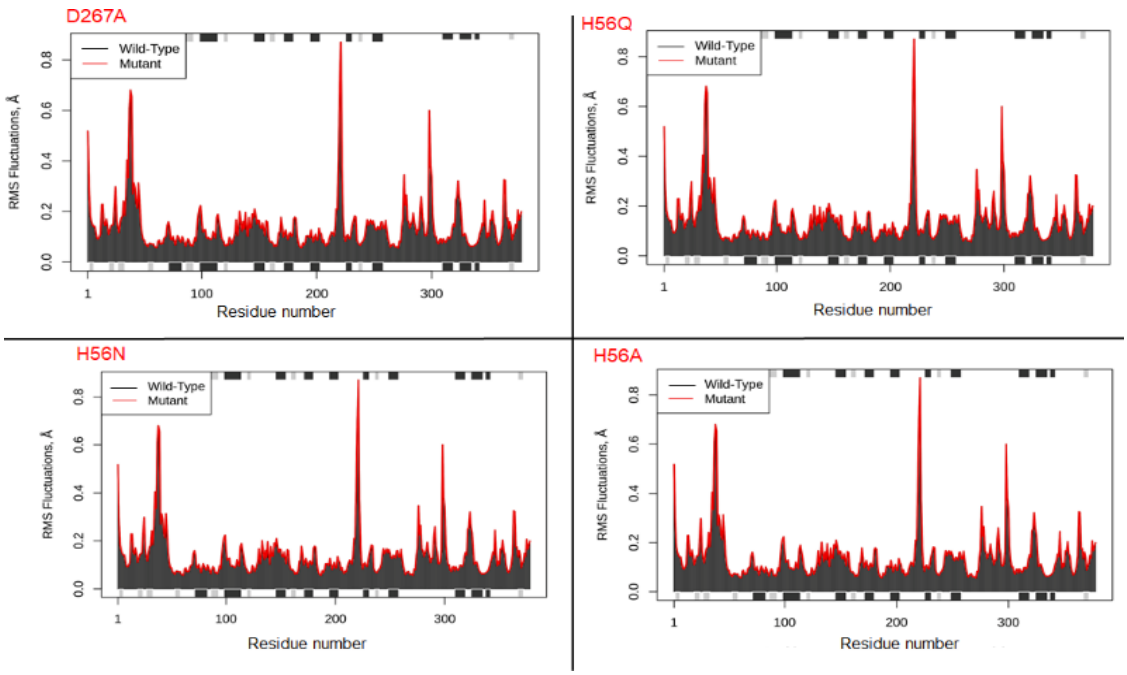


Figure 4.6. Deformation and fluctuations analysis plots of D267A, H56Q, H56N and H56A as predicted by DynaMut.

The interatomic interactions of the wild and mutant protein also showed that the bond interaction was affected because of the mutation, accounting for the destabilization of the mutant protein. The strong hydrogen and disulphide bridge observed in the wild proteins were lost in the mutant proteins (Figure 4.7).

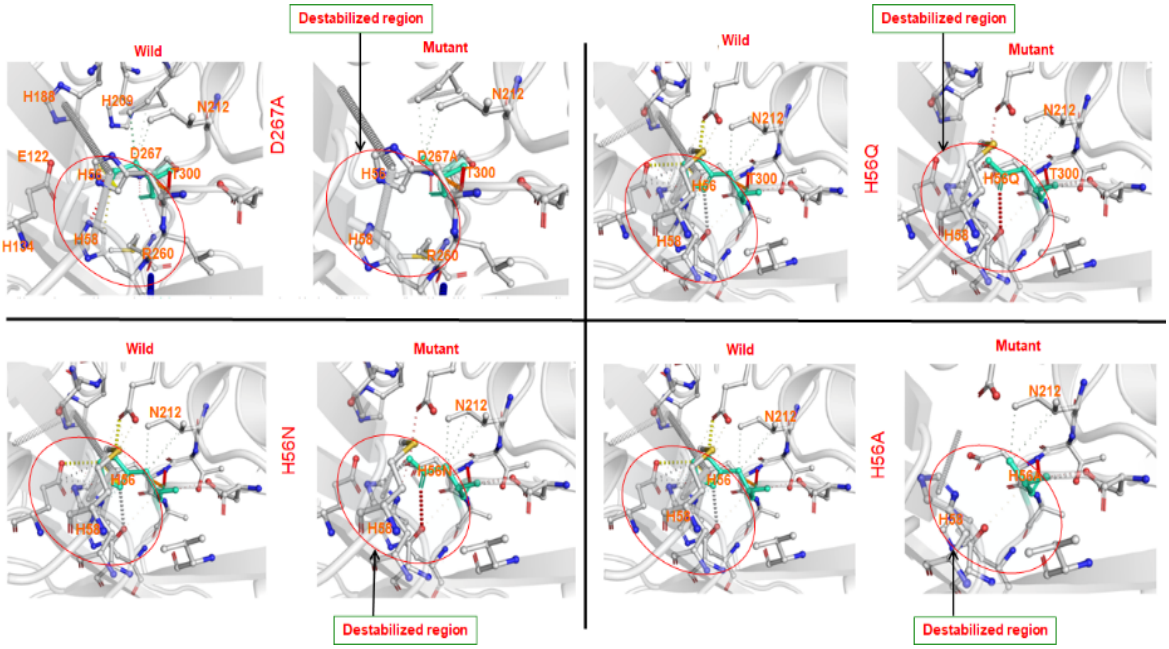


Figure 4.7. Interatomic interaction of wild NagA and mutated NagA predicted by DynaMut.

4.5.6 Structural perturbation of Mycobacterium species NagA protein

Structure and dynamics of protein have been linked to its activity, and the understanding of this can furnish us with mechanistic insights into the pharmacological activities of therapeutic small molecules. We, therefore, elucidated the structural attributes that characterize *NagA* protein in the six *Mycobacterium species*. RMSD was used to estimate the structural stability of the *NagA* system. Figure 4.8 revealed that while *MANagA*, *MSNagA*, and *MtbNagA* systems were relatively stable. *MUNagA* and *MCNagA* attained stability at about 40 ns after their RMSD increased from 1.8 to 3.0 Å and *MMNagA* was stable but until it deviated at about 130 ns (2.5 Å to 3.8 Å) before it eventually stable. *MA-NagA*, *MC-NagA*, *MM-NagA*, *MS-NagA*, *Mtb-NagA*, and *MU-NagA* had an average RMSD value of 2.85 Å, 3.25 Å, 4.25 Å, 2.15 Å, 3.25 Å, and 3.25 Å. The RMSD plots of the active site loop for all the *NagA* systems indicated high stable conformation having almost the same values. RMSF determines the residual fluctuation correlating to the crystallographic B (temperature, *T*) factors (Figure 4.8A). *MCNagA*, *MMNagA*, and *MUNagA* fluctuated throughout but the active site elicited high residual fluctuation across the systems between residues 205-382. The average RMSF

values of *MA-NagA*, *MC-NagA*, *MM-NagA*, *MS-NagA*, *Mtb-NagA*, and *MU*-were 4.8 Å, 8.85 Å, 9.95 Å, 3.5 Å, 3.85 Å, and 4.55 Å respectively.

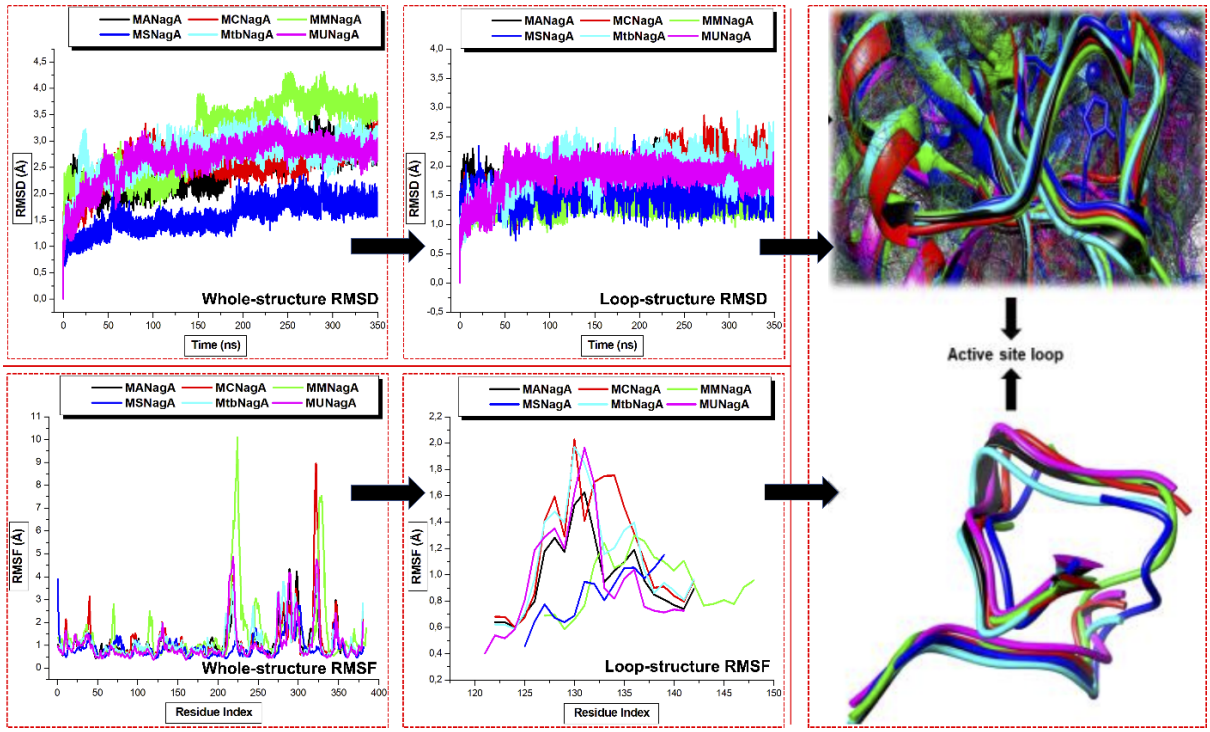


Figure 4.8. Conformational analysis plot showing stability and atomistic motions among NagA proteins and active site flexible loop **(A).** C-α RMSF plot showing the residual fluctuation of the NagA proteins and active site loop **(B)**.

In addition, we superimposed the apo *NagA* structures (Figure 4.9) after the MD simulations to compare their characteristics and mechanistic behaviour. Modelling of the active sites' flexible loop provides an additional clue to the comparative analysis of *NagA*. However, the 350 ns simulation of the *NagA* enzymes showed some variations in the conformation of the active site loop. We observed if the six *NagA* maintains their relative superimposition after relaxation (see Figure 4.S1 supplementary material). The results showed that *Mtb-NagA*, *MS-NagA*, *MM-NagA*, and *MU-NagA* seemed to retain their structural similarity, but *MA-NagA* and *MC-NagA* structures were relatively different. Furthermore, the β_8-strands, specifically, residues 33 – 43, 295 – 303, and 322 – 340 revealed decreased atomistic steric hindrance due to the energy minimization.

Effect of energy minimization on the superimposition of NagA

In the initial apo of *NagA* protein, the flexible loop (Glu122 – Pro136) was conserved and relatively superimposable. However, the flexible loop's positions of the subunit monomers (chain B) changed the post-minimized *NagA*, implying the loss of superimposition. Thus, possibly, this provides different mechanism flexibility regarding each *mycobacterial* species (Figure 4.9). The MD simulations identified a potential mechanism by which the enzyme may structurally function. Based on this information, it is reasonable to deduce that the occluding loop in chain B may have shifted away from the active sites of the ligand after relaxation and provides the potential for the ligand to bind.

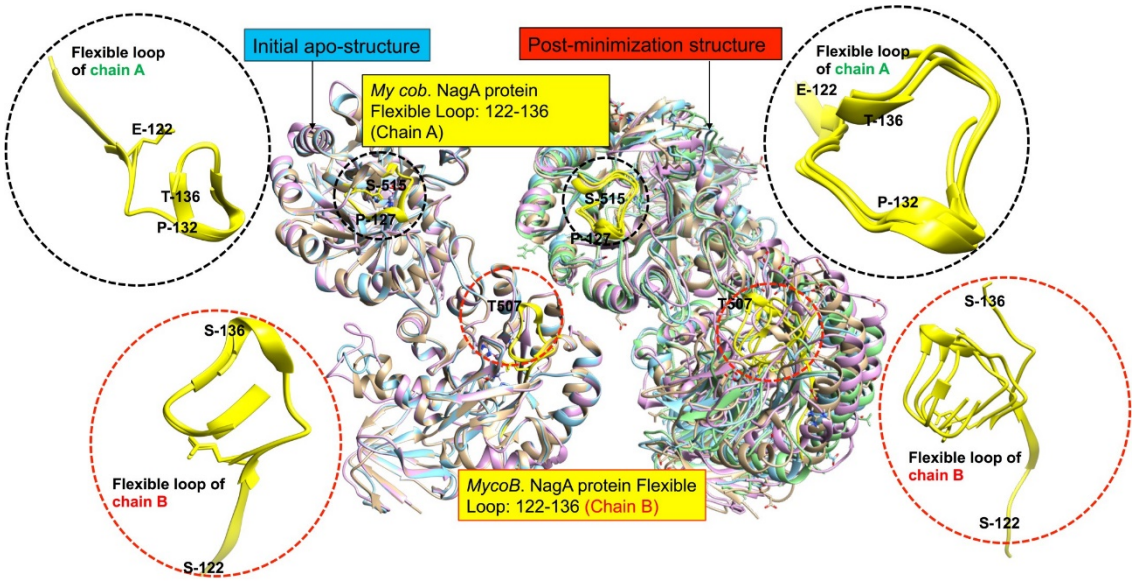


Figure 4.9. Illustration of the Initial and minimized apo-structures of the Mycobacteria NagA enzymes showing the effect of relaxation on the flexible loops (residues: 122-136) that covers the binding site of the ligand in chain B

4.6 Conclusion

Mycobacterial infections, especially tuberculosis, have killed millions of human beings for many decades. This study revealed profound similarities in the amino acid sequence and the structural features of Mycobacteria species using bioinformatics and molecular dynamics techniques. Moreover, it elucidates point mutations that alter the function of the protein. The mutation destroyed the active sites of *NagA*, thus leading to drug resistance. The study

provides structural insights into the architectural setup of *Mycobacterium tuberculosis NagA*. Furthermore, it unravelled the "*NagA*-Achilles's hill" that can be explored in future design and development of antitubercular therapy to circumventing drug resistance that is characteristic of Tuberculosis. Therefore, we considered *NagA* enzyme as a good target that can be explored in the search for a new inhibitor of *Mycobacterium tuberculosis*. Hence, in future, a design of a multifaceted drug using computational methods; molecular docking of inhibitor, pharmacophore model, and MD simulations would be explored.

ACKNOWLEDGEMENT

Authors acknowledge the Centre for High Performance Computing, Cape-Town for providing computational resources.

Compliance with Ethical Standards

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Disclosure of Potential Conflict of Interest

The authors declare no conflict of interest.

Research involving Human participants and/or Animals

This article does not contain any studies with human participants and/or animals performed by any of the authors.

Informed Consent

This study did not require informed consent since the study does not contain any studies with human participant performed by any of the authors.

FUNDING

No funding was received for this study

References

1. Chakraborty, P. & Kumar, A. The extracellular matrix of mycobacterial biofilms: could we shorten the treatment of mycobacterial infections? *Microbial Cell* **6**, 105–122 (2019).

2. Diel, R., Lipman, M. & Hoefsloot, W. High mortality in patients with Mycobacterium avium complex lung disease: A systematic review. *BMC Infectious Diseases* **18**, 1–10 (2018).

3. Khosravi, A. D., Hashemzadeh, M., Shahraki, A. H. & Teimoori, A. Differential identification of mycobacterial species using high-resolution melting analysis. *Frontiers in Microbiology* **8**, 1–7 (2017).

4. Percival, S. L. & Williams, D. W. *Mycobacterium*. *Microbiology of Waterborne Diseases: Microbiological Aspects and Risks* (Elsevier, 2014). doi:10.1016/B978-0-12-415846-7.00009-3

5. Gotuzzo, E. *et al.* Cutaneous Mycobacterial Infections. *Clinical Microbiology Reviews* **32**, 1–25 (2018).

6. Merikanto, I., Laakso, J. T. & Kaitala, V. Invasion ability and disease dynamics of environmentally growing opportunistic pathogens under outside-host competition. *PLoS ONE* **9**, 1–21 (2014).

7. Jason A. Jarzembowski, M. & Young, M. B. Nontuberculous mycobacterial infections. *Medical Clinics of North America* **132**, 361–379 (2008).

8. van Ingen, J. Mycobacteria. *Infectious Diseases* 1645-1659.e2 (2017).

doi:10.1016/B978-0-7020-6285-8.00185-4

9. (WHO), W. H. O. *Global Tuberculosis Report (Executive Summary)*. *Report* (2018).

10. Bruchfeld, J., Correia-Neves, M. & Källenius, G. Tuberculosis and HIV Coinfection. *Cold Spring Harbor perspectives in medicine* **5**, a017871 (2015).

11. Shankar, E. M. *et al.* HIV-*Mycobacterium tuberculosis* co-infection: A 'danger-couple model' of disease pathogenesis. *Pathogens and Disease* **70**, 110–118 (2014).

12. Kurz, S. G., Furin, J. J. & Bark, C. M. Drug-Resistant Tuberculosis: Challenges and Progress. *Infectious Disease Clinics of North America* **30**, 509–522 (2016).

13. Falzon, D. *et al.* Multidrug-resistant tuberculosis around the world: What progress has been made? *European Respiratory Journal* **45**, 150–160 (2015).

14. Wu, X. *et al.* Drug Resistance Characteristics of *Mycobacterium tuberculosis* Isolates From Patients With Tuberculosis to 12 Antituberculous Drugs in China. *Frontiers in Cellular and Infection Microbiology* **9**, (2019).

15. Lee, I. *et al.* Structural and functional studies of the *Mycobacterium tuberculosis* VapBC30 toxin-antitoxin system : implications for the design of novel antimicrobial peptides. **43**, 7624–7637 (2015).

16. Nieto, L. M., Mehaffy, C. & Dobos, K. M. The Physiology of *Mycobacterium tuberculosis* in the Context of Drug Resistance: A System Biology Perspective. *Mycobacterium - Research and Development* (2018). doi:10.5772/intechopen.69594

17. Ahangar Mohd Syed, Furze Christopher M., Guy, Collette S., Fullam, E., Cameron, A. D., Cooper, C., Graham, B. & Maskew, K. S. Structural and functional determination of homologs of the *Mycobacterium tuberculosis* N -acetylglucosamine-6-phosphate deacetylase (NagA). *Journal of Biological Chemistry* **293**, 9770–9783 (2018).

18. Popowska, M., Osińska, M. & Rzeczkowska, M. N-acetylglucosamine-6-phosphate deacetylase (NagA) of Listeria monocytogenes EGD, an essential enzyme for the metabolism and recycling of amino sugars. *Archives of Microbiology* **194**, 255–268 (2012).

19. Liu, A. & Huo, L. *Amidohydrolase Superfamily*. *eLS* (John Wiley & Sons, 2007). doi:10.1002/9780470015902.a0020546.pub2

20. Sugrue, E. *et al.* Evolutionary expansion of the amidohydrolase superfamily in bacteria in response to the synthetic compounds molinate and diuron. *Applied and Environmental Microbiology* **81**, 2612–2624 (2015).

21. Abrahams, K. A. & Besra, G. S. Mycobacterial cell wall biosynthesis: A multifaceted antibiotic target. *Parasitology* **145**, 116–133 (2018).

22. Mouilleron, S. *et al.* The Mechanism of Acetyl Transfer Catalyzed by *Mycobacterium tuberculosis* GlmU . *Biochemistry* **57**, 3387–3401 (2018).

23. McGillewie, L., Ramesh, M. & Soliman, M. E. Sequence, Structural Analysis and Metrics to Define the Unique Dynamic Features of the Flap Regions Among Aspartic Proteases. *Protein Journal* **36**, 385–396 (2017).

24. Hoffmann, K. F., Chalmers, I. W., Padalino, G., Brancale, A. & Ferla, S. Combining bioinformatics, cheminformatics, functional genomics and whole organism approaches for identifying epigenetic drug targets in Schistosoma mansoni. *International Journal for Parasitology: Drugs and Drug Resistance* **8**, 559–570 (2018).

25. Nncube, N. B., Ramharack, P. & Soliman, M. E. S. Using bioinformatics tools for the discovery of Dengue RNA-dependent RNA polymerase inhibitors. *PeerJ* **6**, e5068 (2018).

26. Xiaolin Sun, William T. Jones, and V. N. U. *Applications of Bioinformatics and Experimental Methods to Intrinsic Disorder-Based Protein-Protein Interactions*. *Protein Engineering* (InTech, 2012). doi:10.5772/29246

27. Katsila, T., Spyroulias, G. A., Patrinos, G. P. & Matsoukas, M. T. Computational approaches in target identification and drug discovery. *Computational and Structural Biotechnology Journal* **14**, 177–184 (2016).

28. Brown, D. K. & Tastan Bishop, Ö. Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis: Analyzing Variation at the Protein Level. *Global Heart* **12**, 151–161 (2017).

29. Chen, C., Huang, H. & Wu, C. H. *Chapter 1 Protein Bioinformatics Databases and Resources*. **1558**, (2017).

30. Chetty, S. & Soliman, M. E. S. Possible allosteric binding site on Gyrase B , a key target for novel anti-TB drugs : homology modelling and binding site identification using molecular dynamics simulation and binding free energy calculations. *Medicinal Chemistry Research* **24**, 2055–2074 (2015).

31. Waterhouse, A. *et al.* SWISS-MODEL : homology modelling of protein structures and complexes. 1–8 (2018). doi:10.1093/nar/gky427

32. London, N., Raveh, B. & Schueler-Furman, O. Investigating Protein-Peptide interactions Using the Schrodinger Computational Suite. *Methods in Molecular Biology* **1561**, 235–254 (2017).

33. Shunmugam, L. & Soliman, M. E. S. Targeting HCV polymerase: A structural and dynamic perspective into the mechanism of selective covalent inhibition. *RSC Advances* **8**, 42210–42222 (2018).

34. Pettersen, E. F. *et al.* UCSF Chimera — A Visualization System for Exploratory Research and Analysis. **25**, 1605–1612 (2004).

35. Sievers, F. *et al.* Fast , scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **7**, 1–6 (2011).

36. Ramharack, P. & Soliman, M. E. S. Zika virus NS5 protein potential inhibitors: an enhanced in silico approach in drug discovery. *Journal of Biomolecular Structure and Dynamics* **36**, 1118–1133 (2018).

37. Case, D. A. *et al.* The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **26**, 1668–1688 (2005).

38. Plumbridge, J. An alternative route for recycling of N-acetylglucosamine from peptidoglycan involves the N-acetylglucosamine phosphotransferase system in Escherichia coli. *Journal of Bacteriology* **191**, 5641–5647 (2009).

39. Yadav, V. *et al.* N-acetylglucosamine 6-phosphate deacetylase (nagA) is required for N-acetyl glucosamine assimilation in gluconacetobacter xylinus. *PLoS ONE* **6**, (2011).

40. Malleshappa Gowder, S., Chatterjee, J., Chaudhuri, T. & Paul, K. Prediction and Analysis of Surface Hydrophobic Residues in Tertiary Structure of Proteins. *The Scientific World Journal* 1–7 (2014). doi:10.1155/2014/971258

41. Zou, X., Pham, T. K., Wright, P. C. & Noirel, J. Bioinformatic study of the relationship between protein regulation and sequence properties. *Genomics* **100**, 240–244 (2012).

42. De Ruyck, J., Brysbaert, G., Blossey, R. & Lensink, M. F. Advances and Applications in Bioinformatics and Chemistry Molecular docking as a popular tool in drug design, an in silico travel. 1–11 (2016). doi:10.2147/AABC.S105289

43. Orosz, F. & Ovádi, J. Proteins without 3D structure: Definition, detection and beyond. *Bioinformatics* **27**, 1449–1454 (2011).

44. Kleanthous, C., Wemmer, D. E. & Schachman, H. K. The role of an active site histidine in the catalytic mechanism of aspartate transcarbamoylase. *Journal of Biological Chemistry* **263**, 13062–13067 (1988).

45. Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W. & Huang, R. B. The multiple roles of histidine in protein interactions. *Chemistry Central Journal* **7**, 1 (2013).

46. Hall, R. S., Dao, F. X., Xu, C. & Raushel, F. M. N-acetyl-D-glucosamine-6-phosphate deacetylase: Substrate activation via a single divalent metal ion. *Biochemistry* **46**, 7942–7952 (2007).

47. Bustos-Jaimes, I. *et al.* Allosteric Activation of Escherichia coli Glucosamine-6-Phosphate Deaminase (NagB) In Vivo Justified by Intracellular Amino Sugar Metabolite Concentrations . *Journal of Bacteriology* **198**, 1610–1620 (2016).

48. Hagel, M. *et al.* Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nature Chemical Biology* **7**, 22–24 (2011).

49. Davies, J. S. *et al.* Functional and solution structure studies of amino sugar deacetylase and deaminase enzymes from Staphylococcus aureus. *FEBS Letters* **593**, 52–66 (2019).

50. Bürger, M. & Chory, J. Structural and chemical biology of deacetylases for carbohydrates, proteins, small molecules and histones. *Communications Biology* **2**, (2019).

51. Mei, G., Di Venere, A., Rosato, N. & Finazzi-Agrò, A. The importance of being dimeric. *FEBS Journal* **272**, 16–27 (2005).

52. Shallom, D., Golan, G., Shoham, G. & Shoham, Y. Effect of dimer dissociation on activity and thermostability of the α-glucuronidase from Geobacillus stearothermophilus: Dissecting the different oligomeric forms of family 67 glycoside hydrolases. *Journal of Bacteriology* **186**, 6928–6937 (2004).

53. Rodrigues, C. H. M., Pires, D. E. V. & Ascher, D. B. DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research* **46**, W350–W355 (2018).

54. Adzhubei, I. A. *et al.* HHS Public Access. **7**, 248–249 (2010).

CHAPTER 5

Thompson loop: opportunities for antitubercular drug design by targeting the weak spot in demethylmenaquinone methyltransferase protein

Adewumi T. Adeniyi[a], Opeyemi S. Soremekun[a], Mary B. Ajadi[b], Mahmoud E. S. Soliman[a]

[a]Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

[b]Department of Medical Biochemistry, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Howard Campus, Durban 4000, South Africa

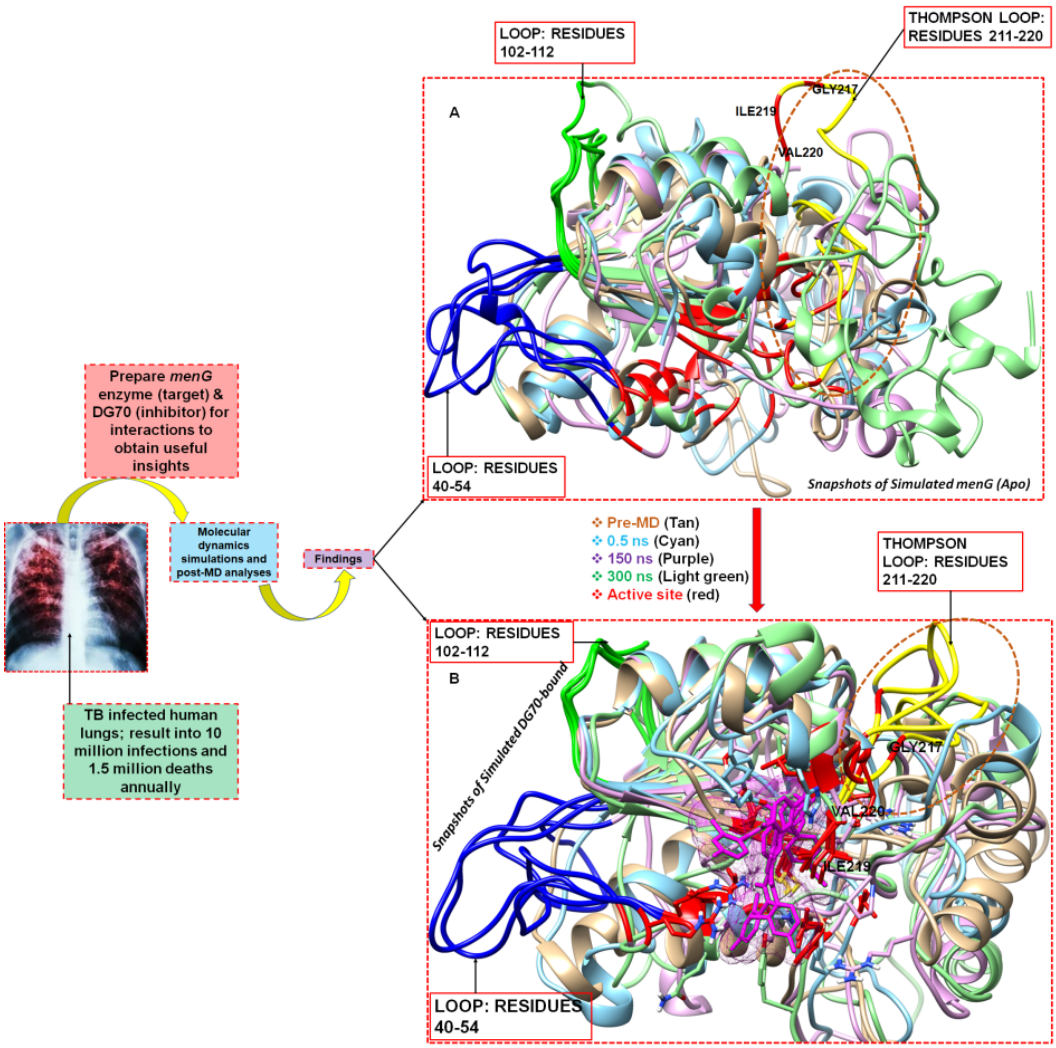
*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 8048

Fax: +27 (0) 31 260 7872

5.1 Graphical Abstract



Graphical superimposed snapshots of Thompson novel loop (yellow) of *menG* protein: apo (**A**) and bound (**B**) systems. Loop switches between open and closed conformations; critical for therapeutics.

5.2 Abstract

Drug-resistant tuberculosis (TB) has remained the top global health challenge, with a yearly estimation of 10 million infections and 1.5 million deaths in humans. Demethylmenaquinone methyltransferase (*menG*) catalyzes demethylmenaquinone conversion to menaquinone (MK) that is implicated in the TB pathogenesis, hence, has become a major drug target. DG70 is a biphenyl amide compound known to be a high binding affinity inhibitor of *menG*. This study investigated the structural and dynamic impacts of DG70 upon binding to *menG* using atom-based dynamic simulation. Our findings revealed that the modeled structure of *menG* possesses some Rossman-like methyltransferase characteristic features including two GXG motif, omega-like loop (residues 210-220) called the Thompson, nine α-helices, five β-strands, etc. Furthermore, atom-based dynamic simulations revealed that the Thompson loop is critical in the therapeutic activity of DG70. The loop assumed an open conformation in the unliganded-*menG* structure. However, in the DG70-*menG*, it assumed a tightly closed conformation. This explains the high binding affinity (-32.48 kcal mol^{-1}) observed in the energy calculations. Interestingly, these findings are further collaborated by the conformational perturbation in the *menG* protein. Conclusively, insights from this study highlights the structural "Achilles heel" in *menG* protein which can be further leverage on by inhibitors tailored to specifically target them.

KEYWORDS: *Mycobacterium tuberculosis*, demethylmenaquinone methyltransferase, drug-resistant Tuberculosis, homology modelling, molecular dynamic simulations

5.3 Introduction

Mycobacterium tuberculosis (*Mtb*) is the world's deadliest infection after the renascent human immunodeficiency virus/AIDS.[1] About two billion latent *Mtb* in man kill 2-3 million people annually.[1] Presently, more than 484 000 new cases of multidrug and extensively resistant (active) TB kill about 1.5 million death of persons every year.[2,3] More disturbing is the mutation-mediated and acquired drug resistance by the newer antitubercular drugs including linezolid, bedaquiline, (BDQ), etc.[4,5] More than one-half of the MDR/XDR-TB drugs have failed due to mutation.[6] In the quest to search for a promising novel drug over the years in this area, researchers have focused on protein-drug and protein-protein interactions for possible tuberculosis cure using various techniques and methodologies through targeting protein biocatalysis.[7,8] While a lot of studies have reported the targeting of constituents of *Mtb* cell wall, most often, the mycolic acid synthesis, however, only a few reports seem to be available on the study of the inhibition of *Mtb* respiratory pathway using experimental and computational tools such as the molecular dynamics (MD). BDQ is the first respiratory inhibitor that acts by binding to the oligomeric and proteolipid subunit C of *Mtb* ATP synthase (*atpE gene*).[5,6] BDQ was very effective against both the susceptible and multi-drug resistant TB.[9] However, there is a high prevalence of mutations in *Mtb* genes, including *atpE*, *atpC*, *mmpR* (Rv0678) to bedaquiline. A whole-genome sequence has revealed a genetic signature of BDQ resistance in a clinical *Mtb* isolate C; a microhetero-resistance found in a targeted deep sequencing analysis.[10,11] Similarly, a previous study shows that *Mtb* demethylmenaquinone methyltransferase, *menG* (rv0558) is a potential target and involves the biosynthesis of menaquinone (MK), the substance required for the cell maintenance of the prokaryotic including *Mtb*.[12,13,14] The amino acids of *Mycobacterium tuberculosis menG* (rv0558) protein are conserved and distributed among *Mycobacterium species* homologs[15] as shown in Figure 5.1.[13]

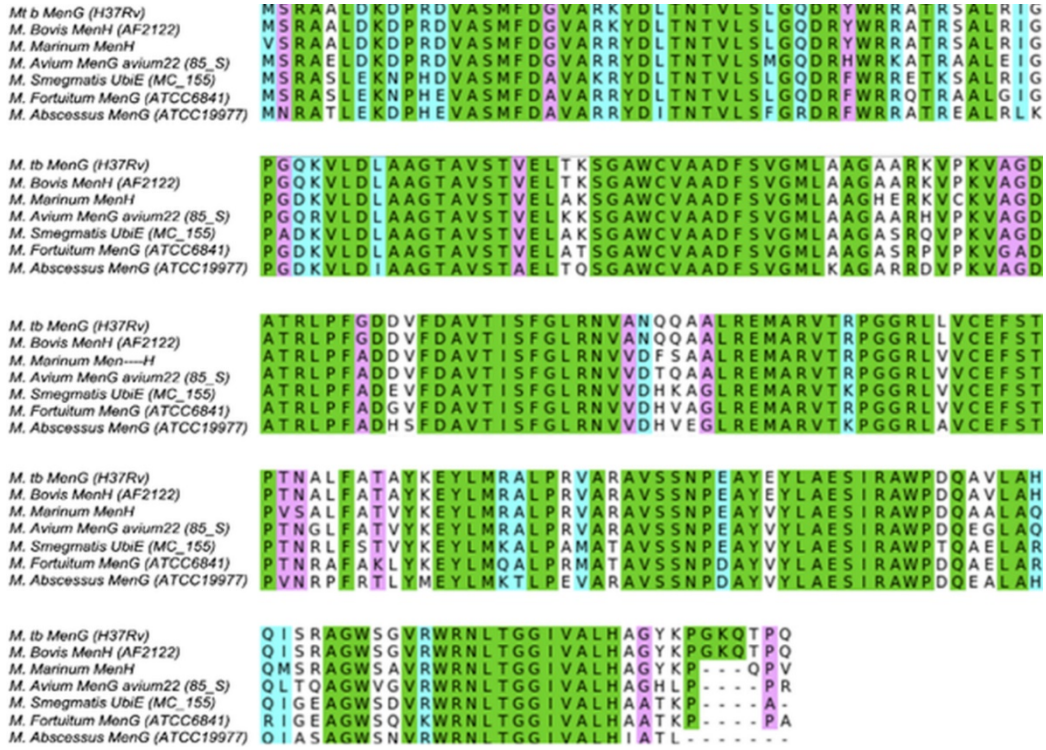


Figure 5.1. Sequence analysis of *Mycobacteria menG* showing ≥ 80 % conserved amino acid residues[16]

MenG is a membrane domain-associated enzyme and belongs to the SAM-dependent methyltransferase (MTase) superfamily comprising of an α/β/α fold structure.[17,18] MTase domain is an important target in drug design. *Mtb menG* encodes S-adenosyl-L-methionine (SAM)-dependent *menG* (Rv0558) that catalyses the biosynthesis of menaquinone [in the form of MK9(H_2)] required for respiration through C-methyl transfer from the SAM.[19] After the methylation catalysis of the *menG* enzyme, SAM converted to S-Adenosyl homocysteine (SAH) which leaves the protein domain .[18] SAM-bound methyltransferases is a Class I enzyme and may possess at least two or more GXG motifs located whereby one is located in the first β-sheet and the second is bound another β-sheet.[20,21] Puffal et al. (2018) also reported that *Mycobacterium smegmatis* (*Msmeg*) *menG* has 234 amino acids and GXG motif.[22] Methyltransferases include Rossmann-like α/β (Class I), TIM α/β-α/β (Class II), tetrapyrrole methylase α/β (Class III), SPOUT α/β (Class IV), etc.[23] *MenG* is a Class I methyltransferase with remarkable structural consistency and a 10 % primary structure similarity.[21]

DG70, a biphenyl amide, is a chemotype compound that inhibits the catalytic methylation of *Mycobacterium tuberculosis* demethylmenaquinone methyltransferase enzymes. The whole cell-based screen of a *Mycobacterium menG* using a putative *PcydAB reporter strain* and HRMS analysis identified was used to determine the potency of DG70 as a *menG* inhibitor.[24] Moreover, this compound is therapeutically active against both the clinical and laboratory drug-susceptible and drug-resistance *Mtb strains*.[13]

In drug design and development, protein dynamics study often provides information that may result to the desired novel target and its mechanism of catalysis (substrate binding) or discovery of an Mtb enzyme inhibitor and its mode of action (MOA).[25,26,27] This study investigates the changes in the conformational structure of *menG* and the inhibition of its catalytic activity with DG70 using computational approaches to understand its potentiality as a pharmacological target to eradicate tuberculosis. We created a homology model of *menG* enzyme and studied its ligand-unbound, substrate-bound (DMK9), and ligand-bound (DG70) systems over 300 ns molecular dynamics (MD) simulations at the atomistic level.

5.3.1 Concise DG70 structural chemistry: probable advantage over the frontline drugs

The study of *menG* inhibitors and specifically, the DG70, a potential drug for the treatment of TB, is relatively new. Hence, it is considered a probable research area by which this disease eradicated. Besides the mutation-mediated resistance in protein residues, including *menG* that militate the cure of TB, the inhibitor activity, which is partly determined by its structure, can also serve as a hindrance. We compared the chemical structures of DG70 and BDQ to provide a probable justification of why DG70 showed a therapeutic advantage over the BDQ drug.

The functional elements of chemical structures (Figure 5.2) determine their potential therapeutic efficacies against infection or disease. They are essential factors in drug design

and development. First, the groups have the potentials to form strong interactions with the active site of *menG*. For instance, high electronegative atoms of DG70 can form strong bonding interactions with the protein interacting residue thereby preventing the compound from moving in and out of the active site. Moreover, the oxygen and hydrogen atoms of the methoxy group and the phenyl rings can form strong conventional (carbon) hydrogen bonds, electrostatic, and pi-sigma interactions. There can also be a formation of alkyl and mixed/pi-hydrophobic interactions between DG70 and the interacting residues.

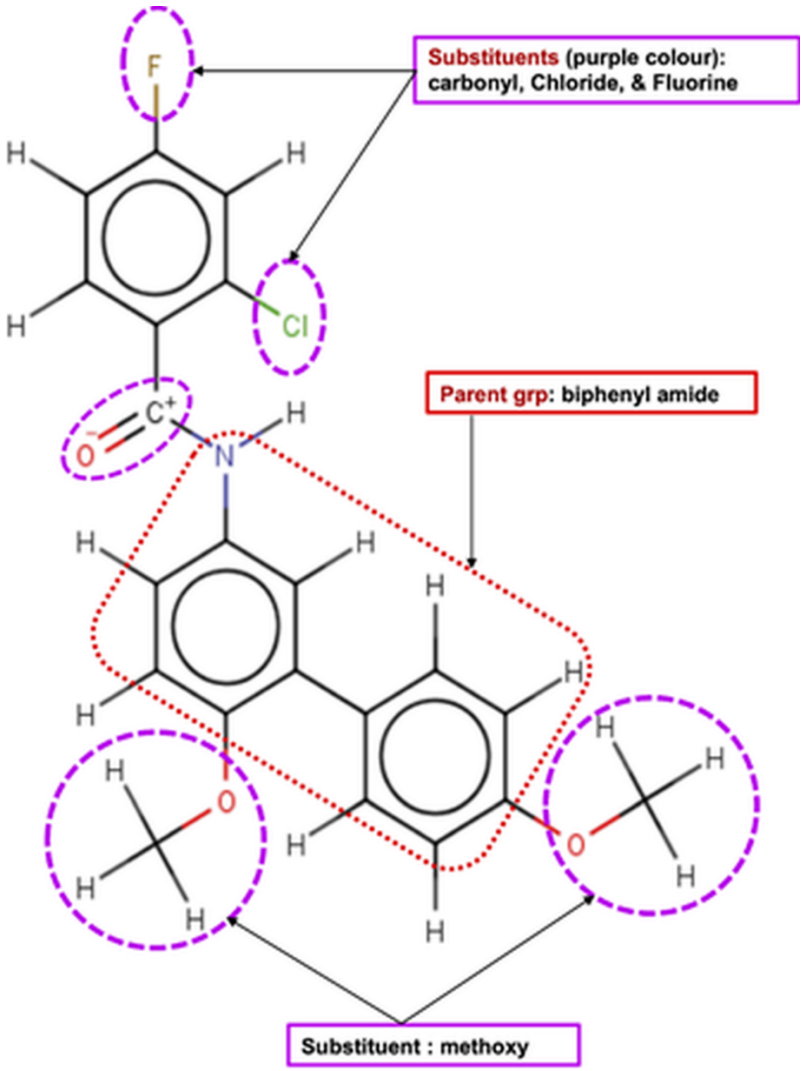


Figure 5.2. *MenG* inhibitor (DG70) and potential interacting functional groups shown in dotted lines.

5.4 Computational Methods

5.4.1 Homology modelling and binding pockets identification

5.4.1.1 *Mycobacterium tuberculosis Men*G Homology modelling

There was no closely related structure of *menG* protein found in the protein data bank (PDB) during this study. Most protein PDB structures have less than 30 % sequence identity in the homologs of the *Mtb menG* sequence.[28] Moreover, the model reported previously by Suhkeja *et al.* (2017) was also not available for use. Hence, we conducted a systematic search for possible structural templates by uploading the *menG* protein sequence (accession code: P9WFR3) obtained from the UniProt database[29] onto the BLAST. The templates of the P9WFR3 sequence were identified from NCBI using the BLASTp (accessed on 28 June 2019).[30,31,32] Table 5.1 showed the sources, the accession codes, and the selection criteria for choosing the template. *Mtb menG* protein templates share some sequence similarities with the SAM-dependent methyltransferases crystal structures.

Table 5.1 Criteria for choosing the templates for modelling the *Mtb MenG* protein

Template Source	PDB code	Query cover (%)	Sequence identity (%)	E-value
Saccharomyces cerevissia sp	*4OBW.A*	91	32.51	0
Lechevalieria aerocolonigenes	*3BUS.A*	92	23.81	0
Saccharomyces cerevissiae spp	*40BW.A*	99	37.26	0
Streptomyces luridus spp	*3OU2.A*	38	21.99	0

Note: The templates also belong to the transferase family.

The MODELLER Software version 9.21, an add-on in UCSF ChimeraTools-1.13.1 was used to create the homology modelling of the target protein in which all the four templates selected to build the model.[33,34] We performed multiple sequence alignments using CLUSTALW online application[35], which uses a BLOSUM matrix and penalties of 10 and 0.1 for gap opening and extension, respectively. The multiple sequence provides evidence that the full length of chain A *Saccharomyces cerevisiae spp* is the best template with the highest identity score. To obtain a 2D secondary structure of *menG* enzyme, we uploaded its sequence to

UCSF Chimera version 1.13.1. where structural alignment and comparison, matchmaking, match-aligning, and modeller (homology) were used to build the model. Five models named A, B, C, D, and E having zDope (kcal mol^{-1}) -0.32, -0.28, -0.14, -0.06 and -0.00 respectively were generated. Model A (zDope = -0.32) was chosen since it is the best one because it has most negative value. The high structural similarity between the homolog and the template used give credence to the modeling strategy used. A Ramachandran plot for the analyses of bond angles and torsional strain was generated using MolProbity.[36] Of all residues, MolProbity results favoured 87.1 % and allowed 96.1 % (223/232), and nine outliers residues. The active sites were obtained using Metapocket2.0 and validated using Raptor-X, GalaxyWeb, and 3DLIGANDSITE resource. The predicted active site residues were not contained in the outliers.

5.4.1.2 Binding pockets identification

Four online resource programs were employed to predict the active sites of *menG* protein structure. Metapocket2.0 server deployed to identify the druggable pockets [37, 38] while cross-validations of the binding sites sorted out using Raptor-X web, [39] 3D-Ligandsite web,[40] and GALAXYWEB server.[41] Metapocket2.0 is a consensus method that incorporates four other methods, including the LIGSITE, PASS, Q-SITEFinder, and SURFNET to generate protein surface for clefts and cavities. Raptor-X is a web server with evolutionary information. It employs a powerful in-house deep learning model Deep Convolutional Neural Fields (DeepCNF) to predict the secondary structure and disorder regions, solvent accessibility.[42] GALAXYWEB server predicts protein structure from a sequence by template-based modeling and refines loop or terminus regions by *ab initio* modeling. It relies on the method tested in the 9th Critical Assessment of techniques for protein Structure Prediction as Seok-server.[43] GALAXYWEB prediction method generates stable core structures from multiple

templates and rebuilds unreliable loops or termini by using an optimized-based refinement method.[44]

5.4.2 `System preparations and docking calculations

DG70 acts against the *menG* catalysis of methylation of demethylmenaquinone (DMK9) to menaquinone (MK9) both *in vitro* and *in vivo*. The catalytic methylation of DMK9 by *Mtb menG* produces MK9 that is required for energy. Hence, we considered docking calculations on the ligands; substrate (DMK9), and inhibitor (DG70). DMK9 and DG70 were drawn in MarvinSketch-17.21 (http://www.chemaxon.com/) and converted to a mol2 format. DMK9 and DG70 were separately assessed in Molegro Molecular Viewer (MMV)[45,46] to ensure that bond angles and hybridization state were corrected. The ligands minimized using the steepest descent method and GAFF force field add-on in Avogadro.[47] The systems were subjected to molecular docking using the predicted active site to refine and reduce the probable false positives due to ligands and to assess the geometrical feasibility at the site. The pose with the lowest negative value was chosen based on the interactions and binding affinities of the ligands. Both DG70 and DMK9 docked using AutoDockTools-1.5.6 graphical user interface[48] to defining the grid box for the active site; the spacing of 5 Å and size 122 × 106 × 86 pointing in x, y, and z directions. Similarly, the grid box for the same active site of DG70 docked defined as having a center (-37.66 × -45.85 × 37.45) and dimensions (120 × 106 × 72) the spacing of 5 Å and size 124 × 106 × 72 and pointing in x, y, and z directions respectively. Docking calculations using the Lamarckian Genetic algorithm carried out with AutoDock Vina.[48] Using UCSF ChimeraTools-1.13.1, atom types were assigned, Gasteiger charges added to the ligands, and merged non-polar hydrogen to carbon atoms. The water molecules were removed from the protein and added polar hydrogen and the ligands docked into the active site of *menG*. Finally, docking was validated based on the lowest energy pose.

5.4.3 Molecular dynamics (MD) simulations

Molecular dynamics simulations find applications in the study of the atomistic motions of biological systems and furnish us with the understanding of the physical movements of atoms and molecules. MD provides the interpretations of molecular processes within the biosystems. [49,50] We performed the MD simulations using the GPU version of the PMEMD.CUDA engine provided with the AMBER package, FF18SB variant of the AMBER force field used to describe the protein.[51] In applying the restrained electrostatic potential (RESP) and the GENERAL AMBER Force Field (GAFF) procedures, the partial charges were added to the ligands using ANTECHAMBER.[18] The LEAP module in AMBER 18 was used to neutralize and solvate the free *menG* and ligand-bound systems by adding hydrogen atoms [H^+], sodium ions [Na^+], and chloride [Cl^-] counter ions.

Furthermore, all atoms explicit solvation was carried out in an orthorhombic TIP3P box of water molecules size 10 Å. The procedure considered initial minimization of 2500 steps with an applied restrained potential of 500 kcal mol^{-1}. Additional 5000 steps of full minimization were carried out by conjugate algorithm without restraining conditions. The systems were gradually heated, starting from 0 to 300 K for 50 psec to maintain a fixed number of atoms and volume considering a canonical ensemble (NVT). Using 1 bar pressure provided by the Barendsen-barostat, the SHAKE algorithm utilized to constrict the hydrogen bond constraint [5253]. The total time for the MD simulation was 300 ns with a time step of 2 fsec using constant pressure of 1 bar, a 300 K, and Langevin thermostat. The coordinates Apo-*menG* and bound-*menG* complexes were each saved every 1ps, and the trajectories were analyzed using the CPTRAJ and PTRAJ module in AMBER 18/GPU. The trajectories analyzed for RMSD, RMSF, DSSP, principal component and analysis, and radius of gyration using CPPTRAJ and PTRAJ. The structural and visual analysis[54] done by employing the graphical user interface of UCSF Chimera and Discovery studio 2019 Client to analyze the binding mechanism of the

ligand-bound systems,[55] while the data plotted with MicroCal Origin 6.0 data analysis software. [56,53]

5.5 Post-dynamics analysis

5.5.1 Binding free energy calculations

Molecular dynamics employs the differences in the free energy to study the mechanism of biological processes.[57] The binding free energy (BFE) method is important in the computational study of protein dynamics to explore their binding mechanism with ligands thermodynamically.[58] The BFE of the bound system calculated using the molecular mechanics/GB surface area method (MM/GBSA) to estimate their binding affinities.[51] The energies considered over 300,000 snapshots from the 300 ns trajectories. This free binding energy (ΔG) calculated for DMK9- and DG70-bounds *menG* and expressed thus:

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{receptor}} - \Delta G_{\text{ligand}}$$

$$\Delta G_{\text{bind}} = E_{\text{gas}} + G_{\text{sol}} - T\Delta S$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}}$$

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{SA}}$$

$$G_{\text{SA}} = \gamma\text{SASA}$$

E_{gas} denotes the gas-phase energy- internal energy: Coulomb (E_{ele}) and van der Waals energies (E_{vdW}). The E_{gas} estimated from the FF18SB force field terms. Polar and non-polar states' energy contributions accounted for the Solvation free energy, G_{sol}. The non-polar solvation energy, SA.G_{SA} was determined from the solvent-accessible surface area (SASA), using a water probe radius of 1.4 A. In contrast, the polar solvation, G_{GB}, was obtained by solving the GB expression. *S* denotes the total entropy and *T*, the temperature of the systems. The contribution of each residue to the total binding free energy obtained at the predicted active

site by carrying out per-residue energy decomposition at the atomic level using MM/GBSA method in AMBER 18.[59]

5.5.2 Receptor-ligand Interactions systems

Analysis of *menG* active amino acids interaction network with DMK9 and DG70 was carried out using the receptor-ligand interaction add-on available on the Discovery studio visualizer 2019 Client.[53] The interactions depicted using the snapshots taken at different periods of 300 ns MD trajectories for DMK9- and DG70-bound systems and subsequently visualized using BIOVIA Discovery Studio software to obtain the molecular forces between the atoms of the ligand and interacting residues. These interactions may provide important information that would help to discover if *menG* is a potential target and contribute to the overall design and development of contributing to the development of drugs. Using this software allows us to depict the interacting residues and the interaction types within the bound systems.

5.5.3 Conformational fluctuations of *MenG* systems

The RMSF is the fluctuations of individual protein residues to their average positions within a given molecular dynamic simulation.[60] It provides an understanding of the flexibility of various regions of *menG* while binding to the ligands. Mathematically, RMSF is as follows:

RMSfi is the RMSF of the ith residue from wh

ich the RMSF of the average is taken and divided by the RMSF's standard deviation [σ(RMSF)] to give resultant standardized RMSF s($RMSF_i$)].

5.5.4 Radius of gyration (RoG)

The radius of gyration of both the Apo and the bound system were determined to further examine the stability of the system through a compactness test. Statistically, the average radius of gyration (RoG) predicts the dimensions of biomolecule as it reflects the molecular

compactness of a system to its shape by describing the RMSD of the atoms from the common center of gravity of a given protein molecule.[61] The compactness of Apo and bound systems assessed along 300 ns MD trajectories by taking the average over 300,000 frames. The equation below expresses the estimation details of RoG is estimated:

The terms in the equations include the position of the i^{th} atom (r_i), mass/weight of each atom (W), and the center mass of atom i (r). The mean value calculated by taking the RoG values over the number of frames in each trajectory.

5.5.5 Principal component analysis (PCA)

The principal component analysis is an important post-MD analysis that describes the atomic displacement and the loop dynamics of a 3D protein. It represents the magnitude (eigenvalues) and direction (eigenvectors) of the motion of protein.[62] Before processing the *menG* PCA, the MD trajectories of free enzyme, DMK9-, and DG70-bound complex enzymes were stripped of solvent and ions using an integrated PTRAJ module in AMBER 18. Considering the dynamics of the C-α atom of *menG*, the first two principal components (PC1 and PC2) were computed and the conformational patterns of the free and bound *menG* systems projected along the first two eigenvectors, i.e. ev1/PC1 vs ev2/PC2 using the C-α atoms Cartesian coordinates.[53]

5.6 Results and Discussion

5.6.1 The Homology Structure of *Mtb MenG*

The preparation of the 3D structure of a protein is the first step in rational drug design.[63,64] However, there was no available crystal structure of menG for this study. Moreover, a previously reported homology model of *Mycobacterium tuberculosis* menG[13] was also not available. Hence, a homology model was built using 4 templates; crystal structures of two

4OBW, 3BUS, and 3OU2 obtained at resolution range between 1.5 – 2.4 Å from the protein data bank (PDB). The criteria (Table 5.1) for choosing the quality templates[64] produce successful menG protein models from which a model with the best zDope (- 0.32 kcal mol-1) was chosen to further this study.

The *menG* (rv0558) sequence consists of 234 amino acids corresponding to the *saccharomyces cerevisiae* transferase amino acid residues (74 – 301). The structural features of a protein are a good indicator of its functions. *Mtb menG* showed two conserved cysteine residues (CYS76 and CYS146) that tend to form a disulphide bridge, thus act as activators in the protein. Site-directed mutagenesis determination and ligand affinity analysis of *menG* could reveal a probable formation of a cysteine bridge. Furthermore, the *MenG* sequence is characterized by hydrophobic amino acids including especially at the active residues.

Topologically, the superimposition of homology *menG* structure and *4OBW* template showed high similarity. The secondary structures of a protein including *menG* are a function of its different conformations. The structural features of protein provide a means of understanding the molecular interactions.[20] In this study, we observed that the *Mtb menG* possesses parallel β-strands and fully formed α-helices (Figure 5.3A); four beta-strands (4_{β}) that are surrounded by nine-alpha (9_{α}) helices. This finding correlates with class I methyltransferase.

Figure 5.3B showed two GXG motifs located on two different loops of *Mtb menG* as against the β-sheet that was previously reported. These motifs are conserved in the homologs of *Mycobacterium tuberculosis menG*. Moreover, motif GLY217 played a binding role at the active site of this protein.

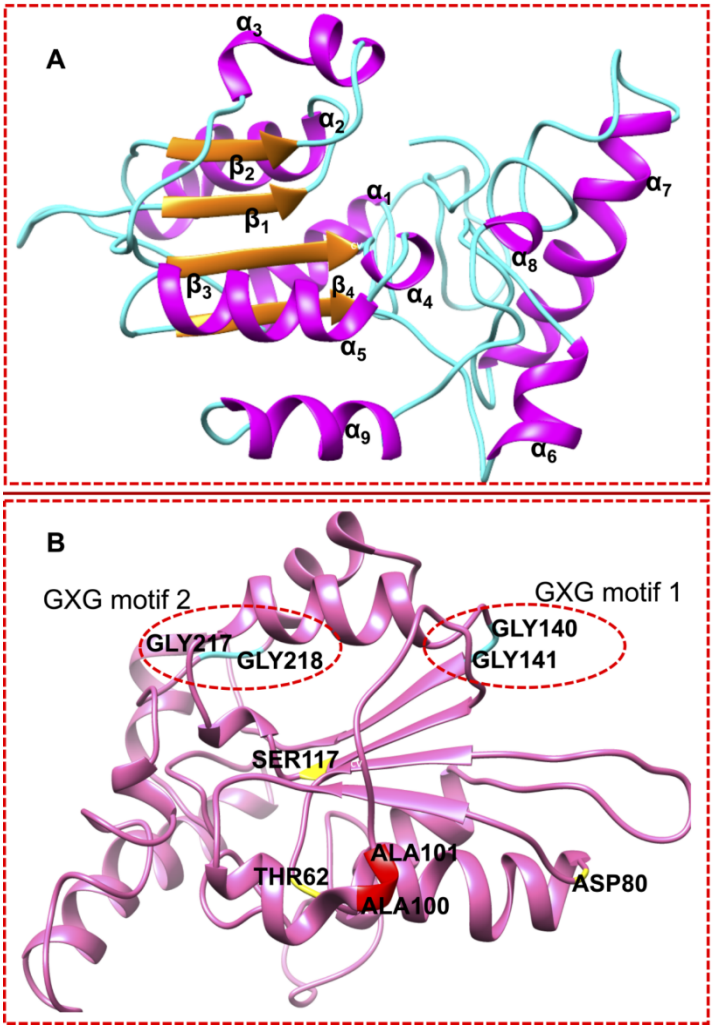


Figure 5.3. A; Secondary structural features of *Mtb menG* showing 4 parallel β-strands (orange) located at the center and surrounded by 9 α-helices (purple). B; GXG motifs residues: GLY140, GLY141, GLY217 (interacting residue), and GLY218 (cyan); UniProt annotated residues of the co-factor binding residues (red) ALA100 and ALA101; and the substrate-binding residues (yellow) THR62, ASP80, and SER117 among the active site residues.

5.6.2 Active site determination and validation

The ligand-binding sites of the model with the highest zDope score (-0.32) predicted to identify the corresponding pocket sites using Metapocket2.0, raptor-X, 3D-LIGANDSITE, and GALAXY (Figure 5.4C). The negative zDope score of -0.32 obtained upon binding the *menG* inhibitor indicates a reasonable accurate structure (native-like structure) and implies that the complex is very stable and reliable such that the DG70 is bound to the protein without moving in and out of the active site. The selected residues (Figure 5.4A) were

considered the active site and used for the molecular docking and modeling of the interaction of *menG* residues with the ligands. The active site determined for the substrate and the *menG* inhibitor. The site validated by generating a Ramachandran plot (Supplementary Figure 5S1) of the active site, which showed that they did not fall within the outlier region.

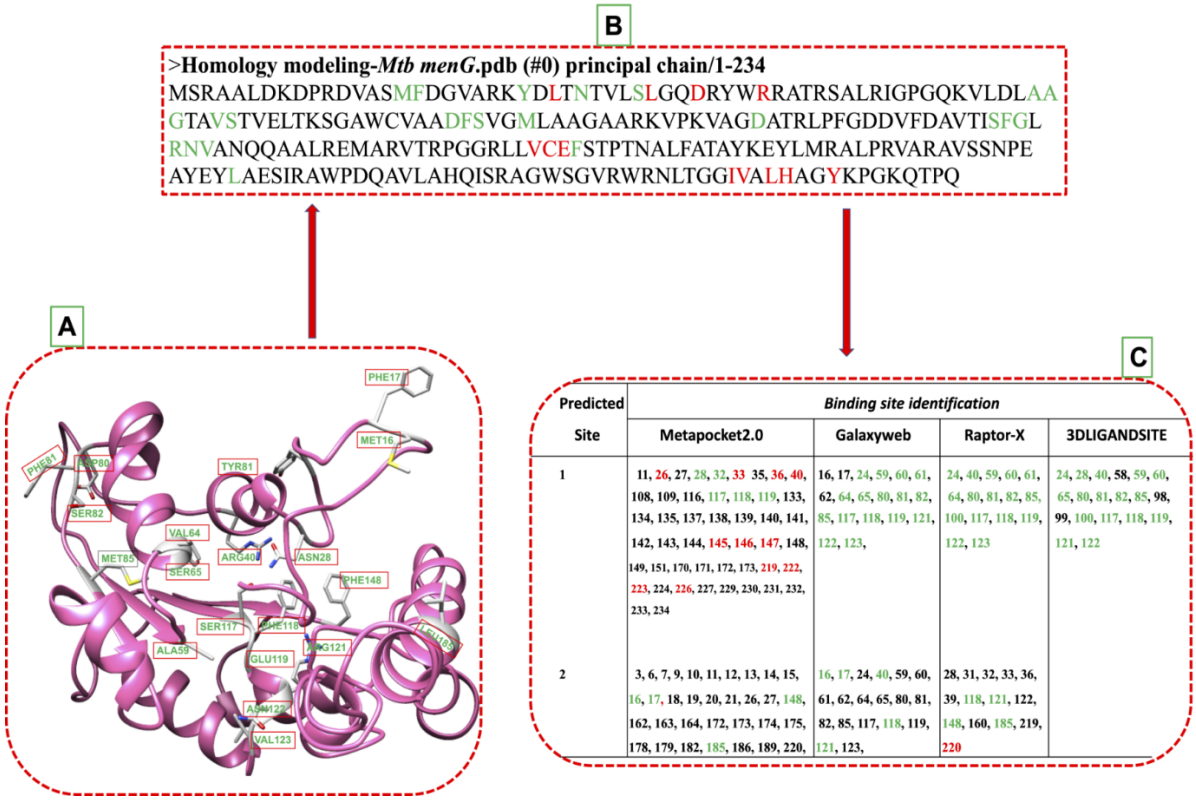


Figure 5.4. *MenG* predicted active sites (green) and the interacting residues (red). **A**; Secondary structure illustration *menG* of the active site (common residues by predictors). **B**; The primary structure displays predicted active residues (green) and the interacting amino acids (red). **C** provides the corresponding numbers of the predicted active amino acids.

5.6.3 Molecular docking of DMK9 and DG70

The Uniprot annotated THR62, ASP80, and SER117 to be among the substrate active site residues and ASP100 and ALA101 to be among the co-factor active site residues.[44] Figure 5.4C showed that the predicted active site residues agreed with UniProt annotation, though it did not differentiate which residues will bind to the substrate or co-factor. Hence, we employed a previously reported criterium for choosing the active site residues.[38] Moreover,

our predicted active site is similar to the one (residues: ARG3, ASP7, VAL20, LYS23, ARG40, ARG48, ALA60, TRP75, GLY99, PHE118, ARG121, VAL136, ALA181, SER188, and ARG190) that was previously reported by Sukheja *et al*., (2017), having residues ARG40, ALA60, PHE118, and ARG121 in common.

Molecular docking is a major computational tool that predicts the multiple orientations of a ligand at a binding site of receptors.[65,66,67] The substrate (DMK9) and DG70 (inhibitor) docked separately in the active pockets of the *menG* and the pose that gave the most favourable conformations (highest negative value) considered for the molecular simulation.

The molecular docking of DMK9 and DG70 on the active pocket of *menG* showed multiple orientations within the binding sites. Although the interacting residues used in this work were like those residues reported by the Sukheja group, our Ramachandran plot showed that ASP7, ASP9, and PRO10 were outliers. Whereas the outliers reported as interacting residues. Rather than using the binding site predicted in the previous work, we chose to predict a new active site of menG protein and validated it before use. DG70 docked on the most favourable binding mode (score = -7.1 kcal mol^{-1}), and DMK9 produced a score of -5.5 kcal mol^{-1} using the same active site. We observed that both ligands interacted with different residues suggesting that one ligand docked at the site of the co-factor site and the second one docked at the substrate's site.

5.6.4 Per-residue energy decomposition (PRED) and Ligand interactions with *MenG*

The per-residue energy decomposition revealed the interactions between the interacting residues and the ligands.[18] Considering the 300 ns MD trajectories, the per-residue decomposition of *menG* residues revealed various contributions of the binding site residues to DMK9 and DG70. The ligands adopted favourable orientations or morphology such that functional groups including chlorine, fluorine, $-OCH_3$, $-(C_6H_5)_2NH-$, and residue groups such

as -NH_2, -COOH, cysteine Sulphur atom interacted with the active site residues formed strong bonds. Explicitly, it observes that the methyl C-atoms of side chains of LEU58, ALA60, ALA91, LYS93, and VAL94 developed stable and strong alky hydrophobic interactions with the methyl C-atom of the prenyl group of DMK9 (Figure 5.5A). Also, there was the formation of a conventional hydrogen bond; the primary chain O atom of GLY206 formed hydrogen (H-bond) with H-2 of DMK9 phenyl group and the oxygen (O-2) of DMK9 created H-bond with the main chain H-bond of VAL210.

The hydrophobic residues that pocket DMK9 predominantly include Ala60, Val91, Lys93, Val94, Ala124, Ala129, GLY206, VAL210, and basic LEU58. Except for GLY206 and VAL210, the interacting residues formed alkyl hydrophobic interactions with the H-atoms of the prenyl group of DMK9. Both GLY206 and VAL210 formed conventional H-bonds (classical); backbone O atom of GLY206 bonded with the one prenyl H-atom while backbone H-atom of VAL210 bonded with the phenyl hydroxyl O atom of DMK9.

The residues that pocket DG70 include TYR24, ARG40, GLN126, CYS146, and GLU147. LEU26, LEU33, VAL145, and CYS146 formed strong and stable bonds with highly electronegative chlorine (Cl) atom of DG70. Figure 5.6A represents the interaction networks of the active site residues with DG70. The systems are characterized by stable hydrophobic, single halogen and hydrogen bonds. DG70 bound to the active residues including LEU26, ILE219, LEU222, and GLN235.

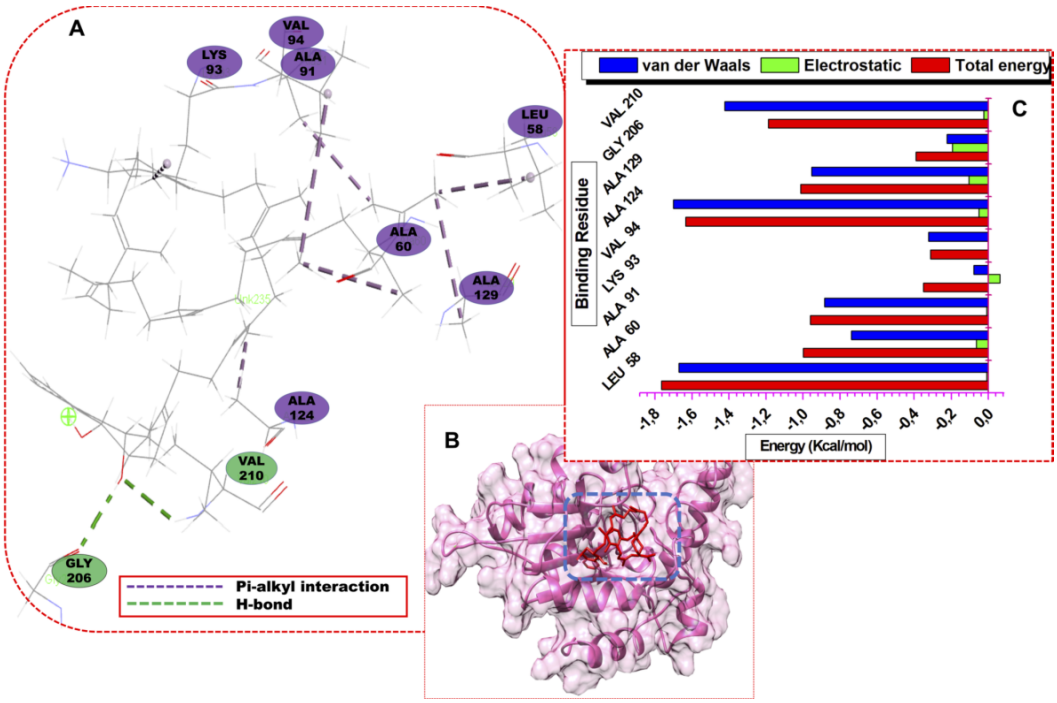


Figure 5.5. A: Visual representation of the DMK-*menG* interaction network (Colour online) and **B**: structural view of DMK that bound to the *menG* the active site. **C**: **e**nergy contributions of the interacting residues of *menG* with DMK9.

The resonating phenyl rings (C13-C18 and C19-C24) of DG70 formed strong pi-alkyl bonds. ASP36 and ARG40 formed pi-anion and pi-cation respectively with the phenyl (C1-C6) ring of DG70. Moreover, LEU26 and VAL145 formed alkyl hydrophobic and pi-sigma interactions with chlorine (Cl-7) and phenyl ring, respectively. Additional interaction is stronger interaction between residue LEU33 and fluorine (F-8) of DG70. Other significant interactions are the formation of conventional and non-classical H-bonds that were formed from the backbones of VAL220 and GLY217 with methyl hydrogen (H-44) and methoxy oxygen (O-27) atoms of DG70, respectively.

Using the MM/GBSA approach, we obtained individual residue-based contributions from the decomposition of the total binding protein energy (BPE) for DG70. We estimated the van der Waals (vdW) and electrostatic (elec.) interactions. The two forces reveal the residue and energy that produced an overall greater impact on the total binding energy. The per-residue decomposition of *menG*-DMK9 and *menG*-DG70 depicted as seen in Figure 5.5C and Figure

5.6C, respectively. The binding strength and the extent of stability of DG70 at the pocket site depend on the intermolecular forces that occurred between the inhibitor and the active site residues. LEU58 (-1.68 kcal mol^{-1}), ALA60 (-1.0 kcal mol^{-1}), ALA91(-0.8 kcal mol^{-1}), ALA124 (-1.58 kcal mol^{-1}), ALA129 (-1.05 kcal mol^{-1}), and VAL210 (-1.25 kcal mol^{-1}) contributed higher energies to the complex, while VAL94 (-0.35 kcal mol^{-1}) and GLY206 (-0.4 kcal mol^{-1}) contributed less energy. Therefore, the van der Waals interactions due to residues 58, 60, 91, 124, and 210 contributed to the high energy interaction of the covalent system (ΔG_{bind} = -31.64 kcal mol-1).

The per-residue interaction energy of the *menG*-DG70 system and the residues that contributed higher energies towards the complex include LEU26 (-1.15 kcal mol^{-1}), VAL145 (-1.9 kcal mol^{-1}), CYS146 (-1.2 kcal mol^{-1}), GLU147 (-1.35 kcal mol^{-1}), ILE219 (-2.8 kcal mol^{-1}), VAL220 (-1.8 kcal mol^{-1}), LEU222 (-1.8 kcal mol^{-1}), HIS223 (-0.5 kcal mol^{-1}), and TYR226 (-0.65 kcal mol^{-1}). Although the per-residue decomposition did not show the formation of a disulphide linkage between CYS146 and CYS76, there was a strong alkyl hydrophobic interaction between Carbon-Sulphur (C-S bond) of CYS and Cl atom of DG70 (Figure 5.6A). In contrast, residues LEU33 (-0.35 kcal mol^{-1}), ARG40 (-0.4 kcal mol^{-1}), PRO151 (-0.25 kcal mol^{-1}), and GLY217 (-0.28 kcal mol^{-1}) contributed less energies while LEU185 did not influence the complex energetically. Therefore, the vdW interactions due to residues 145, 146, 147, 219, 220, 222, and 226 and the elec. interactions due to ASP36 (-1.15 kcal mol^{-1}) contributed to the huge stable covalent system (ΔG_{bind} = -32.48 kcal mol^{-1}). The ΔG_{bind} for the *menG*-inhibitor system showed that DG70 is highly stable. This stable and strong interactions between DG70 and *menG* have a high impact on its inhibitory efficacy. Hence, we can correlate this results with the previous experimental report; the minimum inhibitory concentrations (MIC) of DG70 against drug-susceptible *Mtb* H37Rv (be 4.8 µg mL^{-1}) and drug-resistant *Mtb* strains (1.2-9.6 µg mL^{-1}) respectively.[16] These strong interactions

and high stability of *menG*-DG70 prevent the synthesis of MK9. However, we improve the binding affinity predictions of the complexes by subjecting the systems to 300 ns MD simulations to have more realistic flexible *menG* in an implicit solvent. Protein-ligand systems were analysed using the accurate MMGBSA/free binding calculations to obtain the most favourable pose of DMK9 and DG70 at the *menG* active sites. It observes that DMK9 and DG70 interacted with their respective *menG* active sites throughout the molecular simulations.

5.6.5 Free energy calculations of *MenG*-DG70 binding affinity

The binding free energy provides various distinct energy contributions within the binding pockets and the binding orientations that gave the best intermolecular interactions at protein active sites. The total binding free energy for bound systems; *Apo*DMK9 and ApoDG70 systems obtained using the MM/GBSA approach. We further determine the interactions networks of the DG70-*menG* (denoted ApoDG70Exp) reported by Sukheja *et al*. and depicted its residue-ligand interactions to compare with our results.

The docking pose score (kcal mol^{-1}) for *Apo*-DMK9, *Apo*DG70, and ApoDG70Exp- were -8.1, -7.1, and -5.5, and respectively. The thermodynamic energy contribution of DMK9 and DG70 to their respective complex and the total binding free energy determines their stability at *menG* active site. Table 5.2 summarizes the binding free energies of the complexes contributed by the protein and the ligand.

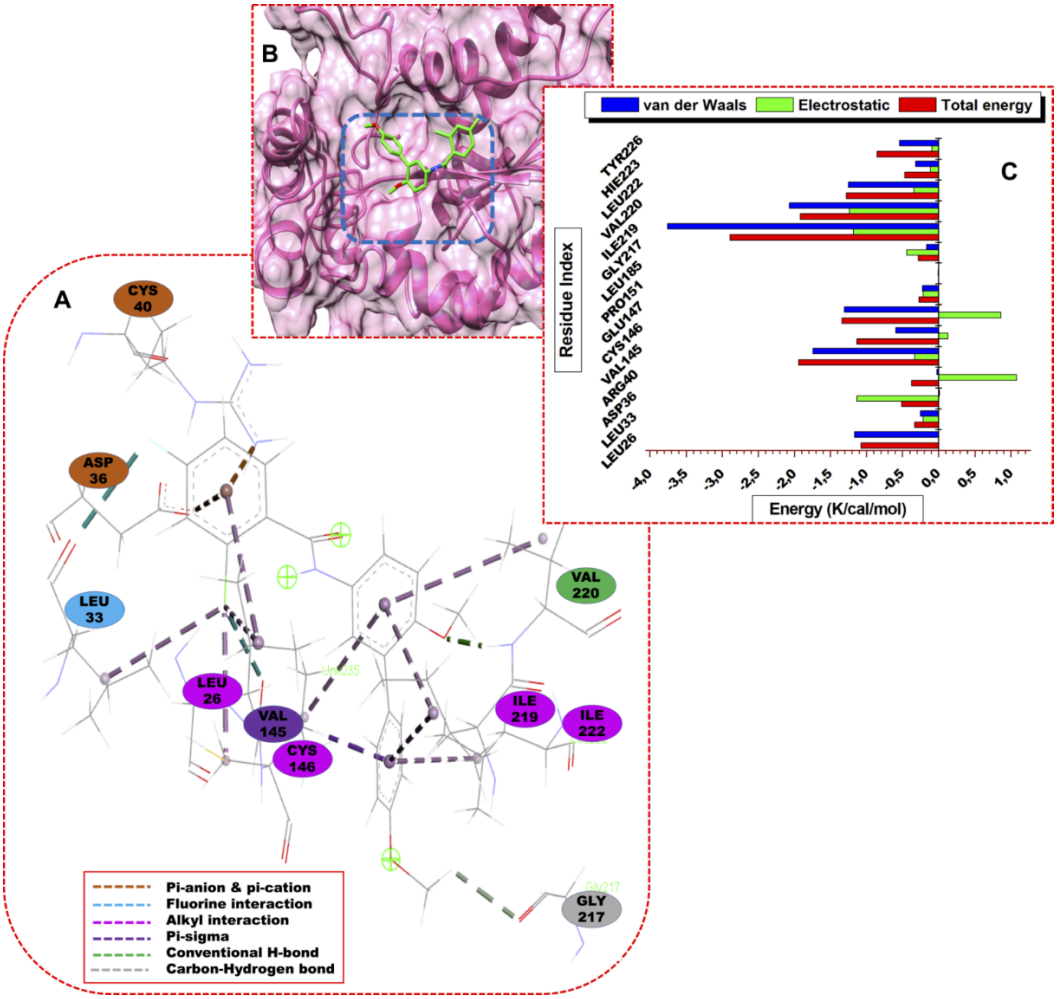


Figure 5.6. Residue-substrate interaction network; energy contributions of the interacting active residues at *menG* active site with DG70.

We docked DG70 on the active site, ran 300 ns MD simulations, and analysed the PRED of ApoDG70Exp complex. Figure 5S2 (supplementary) showed the residues-DG70 interactions and their energy contributions. The residues that contributed high energies towards the complex include THR62 (-0.08 kcal mol^{-1}), ASP80 (-0.12 kcal mol^{-1}), SER117 (-0.05 kcal mol^{-1}), ASN122 (-0.13 kcal mol^{-1}); PHE118 (-0.02 kcal mol^{-1}) and ARG121 (1.10 kcal mol^{-1}) contributed less energies while VAL20, ALA181, and SER188 did not influence the complex energetically. The elec. interactions due to residues 20, 80, 117, and 112 and vdW interactions due to residues contributed to the stable system (ΔG_{bind} = -31.41 kcal mol^{-1}), which the *menG*-inhibitor system is stable. For this active site, interactions were mainly

electrostatics in comparison to our study where the vdW interactions contributed more to the binding of DG70.

Thermodynamically, the analysis of the binding free energy of the DG70-*menG* systems showed proximity with the binding free energy obtained (Table 5.2). Taken together, we can infer that our predicted active site is a potential site to consider for the inhibition of the catalytic activity of *menG* protein by DG70. The most favourable binding free energies (kcal mol^{-1}) were -32.45, -31.41, and -31.64 and these data thus agreed with molecular docking scores.

Table 5.2. Thermodynamics analysis of m*enG*-DMK9 and *menG*-DG70 interactions

	Energy components (Kcal mol^{-1})				
Complex	**Δ_{Evdw}**	**Δ_{Eele}**	**ΔG_{gas}**	**G_{sol}**	**ΔG_{bind}**
ApoDMK9	-16.87	-13.92	-15.61	-30.69	-31.64
ApoDG70	-16.09	-13.91	-15.52	-31.50	-32.48
ApoDG70Exp	-16.13	-13.99	-15.60	-30.45	-31.41

Note: ApoDG70Exp is the same as *menG*-DG70 system reported by Suhkeja *et al*. (2017) whose thermodynamics data was obtained to compare with that of ApoDG70 (inhibitor-bound system) studied in this work. ApoDMK9 is the substrate-bound system.

5.6.6 Conformational stability of *MenG* Apo and bound systems

MD simulations provide useful information about the structures and dynamics of biomolecules that may contribute to the development of a drug in design and the overall achievement of the cure of TB. Specifically, proteins undergo many biological processes to regulate their internal components and maintain cellular functions. For the three studied systems, the conformational changes due to the atomistic deviations estimated using C-α backbone root mean square deviation. This system explains the structural stability and system convergence of the biological systems. For the bound or unbound proteins, attaining the structural stability requires longer timescale MD simulations. The C-α atom RMSD metric

used to estimate the stability of unbound *menG* (Apo), DMK9-bound (ApoDMK9), and DG70-bound (ApoDG70) systems.

The RMSDs of the alpha carbon (C-α) backbone of both Apo and bound systems determined throughout 300 ns MD simulations. Figure 5.7B is a graphical representation of C-α atoms RMSD of the systems. The C-α RMSD backbone atoms of all the arrangements were relatively stable. However, the RMSD of the Apo (≈ 8.0 Å) was quite higher than the RMSD of the bound systems which converged (RMSD = ≈ 4.5 Å) from 0 ns up to 110 ns before the RMSD of ApoDG70 rose to 4.7 Å. The average RMSDs (Å) of Apo, ApoDMK9, and ApoDG70 systems were 6.83, 4.66, and 4.97 respectively, and the order of increasing stabilities follows Apo < ApoDG70 < ApoDMK9. The high stability of the bound systems compared with the Apo system could be because of inward pulling interactions of the residues with the bound DG70. This difference leads to structural activation and conformational transformations.

5.6.7 Conformational fluctuation of Mtb *MenG* protein

The RMSF analysis furnishes us with the significance of molecular dynamics at the atomistic level and a better understanding of the conformational changes that occur upon ligand binding. The average RMSF trajectories of MD simulations demonstrate the differences in the flexibilities of the residues with or without a ligand. The RMSF of the C-α backbone for both Apo and bound systems monitored throughout 300 ns simulations. The C-α backbone atoms RMSF of each residue in Apo, ApoDMK9, and ApoDG70 calculated and represented graphically (Figure 5.7C).

The average RMSF (Å) values of Apo, ApoDMK9, and ApoDG70 systems are 2.30, 2.00, and 2.40, respectively. The average RMSF of the C-α backbone in ApoDMK9 was lower than the RMSF of the Apo and ApoDG70. The fluctuations in the ApoDMK9 system

occurred between residues 180-205, which correspond to the ALA91, LYS93, and VAL94 interacting regions. Similarly, distinct high fluctuations occurred between residues 1-120, which contain residues 26, 33, 36 and 40 that formed hydrogen bond, hydrophobic, and salt bridge interactions with DG70 at the active site. Moreover, the longest loop (residue 40-54) found within this region. The flexibility greatly extended in comparison with Apo, and thus confirms the loop flexibility upon DG70-binding. Another fluctuation occurred between 225-234, which may also suggest the binding effect of DG70.

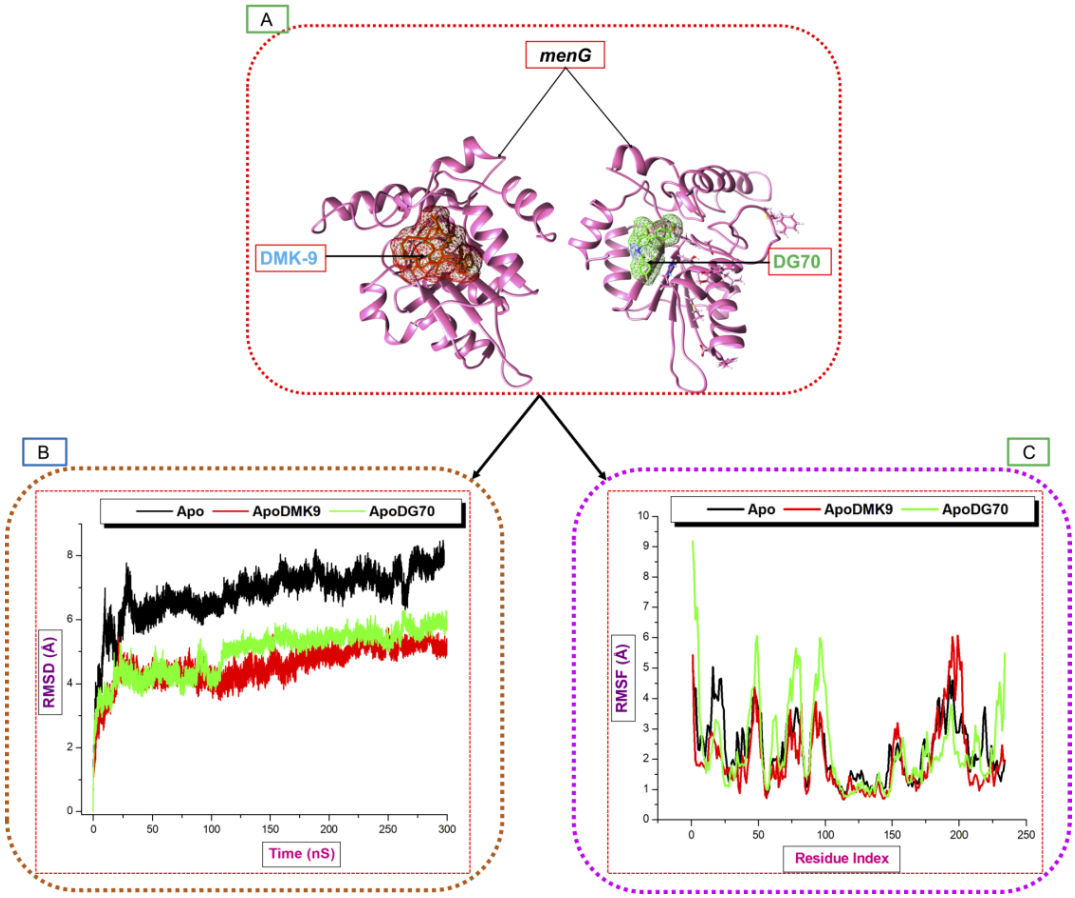


Figure 5.7. C-α backbone RMSD and RMSD plot of Apo and the complexed systems based on the residues ran for 300 ns simulation.

5.6.8 Distribution of atoms (The radius of gyration) around the *MenG* backbones

The shape and folding of the Apo and bound-protein systems could further explore by measuring the radius of distribution/gyration (RoG) around the residue C-α atoms from the

center of mass. Hence, this analysis provides the compactness of a protein that undergoes some dynamic forces. High RoG values depict a less tight structure and increased mobility. The radius of gyration of C-α atoms of *menG* measured before and after binding to DMK9 and DG70 in separate Apo system over 300 ns simulations. The average RoG (Å) of Apo, ApoDMK9, and ApoDG70 systems were 18.14, 18.50, and 18.27 respectively, hence implies that all the three systems showed very similar compactness. However, Figure 5.8B showed the atomic distributions in ApoDG70 from 115-150 ns (low RoG, high stability) and 255-280 ns (high RoG, low stability). These correlate with the high RMSD observed between 110-160 ns and 260-300 ns.

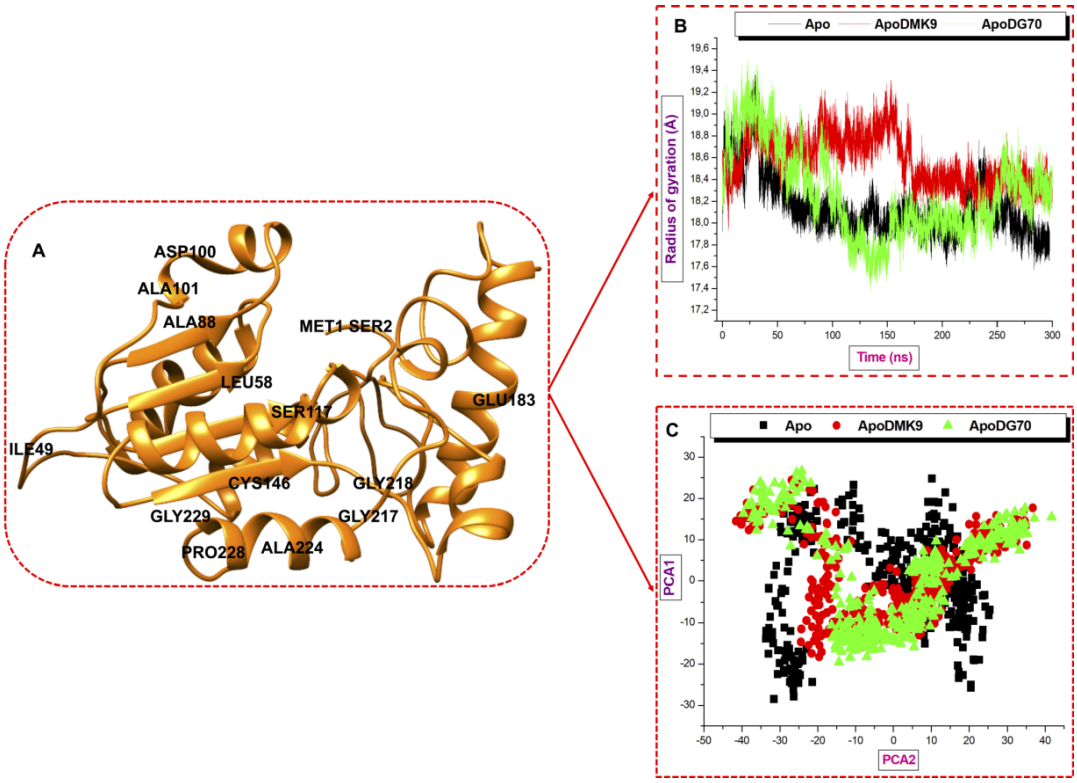


Figure 5.8. A illustrates whole *Mtb menG* structure showing some residues. **B:** Radii of gyrations plot of Apo and ligands-bound systems measured over a 300 ns simulation showed the differences arising in radius deviation. **C:** Projection of Eigenvalues of the C-α backbone, during 300 ns simulation, for Apo, DMK9- and DG70-bound conformations of menG protein along the PC1 and PC2 principal components.

5.6.9 Principal component analysis (PCA) of structural dynamics and motion

PCA is an essential post-dynamics analysis used to measure the conformational transitions and variations of the free (Apo) and ligand-bound protein complexes. It has applied extensively, to study the experimental data and the trajectory of the changes of simulated proteins over a time range. Moreover, the PCA tool is a robust quantitative technique that uses the covariance matrix obtained from the averaged structure using Cartesian coordinates to get the convergence of MD simulation in the space. It provides the means of determining the modes of motion and compares the positional fluctuation in the dynamically simulated structures. Lastly, PCA demonstrated the atoms' displacement and the loop dynamics of a protein and was performed on the *menG* C-α atoms using the CPPTRAJ in AMBER18 GPU to compute the first two components; PC1 and PC2 that are shown graphically in Figure 5.8C. The PC1 (X-axis) and PC2 (Y-axis) represent a covariance matrix after the elimination of eigenvectors. During the simulation, each point between the single-directional motions is a unique orientation due to the overlapping of similar structural conformations.

The eigenvectors computed from the MD trajectories of the Apo and the bound systems varied greatly. The Apo system showed extensive, restricted structural motions of C-α atoms while both DMK9- and DG70-bound have broad spatial coverage which consequently showed that the free system is very rigid.[26] Once again, this result corroborates with the stability of the systems showing the distribution pattern of the around the mass and the deviations of the system stability for the substrate- and inhibitor-bound systems. This result further demonstrates the structural loop flexibility while DG70 bound to the active site of the *menG* enzyme.

5.6.10 Loop dynamics of *MenG* protein and distance metrics

The evolution of enzymes often involves sequence changes in loop regions. The study of loops can provide useful scientific significant reports in determining the protein functions,

including shape, dynamics, binding properties, and physicochemical properties of proteins. They are located on the surface of the protein and often, follow a regular pattern. They are instrumental in protein study whereby loops function as connecting ends for secondary structure (α-helices and β-sheet), protein-ligand interactions, protein-protein interactions, enzyme catalysis, and recognition of sites.[68] The biological function of loops depends on the determinants of their plasticity and the time scale of motions.

There seemed to be little bioinformatics and the experimental findings of methyltransferase protein superfamily, especially the loops of the *menG* enzymes. Generally, the loops are not readily observable. However, we found three loops around the active site and studied them to determine their probable roles in the *menG* conformation and the binding properties. The three *menG* loops contain residues 40-54 (blue), 102-112 (green), and 211-220 (yellow) as shown in Figure 5.9. The residues 40-54, 102-112 and 211-220 visually looked omega-like loop (Ω loop) and showed that the loops involved in the *menG* molecular recognition and regulatory functions.[68] Essentially, residue 211-220 may be the common K-loop in the methyltransferase family that changes conformation upon activation to block the S-adenosyl-L-methionine-binding. Hence, we presumed that *menG* inhibitor bound to the active site of co-factor (SAM) against the substrate's site. Subsequently, residues 211-220, the Ω loop is referred to as the Thompson loop and used interchangeably. To determine the potential roles of these loops, we studied the visual static loop structures to observe possible conformational changes in both the Apo and inhibitor-bound systems. Figure 5.9A and Figure 5.9B showed significant conformational changes in the three loops of the Apo and the DG70-bound structures, respectively, as the simulation time increases (Pre-MD (0), 0.5, 150, and 300 ns).

Thompson loop is essential in the binding of DG70 because it contains at least three most contributing interacting residues, including GLY217, ILE219, and VAL220 of the active site and loop. The active site residues GLY217, ILE219, and VAL220 form the part of the

Thompson loop that cause open conformation in the Apo system as simulation time increases. And the residues GLY217, ILE219, and VAL220 of the Thompson loop cause closing conformational changes in the DG70-bound complex as the simulations time increases.

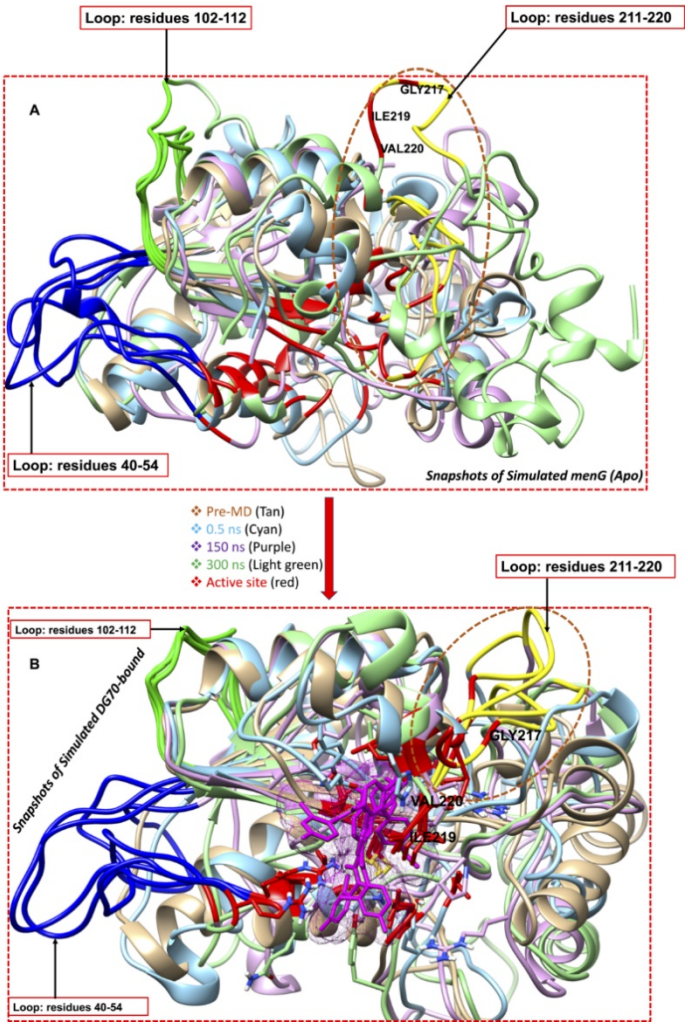


Figure 5.9. Illustration of the opening and closing conformational change of the unbound and bound *menG* protein. **A**; Unbound systems superimposed snapshot at 0, 0.5, 150, 300 ns showing the active site opening of Thompson loop (yellow) and **B**; bound systems superimposed picture at 0, 0.5, 150, 300 ns showing the active site closing conformation by Thompson loop (yellow).

Therefore, the close conformation of the *menG* loop (Thompson loop) enhances the binding of DG70 and subsequently, its therapeutic effectiveness. Specifically, the changes in the formation are due to the loop residues ARG40, GLY217, ILE219, and VAL220 which control the opening and closing of the active site.

Second, the RMSD and RMSF were estimated to explore the dynamics of the stability and flexibility of the enzyme loops, respectively. According to the Figures 5.10B, 5.11D, and 5.11F plots), the unbound systems of the loops maintained a stable structure throughout the simulation time (RMSD < 6.5 Å) while the binding of DMK9 and DG70 induced conspicuous deviations in *menG*. The fluctuations occurred between residues 115-223, which contained GLY217, ILE219, and VAL220 that form hydrophobic and hydrogen bond interactions.

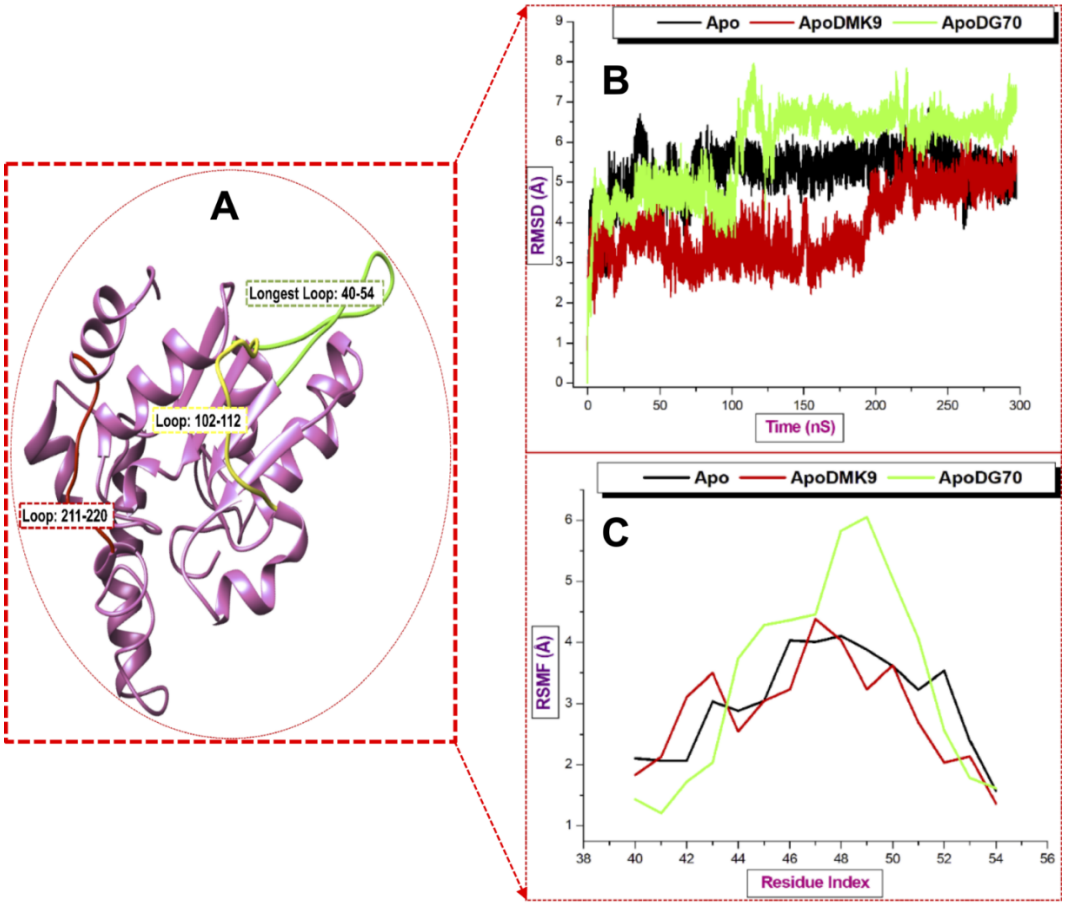


Figure 5.10. A; illustrates three loops in *menG* protein: residues 40-54 (green), 102-112 (yellow), and Thompson loop (red) along the trajectory. RMSD (**B**) and RMSF values (**C**) of the loops.

The atomistic deviations of < 4.5 characterized the residues 102-112 and 211-220 were higher RMSD (~ 6.5 Å) in the region of loop 40-54. Thompson loop was stable throughout while loop 102-112 deviated (2.0 to 4.5 Å). DMK9 and un-liganded *menG* systems had

RMSF values of ~ 4.25 Å and 4.00 Å respectively in the first loop (residues 40-54) and second (residues 102-112) loop while the DG70-bound complex had a higher RMSF value of ~ 6 Å. Hence, the binding of DG70 was not felt around these loops although the RMSF significantly decrease to ~ 3.2 A in the second loop and remained lower (below 3.0 Å) in residues 211-220 than the RMSF of the Apo structure. This finding showed that Thompson loop has a profound impact in the binding of DG70 to *menG* and this correlates with the PRED which depicted that three of the interacting residues (GLY217, ILE219, and VAL220) found in the third loop.

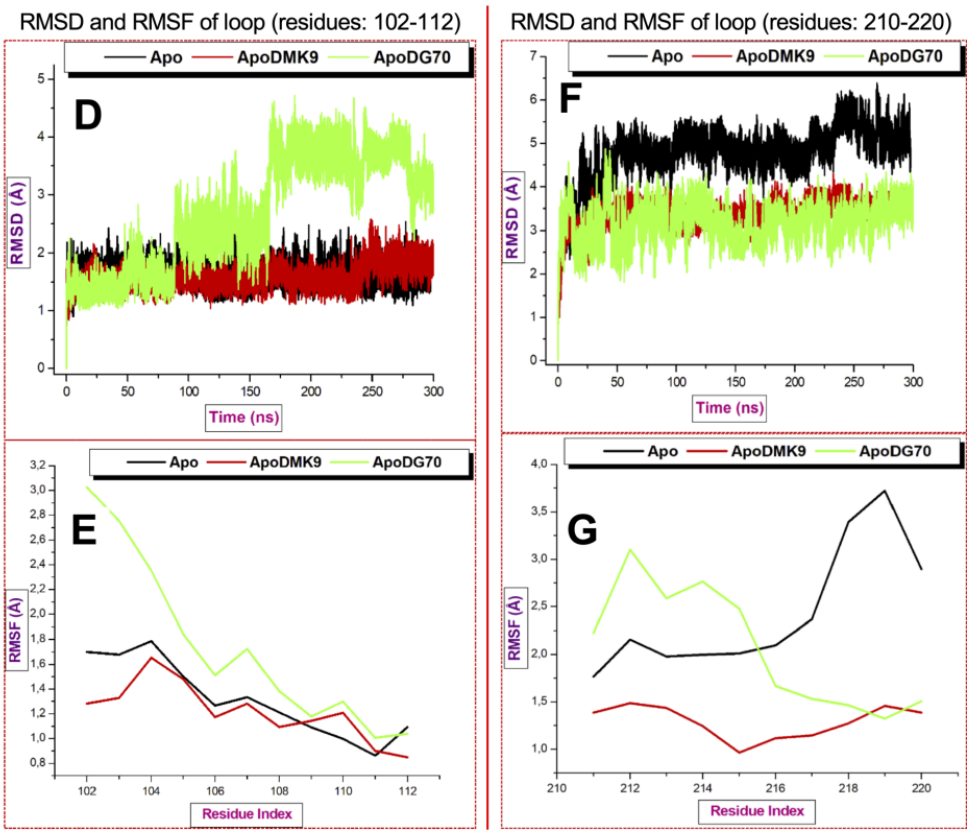


Figure 5.11. Conformational dynamics of Loop 102-112 and 210-220. **D** and **E** are the RMSD and RMSF for loop 102-112. **F** and **G** are the RMSD and RMSF for Thompson loop.

Considering the averaged data, the DG70-bound *menG* had the highest mean RMSD value of 5.84 Å and mean RMSD values 3.34 Å in the loop 40-54 region. The mean RMSD value

(5.84 Å) of Apo across the loops except for residues 102-112. Therefore, we can infer that the binding of DG70 caused instability in the *menG* structure because of inhibition.

Figure 5.10C, 11E, and 11G plots showed the RMSF values of the three loops. The highest average RMSF values of Apo, DMK9-bound, and DG70-bound were observed in the loop 40-54 region. In comparison, DG70-bound *menG* and Apo had RMSF values of 3.34 and 3.04, respectively, which implies similarities in their flexibilities. Table 5S1 (Supplementary) detailed the loops averaged RMSD and RMSF estimated for the unbound and bound-*menG* system.

To further probe into the inhibitory efficacy of DG70 on the inter-residual dynamics and motions of *menG*, the distance, D analysis was estimated from the snapshots of the 300 ns MD simulations of the Apo and bound-*menG* systems. The distances measured at 0, 150, and 300 ns between the interacting residues VAL145 and ILE219 of the bound and unbound *menG*. Figure 5.12A, 5.12C and 5.12E showed that the distance between the two residues in the unbound *menG* decreases with increasing simulation time. At 0 ns, the distance was 12.356 Å and rose to 18.91 Å after 150 ns, and 20.49 Å after 300 ns. On the contrary, the distance between the two residues in the DG70 bound system decreases; between 0 ns and 150 ns, the distance decreased from 11.55 Å to 8.12 Å and again decreased to 7.36 Å after 300 ns Figure 5.12B, 5.12D and 5.12F.

The distance increased in the Apo structure and decreased in the DG70-bound in the same proportion as the simulation time. The decrease in the distance enhanced the DG70-bound *menG* interactions and the stability of the system. In other words, the decrease in the distance caused stronger interactions between the DG70 and the interacting residues at the active site, thus, promoting effective inhibition of the methylation activity of *menG* enzyme.

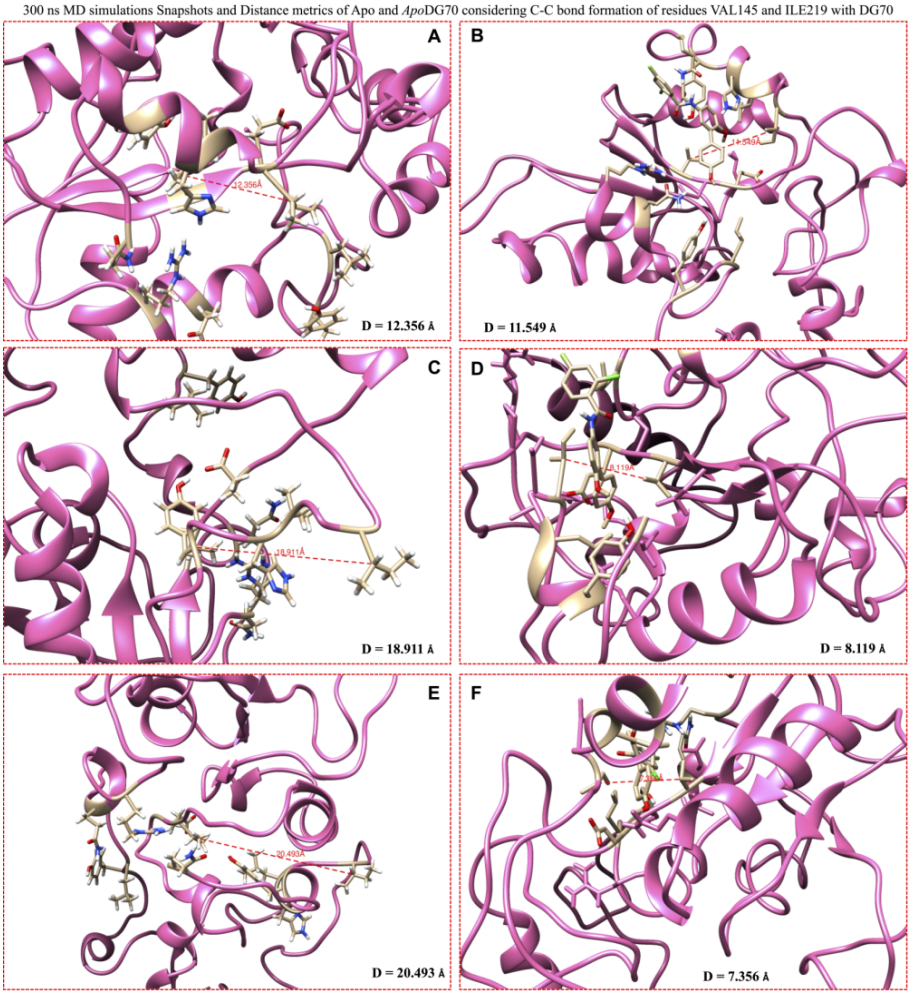


Figure 5.12. Distance (D) between VAL145 and ILE219 of the Apo and DG70-bound *menG* measured using the snapshots taken at 0, 150, 300 ns MD simulations. The D (Å) measured from Apo snapshots (ns) **A** (0), **C** (150), and **E** (300) are 12.356, 18.911, and 20.493 respectively. The D measured from DG70-bound snapshots **B**, **D**, and **F** are 11.549, 8.119, and 7.356, respectively.

5.6.11 Secondary structure (DSSP) analysis of *MenG* systems

A secondary structure assignment analysis carried out to study the prominent secondary structure elements, including alpha helices, beta sheets, loops, and coils in the *menG* protein during the whole 300 ns simulation time. There were 9α-helices, 4β-sheets, and three long conspicuous loops in the *menG* structure discussed above before simulations (supplementary Figure 5S3A). However, after post-MD simulations, (supplementary Figure 5S3B) revealed the visual transformations in the secondary structure of *menG* such as conformational

changes in the helices and β-sheets, perhaps, changes to the loop structures because there seemed to be helices and beta sheets that transformed to loop.

Therefore, we did a comparative defined secondary structure of the un-liganded, DMK9-, and DG70-bound *menG* systems. The trajectories of the snapshots after at 300 ns were obtained and plotted using the in-house protocol. Supplementary Figure 5S4 depicts the secondary structures of *menG* after 300 ns MD simulations. Supplementary Figure 5S4A is the unbound *menG* after 300 ns. We discovered that there were changes from the helices to loop and beta to loop after the 300 ns that led to the increased number of loops in the bound systems; there was a drastic change from alpha to bend (residues 1-101 and 142-178) and turn (residues 181-201) in DG70-bound *menG* (Supplementary Figure 5S4C) at 300 ns. There were transitions in the residues 31-39, 154-160, and 225-229 from the helices to bend upon DG70 binding. There were also transitions in the beta residues 85, 117, and 142-146 to bend. On the other hands, there was a change from alpha to bend upon the binding of DMK9 to *menG* (Supplementary Figure 5S4B) in residues 1-31 after the 300 ns. These changes in the elements of *menG* secondary structure were the cause of the instability[69] that the DG70-bound *menG* displaced in the RMSF values of the whole and the loop regions stated in sections 5.7 and 5.10.

5.6 Conclusion

The detailed molecular modelling study and the dynamics analyses provided in this report sets out the unique structural features and the conformational changes in *menG* whole structure and the omega-like loops especially the Thompson loop (residues 211-220) region. First, the Uniprot annotated residues were among the predicted active site residues. Nevertheless, the per-residue decomposition analysis showed that the none of the annotated residues did interact with DMK or DG70. On the other hand, PRED revealed that the substrate and the inhibitor bound to interacted with different residues which separates

different active sites. Second, the predicted active site showed similarity to the previous prediction, however, the Ramachandran plot revealed that residues 7, 9, and 10 considered as interacting residues in the previous work are outliers.

Homology modelling revealed the secondary structure of the *menG* enzyme and some static features including the loops, α-helices, β-strands and GXG motif that showed similarities with the superfamily of methyltransferase. MD simulations showed motional deviations in the *menG* whole and its loop at the binding site. The stability of the DG70-bound system showed the consistency of the bound inhibitor at the active site throughout the 300 ns simulations. This complex stability correlates with the high binding free energy calculated and the residue-ligand interaction networks characterized by the strong hydrophobic, halogen, and hydrogen bond.

The graphical investigation of the loop dynamics at different time intervals confirmed the strong flexibilities that occurred in the whole structures of the free and the bound *menG* systems revealed by the RMSF analysis. The decreased distances between the interacting residues VAL145 and ILE219 in the DG70-bound system as against the increased distances between the same residues in the un-liganded system throughout a 300 ns MD further confirmed the binding impact of DG70 on inhibition of *menG* catalytic activity. The PCA eigenvectors obtained from the simulations showed clear variations among the systems which confirmed the dynamic conformational changes from the free to ligand-bound protein. Visually, there were some transitional changes in the secondary structure of *menG* to the loops hence, the free and the bound systems were analysed using DSSP analysis which confirmed some conformational transformations such as helices to the loop.

Perhaps, it is important to carryout experimental analysis such as crystallographic studies to further explore the *menG* loops in relation to their functions in the ligand binding. The homology modelling insights will serve as aid to the researcher to synthesize the crystallized

structure of this protein. Moreover, the demonstration of the above binding landscape of *menG* protein provides the basis that can utilize pharmacophore models in screening for better effective drugs with tolerable or no toxicity.

Acknowledgment

First, the authors acknowledge THE ALMIGHTY GOD who created all things including lives, wisdom, and knowledge for every human being. Second, we acknowledge the Centre for High-Performance Computing, Cape-Town for providing computational resources, University of KwaZulu-Natal, Durban, South Africa for the free Tuition remission fees, and College of Health Sciences for their support.

Compliance with Ethical Standards

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Disclosure of Potential Conflict of Interest

The authors declare no conflict of interest.

Research involving Human Participants and/or Animals

Informed Consent

This study did not require informed consent since the study does not contain any studies with human participants performed by any of the authors.

FUNDING

No funding was received for this study

References

1. Ahmad, N. M. R. *et al.* Analyzing Policymaking for Tuberculosis Control in Nigeria.

Complexity **2018**, 1–13 (2018).

2. Sharma, A. *et al.* Estimating the future burden of multidrug-resistant and extensively drug-resistant tuberculosis in India, the Philippines, Russia, and South Africa: a mathematical modelling study. *The Lancet Infectious Diseases* **17**, 707–715 (2017).

3. Lee, E. H. *et al.* Improved Fluoroquinolone-Resistant and Extensively Drug-Resistant Tuberculosis Treatment Outcomes. *Open Forum Infectious Diseases* **6**, 1–7 (2019).

4. Mokrousov, I., Akhmedova, G., Polev, D., Molchanov, V. & Vyazovaya, A. Acquisition of bedaquiline resistance by extensively drug resistant *Mycobacterium tuberculosis* strain of Central Asian Outbreak clade. *Clinical Microbiology and Infection* 6–8 (2019). doi:10.1016/j.cmi.2019.06.014

5. Ghajavand, H. *et al.* High prevalence of bedaquiline resistance in treatment-naive tuberculosis patients and verapamil effectiveness. *Antimicrobial Agents and Chemotherapy* **63**, 1–5 (2019).

6. Olaru, I. D., Heyckendorf, J., Andres, S., Kalsdorf, B. & Lange, C. Bedaquiline-based treatment regimen for multidrug-resistant tuberculosis. *The European respiratory journal* **49**, (2017).

7. Eyers, C. E., Vonderach, M., Ferries, S., Jeacock, K. & Eyers, P. A. Understanding protein–drug interactions using ion mobility–mass spectrometry. *Current Opinion in Chemical Biology* **42**, 167–176 (2018).

8. Tabei, Y., Kotera, M., Sawada, R. & Yamanishi, Y. Network-based characterization of drug-protein interaction signatures with a space-efficient approach. *BMC Systems Biology* **13**, 1–15 (2019).

9. Andries, K. *et al.* Acquired resistance of *Mycobacterium tuberculosis* to bedaquiline.

PLoS ONE **9**, 1–11 (2014).

10. Margaretha de Vos, Kristin B. Wiggins, Brigitta Derendinger, Anja Reuter, Tania Dolby, Scott Burns, Marco Schito, David M. Engelthaler, John Metcalfe, Grant Theron, Annelies van Rie, James Posey, Rob Warren, H. C. Bedaquiline Microheteroresistance after Cessation of Tuberculosis Treatment. *New Engl Journal of Medicine* **380**, 2178–2180 (2019).

11. Chawla, K., Martinez, E., Kumar, A., Shenoy, V. P. & Sintchenko, V. Whole-genome sequencing reveals genetic signature of bedaquiline resistance in a clinical isolate of *Mycobacterium tuberculosis*. *Journal of Global Antimicrobial Resistance* **15**, 103–104 (2018).

12. Johnston, J. M., Arcus, V. L., Morton, C. J., Parker, M. W. & Baker, E. N. Crystal Structure of a Putative Methyltransferase from m *Mycobacterium tuberculosis*: Misannotation of a Genome Clarified by Protein Structural Analysis. *Society* **185**, 4057–4065 (2003).

13. Sukheja, P. *et al.* A Novel Small-Molecule Inhibitor of the *Mycobacterium tuberculosis* Demethylmenaquinone Methyltransferase MenG Is Bactericidal to Both Growing and Nutritionally Deprived Persister Cells. *American Society for Microbiology* **8**, 1–15 (2017).

14. Iqbal, I., Bajeli, S., Akela, A. & Kumar, A. Bioenergetics of Mycobacterium: An Emerging Landscape for Drug Discovery. *Pathogens* **7**, 1–30 (2018).

15. Lee, P. T., Hsu, A. Y., Ha, H. T. & Clarke, C. F. A C-methyltransferase involved in both ubiquinone and menaquinone biosynthesis: Isolation and identification of the Escherichia coli ubiE gene. *Journal of Bacteriology* **179**, 1748–1754 (1997).

16. Paridhi Sukheja, Pradeep Kumar, Nisha Mittal, Shao-Gang Li, Eric Singleton, A., Riccardo Russo, Alexander L. Perryman, Riju Shrestha, Divya Awasthi, B., Seema Husain, Patricia Soteropoulos, Roman Brukh, Nancy Connell, A. & Joel S. Freundlich, D. A. A Novel Small-Molecule Inhibitor of the *Mycobacterium tuberculosis* Demethylmenaquinone Methyltransferase MenG Is Bactericidal to Both Growing and Nutritionally Deprived Persister Cells Paridhi. **8**, 1–15 (2017).

17. Zou, X. W. *et al.* Structure and mechanism of a nonhaem-iron SAM-dependent C-methyltransferase and its engineering to a hydratase and an O-methyltransferase. *Acta Crystallographica Section D: Biological Crystallography* **70**, 1549–1560 (2014).

18. Ramharack, P. & Soliman, M. E. S. Zika virus NS5 protein potential inhibitors: an enhanced in silico approach in drug discovery. *Journal of Biomolecular Structure and Dynamics* **36**, 1118–1133 (2018).

19. Kwon, O. & Meganathan, R. Biosynthesis of Menaquinone (Vitamin K2) and Ubiquinone (Coenzyme Q). *EcoSal Plus* **3**, (2009).

20. Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransfer: A chronicle of convergence. *Trends in Biochemical Sciences* **28**, 329–335 (2003).

21. Struck, A. W., Thompson, M. L., Wong, L. S. & Micklefield, J. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **13**, 2642–2655 (2012).

22. Puffal, J., Mayfield, J. A., Moody, D. B. & Morita, Y. S. Demethylmenaquinone Methyl Transferase Is a Membrane Domain-Associated Protein Essential for Menaquinone Homeostasis in Mycobacterium smegmatis. *Frontiers in Microbiology* **9**, 1–12 (2018).

23. Wlodarski, T. *et al.* Comprehensive structural and substrate specificity classification of the saccharomyces cerevisiae methyltransferome. *PLoS ONE* **6**, e23168 (2011).

24. Zimenkov, D. V. *et al.* Examination of bedaquiline- and linezolid-resistant *Mycobacterium tuberculosis* isolates from the Moscow region. *Journal of Antimicrobial Chemotherapy* **72**, 1901–1906 (2017).

25. Sherlin, D. & Anishetty, S. Mechanistic insights from molecular dynamic simulation of Rv0045c esterase in *Mycobacterium tuberculosis*. *Journal of Molecular Modeling* **21**, 1–8 (2015).

26. Ramharack, P., Oguntade, S. & Soliman, M. E. S. Delving into Zika virus structural dynamics-a closer look at NS3 helicase loop flexibility and its role in drug discovery. *RSC Advances* **7**, 22133–22144 (2017).

27. Grāve, K., Bennett, M. D. & Högbom, M. Structure of *Mycobacterium tuberculosis* phosphatidylinositol phosphate synthase reveals mechanism of substrate binding and metal catalysis. *Communications Biology* **2**, 1–11 (2019).

28. Chen, C., Huang, H. & Wu, C. H. *Chapter 1 Protein Bioinformatics Databases and Resources*. **1558**, (2017).

29. Morgat, A. *et al.* Enzyme annotation in UniProtKB using Rhea. *Bioinformatics (Oxford, England)* **36**, 1896–1901 (2020).

30. Brister, J. R., Ako-Adjei, D., Bao, Y. & Blinkova, O. NCBI viral Genomes resource. *Nucleic Acids Research* **43**, D571–D577 (2015).

31. Chetty, S. & Soliman, M. E. S. Possible allosteric binding site on Gyrase B, a key target for novel anti-TB drugs: Homology modelling and binding site identification using molecular dynamics simulation and binding free energy calculations. *Medicinal*

Chemistry Research **24**, 2055–2074 (2015).

32. Hu, G. & Kurgan, L. Sequence Similarity Searching. *Current Protocols in Protein Science* **95**, 1–19 (2019).

33. Pettersen, E. F. *et al.* UCSF Chimera — A Visualization System for Exploratory Research and Analysis. **25**, 1605–1612 (2004).

34. Zheng Yanga, Keren Laskerb, Dina Schneidman-Duhovny, Ben Webb, Conrad C. Huang, Eric F. Pettersen, Thomas D. Goddard, Elaine C. Meng, Andrej Sali, and T. E. F. UCSF Chimera, MODELLER, and IMP: an Integrated Modeling System. *Journal of Structural Biology* **23**, 1–7 (2012).

35. Sievers, F. *et al.* Fast , scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **7**, 1–6 (2011).

36. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **66**, 12–21 (2010).

37. Huang, B. MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. *OMICS: A Journal of Integrative Biology* **13**, 325–330 (2009).

38. Adeniji, E. A., Olotu, F. A. & Soliman, M. E. S. Exploring the Lapse in Druggability: Sequence Analysis, Structural Dynamics and Binding Site Characterization of K-RasG12C Variant, a Feasible Oncotherapeutics Target. *Anti-Cancer Agents in Medicinal Chemistry* **18**, 1540–1550 (2018).

39. Wang, S., Li, W., Liu, S. & Xu, J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic acids research* **44**, W430–W435 (2016).

40. Wass, M. N., Kelley, L. A. & Sternberg, M. J. E. 3DLigandSite: Predicting ligand-

binding sites using similar structures. *Nucleic Acids Research* **38**, 469–473 (2010).

41. Heo, L., Shin, W. H., Lee, M. S. & Seok, C. GalaxySite: Ligand-binding-site prediction by using molecular docking. *Nucleic Acids Research* **42**, 210–214 (2014).

42. Wang, H. *et al.* Template-based protein structure modeling using the RaptorX web server. *HHS Public Access* **7**, 1511–1522 (2016).

43. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research* **40**, 294–297 (2012).

44. Desai, S., Tahilramani, P., Patel, D., Meshram, D. & Patel, P. in Silico Prediction and Docking of Tertiary Structure of Protein X, Multifunctional Protein of Hepatitis B Virus (Hbv). *Indo American Journal of Pharmaceutical Research* **3**, 7–9 (2017).

45. Rene Thomsen and Mikael H. Christensen. MolDock: A New Technique for High-Accuracy Molecular Docking. *J. Med. Chem.* **663**, 3315–3321 (2006).

46. Subramanian, U., Sivapunniyam, A., Pudukadu Munusamy, A. & Sundaram, R. An in Silico Approach towards the Prediction of Druglikeness Properties of Inhibitors of Plasminogen Activator Inhibitor. *Advances in Bioinformatics* **2014**, 1–7 (2014).

47. Marcus D Hanwell, DonaldECurtis, David C Lonie, Tim Vandermeersch, E. Z. and G. R. H. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* **4**, 1–17 (2012).

48. Allouche, A. Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares. *Journal of computational chemistry* **32**, 174–182 (2012).

49. Assmus, H. E., Herwig, R., Cho, K. H. & Wolkenhauer, O. Dynamics of biological systems: Role of systems biology in medical research. *Expert Review of Molecular*

Diagnostics **6**, 891–902 (2006).

50. O'Hara, L. *et al.* Modelling the Structure and Dynamics of Biological Pathways. *PLoS Biology* **14**, 1–16 (2016).

51. Shunmugam, L. & Soliman, M. E. S. Targeting HCV polymerase: A structural and dynamic perspective into the mechanism of selective covalent inhibition. *RSC Advances* **8**, 42210–42222 (2018).

52. 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. *Journal of Computational Physics* **23**, 327–341 (1977).

53. Soremekun, O. S., Olotu, F. A., Agoni, C. & Soliman, M. E. S. Recruiting monomer for dimer formation: resolving the antagonistic mechanisms of novel immune check point inhibitors against Programmed Death Ligand-1 in cancer immunotherapy. *Molecular Simulation* **45**, 777–789 (2019).

54. William Humphrey, Andrew Dalke, and K. S. T. VMD: Visual Molecular Dynamics William. *Journal of Molecular Graphics* **14**, 33–38 (1996).

55. Kalathiya, U., Padariya, M. & Baginski, M. Structural, functional, and stability change predictions in human telomerase upon specific point mutations. *Scientific Reports* **9**, 1–13 (2019).

56. Seifert, E. OriginPro 9.1: Scientific data analysis and graphing software - Software review. *Journal of Chemical Information and Modeling* **54**, 1552 (2014).

57. Vytautas Gapsys, Servaas Michielssens, Jan Henning Peters, Bert L. de Groot, and H. L. *Calculation of Binding Free Energies*. *Methods Mol. Biol.* **1215**, (2015).

58. Schauperl, M. *et al.* Binding Pose Flip Explained via Enthalpic and Entropic

Contributions. *Journal of Chemical Information and Modeling* **57**, 345–354 (2017).

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

60. Martínez, L. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS ONE* **10**, 1–10 (2015).

61. Lobanov, M. Y., Bogatyreva, N. S. & Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Molecular Biology* **42**, 701–706 (2008).

62. Post, M., Wolf, S. & Stock, G. Principal component analysis of nonequilibrium molecular dynamics simulations. *Journal of Chemical Physics* **150**, 1–11 (2019).

63. Ramírez, D. Computational Methods Applied to Rational Drug Design. *The Open Medicinal Chemistry Journal* **10**, 7–20 (2016).

64. Batool, M., Ahmad, B. & Choi, S. A structure-based drug discovery paradigm. *International Journal of Molecular Sciences* **20**, 1–18 (2019).

65. Guedes, I. A., de Magalhães, C. S. & Dardenne, L. E. Receptor-ligand molecular docking. *Biophysical Reviews* **6**, 75–87 (2014).

66. Ferreira, L. G., Dos Santos, R. N., Oliva, G. & Andricopulo, A. D. *Molecular docking and structure-based drug design strategies*. *Molecules* **20**, (2015).

67. Ramírez, D. & Caballero, J. Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules* **23**, 1–17 (2018).

68. Papaleo, E. *et al.* The Role of Protein Loops and Linkers in Conformational Dynamics

and Allostery. *Chemical Reviews* **116**, 6391–6423 (2016).

69. Tanwar, Himani and Doss, C. G. P. An Integrated Computational Framework to Assess the Mutational Landscape of α-L-Idurondase IDUA Gene. *Journal of Cellular Biochemistry* 555–565 (2018). doi:10.1002/jcb.26214

CHAPTER 6

Weak spots inhibition in the *Mycobacterium tuberculosis* antigen 85C target for Antitubercular drug design through selective irreversible covalent inhibitor-SER124

Adeniyi T. Adewumi[a], Ahmed Elrashedy[a], Soremekun S. Opeyemi[a], Mary B. Ajadi[b], and Mahmoud E.S. Soliman a*

[a]Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

[b]Department of Medical Biochemistry, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Howard Campus, Durban 4000.

*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 8048

Fax: +27 (0) 31 260 7872

6.1 Abstract

Mycobacterium tuberculosis (*Mtb*) encoded secreted antigen 85 enzymes (Ag85A/Ag85B/ Ag85C) play that critical roles in the virulence, survival, and drug-resistant TB of the pathogen. Ag85 proteins are potential antitubercular drug targets because they are essential in the catalytic synthesis of trehalose moieties and mycolic acid attachment to the *Mtb* cell wall. Recently, experimental protocols led to the discovery of a selective covalent Ag85 inhibitor, β-isomer monocyclic enolphosphorus Cycliphostin ($CyC_{8\beta}$) compound, which targets the Ag85 serine 124 to exhibit a promising therapeutic activity. For the first time, our study unravelled the structural features among *Mtb* Ag85C homologs and motions and dynamics of Ag85C when the $CyC_{8\beta}$ bound covalently and in open model conformations to the protein using bioinformatics tools and integrated Molecular dynamics simulations. Comparative Ag85C sequence analysis revealed conserved regions; 70 % active site, 90 % Adeniyi loop L1, and 50 % loop L2, which acts as a switch between open and closed conformations. The average C-α atoms RMSD (2.05 Å) and RMSF (0.9 Å) revealed instability and high induced flexibility in the $CyC_{8\beta}$ covalent-bound compared to the apo and open model systems, which displayed more stability and lower fluctuations. DSSP showed structural transitions of α-helices to bend and loops to 3_{10}-helices in the bound systems. SASA of $CyC_{8\beta}$ covalent bound showed active site hydrophobic residues exposure to huge solvent. Therefore, these findings present the potential opportunity hotspots in Ag85C protein that would aid the structure-based design of novel chemical entities capable of resulting in potent antitubercular drugs.

Keyword: Antigen 85, Adeniyi Loops, drug-resistant TB, Molecular dynamics simulations, β-isomer enolphosphorus Cycliphostin ($CyC_{8\beta}$).

6.2 Introduction

Tuberculosis (TB) is a lung infection caused by the pathogen known as *Mycobacterium tuberculosis* (*Mtb*) that has "blown away" human generations. About 10 million new TB infections, which result in approximately 1.3 million deaths[1] are reported annually in humans.[2] Although there have been various strategies, including diagnosis, vaccine program, and curative therapy channeled toward TB eradication, high morbidity and mortality rate of TB has continued for many decades. More challenging TB is the metamorphosising active form caused by the evolving *Mtb strains due to* mutation-mediated resistance such as the non-synonymous single-nucleotide polymorphisms (nsSNP)[3, 4] in Ag85 protein by insertions or deletions in the genes encoding from drug target/drug metabolic pathways. Besides, the induced-mutation resistance could also occur due to clonal expansion and independently acquired drug resistance.[5]

Moreover, the slow growth rate of bacilli and the thick, greasy drug-impermeable nature of the cell wall contributes to the success of the *Mtb* drug resistance. The increasing rate of active TB (or extensively/multidrug) resistant TB is quite worrisome. Some of the frontline antitubercular drugs exhibiting resistance include isoniazid, rifampicin. Moreover, literature has shown that new drugs like bedaquiline exhibit frameshift and phenotypic drug-resistance.[6,7] Although some of these drugs are still therapeutically relevant, especially as combined anti-TB therapies, an associated severe problem is the many drug dose and dosage regimen, requiring two years treatment period, which results in an inherent difficulty in the patients' compliance.[8]

The cell of mycobacteria contains different proteins that are responsible for the successes of the microorganisms in pathogenesis, survival, and resistance to many drugs.[8,9] Among these proteins, *Mtb* antigen (Ag) 85 protein complex (Ag85A, Ag85B, and Ag85C) bind to the human fibronectin via ether heparin and cellular-binding domains or a collagen-binding domain as it escapes the host immune system.[10,11] Ag85 proteins are potential targets for the

treatment of drug-susceptible and drug-resistant TB because they catalyse the attachment of mycolic acids in the cell wall and biosynthesis pathway of TMM and TDM.[12] *Mycobacterium tuberculosis* cell envelope comprises macromolecules, the mycolyl-arabinogalactan-peptidoglycan. In other words, the macromolecules include mycolic acids, arabinogalactan (AG), and peptidoglycan (PG), (mAGP complex). PG is found outside the plasma membrane covalently linked to the AG. The mycolic acids are long-chain α-alkyl-β-hydroxyl fatty acids (C70-90) that occupy about 60 % of the whole cell wall and are attached to the AG. They are esterified to the monomycolate (TMM) or dimycolate (TDM) of trehalose, which serve as donors in the metabolism of mycolic acids. Moreover, the mycolic acids interspersed with non-covalent glycolipids, proteins, and outer polysaccharide capsules.[9]

Mtb antigen 85 enzymes are quite crucial in the attachment of mycolic acids to the cell wall and synthesis of trehalose moieties. Besides, this protein involves the efflux mechanism, which may result in the phenotypic resistance.

The enzymes share common structural and functional similarities such as mycolic acid donor TMM and 68-79 % conserved catalytic site sequence identity. Additionally, they have serine mycolyl esterase activity, carboxylesterase consensus sequence (GXSXG), among others.[13] Therefore, a single removal of one of the enzymes affects the complex containing the three proteins. Precisely, the alteration of the *fbpC2* gene result into 40% reduction in the AG-attached mycolic acids and the reductive impact on either *fbpA* or *fbpB* genes leads to TDM reduction, which indicates low functional activity [14, 15]

Figure 6.1 illustrates structural features of Ag85C protein. *Mycobacterium tuberculosis* antigen 85C possesses α/β fold arrangements that are typical of the hydrolase family.[13] This feature is an implication of similarities in the sequence, mechanism, and structural studies for the antigen enzymes. The Ag85C is a 31.5 kDa[10] secreted protein comprising of about 274 amino acids residues (PDB code 5OCJ)[12]. The crystallized structure

consists of Chain A (8-282 residues) and Chain B (6-282 residues) located in the C-terminus except for the FSRPG residues, which are found in the N-terminus.[10] The inhibition of the catalytic mechanism of the Ag85C enzyme involves the attack of its nucleophilic serine by a selective covalent inhibitor. The proteins exhibit a conserved catalytic triad (SER124, GLU228, HIS260) at the active site, about nine α-helices arranged in up and down circular fashion surrounded by parallel and one opposite β-sheets localised at the centre.

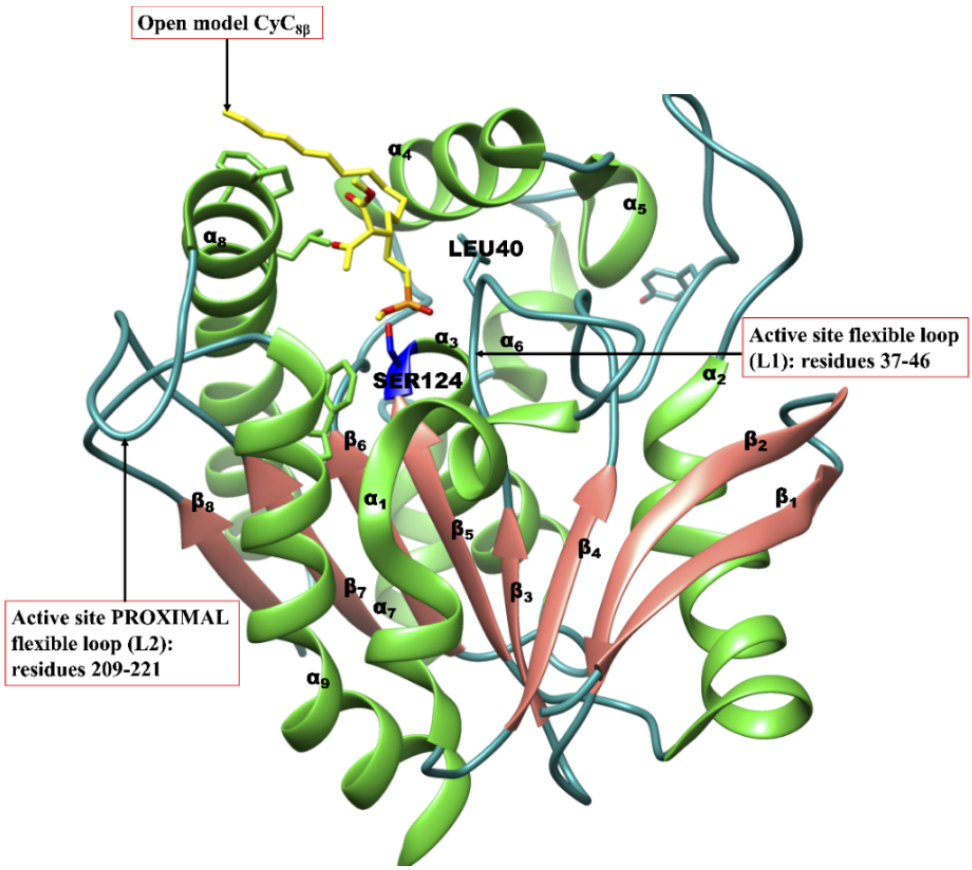


Figure 6.1. The 3D structure of Ag85C protein detailed some of its structural features; α-helices (lime green), β-strands (Salmon), and loops (dark cyan); Serine124 (blue) in complex with open model $CyC_{8\beta}$ (yellow).

Drug-rug-resistance TB, "The wicked monster" and has rendered many novel efficacious antimicrobial drugs ineffective for decades. Although some of the current antitubercular agents against the cell wall proteins are still relatively potent, as combined therapy, many challenges are confronting the long-term effectiveness and efficacy of these drugs. First, poor

patient compliance often due to massive doses and lengthy dosage regimen eventually results in drug resistance.[16] For instance, the inhibitor, including promising ebselen (selective covalent inhibitor) bonded with Cysteine-209 (proximal residue to the active site) of Ag85C, in synergy with p-chloromercuribenzoic acid and iodoacetamide to inhibit the enzyme's catalytic activity.[17] However, the studies also revealed that the *Mtb* Ag85C mutants disrupted the essential interactive hydrogen-bond network at the active site linked to selective covalent inhibition.[17] Second, TB and HIV/AIDS co-infection in human also contributes to the success of drug-resistant *Mycobacterium strains*.[18, 19, 20] And again, the impact of the outbreak COVID-19 infection caused by new novel severe acute respiratory syndrome coronavirus (SAR-CoV-2), was anticipated and found to increase the annual human TB deaths[21, 22]. Moreso, the infections are pulmonary infections.

The design of small molecules against TB recently led to the discovery of some cyclipostins and cyclophostin phosphonate analogues (CyCs) and cyclic enolphosph(on)ate compounds. The compounds exerted their therapeutic activities by inhibiting the catalytic synthesis of Ag85 through the selective covalent bond formation with Ag85 SER124. This intentive approach showed potent antimycobacterial activity with low toxicity towards the host cells.[23] The compounds have a strong affinity for lipolytic enzymes bearing a catalytic serine (SER) or cysteine (CYS) residues to impair enzyme activity involving a wide range of physiological processes.[8] The mechanism of action of the compounds follows a putative selective covalent inhibition; nucleophilic attack by a catalytic SER or CYS to form a covalent bond between the enol-phosphorus atom and the catalytic residue.[8] In a previous study, among eight members that exhibited antitubercular activities, $CyC_{7\beta}$ and $CyC_{8\beta}$ showed strong *in vivo* potencies (MIC_{50}, μM IC_{50} = 3.1 and ≈11.6 respectively). In contrast, CyC_{17} showed *in vitro* (MIC_{50} = 0.5 μM but lack intracellular activity. Again, another study confirmed $CyC_{8\beta}$ as the most potent inhibitor with IC50 =55 ±5 μM) against Ag85C protein, while $CyC_{7\beta}$ and

CyC_{17} followed.[12] The $CyC_{8\beta}$ inhibited the Ag85C activity onto trehalose, which prevents the mycolic acid transfer to AG in the cell wall and the TDM/TMM synthesis. Therefore, we considered cis (β-isomer)-monocyclic enolphosphonate cycliphostin ($CyC_{8\beta}$). Figure 6.2 is the 2D chemical structure of $CyC_{8\beta}$ that was cleaned with 3D add on in the MarvinSketch software-17.21. (http://www.chemaxon.com/).

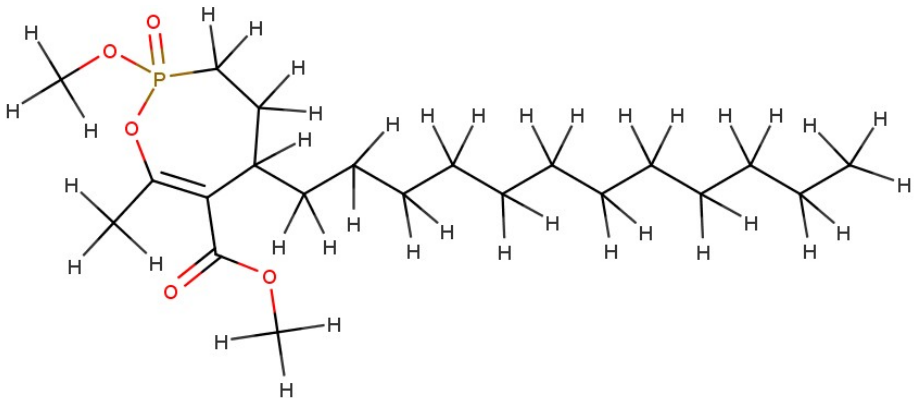


Figure 6.2. 2D structure of Ag85C inhibitor, a monocyclic enolphosphonate cyclophostin.

Here, the protein sequence alignment analysis among the homologs of *Mtb* Ag85C protein and selective covalent inhibition of *Mtb* Ag85C using the molecular dynamic (MD) simulations of the free and inhibitor-bound in two docked mode studied for the first time. Therefore, our study investigated the interactions between the Ag85C protein and inhibitor ($CyC_{8\beta}$) at the molecular level using bioinformatics Tools and molecular dynamics methods. This study unravelled Ag85C structural features using bioinformatics Tool and integrated molecular dynamics (MD). Furthermore, we investigate the dynamics and motions of Ag85C when $CyC_{8\beta}$ bound covalently and noncovalently to the protein.

6.3 Bioinformatics Tools and Computational Methodology

6.3.1 Sequence alignment of the homologs of *Mycobacterium tuberculosis* Ag85C protein

Bioinformatic Tools are useful in the drug design research, and we had used it in our previous studies to comparative sequence alignments among the protein homologs.[4] The unique regions in the primary structures of five (5) homologs of the *Mycobacterium tuberculosis* antigen 85C enzymes investigated using the sequence alignment tools adds-on available through UCSF ChimeraTool-1.13.1 with the default setting.[24]

Table 6.1 showed the respective Uniprot accession codes and the number of amino acid residues of the homologs of Mycobacteria Ag85C protein including *M. tuberculosis*, *M. smegmatis*, *M. marinum*, *M. leprae*, and *M. ulcerans* used for the multiple sequence analysis.[25] The mycobacteria species were chosen based on the pathogenesis, virulence, survival, and resistance success.

Table 6.1. UniProt/PDB information of five mycobacteria antigen 85C homologs

Mycobacterium species antigen 85 C	**Uniprot Accession code**	**Number of Amino acid residue**
M. tuberculosis	5OCJ (PDB chain B)	302
M. smegmatis	A0A0D6IE52	335
M. marinum	B2HLP6	346
M. leprae	Q05862	333
M. ulcerans	A0A1B4Y981	346

The conserved regions can be depicted as identical, partially conserved, or poorly conserved depending on the similarities across the amino acid residues. The active site was taken from the previous studies and validated by the identifying the interacting residues including the SER124, ASP38, LEU40, ARG41, ILE222, PRO223, PHE226, LEU227, GLU228, and HIS260 with selective covalent inhibitor ($CyC_{8\beta}$) of Ag85C enzyme using the crystallised structure obtained from the Protein Data Bank (PDB code: 5OCJ).[12] Sequence analysis completed using the Chimera Tool visualisation software.[26] Two Loops of the crystallised structure envisaged to be crucial to ligand-binding/inhibitory mechanism. These two Ag85C

loops interchangeably call Adeniyi Loops L1 an L2. Therefore, we determined the sequence conservation of the flexible Loops LI and L2.

Furthermore, we carried out the homology modelling of the homologs of *Mycobacterium tuberculosis* Ag85C enzymes using the amino acid sequence of each organisms to determine the 2D protein structures of these Mycobacteria for probable future studies. Table 6.**2** summarises the template quality criteria obtained from the SWISS-MODEL online server used for choosing the templates; sequence identity and sequence similarity, Global Model Quality Estimate (GMQE)[27], and Quartenary Structure Quality Estimate (QSQE)[28]. The server models the target with top-ranked templates from the several similar quality templates considering the quality criteria. The top-ranked templates and alignments are compared to verify alternative conformational states or cover different regions of the target protein. The GMQE provides information about the stability of the model, and it scales from 0 to 1. The GMQE values of the study models are close to 1 (ranges between 0.63 and 0.83), which showed that the models were stable. The Quality Model Energy Analysis (QMEAN) is a composite scoring function (or Z scores) describing the major geometrical aspects of the protein models,[29] which is critical to the model quality. The sequence identity (%) of *Mtb*, *M. smegmatis*, *M. marinum*, *M. leprae*, and *M. ulcerans* were 100, 78.09, 92.47, 71.79, and 92.47, respectively.

Table 6.2. Template and Quality criteria for modelling the homologs of mycobacteria Ag85C protein obtained from the Swiss-model resource.[30, 31]

Ag85C	Template	Seq. (%)	Seq. (%)	QSQE	QMEAN	GMQE

Model	Identity	Similarity				
M. tuberculosis	5kwj (Ag85C)	100.00	1.00	1.49	0.81	0.63
M. smegmatis.	1va5 (Ag85C)	78.09	0.56	0.00	-0.00	0.76
M. marinum.	1va5 (Ag85C)	92.47	0.61	0.00	1.07	0.78
M. leprae.	4qek (Ag85C)	71.79	0.59	0.00	1.02	0.83
M. ulcerans.	1va5 (Ag85C)	92.47	0.61	0.00	1.07	0.77

Note: *M.* = *Mycobacterium*, QSQE = Quartenary Structure Quality Estimate, QMEAN = Quality Model Energy Analysis (or Z-scores), GMQE = Global Model Quality Estimate.

6.3.2 Systems preparation

The X-ray crystal structure obtained from the RSCB Protein Data Bank (PDB code: 5OCJ) [12] consists of two chains of *Mtb* Ag85C. Chain A contained partial 3D-structure of the inhibitor ($CyC_{8\beta}$) and chain B included a full 3D-structure of the inhibitor with both modelled in an open conformation and other co-crystallised molecules including DMS and water. However, since both monomers contained the active residues, we chose chain B because it saves the stress and time of extracting the inhibitor from chain A or drawing it. Another reason for choosing chain B was because it is co-crystallized with a full structure of Compound $CyC_{8\beta}$ while Chain A was co-crystallized with a partial structure of the $CyC_{8\beta}$. We prepared the selective covalent system in section 6.3.1 and set up noncovalent inhibition (open model system) to compare the activities. The preparation of the unliganded and $CyC_{8\beta}$-bound Ag85C involves the removal of co-crystallised molecules carried out using UCSF ChimeraTool-1.13.114 to obtain Ag85C monomer and the $CyC_{8\beta}$ for molecular dynamics (MD) simulations.

6.3.3 Molecular dynamics (MD) simulations

The applications of computational methods provide comprehensive insights into the structural dynamics and motions of biological molecules, using and the molecular dynamics (MD) simulations methods are extensively useful to study proteins. In other words, the methods

assist in the drug discovery and development to explore the biophysical properties of the molecules up to the atomistic level.[32] The molecular dynamics simulations of the unbound and bound systems was set up using the graphic processor unit (GPU)-Compute Unified Device Architecture (CUDA) accelerated version of AMBER18 package with integrated Particle Mesh Ewald Molecular (PMEMD) engine and LEAP modules.[33] Partial atomic charges were added to the $CyC_{8\beta}$ compound using antechamber module by applying a ff14SB/General Amber Force Field [34, 35] and restrained electrostatic potential procedures.[32] Before running the LEAP module, the execution of *pdb4amber* script added hydrogens to the histidine residues at a constant pH, which automatically modify the protein system for use with tleap. Again, LEAP module was applied to neutralise Ag85C systems and solvate all atoms explicit in an orthorhombic TIP3P box of solvent (water) molecules sized 10 Å by adding the hydrogen atoms, sodium, and chloride ions.

An *in silico* in-house protocol devised to enhance the selective covalent inhibition reported by Khan *et al*. [36] finds application to set up an irreversible bond system between Ag85C Serine124 and $CyC_{8\beta}$. Dabble interfaces Tool to generate the input files containing the topology and coordinates for the simulation of the covalent system.[37] The apo and bound Ag85C were partially minimised (with restrained with a potential of 10 Å) in two steps sequence using 500 steepest descent and full minimisation (without restraint) of 2500 and 5 steps of a conjugate gradient. Moreover, the study systems underwent a three-step gradual heating process from 0–300 K and further restrained with a potential of 10 Å. Finally, a potential harmonic restraint of 10 kcal/mol*$Å^2$ for solute atoms and a Langevin thermostat with a collision frequency of 1ps were applied to the systems before the stabilisation of the system. By keeping pressure and the atom number constant, eight steps equilibrations were executed for 2.5 ns while maintaining the pressure of 1 bar using Berendsen-Barostat and the

SHAKE algorithm for the hydrogen bond constraint before running 350 ns MD simulations.[38, 39]

The post-MD coordinates and trajectories obtained were analysed using the integrated CPPTRAJ and PTRJ modules containing in the AMBER 18 package.[40, 41] Subsequently, some post-MD analyses including root-mean-square fluctuation (RMSF), the radius of gyration (RoG), Dynamic cross-correlation matrix (DCCM), definition secondary structure protein (DSSP), others used to investigate the molecular systems in this study. DCCM is widely used as a post-MD analysis to estimate the correlation coefficient of inter-residue motions and dynamics of a protein; mostly used to compare the dynamic movements of a free protein and ligand-residue interactions profile.[32] Lastly, the structural and visual analyses obtained with the graphic interface of the UCSF Chimera Tool-1.13.1[26, 42] The data plotted with MicroCal Origin 6.0 data analysis software.[43]

6.4 Post MD analysis

6.4.1 Root mean square fluctuations (RMSF)

The root mean square fluctuation (RMSF) analysis provides the estimation of the properties of biological molecules in drug design study, the interactions of protein-protein or protein-ligand at the atomistic level. It indicates the changes in the fluctuations of a protein structure upon the binding of a ligand [36]; ligands in this study, selective covalent $CyC_{8\beta}$, and open model $CyC_{8\beta}$-bound. RMSF is the fluctuation of the individual residue of an enzyme or their average position in an obtainable MD simulation trajectory. This analysis provides insights into the flexibility differences in different regions of Ag85C enzymes in the unbound, bound/covalent systems. RMSF is mathematically represented and calculated as follows:

RMSFi is the RMSF of the ith residue from which the RMSF of the average is taking and divided by the RMSF's standard deviation [σ(RMSF)] to yield the resultant standardised RMSF [s(RMSFi)].

6.4.2 Dynamic correlation coefficient matrix and Principal component analyses

The Dynamic correlation coefficient matrix (DCCM)[44] and the Principal component analysis (PCA)[45] are another two important post-MD simulations analyses that have been used extensively in the study of protein-ligand interactions. C-α atoms DCCM, also known as dynamic cross-correlation, measures the strength and similarities (i.e., correlation coefficient) of the linear relationship in motion between the inter-residues of a protein.[46] We calculated the dynamic cross-correlation based on the Ag85C fluctuations upon ligand binding to compare with the apo system during the simulations using the AMBER18 incorporated CPPTRJ module. DCC scales from -1 (if correlation is perfect negative), 0 (no correlation) or +1 (if correlation is perfect positive).[32] Mathematically, this analysis is as follows:

Where Cy = correlation coefficient, i = ith residue, j = jth residue, while ith and jth are both the corresponding displacement vectors, respectively.[44]

The PCA is another essential post-dynamic analysis used with MD simulations with a trajectory to study the protein structures. PCA identified the most significant modes of protein fluctuation with the system's motion about the eigenvectors (direction) and eigenvalues (magnitude).[44] However, the PCA provides insights into the direction and extent of the movement of each residue in a protein to understand the principal fluctuation modes in the conformational and dynamic changes occurring during simulations. The relative changes that occurred in Ag85C protein upon the covalently binding of $CyC_{8\beta}$ and open model $CyC_{8\beta}$

conformation were studied using the PCA. The magnitude and vector of the motion of enzymes depicted by eigenvalues and eigenvectors and computed using the integrated AMBER18 CPPTRJ module.[47] The first two principal components, PC1 and PC2, were calculated from the dynamics of the enzyme C-α atoms, and the conformational changes of the apo and bound Ag85C systems viewed along ev1/PC1 plotted against ev2/PC2 with C-α atom Cartesian coordinates.

6.4.3 Thermodynamic calculations

The molecular mechanics/generalised-born surface area (MM/GBSA) method was used to the calculations of binding free energy (BFE) of the protein system.[41] It is an imperative method used to determine the in-depth binding mechanism of the receptor-ligand interacting systems. Here, we used the BFE with the MM/GBS technique to estimate the binding affinity within the docked Ag85C-$CyC_{8\beta}$ system to gain insights into their binding landscape.[48] Thus, we employed this approach to calculate the binding free energies of the inhibitor $CyC_{8\beta}$ modelled in the open conformation using 35 000 snapshots generated from the 350 ns trajectories. The MM/GBSA protocol requires that the used explicit solvent be discarded and substituted with a dielectric continuum. The binding free energy (ΔG) calculated for the open modelled $CyC_{8\beta}$ and covalently bonded systems expressed thus:

$$\Delta G_{bind} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand}$$

$$\Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S$$

$$E_{gas} = E_{int} + E_{vdW} + E_{ele}$$

$$G_{sol} = G_{GB} + G_{SA}$$

$$G_{SA} = \gamma SASA$$

where ΔG_{bind} is the gas-phase summation, E_{gas} denotes the gas-phase energy and solvation energy (G_{sol}) < the entropy ($T\Delta S$), E_{int} = internal energy, Coulomb (E_{ele}), and van der Waals energies (E_{vdW}). The E_{gas} estimated from the AMBER FF14SB force field terms. Polar and non-polar states' energy contributions accounted for the Solvation free energy, G_{sol}. The non-polar solvation energy, SA. G_{SA} was calculated using the solvent-accessible surface area (SASA) obtained by a water probe radius of 1.4 Å. In contrast, the polar solvation, G_{GB}, was obtained by solving the GB expression. The *S* denotes the total entropy of the system, which was calculated using the normal mode analysis incorporated in AMBER Tools [49] and *T* is the temperature. The interaction entropy ($-T\Delta S$) was estimated using 40 frames due to the computational cost of calculating the change in conformational energy for large frames using normal mode method. The surface tension constant, γ, was set at 0.0072 kcal/ mol*Å^2. The contribution of each residue to the total binding free energy obtained at the predicted active site by carrying out per-residue energy decomposition at the atomic level using MM/GBSA method in AMBER 18.[40, 41]

6.4.4 Per-residue energy decomposition (PRED) analysis

The PRED performed to obtain the contribution of each residue to the total binding free energy profile between the inhibitor $CyC_{8\beta}$ modelled in the open conformation Ag85C enzyme. PRED analysis achieved using the MM/GBSA method in AMBER18 GPU.

6.5 Results and Discussions

6.5.1 Amino acids sequencing of homologs of Mycobacteria Ag85C protein

Comparative amino acid sequence alignment of the *Mycobacterium tuberculosis* Ag85C enzyme and its four homologs (Figure 6.3). The sequence analysis showed that the active site of the protein is 70 % fully conserved (red), including residue SER124 (light blue region) that forms a covalent bond with selective covalent inhibitor $CyC_{8\beta}$ and 20 % partially conserved.

Moreover, we observed that a leucine and aspartic residue-rich flexible loop, L1 (green arrow region) comprising of residues LEU36, LEU37, ASP38, GLY39, LEU40, ARG41, ALA42, GLN43, ASP44, ASP45, and TYR46 in *Mtb* Ag85C enzyme. L1 residues are more than 90 % conserved among the homologs of the *Mtb* Ag85C enzyme. ASP44 of Ag85C is the only partially conserved residue because glutamine (GLU) replaced it in the *M. leprae*. This loop contains three of the active site residues; ASP38, LEU40, ARG41. Thus, we investigated the static and dynamical motions of loops L2 to explore its potential as a weak hotspot for designing antitubercular drugs.

Similarly, the second flexible loop, L2 (blue arrow region) is a glycine-rich, proximal to the active site, implying possible impactful structural conformational activities upon ligand binding. This loop contains residues including CYS209, GLY210, ASN211, GLY212, THR213, PRO214, SER215, and ASP216. Loop L2 is 50 % fully conserved and 40 % partially conserved among these mycobacteria. Notably, the glycine-rich residues conserved in all the five mycobacteria species except *M. leprae*. Also, the sequence alignment showed catalytic triad residues SER124, GLU228, HIS260 (purple downward arrow). Both SER124 and HIS260 are 100 % conserved among the mycobacteria while GUY228 is conserved in all the *M.* homologs except *M. leprae* Ag85C, which has the catalytic glycine replaced by serine. In this study, the structural similarities and differences in the sequence discussed among the *M. tuberculosis*, *M. smegmatis*, *M. marinum*, *M. leprae*, and *M. ulcerans* are potential hotspots that may be explored in the identification of a promising antimycobacterial compound.

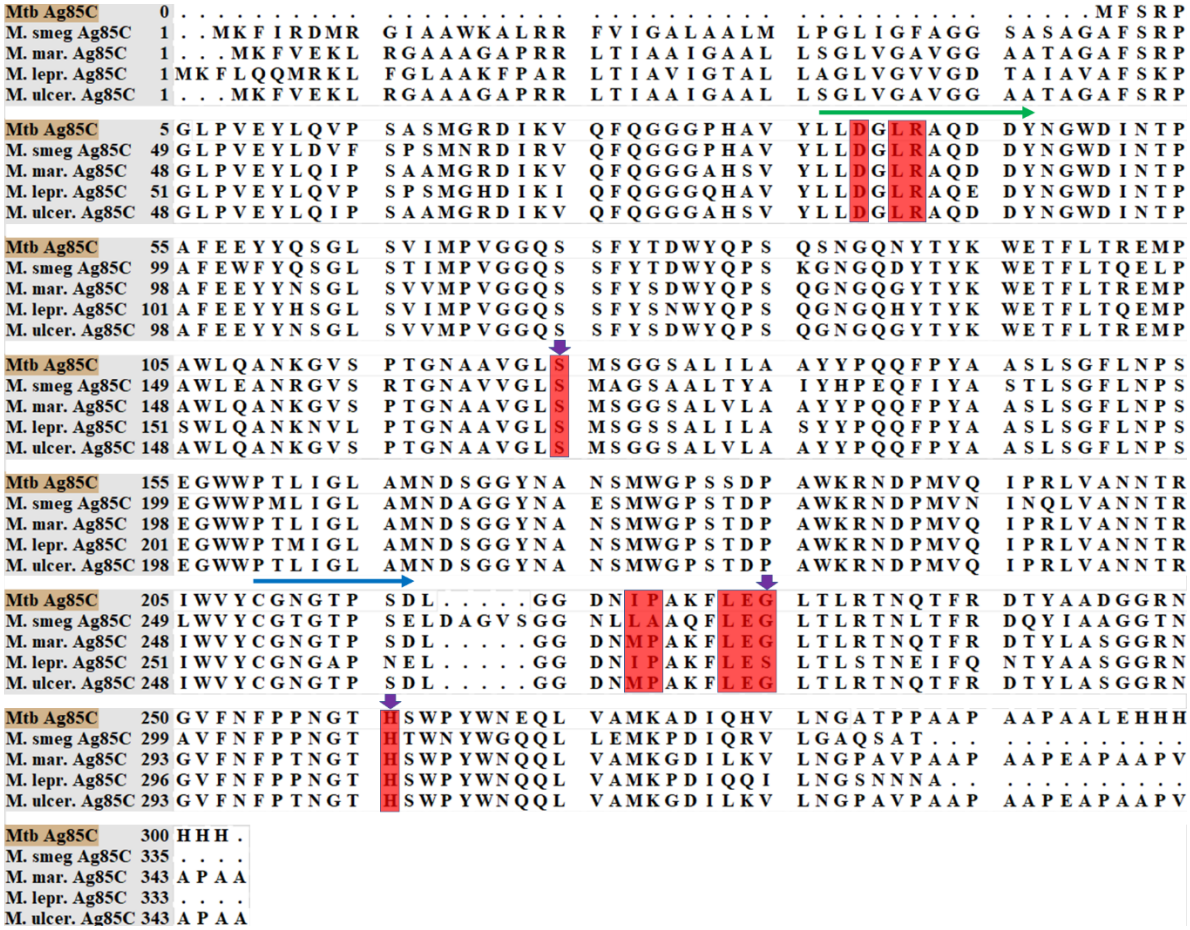


Figure 6.3. Comparative sequence analysis of five Mycobacteria Ag85C enzymes; *tuberculosis, smegmatis; marinum*, *leprae*, and *ulcerans*. The active site residues of the Ag85C proteins, including SER124, are 70 % fully conserved (red) and 20 % partially conserved. Catalytic triad residues (purple downward arrow region). Aspartic acid-rich loop (L1) is > 90 % fully conserved (green arrow region) while the glycine-rich loop L2 (blue arrow region) is 50 % fully conserved and 40 % partially conserved.

6.5.2 Pre- and post-covalent bond formation between Ag85C and inhibitor $CyC_{8\beta}$

Figures 6.4A showed the catalytic serine residue (SER124) of Ag85C protein before bonding to the selective covalent inhibitor ($CyC_{8\beta}$), and Figure 6.4B presented the putative covalent bond between the protein and the ligand. The covalent bond created using the dabble tool, which also generated the parameter and topology files we used for running molecular dynamics analysis of the protein-ligand system. The post-MD study carried out in comparison with the apo and noncovalent or open model ($CyC_{8\beta}$) modelled in an open

conformation) systems. Moreover, the figures identified the two loops concisely discussed in section 6.5.1: Adeniyi loops L1 and L2 before and after the covalent bond formation.

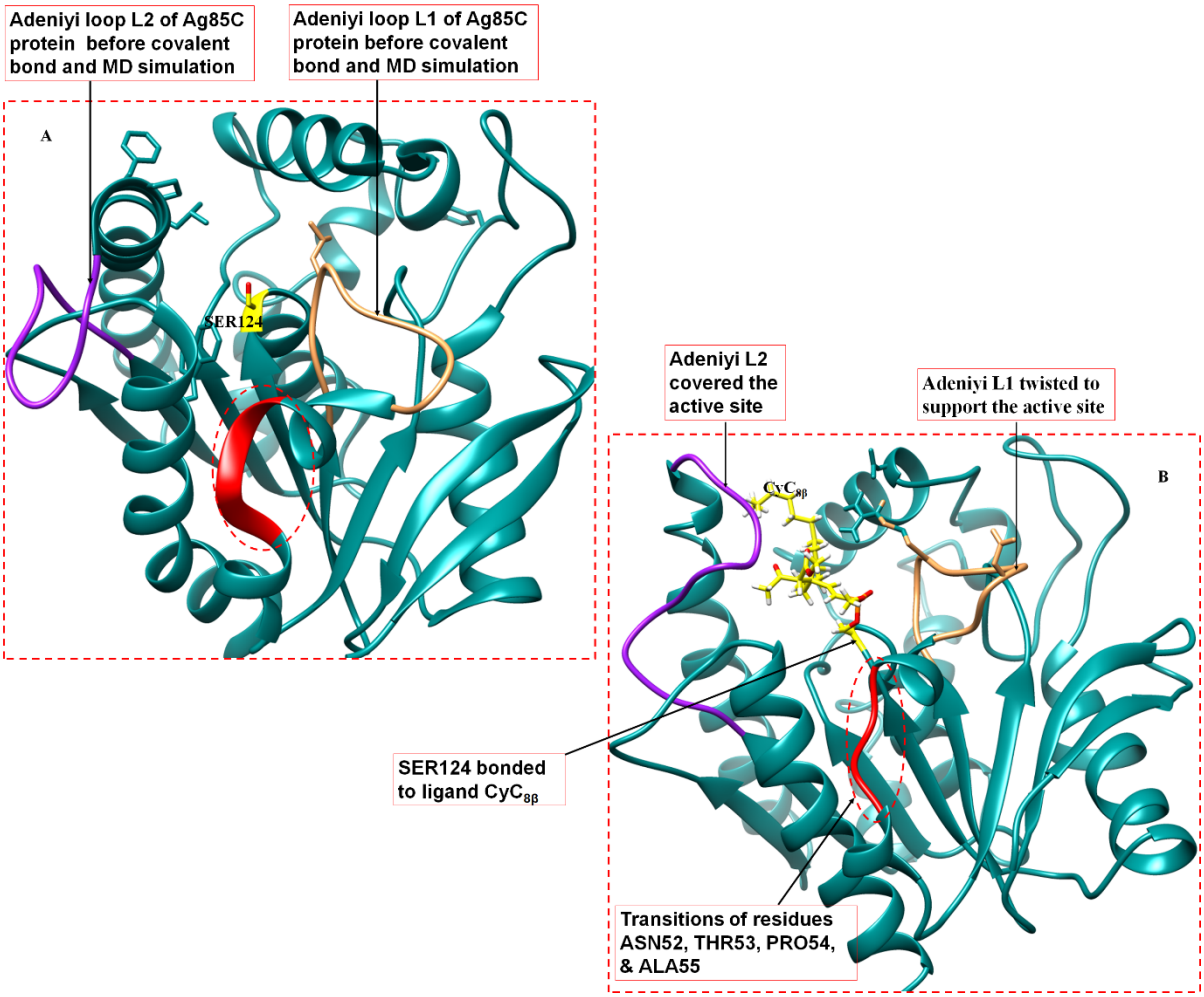


Figure 6.4. Structure of Ag85C protein showing the catalytic SER124 (yellow) before the covalent bond formation (**A**) and post-covalent bond creation (**B**). Studied loops L1 (purple) and L2 (sandy brown), also named Adeniyi loops at 50 ns MD simulations.

6.5.3 Ligand-residues profile and per-residue energy decomposition (PRED)

Here, we ignored the interactions of the other residues with the ligand ($CyC_{8\beta}$) and assumed the bonding with SER124 while there might be other interactions in the selective covalent inhibition system. Whereas, in the open model ligand-protein system, Figure 6.5**A** illustrates the interaction network between the ligand ($CyC_{8\beta}$) and Ag85C residues after 350 ns MD

simulations. Figure 6.5**B** showed the $CyC_{8\beta}$ docked in the open conformation within the active site of the Ag85C enzyme.

PRED is a reliable, rational drug design approach.[41] To obtain the interaction profile between $CyC_{8\beta}$ and Ag85C interface, we performed residues binding free energy (BFE) decomposition analysis and determined the individual binding free energy. The total BFE for the $CyC_{8\beta}$ compound was decomposed into individual residue-based contributions using the MM/GBSA approach. Figure 6.5**C** showed the van der Waals (vdW) and electrostatic (elec) interaction contributions based on the BFE of $CyC_{8\beta}$-Ag85C were estimated to obtain insights into the residues and energy constituents having more significant impacts on the total energy. The interactions (intermolecular) between the hydrophobic active site residues enhance the binding and stability of the $CyC_{8\beta}$ compound in the pocket. The per-residue decomposition energies of the $CyC_{8\beta}$-Ag85C complex range between -178 and 19 (kcal/mol). The interacting residues and the most energy contributions towards the complex are; ARG41 (-178 kcal/mol), ASN221 (-94 kcal/mol), PRO223 (-56 kcal/mol), ALA (-80 kcal/mol), PHE226 (-74 kcal/mol), HIS260 (-74 kcal/mol), and TRP262 (-80 kcal/mol). Each residue contributed high energy towards the complex, but TRP262 contributed the highest total energy. However, ARG41 and ASN221 have the highest vdW energy contribution to the binding of the inhibitor. Whereas, all the active site residues showed high electrostatic energy contributions towards $CyC_{8\beta}$-Ag85C complex. These findings can be useful in the search for potential identification of antitubercular compounds in the future study of Ag85C protein as a target.

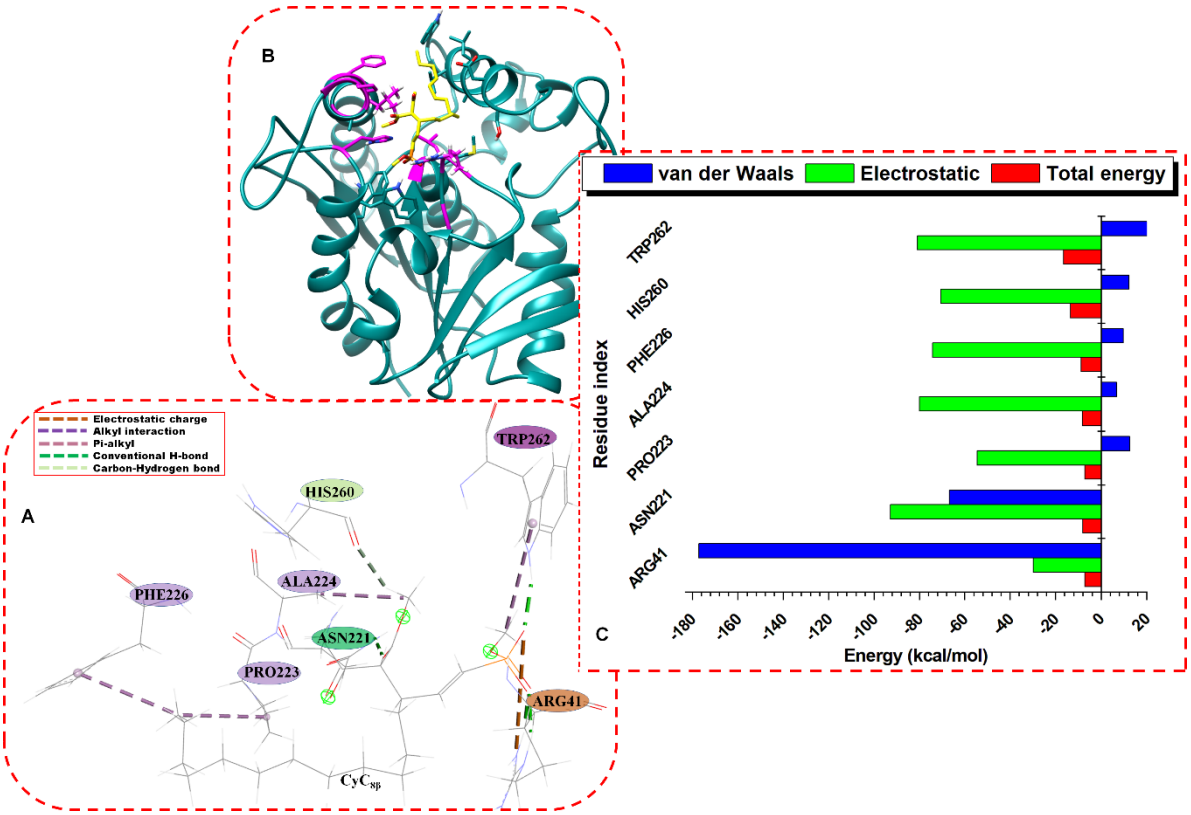


Figure 6.5. (Colour online) Visual $CyC_{8\beta}$-Ag85C interaction networks (**A**). Open model $CyC_{8\beta}$-Ag85C system (**B**) per residue energy contributions of the protein.

Binding free energy calculations of $CyC_{8\beta}$ open model-Ag85C binding affinity

The total binding free energy for the $CyC_{8\beta}$ open model-bound system obtained using the MM/GBSA approach. Table 6.2 showed the binding free energy (kcal/mol) for the open model-Ag85C system. The electrostatic force (ΔE_{elec} = -975.05 ± 42) contributed almost 9 times the energy from the van der Waals force (ΔE_{vdW} = -117.67 ± 11). Hence, the binding showed a high binding affinity of the $CyC_{8\beta}$ open model for the interacting residues at the active site. Moreover, the binding free energy at gas-phase (ΔG_{gas} = -145.81 ± 42) is useful information, which showed that the interactions in the open model-bound system favoured occurred at the gas-phase than in the water (G_{solv} = -127.84 ± 30). Additionally, the averages of bond, angle and dihedral obtained were 832.99 ± 23.97, 2124.04 ± 35.45, and 3640.66 ± 22.10 respectively.

Table 6.2. Thermodynamics analysis; summary of MM/GBSA-based binding free energy contributions to the $CyC_{8\beta}$ open model-Ag85C complex[a]

	Energy components (kcal mol^{-1})							
Complex	**ΔE_{vdW}**	**ΔE_{elec}**	**ΔG_{gas}**	ΔG_{GB}	ΔG_{SA}	$-T\Delta S$	**ΔG_{solv}**	**ΔG_{bind}**
$CyC_{8\beta}$-Ag85C	-117.67 (± 1.1)	-975.05 (± 4.4)	-145.81 (± 4.4)	-122.80 (± 2.9)	3.85 (± 0.1)	-24.28 (± 3.3)	-127.84 (± 2.9)	-131.70 (± 2.9)

[a]ΔE_{elec} (electrostatic), ΔE_{vdw} (van der Waals), ΔG_{bind} (calculated total free binding energy) ΔG_{gas} (gas-phase energy) and ΔG_{solv} (solvation free energy), $-T\Delta S$ (entropy contribution term), T =temperature (298.15 K)

The binding free energy is very difficult to calculate *in silico* because of the challenge in entropy estimation. The solution to all the possible solvent conformations and solute entropy requires computing the range of values of the underlying degree of freedoms that correspond to different conformations with the same potential energy.[50] The estimates of the interaction entropy changes during the formation $CyC_{8\beta}$-Ag85C complex were obtained from a normal mode analysis of the energy-minimized structures. The average entropy terms obtained from the analysis are translational (-13.0909 ± 0.0), rotational (-11.2108 ± 0.1), and vibrational (0.0198 ± 3.3) The conformational entropy is – TΔS is -24.28 ± 3.3 upon the noncovalent binding of compound $CyC_{8\beta}$ to Ag85C.

Figure 6.6 illustrates the trajectory images of the apo, covalently bound $CyC_{8\beta}$-Ag85C, and open model $CyC_{8\beta}$-Ag85C systems at 0, 100, 200 and 350 ns MD simulations to visualise the magnitude of structural changes occurred.

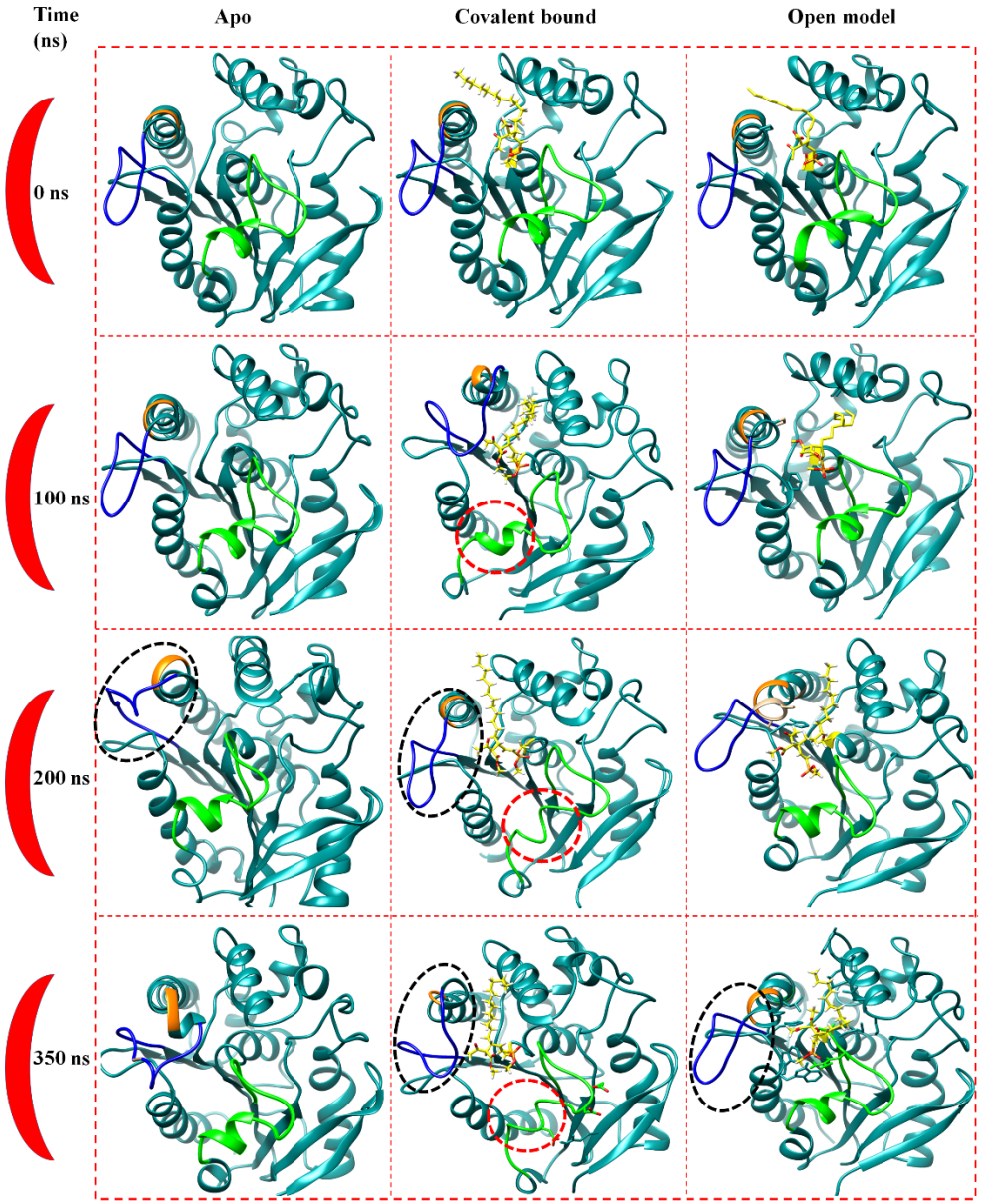


Figure 6.6. Representative snapshots comparison among the Apo, covalent bound system, and open models at 0, 100, 200 and 350 ns showing the magnitude of structural changes in the Ag85C enzyme.

The two notable changes in apo and covalent bound systems occurred, whereby loops L1 and L2 exhibited twisting and transitions, respectively. Loop L2 (enclosed in black oval shapes) twisted and shifted towards the active site of Ag85C at 200 and open again at 350 ns MD simulations. Similarly, the observable changes in the covalent bound system occurred at 100, 200, and 350 ns. Moreover, loop L1 (enclosed in red circles) extended because of the structural transition of a helix (residues GLY44, TRP45, ASP46, and ILE47). The active site

remains closed upon inhibitor binding. Therefore, the active site switches between open and closed conformations in the apo system compared to covalent bound and open model systems, which exhibited closed and opened binding site structures, respectively.

6.5.4 Stability and fluctuation patterns of Ag85C systems

The average alpha carbon (C-α) atom root-mean-square deviation (RMSD) of a protein can be calculated from the MD simulations trajectories to deduce the stability status of such protein. For a protein system to attain stability, it must be subjected to a long-time MD simulation. Hence, the structural examination and dynamics analysis were determined based on the time scale of the simulations. The C-α atoms RMSD of the apo, covalently bound, and open model systems were observed for 350 ns and graphically monitored for convergence and stability. Figure 6.7A showed that the $CyC_{8\beta}$-bound covalent system and $CyC_{8\beta}$-bound open model did not converge with the unbound Ag85C. The RMSD of the Ag85C C-α atoms displayed high instability upon the binding of the ligand covalently compared to the apo and open model systems.

Moreover, the covalent bound system became stable at 150 ns compared to the noncovalent system, which maintained a stable structure throughout 350 ns. On the average, the apo, covalent $CyC_{8\beta}$-bound, and open model $CyC_{8\beta}$-Ag85C systems had C-α atoms RMSD values of 1.59 Å, 2.05 Å, and 0.91 Å respectively. The mean RMSD difference for the $CyC_{8\beta}$ covalently bound system is significant compared to the apo and the open model bound-$CyC_{8\beta}$ systems. These findings agreed with the visual structural changes obtained from the trajectory snapshots (Figure 6.6A).

The induced structural instability induced in Ag85C upon the covalently bound inhibitor highly favours the inhibition of protein catalytic function and may indicate conformational inactivity of the Ag85C. The variability may also mean the mechanism through which the $CyC_{8\beta}$ covalently bound inhibits the Ag85C function. Moreover, the notable deviations may

be associated with the hydrophilic characters of SER124 residue and arginine within the active site upon ligand binding. Polar amino acids often exposed to the solvent, which results in changes because of charge difference. This behaviour affects the hydrophobic properties of the protein, and consequently, the function of the protein.[51]

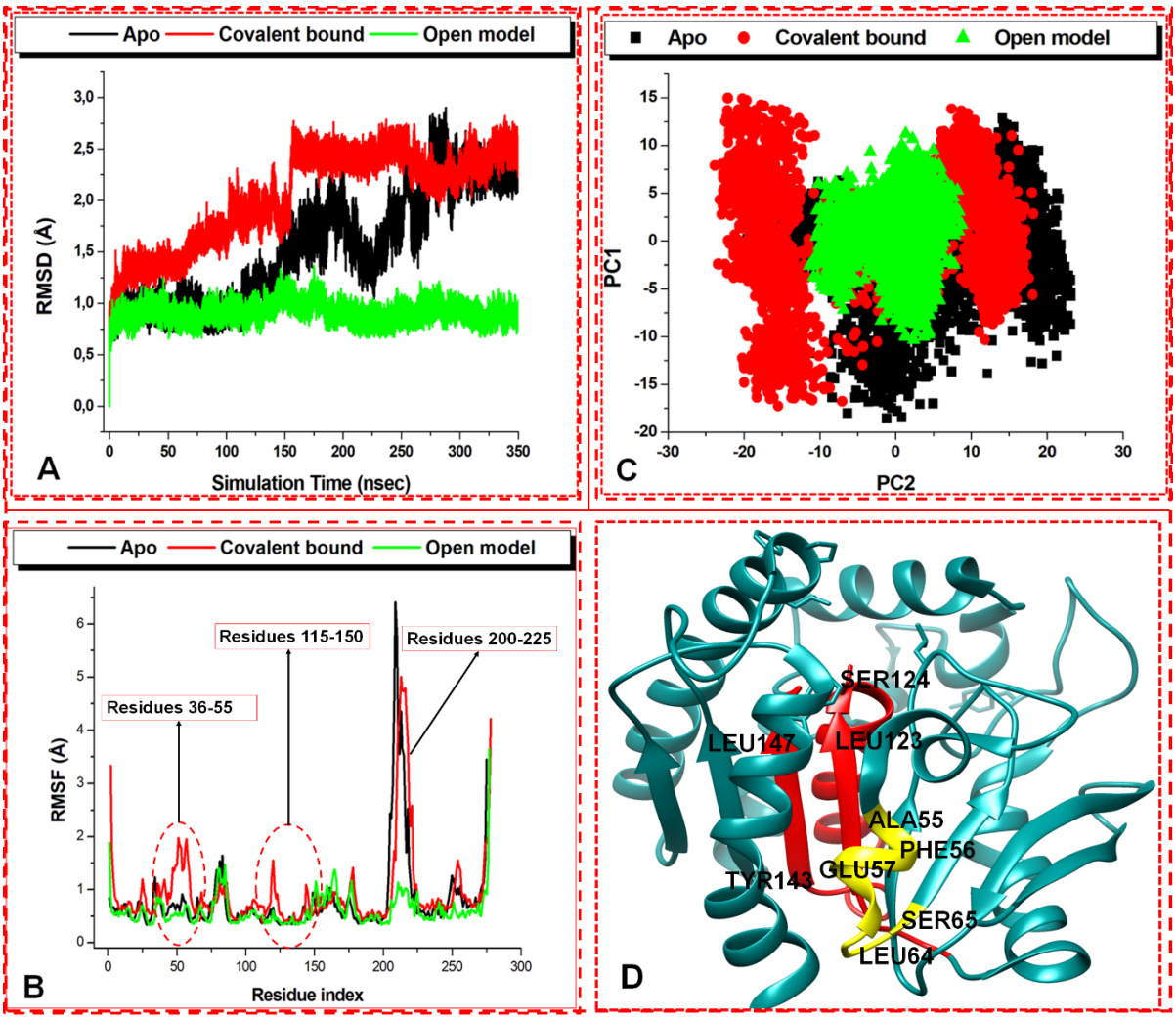


Figure 6.7. Comparative RMSD (**A**), RMSF (**B**), and PCA (**C**), for the C-α backbone atoms of Ag85C whole system of the unliganded (black), covalently bound system (red), and open model-bound (green) throughout the 350 ns MD simulations. PCA plot showed projections of conformational behaviours of the protein systems along the PC1 and PC2 are the first two principal eigenvalues. **D** show residues 115-150 (red) and residues 55-65 (yellow).

To accurately estimate the degree of structural instability induced by the putative covalent inhibitor in both bound modes, we carried out investigations into the dynamic occurrences at the active sites and on the ligand in the covalently bound and open model systems throughout

350 ns. Figures 6.8A and 6.8B showed the RMSD plots for the active sites of the three protein systems. However, the results did not show any notable difference in the motions and dynamics, which imply the similarities in the structural behaviour compared to the whole protein system.

Furthermore, RMSF values of the apo and bound systems were estimated to determine the impact of the binding of compound $CyC_{8\beta}$ on the dynamic behaviour of the protein. The role of amino acids is critical to conformational features of proteins. However, there may be a conformational change in protein due to mechanical events or chemical reactions. Hence, direct interactions of the active site of Ag85C with the inhibitor, $CyC_{8\beta}$, should induce changes in the conformation of the protein by ligand-induced motion. Higher RMSF values indicate more flexible movements, considering the average position of the protein residues and lower fluctuation values signifies the restriction of conformational changes during the simulation. Figure 6.7B presented the C-α atoms RMSF plots for Apo protein, covalently $CyC_{8\beta}$-bound, and $CyC_{8\beta}$ open model-bound. The C-α atom RMSF of apo exhibited a higher degree of flexibility compared to the open model system, which might be responsible for the Ag85C catalytic role of mycolic acid attachment and the synthesis of trehalose moieties. Moreover, the lowest RMSF value for the open model $CyC_{8\beta}$-bound system compared to the apo system may be an indication of the system's compactness. The fluctuation of the Ag85C reduced upon the binding of the ligand in open model pose compared to the apo system.

On the average, the C-α atoms RMSF values of the apo, covalent system, and open model system were 0.8 Å, 0.9 Å, and 0.6 Å respectively. The covalently bound system experienced a high degree of flexibility due to various conformational changes that occurred while the active site was closing. The covalent system showed elevated C-α atoms fluctuations throughout the 350 ns simulations and the notable variations in the regions comprising of residues 55-65, 110-140, and 210-225. Moreover, a high degree of conformational changes

occurs near the loop L1 (residues 38-55) and residues 115-150 in the covalent bound system. However, the most top fluctuation occurs in the apo and the covalent bound systems between residues 200 and 225, which contains two of the active site residues (223 and 224) and very close to loop L2 region (residues 209-220).

Residues 55-65 and 115-150 are non-active site residues except the catalytic SER124, yet the residues showed similar *B`-factor* values compared to the RMSF values of the active site residues including SER124. The findings imply high flexible characteristic for the residues (55-65 and 115-124) upon ligand binding. Inferentially, these non-active site residues have high *B`-factor* tendency, which is an indication that the vibrational and the fast motions of the Ag85C protein C-α atoms. Subsequently, this implies clear biological significance like the active site residues.

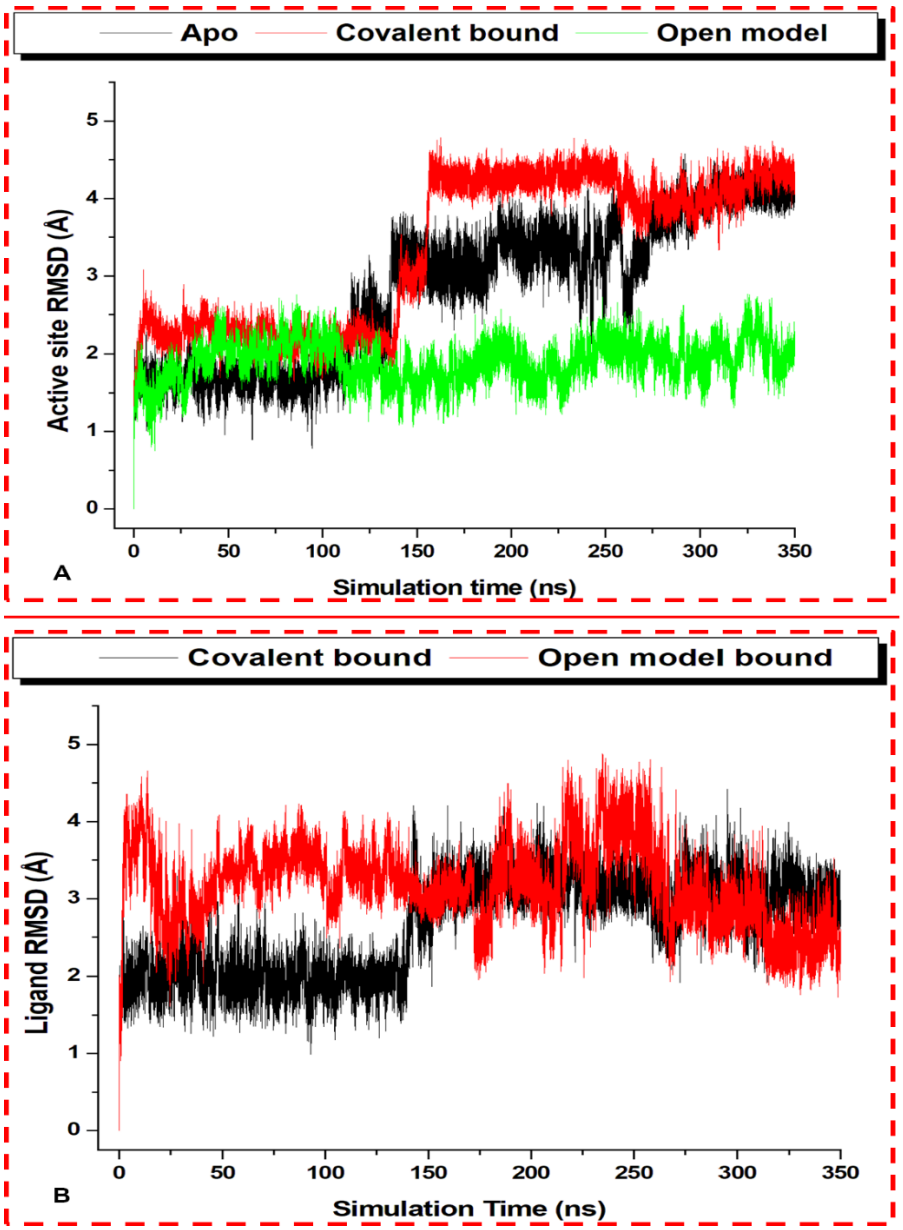


Figure 6.8. C-α atom RMSD of the Ag85C protein active site (**A**) for the unbound (black), covalent bound (red), and noncovalent bound ligand system (green). RMSD (**B**) for the covalently docked ligand (black) and noncovalently or open model bound (red) systems.

Perhaps, the elevated flexibility reflected the mechanism by which the covalent ligand selectively inhibits the functions of the protein whereby the system interactions occurred due to conformational adjustments.[52] Hence, the motion behaviours would hinder the enzyme from functioning at the optimum level, leading to the interruption of the chain of proteins that

involve in the synthesis of TDM/TMM and covalent attachment of mycolic acids in the mycobacterial cell walls.

Moreover, the RMSF values of the covalent bound system indicate a strong interacting propensity between the selective covalent inhibitor and SER124 residues. The binding residues of Ag85C-$CyC_{8\beta}$ complex include ASP38, LEU40, ARG41, SER124, ILE222, PRO223, PHE226, LEU227, GLU228, and HIS260. The catalytic SER124 is in-between the α_3-helix and the β_5-strand in the active site. The approach of irreversible covalent inhibition of biological targets using covalent compounds to cure pathogenic diseases has remained promising in the drug design area[53], and the dynamic perspectives in this study corroborated with the experimental protocol reported on the putative covalent inhibition of the Ag85C role through the interaction of SER124 residue with the phosphorus oxygen atom of the Cis (β-isomer)-monocyclic enolphosphorus Cycliphostin ($CyC_{8\beta}$) compounds.

6.5.5 Principal component analysis

To further explore the dynamics and motions of Ag85C, the Principal correlation analysis (PCA) was carried out on the C-α atoms of Ag85C using the CPPTRJ module in the AMBER18 GPU. The conformational patterns of the free and bound systems were projected along with the first two principal components (PC1 and PC2) of the individual vector (ev1 vs ev2). Figure 6.7C shows the general motions shifts for the unliganded and liganded systems along PC1 and PC2. The plot showed distinct conformations and clear separation of motions along the essential subspaces of two principal components. The apo and $CyC_{8\beta}$ covalently bound systems exhibited highly dispersed movements along both PC1 and PC2, and the $CyC_{8\beta}$ open model system displayed a compact structure.

Though the C-α atoms of open model $CyC_{8\beta}$ system did not move like the selective covalent inhibition system, it did not move like the free protein either. Therefore, it could imply that the binding of the open model $CyC_{8\beta}$ is different from the covalent system. The induced

dispersions in the bound protein systems may simply indicate structure-activity in comparison with free Ag85C. The apo structure showed dispersed motion shifts; the movement shifts of $CyC_{8\beta}$ covalently bound is more dispersed while $CyC_{8\beta}$ open modelled showed highly compact movements. These findings still agreed with the above-detailed results, which, taken together, imply that the binding of the $CyC_{8\beta}$ compound actively favourable for the inhibition of TDM/TMM and the prevention of mycolic acid attachment.

6.5.6 Dynamic cross-correlation matrix (DCCM)

The dynamic cross-correlation matrix was determined during 350 ns MD simulations to obtain the position of the C-α atoms of the unbound and bound systems.[54] This method gave us the insights into the inhibitory impact of $CyC_{8\beta}$ in the Ag85C systems by estimating the inter-residual motions and dynamics of Ag85C protein in the $CyC_{8\beta}$ covalently bonded and $CyC_{8\beta}$ open model systems compared with the apo system. Figure 6.9A-C showed the C-α atom motions indicating highly positive correlated motions (red to yellow), highly negative/anti-correlated movements (cyan to black), and similar correlations (green).

The Apo Ag85C displayed the highest degree of positively correlated movements, followed by the $CyC_{8\beta}$ covalently bound system (Figure 6.9A**)**. The decreasing magnitude of positive correlation follows the Apo > covalent bound > open model. However, Figure 6.9C showed significantly reduced a positive relationship in the $CyC_{8\beta}$ open model, indicting fluctuation reduction in the Ag85C C-α atoms, and low average RMSF value. The SER124 selective covalent inhibition system induced a tremendous increase in the fluctuations of the C-α atom Ag85C; these results agreed with the high average RMSF findings. The Dashed oval shape (Figure 6.9A) indicates SER124 residue to which $CyC_{8\beta}$ is selectively bound showed decreased positively correlated motions (red) to anti-correlated movements. Inferentially, these properties differentiate the inhibitory action/mechanism of the ligand when it bounds covalently to the protein from open model bound. In other words, $CyC_{8\beta}$ covalently bonded to

Ag85C SER124 residue inhibited its catalytic functions differently when the $CyC_{8\beta}$ bound to openly modelled to the protein. Compared to the free Ag85C, the higher fluctuations in the covalent system or the lower in the $CyC_{8\beta}$ open modelled system may imply the enzyme inactivity to function efficiently in the TDM/TMM synthesis and the mycolic acids attachment to the cell wall.

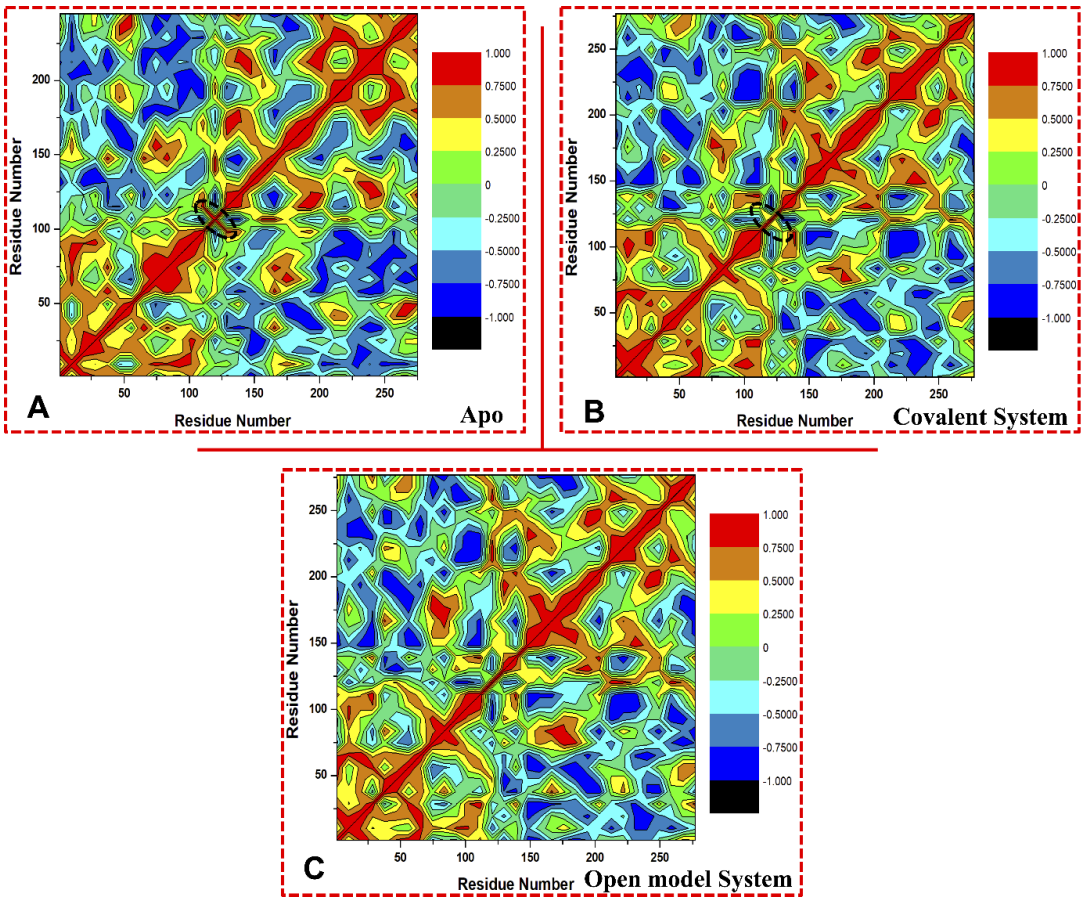


Figure 6.9. Dynamic cross-correlation matrix presenting the correlations of residues in the Apo (**A**), Covalently bound $CyC_{8\beta}$ system (**B**), and Open model bound $CyC_{8\beta}$ (**C**). The colour legends on the right of each plot indicate the status of correlated motions.

6.5.7 Comparative Ag85C protein loop dynamics and distance matrices

Loops are highly mobile structures that may occur in stable and unstable protein structures, and their variability due to solvent exposure often influences the folding, twisting, stability, and function of proteins.[55] They may serve as mediators in biological processes, including protein-ligand, protein-protein, and signalling.[50] The characterisation of protein using wet

techniques such as X-ray crystal or NMR is often challenging due to their amphipathic nature and the ligand-binding requirements.[56] During the protein crystallisation, secondary structures such as helix and strand may unfold/transition directly to the loop due to the hydrogen bonding reduction and other reasons.[57] The study of loop dynamics and motions uncovers important useful information in understanding protein conformational activity. Targeting weak spots such as active loops in the Ag85C enzyme is one of the opportunities by which we can develop novel efficacious antitubercular agents through drug design research.

The X-ray crystal structure of Ag85C resolved with some loops, and we investigated the static properties, dynamics, and motions of the two essential loops: loop 1, L1 (residues 36-46); located within the active site and loop 2, L2 (residues 209-216) because it is proximal to the active site. Figures 6.10A-C showed both loops in apo, selective covalent inhibitor bound, and the open model systems. There was no notable change in the two loops of the apo structure compared to the bound systems which showed conformational transitions that can be attributed to the dynamic interaction with $CyC_{8\beta}$.

Ag85C Loop (L1) is hydrophobic but also characterised by aspartic-rich and aromatic side-chain position of TYR46 that can allow high variability in the protein's function and inhibitor binding. The covalent system of Ag85C-$CyC_{8\beta}$ showed that this loop interacted with the inhibitor to a greater extent when compared to the open model system. Therefore, the loop structure is significant for the binding of the ligand to Ag85C. Again, this is evident in the twisting that we observed as the ligand interacted during the simulations. Interestingly, we discovered that the conformational unfolded α-helix residues ASN47, GLY48, TRP49, ASP50, ILE51, ASN52, THR53, PRO54, and ALA55 to loop and extend the length of loop L1 in the covalent bound system. Whereas $CyC_{8\beta}$ open model system only lost α-helix residues THR53 and PRO54 to transition into a loop structure.

Similarly, Ag85C Loop (L2), is glycine-rich and is also characterised by hydrophobic residues. Figure 6.10B showed loss of the L2 loop residues ASP209, LEU210, and GLY214 (proximal to the active site) to 3_{10}-helix and a turn structure in the selective covalent $CyC_{8\beta}$ system. These changes did not occur in both apo and $CyC_{8\beta}$ open model systems. This conformational change was simply due to the interaction of the selective covalent inhibitor. Objectively, loop L2 and the transitioned residues were involved in the binding of the ligand. Besides, loop L2 remained closed throughout the 350 ns simulations in the selective covalent inhibition system. At the same time, it exhibited opening and closing in the unbound Ag85C but actively opened in the $CyC_{8\beta}$ open model system.

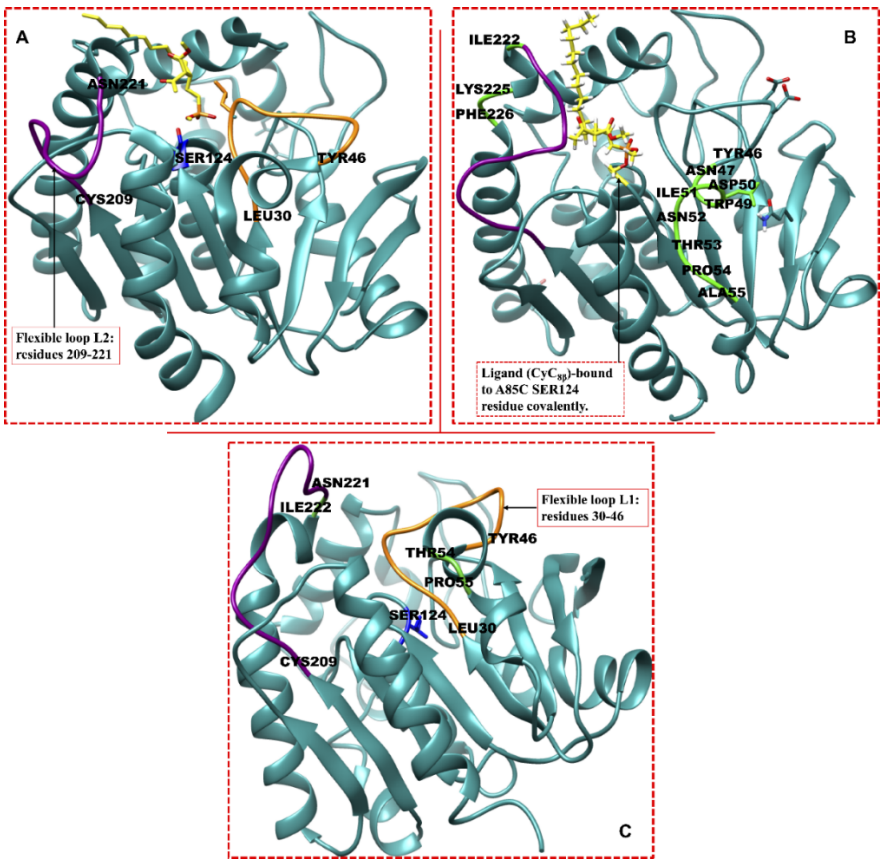


Figure 6.10. Visual conformational transitions in the secondary structures of the post 350 ns MD simulations; potential loops in Ag85C protein in the Apo (**A**), selective covalent $CyC_{8\beta}$ (**B**), and open model $CyC_{8\beta}$ (**C**) bound systems. Loop (L2) (**purple**) and Loop (L1) (**sandy brown**)

We, therefore, measured the dynamical events involved in these conformation transitions of the Ag85C enzyme as induced by the various $CyC_{8\beta}$-binding modes to investigate further the

role of the loops in the function of the protein and $CyC_{8\beta}$ binding. To understand the dynamics and motions regarding the inhibitor's binding, we calculated the C-α RMSD and C-α RMSF for loop L1 and L2 to determine their stabilities and deviations over a simulation period of 350 ns MD. Respectively, the C-α RMSD and RMSF plots for apo, selective covalent inhibition, and $CyC_{8\beta}$ open model systems of Loop L1 and L2 are presented in Figures 6.11 and 6.12.

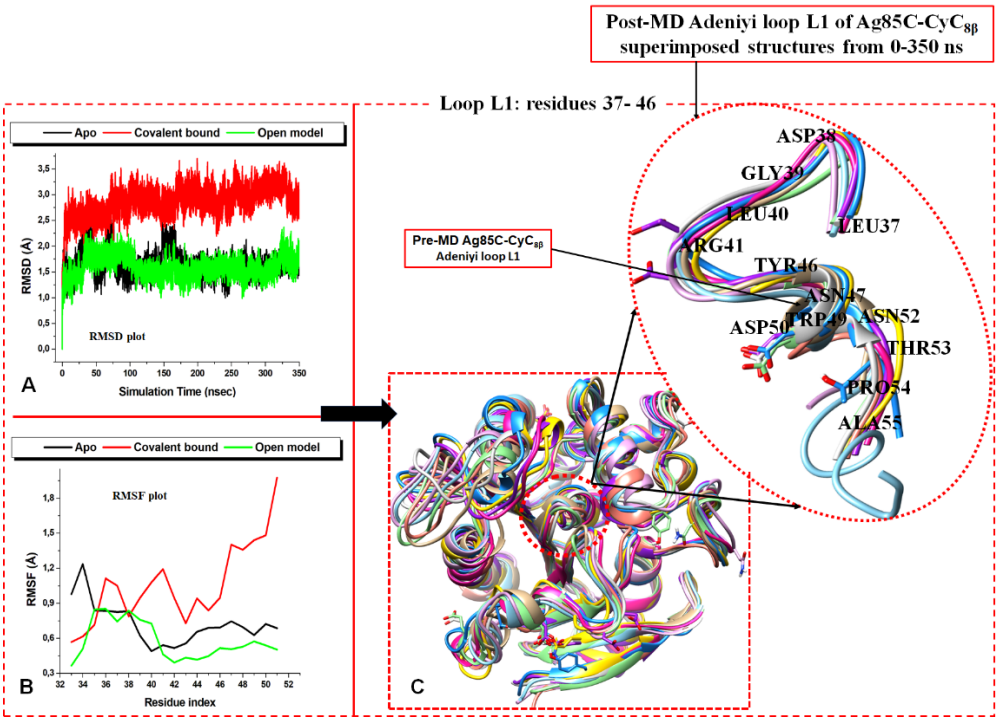


Figure 6.11. C-α RMSD (**A**) and RMSF (**B**) plots of Loop L1 of unbound and bound systems; Apo (black), Covalent bond system (red), and open model system (green). **C** illustrated the twisting and folding of loop L1 as it interacted with the protein active site as simulation time increases.

The RMSD value of the apo L1 converged with the open model system but not with the covalent system (Figure 6.11A). Instead, RMSD of apo L2 converged with the covalent inhibition system but not an open model system. The average C-α RMSD values of L1 in apo, selective covalent inhibition and open model $CyC_{8\beta}$ systems were 3.00 Å, 2.79 Å, and 1.51 Å and the average C-α RMSD values of L2 were 1.64 Å, 2.86 Å, and 1.56 Å respectively for the same systems. In the $CyC_{8\beta}$ open model system, L1 and L2 are

consistently stable as compared to the unbound system, the low average RMSD values. Loop L1 was unstable in the covalent bound system, while loop L2 was stable. The average C-α RMSF values of L1 for apo, covalent and $CyC_{8\beta}$ open model systems were 3.26 Å, 3.16 Å, and 0.82 Å respectively and the average C-α RMSF values of L2 were respectively 0.73 Å, 1.06 Å, and 0.58 Å for the same systems. Such findings matched and may indicate inhibition-induced instability in the Ag85C structure due to the binding of ligand $CyC_{8\beta}$.

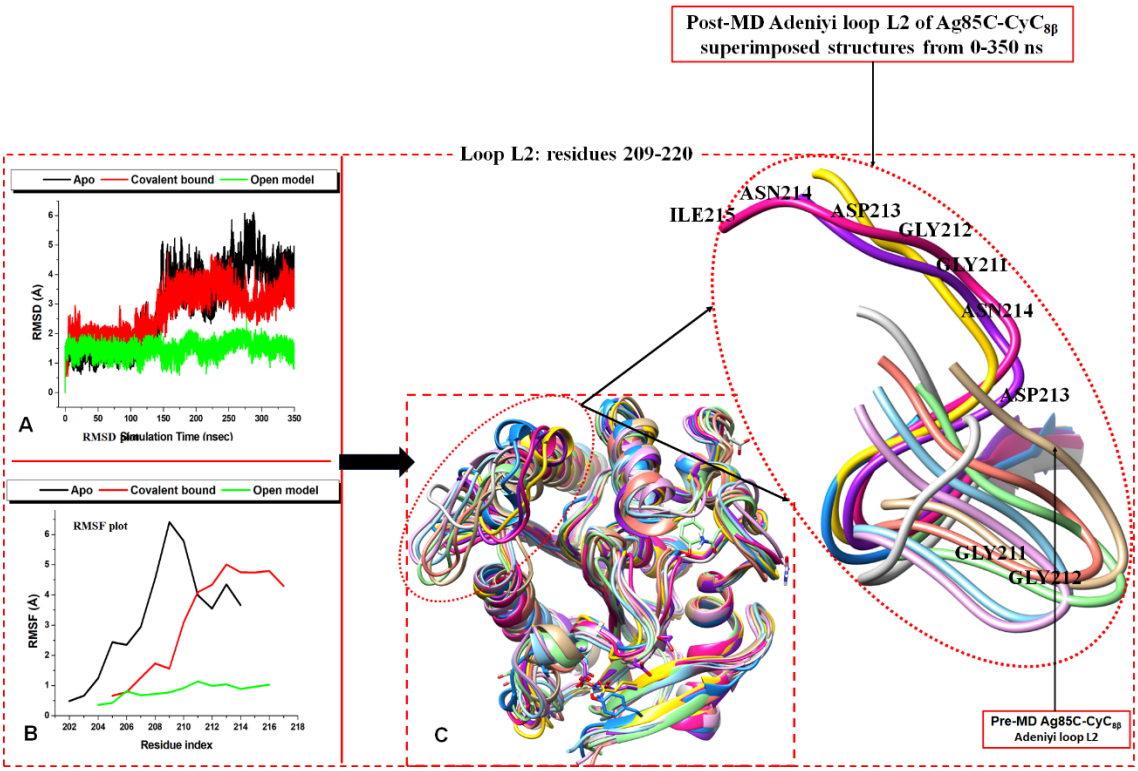


Figure 6.12. C-α RMSD (**A**) and RMSF (**B**) plots of Loop L2 of unbound and bound systems; Apo (black), Covalent bond system (red), and open model system (green). **C** illustrated the movement of loop L2 to cover the protein active site as simulation time increases.

We used the distance (D) metric to investigate the inhibitory activity of this protein on the two different poses of binding of the ligand.[39] The measurement of distance (D) is the most frequently used metric for describing the curling and twisting motions in proteins.[58, 59] Finally, the distance between two different residues and a loop residue covering the active site, causing the opening and closing of the active site residue, was considered in this study. Distances (D) between the opposite residues LEU164 and PRO223 located in the α_4- and α_8-

helix, respectively, estimated in the selective covalent and open model $CyC_{8\beta}$. The snapshots taken at times (ns) 0-49, 50-99, 100-149, 150-249, and 250-350 used to determine the distances between the C-α residues LEU164 and PRO223 of bound and unbound systems. D (Å) between LEU164 and PRO223 in apo measured were 12.49, 11.92, 12.19, 12.79, and 10.78, an indication that the active site underwent open and closed conformation and displayed similarities with $CyC_{8\beta}$ open model system. Figure 6.13 showed that distance (D) between the C-α residues LEU164 and PRO223 decreased with increasing simulation time until the selective covalent $CyC_{8\beta}$ system active site was closed. Hence, the closed conformation of the active of Ag85C upon the binding of the selective covalent inhibitor could only enhance the protein's conformational activity and the inhibitory action, and these results also support the findings as mentioned earlier.

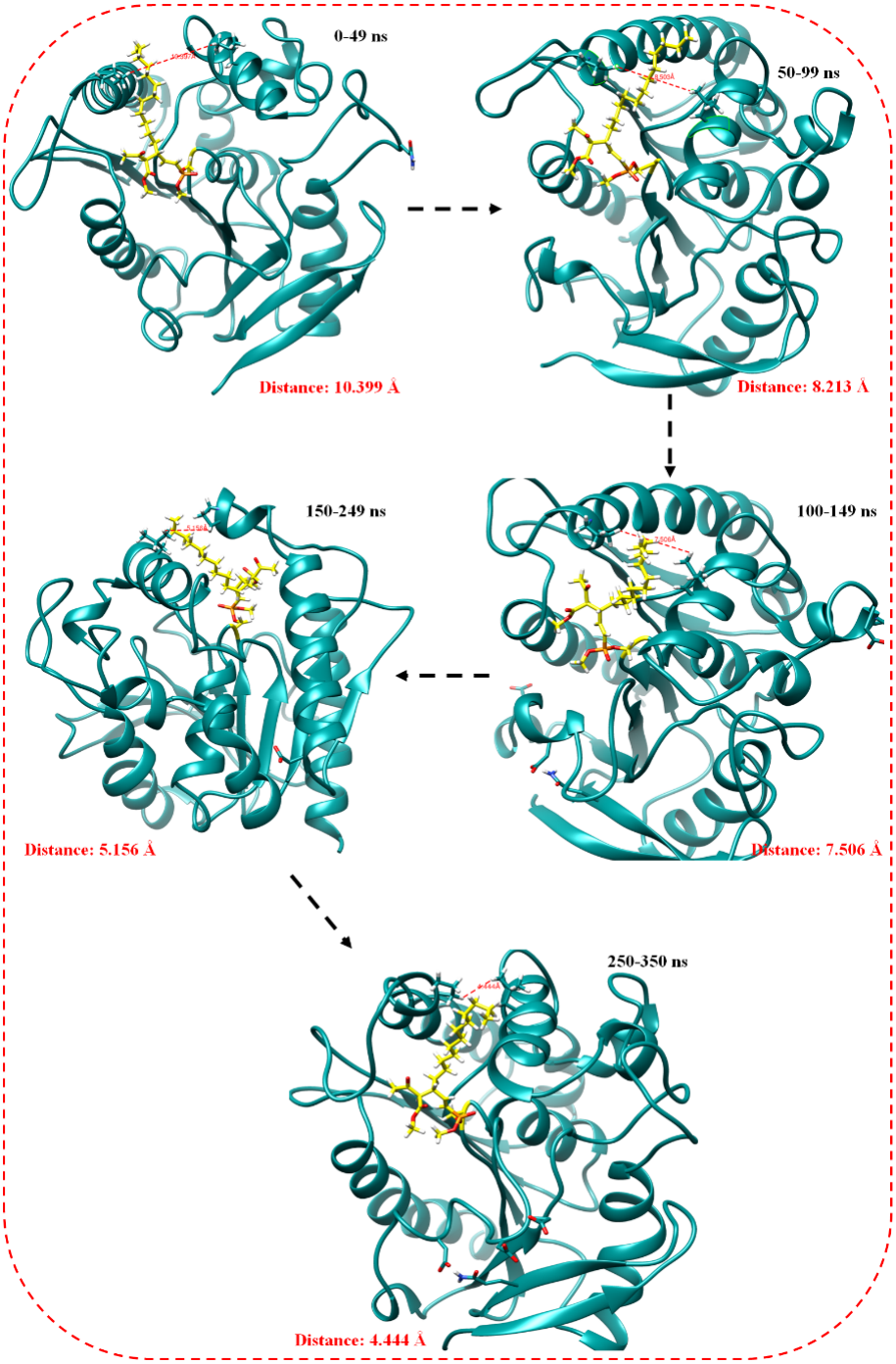


Figure 6.13. Estimated distance (D) between LEU164 and PRO223 of the CyC8β covalently bound Ag85C using the snapshots taken at 0-49, 50-99, 100-149, 150-249, and 250-350 ns MD simulations. The D values were 10.399 Å, 8.213 Å, 7.506 Å, 5.156 Å, and 4.444 Å.

6.5.8 Defined Secondary structure (DSS) of *Mtb* antigen 85C protein

Defined Secondary structure protein (DSSP) analysis used to characterise/assign any protein elements, including loops, α-helices, coil, β-sheets, among others that do not unfold directly or adequately.[51] The most common DSSP element assignors are None, Para, Anti, 310, Alpha,

Turn, and Bend indicated as legends on the far right-hand side of each chart scaled from 0-7 except the None. Having observed the static Ag85C protein, especially the loops proximal to the active site, we monitored the visual transition or unfolding behaviour of these upon ligand binding in comparison with free Ag85C structure. In other words, a defined secondary structure protein analysis carried out to further gain insights into the inhibitory activity of $CyC_{8\beta}$ bound covalently and $CyC_{8\beta}$ open model against the Mtb enzyme catalytic functions using the 350 ns trajectories obtained during MD simulation. Figure. 6.14(A-C) represents the DSSP plots for the unbound (A), covalent system (B), and open model system (C).

The Apo and $CyC_{8\beta}$ open model DSSP charts showed insignificant changes in their secondary structure elements except around the residues 41-61, the region where THR53 and PRO54 in the $CyC_{8\beta}$ open model-Ag85C system transitioned to the 3_{10}-helix from 70-80 ns. Whereas there were notable changes in the structural elements of Ag85C L1 upon the binding of the selective covalent inhibitor as revealed by the visual molecular dynamic tool. Similarly, in the same 41-61 region of covalently bound $CyC_{8\beta}$-Ag85C system, Figure 6.13B showed that residues ASN47, GLY48, TRP49, ASP50, ILE51, ASN52, THR53, PRO54, and ALA55 were transitioned from alpha to mixture Turn and Bend throughout the 350 ns simulations. These losses of α-helices could be due to the reduction in the hydrogen bonds compared to the unbound protein.

Figure 6.14C chart showed that no changes observed in the L2 of $CyC_{8\beta}$ open model-Ag85C and unliganded Ag85C systems. Whereas there were significant losses of loop residues to 3_{10}-helix among the Ag85C residues 219-226 upon the covalent $CyC_{8\beta}$-bound from 160-220 ns and 330-350 ns. Interestingly, this region also contained the residues ILE222, LYS225, and PHE226 observed to undergo visual losses of alpha to loop.

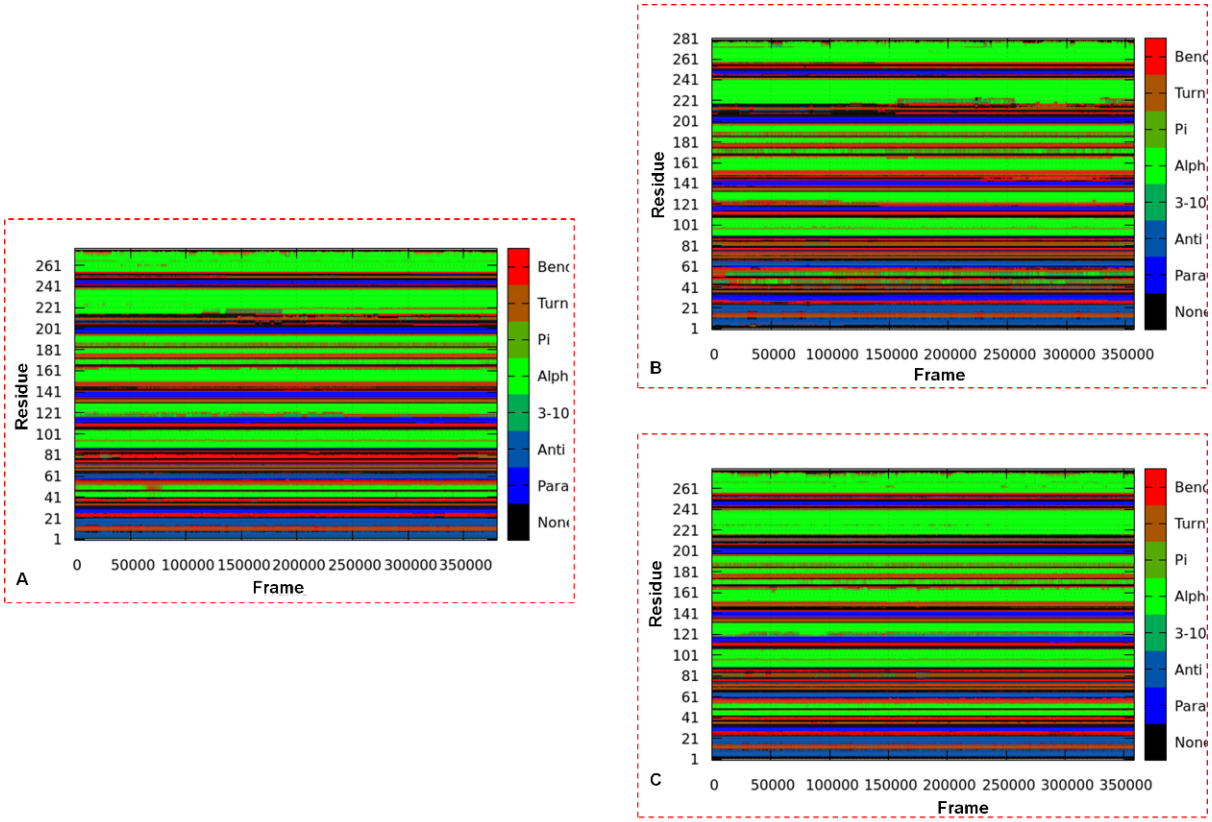


Figure 6.14. The DSSP charts for Ag85C protein systems taken throughout 350 ns MD simulations; unliganded Ag85C (**A**), covalently bound $CyC_{8\beta}$-Ag85C (**B**), and open model $CyC_{8\beta}$ (**C**)

6.5.9 Solvent accessible surface area (SASA)

Solvent accessible surface area (SASA) measures the area of the protein exposed to the solvent molecules.[60] SASA is estimated from the coordinates of thermodynamic interactions between protein and water. It provides insights about the protein residues that exposed to water and those that buried in the active site to contribute to the hydrophobic stabilisation of protein structure. In this study, we determined changes in the accessibility of protein to solvent by computing SASA using the "surf" command in the AMBERTOOLS 18 GPU CPPTRAJ module that implemented a linear combination of pairwise overlaps (LCPO) algorithm. [61] Figure 6.15 calculable SASA of the unbound Ag85C and bound systems obtained from set trajectory coordinates of between Ag85C protein systems and water. The SASA calculations showed that the residual $_{exposure}$ of apo and open model systems was

consistent throughout the 350 ns simulation. At the same time, there were fluctuations in the residual exposure of the Covalent bound $CyC_{8\beta}$-Ag85C.

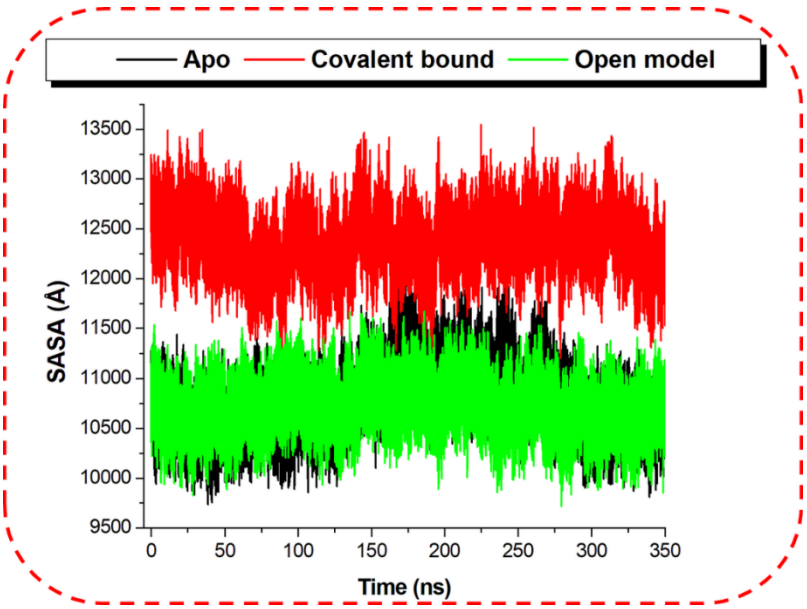


Figure 6.15. Schematic representation of the solvent-accessible surface area of Apo (black), $CyC_{8\beta}$ covalent bound (red), and Open bound (green).

The SASA of the covalent system was high compared to the apo and open model systems. The high SASA values could simply be due to the exposure of the active site hydrophobic residues to enclosing solvent[62], which explained why the instability (RMSD average) was also high.[63] On the other hand, the increase in SASA value in the covalently-bound $CyC_{8\beta}$ indicated a restructured active site. The exposure of the active site hydrophobic residues to enclosing solvent may be associated with the selective covalent inhibition mechanism of this compound and eventual loss of protein functions. Finally, this SASA increase value may perhaps be due to the compressibility of the protein resulting from high induced-pressure.[64]

6.6 Conclusion

This scientific idea of studying the molecular mechanism of Ag85C and the putative covalent inhibitory mode of action of cis (β-isomer)-monocyclic enolphosphorus cycliphostin ($CyC_{8\beta}$) using combined bioinformatics tools and computational methods provides a novel approach to design new antitubercular drugs. These findings obtained at the molecular level furnished us with the structural and dynamic behaviour of Ag85C protein and its interactions with the inhibitor that would assist in the experimental study. The detailed bioinformatics and molecular dynamics analyse, and the findings presented in this work are innovative. Sequence alignment of the homologs of Ag85C including *Mtb*, *M. smegmatis*, *M. marinum*, *M. leprae*, and *M. ulcerans* showed that the unique structural features among the amino acids of these mycobacteria are conserved. The individual active site features exhibited different levels of conservation; catalytic triad residues SER124, GLU228, HIS260 is 70 % conserved, Adeniyi loop (L1) is 90 % conserved, and Adeniyi loop L2 was 50 % entirely and 30 % partially conserved. On this note, these structural similarities among the Ag85C are weak spots to employ in targeting the protein by a selective covalent inhibition compound that can cure mycobacterial infections. The interacting residues obtained from the PRED analysis for the open model $CyC_{8\beta}$-bound system provides a basis for the identification of potential drugs. Also, Adeniyi loop L1 is critical for the therapeutic activity because it acts as a switch between the open conformation and closed conformation. The transition of the residues also occurred upon binding of the $CyC_{8\beta}$ open model system but was not with many residues, unlike the covalent system.

These explained the decreased distance between an active site residue proximal to loop L2-amino acid (opposite) and the high-energy transitioned primary structure of Ag85C upon binding of the selective covalent inhibitor and the high binding affinity (-26.03 kcal/mol) of open model $CyC_{8\beta}$ system observed in the energy calculations. The binding of the $CyC_{8\beta}$-bound covalent inhibitor prompted the conformational shifts resulting from the instability,

flexibility, secondary structure transitions, correlated dynamics, and decreased distance. The results obtained for the selective covalent inhibitor bound system are quite promising. However, the findings obtained for the open model $CyC_{8\beta}$ especially the high free binding energy, consistent stability, and flexibilities are potential hotspots for the present/future drug identification that would facilitate drug design and development. Based on the free binding energy calculations and the characteristic hydrophobic interacting residues including PRO223, ALA224, PHE226, TRP262, and catalytic nucleophile HIS260 obtained from $CyC_{8\beta}$ open model bound Ag85C, I would suggest a virtual screening study to identify possible inhibitors of Ag85C as a molecular target.

Acknowledgment

First, the authors acknowledge the CREATOR of ALL THINGS, ALMIGHTY GOD, who gives lives, wisdom, and knowledge to all human races. Second, we recognise the Centre for High-Performance Computing, Cape-Town for providing computational resources, University of KwaZulu-Natal, Durban, South Africa for the free Tuition remission fees, and also the College of Health Sciences for their unquantifiable supports.

Compliance with Ethical Standards

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Disclosure of Potential Conflict of Interest

The authors declare no conflict of interest.

RESEARCH INVOLVING HUMAN PARTICIPANTS AND/OR ANIMALS

Informed Consent

This study did not require informed consent since the research work does not contain any studies with human participants performed by any of the authors.

FUNDING

No funding was received for this study

References

1. Backus, K. M. *et al.* The three *Mycobacterium tuberculosis* antigen 85 isoforms have unique substrates and activities determined by non-active site regions. *J. Biol. Chem.* **289**, 25041–25053 (2014).

2. Meehan, C. J. *et al.* Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat. Rev. Microbiol.* **17**, 533–545 (2019).

3. Jiang, Y. *et al.* Single nucleotide polymorphism in Ag85 genes of *Mycobacterium tuberculosis* complex: Analysis of 178 clinical isolates from China and 13 BCG strains. *Int. J. Med. Sci.* **12**, 126–134 (2015).

4. Adewumi, A. T., Ramharack, P., Soremekun, O. S. & Soliman, M. E. S. Delving into the Characteristic Features of " Menace " *Mycobacterium tuberculosis* Homologs : A Structural Dynamics and Proteomics Perspectives. *Protein J.* **39**, 118–132 (2020).

5. Iketleng, T. *et al.* *Mycobacterium tuberculosis* Next-Generation Whole Genome Sequencing: Opportunities and Challenges . *Tuberc. Res. Treat.* **2018**, 1–8 (2018).

6. Chawla, K., Martinez, E., Kumar, A., Shenoy, V. P. & Sintchenko, V. Whole-genome sequencing reveals genetic signature of bedaquiline resistance in a clinical isolate of *Mycobacterium tuberculosis*. *J. Glob. Antimicrob. Resist.* **15**, 103–104 (2018).

7. Veziris, N. *et al.* Rapid emergence of *Mycobacterium tuberculosis* bedaquiline resistance: lessons to avoid repeating past errors . *Eur. Respir. J.* **49**, 1601719 (2017).

8. Nguyen, P. C. *et al.* Cyclipostins and Cyclophostin analogs as promising compounds in the fight against tuberculosis. *Sci. Rep.* **7**, 1–15 (2017).

9. Ahangar Mohd Syed, Furze Christopher M., Guy, Collette S., Fullam, E., Cameron, A. D., Cooper, C., Graham, B. & Maskew, K. S. Structural and functional determination of homologs of the *Mycobacterium tuberculosis* N -acetylglucosamine-6-phosphate deacetylase (NagA). *J. Biol. Chem.* **293**, 9770–9783 (2018).

10. Wiker, H. G. & Harboe, M. The antigen 85 complex: A major secretion product of *Mycobacterium tuberculosis*. *Microbiol. Rev.* **56**, 648–661 (1992).

11. Kremer, L., Maughan, W. N., Wilson, R. A., Dover, L. G. & Besra, G. S. The M. tuberculosis antigen 85 complex and mycolyltransferase activity. *Lett. Appl. Microbiol.* **34**, 233–237 (2002).

12. Viljoen, A. *et al.* Cyclipostins and cyclophostin analogs inhibit the antigen 85C from *Mycobacterium tuberculosis* both in vitro and in vivo. *J. Biol. Chem.* **293**, 2755–2769 (2018).

13. Ronning, D. M., Vissa, V., Besra, G. S., Belisle, J. T. & Sacchettini, J. C. *Mycobacterium tuberculosis* antigen 85A and 85C structures confirm binding orientation and conserved substrate specificity. *J. Biol. Chem.* **279**, 36771–36777 (2004).

14. Jackson, M. *et al.* Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the *Mycobacterium tuberculosis* cell envelope. *Mol. Microbiol.* **31**, 1573–1587 (1999).

15. Nguyen, L., Chinnapapagari, S. & Thompson, C. J. FbpA-dependent biosynthesis of trehalose dimycolate is required for the intrinsic multidrug resistance, cell wall

structure, and colonial morphology of Mycobacterium smegmatis. *J. Bacteriol.* **187**, 6603–6611 (2005).

16. Jimmy, B. & Jose, J. Patient medication adherence: Measures in daily practice. *Oman Med. J.* **26**, 155–159 (2011).

17. Favrot, L., Lajiness, D. H. & Ronning, D. R. Inactivation of the *Mycobacterium tuberculosis* antigen 85 complex by covalent, allosteric inhibitors. *J. Biol. Chem.* **289**, 25031–25040 (2014).

18. Lawn, S. D., Bekker, L.-G., Middelkoop, K., Myer, L. & Wood, R. Impact of HIV Infection on the Epidemiology of Tuberculosis in a Peri-Urban Community in South Africa: The Need for Age-Specific Interventions. *Clin. Infect. Dis.* **42**, 1040–1047 (2006).

19. Shankar, E. M. *et al.* HIV-*Mycobacterium tuberculosis* co-infection: A 'danger-couple model' of disease pathogenesis. *Pathog. Dis.* **70**, 110–118 (2014).

20. Nglazi, M. D., Bekker, L. G., Wood, R. & Kaplan, R. The impact of HIV status and antiretroviral treatment on TB treatment outcomes of new tuberculosis patients attending co-located TB and ART services in South Africa: A retrospective cohort study. *BMC Infect. Dis.* **15**, 1–8 (2015).

21. Escalera-Antezana, J. P. *et al.* Clinical features of the first cases and a cluster of Coronavirus Disease 2019 (COVID-19) in Bolivia imported from Italy and Spain. *Travel Med. Infect. Dis.* 1–6 (2020). doi:10.1016/j.tmaid.2020.101653

22. Togun, T., Kampmann, B., Stoker, N. G. & Lipman, M. Anticipating the impact of the COVID-19 pandemic on TB patients and TB control programmes. *Ann. Clin. Microbiol. Antimicrob.* **19**, 1–6 (2020).

23. Madani, A. *et al.* Cyclipostins and Cyclophostin Analogues as Multitarget Inhibitors That Impair Growth of Mycobacterium abscessus. *ACS Infect. Dis.* **5**, 1597–1608 (2019).

24. Pettersen, E. F. *et al.* UCSF Chimera — A Visualization System for Exploratory Research and Analysis. **25**, 1605–1612 (2004).

25. Morgat, A. *et al.* Enzyme annotation in UniProtKB using Rhea. *Bioinformatics* **36**, 1896–1901 (2020).

26. Zheng Yanga, Keren Laskerb, Dina Schneidman-Duhovny, Ben Webb, Conrad C. Huang, Eric F. Pettersen, Thomas D. Goddard, Elaine C. Meng, Andrej Sali, and T. E. F. UCSF Chimera, MODELLER, and IMP: an Integrated Modeling System. *J. Struct. Biol.* **23**, 1–7 (2012).

27. Biasini, M. *et al.* SWISS-MODEL : modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).

28. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. Modeling protein quaternary structure of homo- and hetero- oligomers beyond binary interactions by homology. *Sci. Rep.* **7**, 1–15 (2017).

29. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350 (2011).

30. Huang, B. MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. *Omi. A J. Integr. Biol.* **13**, 325–330 (2009).

31. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

32. Shunmugam, L. & Soliman, M. E. S. Targeting HCV polymerase: A structural and

dynamic perspective into the mechanism of selective covalent inhibition. *RSC Adv.* **8**, 42210–42222 (2018).

33. Lee, T. S. *et al.* GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **58**, 2043–2050 (2018).

34. Zheng, S. *et al.* VFFDT: A New Software for Preparing AMBER Force Field Parameters for Metal-Containing Molecular Systems. *J. Chem. Inf. Model.* **56**, 811–818 (2016).

35. Ramharack, P., Oguntade, S. & Soliman, M. E. S. Delving into Zika virus structural dynamics-a closer look at NS3 helicase loop flexibility and its role in drug discovery. *RSC Adv.* **7**, 22133–22144 (2017).

36. Bjij, I., Khan, S., Betz, R., Cherqaoui, D. & Soliman, M. E. S. Exploring the Structural Mechanism of Covalently Bound E3 Ubiquitin Ligase: Catalytic or Allosteric Inhibition? *Protein J.* **37**, 500–509 (2018).

37. Khan, S., Bjij, I., Betz, R. M. & Soliman, M. E. S. Reversible versus irreversible inhibition modes of ERK2: A comparative analysis for ERK2 protein kinase in cancer therapy. *Future Med. Chem.* **10**, 1003–1015 (2018).

38. 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).

39. Adewumi, Adeniyi T, Ajadi, Mary B, Soremekun, Opeyemi S. and Soliman, M. E. S. Thompson loop : opportunities for antitubercular demethylmenaquinone methyltransferase protein †. *RSC Adv.* **10**, 23466–23483 (2020).

40. Joseph M. Hayes and Georgios Archontis. *MM-GB(PB)SA Calculations of Protein-Ligand Binding Free Energies*. (InTech Europe, 2012). doi:DOI: 10.5772/37107

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

42. Zhang, C. *et al.* Protein structure and sequence re-analysis of 2019-nCoV genome does not indicate snakes as its intermediate host or the unique similarity between its spike protein insertions and HIV-1. **2**, 1–11 (2019).

43. Seifert, E. OriginPro 9.1: Scientific data analysis and graphing software - Software review. *J. Chem. Inf. Model.* **54**, 1552 (2014).

44. Soremekun, O. S., Olotu, F. A., Agoni, C. & Soliman, M. E. S. Recruiting monomer for dimer formation: resolving the antagonistic mechanisms of novel immune check point inhibitors against Programmed Death Ligand-1 in cancer immunotherapy. *Mol. Simul.* **45**, 777–789 (2019).

45. Hassan, S. *et al.* Computational approach identifies protein off-targets for Isoniazid-NAD adduct: hypothesizing a possible drug resistance mechanism in *Mycobacterium tuberculosis* . *J. Biomol. Struct. Dyn.* **0**, 1–14 (2019).

46. Khan, S. *et al.* Covalent simulations of covalent/irreversible enzyme inhibition in drug discovery: A reliable technical protocol. *Future Med. Chem.* **10**, 2265–2275 (2018).

47. Post, M., Wolf, S. & Stock, G. Principal component analysis of nonequilibrium molecular dynamics simulations. *J. Chem. Phys.* **150**, 1–15 (2019).

48. Devnarain, N. & Soliman, M. E. S. A panoptic uncovering of the dynamical evolution of the Zika Virus NS5 methyltransferase binding site loops—zeroing in on the molecular landscape. *Chem. Biol. Drug Des.* **92**, 1838–1850 (2018).

49. Sun, H. *et al.* Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* **20**, 14450–14460 (2018).

50. Shehu, A. & Kavraki, L. E. Modeling structures and motions of loops in protein molecules. *Entropy* **14**, 252–290 (2012).

51. Tanwar, Himani and Doss, C. G. P. An Integrated Computational Framework to Assess the Mutational Landscape of α-L-Idurondase IDUA Gene. *J. Cell. Biochem.* 555–565 (2018). doi:10.1002/jcb.26214

52. Závodszky, P., Kardos, J., Svingor, Á. & Petsko, G. A. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7406–7411 (1998).

53. Bjij, I., Ramharack, P., Khan, S., Cherqaoui, D. & Soliman, M. E. S. Tracing potential covalent inhibitors of an E3 ubiquitin ligase through target-focused modelling. *Molecules* **24**, 1–15 (2019).

54. 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).

55. Gupta, V. *et al.* Structural ordering of disordered ligand-binding loops of biotin protein ligase into active conformations as a consequence of dehydration. *PLoS One* **5**, 1–10 (2010).

56. Montenegro, F. A., Cantero, J. R. & Barrera, N. P. Combining mass spectrometry and X-ray crystallography for analyzing native-like membrane protein lipid complexes. *Front. Physiol.* **8**, 1–8 (2017).

57. Debenedictis, E. P. & Keten, S. Mechanical unfolding of alpha- and beta-helical protein motifs. *Soft Matter* **15**, 1243–1252 (2019).

58. Mcgillewie, L. & Soliman, M. E. Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses. *Proteins Struct. Funct. Bioinforma.* **83**, 1693–1705 (2015).

59. McGillewie, L., Ramesh, M. & Soliman, M. E. Sequence, Structural Analysis and Metrics to Define the Unique Dynamic Features of the Flap Regions Among Aspartic Proteases. *Protein J.* **36**, 385–396 (2017).

60. Yang, L. Q. *et al.* Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms. *J. Biomol. Struct. Dyn.* **32**, 372–393 (2014).

61. Zhang, D. & Lazim, R. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Sci. Rep.* **7**, 1–12 (2017).

62. Malleshappa Gowder, S., Chatterjee, J., Chaudhuri, T. & Paul, K. Prediction and Analysis of Surface Hydrophobic Residues in Tertiary Structure of Proteins. *Sci. World J.* 1–7 (2014). doi:10.1155/2014/971258

63. Mitusińska, K., Raczyńska, A., Bzówka, M., Bagrowska, W. & Góra, A. Applications of water molecules for analysis of macromolecule properties. *Comput. Struct. Biotechnol. J.* **18**, 355–365 (2020).

64. Chai, C. C. & Jhon, M. S. Molecular dynamics study on protein and it's water structure at high pressure. *Mol. Simul.* **23**, 257–274 (2000).

CHAPTER 7

7.1 Conclusion

There have been highly devoted unrelenting interdisciplinary efforts in a quest to develop an anti-TB drug in over a century that could put a lasting end to the menace of tuberculosis. Many drugs have been very effective at sometimes but later lost their efficacies because of the continuous evolution of the active *Mycobacterium tuberculosis strains*. These new *Mtb strains* develop mutation-mediated or acquired drug resistance against the current antitubercular drugs; the condition that has remained the top global challenge in the two decades.

The present study was intended to gain a complex understanding that describing the structural and molecular behaviours of the weak spots in the protein, specifically, the loops covering the active site and the binding residues of the three studied targets crucial to the drug resistance, survival, and virulence of the *Mtb*. The project incorporated the application of bioinformatics tools to obtain the structural sequence (for *NagA* and Ag85C). Moreover, a homology modelling was carried out to create an *Mtb menG* 3D model and its homologs in *NagA* and Ag85C for sequence analysis. An investigation of unique similarities and differences among the homologs of the Mtb proteins were determined for information on the drug design of a polypharmacological agent.

The MD analyses and findings presented here are innovative. Sequence alignment of homologs of Mtb *NagA* among the other selected mycobacterial species showed that unique primary structures were conserved. The interacting residues and loops of Ag85C among the mycobacteria were conserved. The effectiveness of menG inhibitors DG70 and $CyC_{8\beta}$ for menG and Ag85C respectively was examined based on the per-residue energy decomposition

analysis and the overall binding free energy contributions. The findings showed the most integral chemical frame of the inhibitors. This information ascertains the inhibitory. The Thompson loop in the *menG* assumed an open conformation in the unbound structure and a tightly closed conformation in DG70-menG complex.

Similarly, Adeniyi loop in the Ag85C also opened in the unbound protein and closed upon the ligand binding. The binding of the selective covalent inhibitor ($CyC_{8\beta}$) to Ag85C revealed some weak hotspots in the protein that can be explored. Mutation analysis of the *NagA* D267A, H56Q, H56N, and H56A mutants revealed destabilizing effects at the binding site of the protein.

However, two of the top challenges and shortcomings we encountered during this study including the problem of parameterization of Zinc and cadmium metals critical to the catalytic functions of the target *NagA*, optimization issue of the transition state inhibitor of *NagA*, etc.

More than six months, it was difficult running the molecular dynamics (MD) simulations and MMGB(PB)SA calculations for the ligand-bound target because the metals situated at the *NagA* protein active site were causing intraatomic bonding of the ligand (N-methylhydroxyphosphinyl-D- glucosamine-6-phosphate). Therefore, the dynamics and motion characterization data of the ligand-protein interactions could be achieved in this study. It is our intention to obtain the parametrization of these metals in future to understand the characteristic behaviour of the interactions of NagA upon the binding of N-methylhydroxyphosphinyl-D- glucosamine-6-phosphate.

Altogether, this study presents requisite findings into the binding terrains and mechanism of inhibition of *NagA*, *menG*, and Ag85C protein towards the design and development of a drug for the cure of tuberculosis or other mycobacterial infection.

7.1 Future perspectives

Considering the vital roles of the three studied targets (*NagA*, *menG*, Ag85C) and our findings regarding the dynamics and structural characterisation findings obtained in this work, further study like searching for a promising lead (Pharmacophore) using Structure-based Computer-aided drug design (CADD) strategy can performed. Presently, the only known inhibitor of *NagA* is the transition state of its natural substrate. There are two metals (Zn and Cd) at the active site of *NagA* hthat ave catalytic significance. However, running molecular dynamics simulations for the target was very complicated, which could not be achieved during the short time of this study. Optimization and characterization of the structural dynamics and functional motion in the presence of the metals in future could be integrated into the design of potential *NagA* inhibitor.

Second, the *menG* crystallized structure and its binding site have not been; it does not seem to have been deposited into the Protein data bank (PDB). It would be interesting to characterize the motion and dynamics feature of the protein. Moreover, the target can be used to explore an effective inhibitor in pharmacophore modelling.

Other Future research focus include:

1. building Mesoporous silica nanoparticles (MSN) as delivery system and studying the interactions of a few drugs with the MSN to understand controlled release mechanism and sustainable delivery,
2. apply NAMD and DFT to molecular dynamics to gain dynamic and motion insights into some targets of *Mycobacteriun tuberculosis* and *Plasmodium falciparum*, and

3. possibly do pharmacophore and compound screening works of the ligands and target studied in this project research using the structure-based virtual screening(SBVS).

Supplementary Information

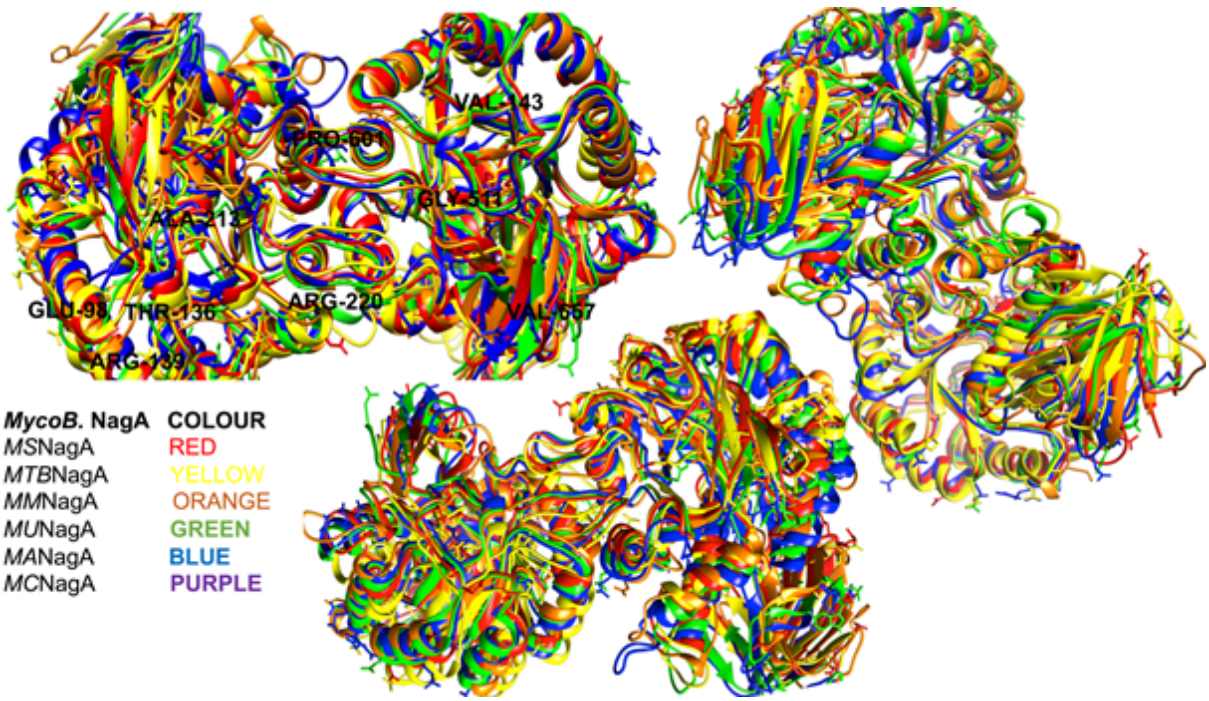


Figure. 4.S1. Different views of post-minimized apo-structures of Mycobacteria NagA enzymes demonstrated the effect of relaxation on the crystal structures. *MS*-NagA, *Mtb*-NagA, *MM*-NagA, *MU*-NagA, *MA*-NagA, and *MC*-NagA enzymes are represented in *red*, *yellow*, *orange*, *green*, *blue,* and *purple.*

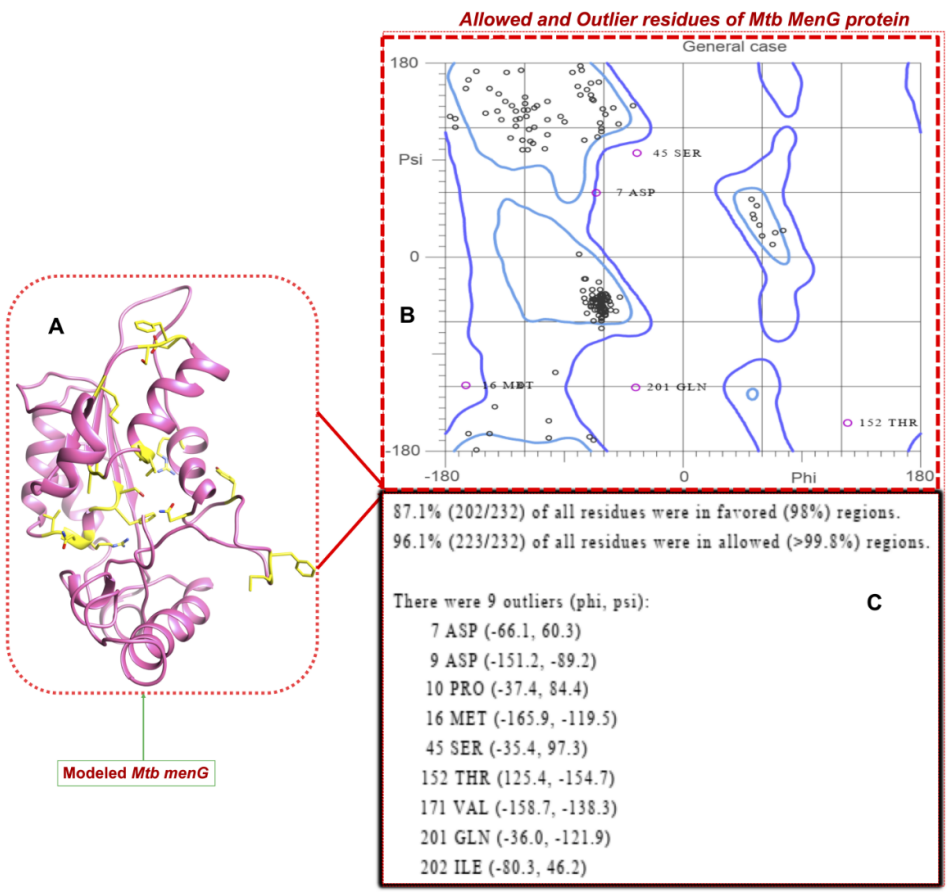


Figure 5S1. Ramachandran plot validation of the active site of the modeled *Mtb menG*.

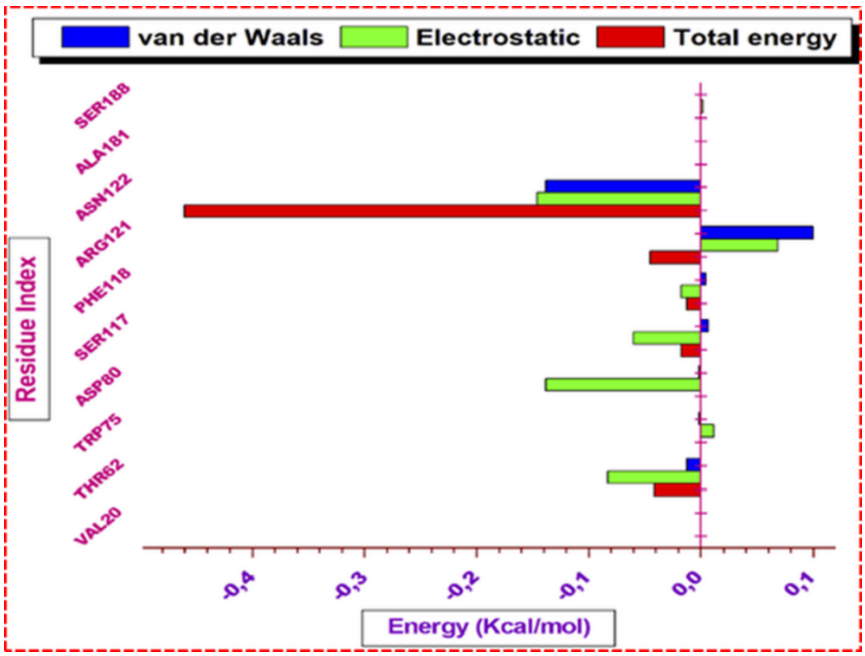


Figure 5S2. Per-residue decomposition of a *menG*-DG70 complex using the previous predicted active site that reported.

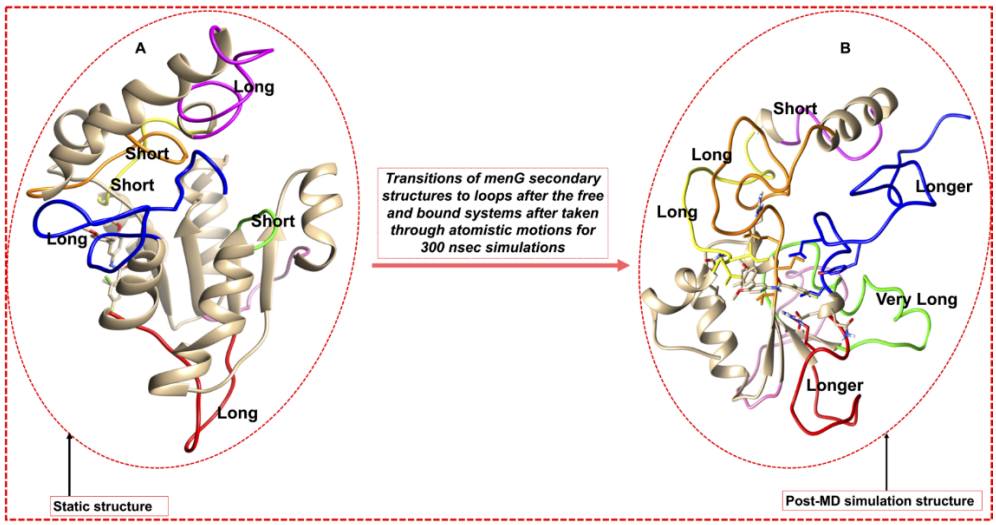


Figure 5S3 Secondary structure (α-helices and β-sheets) transformation to Loop structures over a 300 ns trajectory.

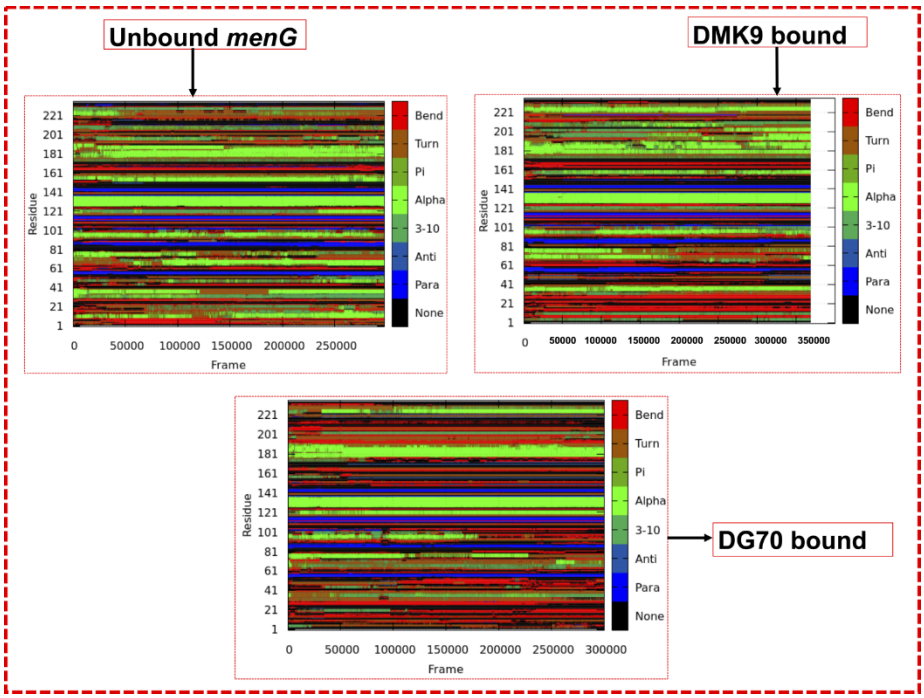


Figure 5S4. Illustrations of DSSP for *Apo* (**A**), ApoDMK9 (**B**), and ApoDG70 (**C**) systems over 300 ns MD simulation.

Table 5S1. Mean RMSD and RMSF of the Apo, DMK9-, and DG70-bound

System	Loop residues 40-54	102-112	211-220
	Average RMSD (Å)		
Apo	5.26 ± 0.63	1.62 ± 0.17	4.77 ± 0.63
ApoDMK9	4.03 ± 0.84	1.60 ± 0.17	3.45 ± 0.30
ApoDG70	5.84 ± 1.03	2.69 ± 0.98	3.19 ± 0.43
	Average RMSF (Å)		
Apo	3.04 ± 0.80	1.32 ± 0.31	2.44 ± 0.67
ApoDMK9	2.86 ± 0.85	1.22 ± 0.23	1.29 ± 0.17
ApoDG70	3.34 ± 1.67	1.74 ± 0.69	2.07 ± 0.64

APPENDICES

Appendix A

The Protein Journal
https://doi.org/10.1007/s10930-020-09890-4

Delving into the Characteristic Features of "Menace" *Mycobacterium tuberculosis* Homologs: A Structural Dynamics and Proteomics Perspectives

Adeniyi T. Adewumi[1] · Pritika Ramharack[1] · Opeyemi S. Soremekun[1] · Mahmoud E. S. Soliman[1]

Abstract

The global increase in the morbidity/mortality rate of Mycobacterial infections, predominantly renascent tuberculosis, leprosy, and Buruli ulcers have become worrisome over the years. More challenging is the incidence of resistance mediated by mutant *Mycobacterium strains* against front-line antitubercular drugs. Homologous to all *Mycobacteria* species is the GlcNAc-6-phosphate deacetylase (NagA) which catalyzes essential amino sugars synthesis required for cell wall architecture, hence, metamorphosing into an important pharmacological target for curtailing virulence and drug-resistance. This study used integrated bioinformatics methods, MD simulations, and *DynaMut* and *PolyPhen2* to; explore unique features, monitor dynamics, and analyze the functional impact of non-synonymous single-nucleotide polymorphisms of the six NagA of most ruinous Mycobacterium species; *tuberculosis* (*Mtb*), *smegmatis* (*MS*), *marinum* (*MM*), *ulcerans*, *africanum*, and *microti respectively*. This approach is essential for multi-targeting and could result in the identification of potential polypharmacological antitubercular compounds. Comparative sequential analyses revealed ≤ 50% of the overall structure, including the catalytic Asp267 and reactive Cys131, remained conserved. Interestingly, *MS*-NagA and *MM*-NagA possess unique hydrophobic isoleucine (Ile) residues at their active sites in contrast to leucine (Leu) found in other variants. More so, unique to the active sites of the NagA is a 'subunit loop' that covers the active site; probably crucial in binding (entry and exit) mechanisms of targeted NagA inhibitors. Relatively, nsSNP mutations exerted a destabilizing effect on the native NagA conformation. Structural and dynamical insights provided, basically pin-pointed the "Achilles' heel" explorable for the rational drug design of target-specific 'NagA' inhibitors potent against a wide range of mycobacterial diseases.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10930-020-09890-4) contains supplementary material, which is available to authorized users.

✉ Mahmoud E. S. Soliman
soliman@ukzn.ac.za

[1] Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

Published online: 11 March 2020

Appendix B

RSC Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2020, **10**, 23466

Thompson loop: opportunities for antitubercular drug design by targeting the weak spot in demethylmenaquinone methyltransferase protein†

Adeniyi T. Adewumi, [a] Opeyemi S. Soremekun, [a] Mary B. Ajadi [b] and Mahmoud E. S. Soliman [*a]

Drug-resistant Tuberculosis (TB) has remained the top global health challenge, with a yearly estimation of 10 million infections and 1.5 million deaths in humans. Demethylmenaquinone methyltransferase (*menG*) catalyzes demethylmenaquinone conversion to menaquinone (MK) that is implicated in the TB pathogenesis, hence, it has become a major drug target. DG70 is a biphenyl amide compound known to be a high binding affinity inhibitor of *menG*. This study investigated the structural and dynamic impacts of DG70 upon binding to *menG* using atom-based dynamic simulation. Our findings revealed that the modeled structure of *menG* possesses some Rossman-like methyltransferase characteristic features including two GXG motifs, an omega-like loop (residues 210–220) called the Thompson loop, nine α-helices, five β-strands, *etc.* Furthermore, atom-based dynamic simulations revealed that the Thompson loop is critical in the therapeutic activity of DG70. The loop assumed an open conformation in the unliganded-*menG* structure. However, in the DG70-*menG*, it assumed a tightly closed conformation. This explains the high binding affinity (−32.48 kcal mol^{-1}) observed in the energy calculations. Interestingly, these findings are further collaborated by the conformational perturbation in the *menG* protein. Conclusively, insights from this study, highlight the structural "Achilles heel" in *menG* protein which can be further leveraged by inhibitors tailored to specifically target them.

Received 9th April 2020
Accepted 15th June 2020

DOI: 10.1039/d0ra03206a

rsc.li/rsc-advances

1 Introduction

Mycobacterium tuberculosis (*Mtb*) is the world's deadliest infection after the renascent human immunodeficiency virus/AIDS.[1] About two billion latent *Mtb* in man kill 2–3 million people annually.[1] Presently, more than 484 000 new cases of multidrug and extensively resistant (active) TB cause about 1.5 million human deaths every year.[2,3] More disturbing is the mutation-mediated and acquired drug resistance by the newer antitubercular drugs including linezolid, bedaquiline, (BDQ), *etc.*[4,5] More than one-half of the MDR/XDR-TB drugs have failed due to mutation.[6] In the quest for a promising novel drug over the years in this area, researchers have focused on protein–drug and protein–protein interactions for possible tuberculosis cures using various techniques and methodologies through targeting protein biocatalysis.[7,8] While a lot of studies have reported the targeting of constituents of *Mtb* cell walls, most often, the mycolic acid synthesis, only a few reports seem to be available on the study of the inhibition of the *Mtb* respiratory pathway using experimental and computational tools such as molecular dynamics (MD). BDQ is the first respiratory inhibitor that acts by binding to the oligomeric and proteolipid subunit C of *Mtb* ATP synthase (*atpE gene*).[5,6] BDQ was very effective against both the susceptible and multi-drug resistant TB.[9] However, there is a high prevalence of mutations in *Mtb* genes, including *atpE*, *atpC*, *mmpR* (Rv0678) to bedaquiline. A whole-genome sequence has revealed a genetic signature of BDQ resistance in a clinical *Mtb* isolate C; a microhetero-resistance found in a targeted deep sequencing analysis.[10,11] Similarly, a previous study shows that *Mtb* demethylmenaquinone methyltransferase, *menG* (rv0558) is a potential target and involves the biosynthesis of menaquinone (MK), the substance required for the cell maintenance of the prokaryotic including *Mtb*.[12–14] The amino acids of *Mycobacterium tuberculosis menG* (rv0558) protein are conserved and distributed among *Mycobacterium species* homologs[15] as shown in Fig. 1.[13]

MenG is a membrane domain-associated enzyme and belongs to the SAM-dependent methyltransferase (MTase) superfamily comprising of an α/β/α fold structure.[17,18] MTase domain is an important target in drug design. *Mtb menG* encodes *S*-adenosyl-L-methionine (SAM)-dependent *menG* (Rv0558) that catalyzes the biosynthesis of menaquinone [in the

[a]*Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa. E-mail: soliman@ukzn.ac.za; Fax: +27 31 260 7872; Tel: +27 31 260 8048*

[b]*Department of Medical Biochemistry, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Howard Campus, Durban 4000, South Africa*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0ra03206a

Appendix C

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1844061

Taylor & Francis
Taylor & Francis Group

Weak spots inhibition in the *Mycobacterium tuberculosis* antigen 85C target for antitubercular drug design through selective irreversible covalent inhibitor-SER124

Adeniyi T. Adewumi[a], Ahmed Elrashedy[a], Opeyemi S. Soremekun[a], Mary B. Ajadi[b] and Mahmoud E. S. Soliman[a]

[a]Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Durban, South Africa; [b]Department of Medical Biochemistry, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

Communicated by Ramaswamy H. Sarma

ABSTRACT

Mycobacterium tuberculosis (*Mtb*) encoded secreted antigen 85 enzymes (Ag85A/Ag85B/Ag85C) play that critical roles in the virulence, survival and drug-resistant TB of the pathogen. Ag85 proteins are potential antitubercular drug targets because they are essential in the catalytic synthesis of trehalose moieties and mycolic acid attachment to the *Mtb* cell wall. Recently, experimental protocols led to the discovery of a selective covalent Ag85 inhibitor, β-isomer monocyclic enolphosphorus Cycliphostin ($CyC_{8\beta}$) compound, which targets the Ag85 serine 124 to exhibit a promising therapeutic activity. For the first time, our study unravelled the structural features among *Mtb* Ag85C homologs and motions and dynamics of Ag85C when the $CyC_{8\beta}$ bound covalently and in open model conformations to the protein using bioinformatics tools and integrated Molecular dynamics simulations. Comparative Ag85C sequence analysis revealed conserved regions; 70% active site, 90% Adeniyi loop L1 and 50% loop L2, which acts as a switch between open and closed conformations. The average C-α atoms RMSD (2.05 Å) and RMSF (0.9 Å) revealed instability and high induced flexibility in the $CyC_{8\beta}$ covalent-bound compared to the apo and open model systems, which displayed more stability and lower fluctuations. DSSP showed structural transitions of α-helices to bend and loops to 3_{10}-helices in the bound systems. SASA of $CyC_{8\beta}$ covalent bound showed active site hydrophobic residues exposure to huge solvent. Therefore, these findings present the potential opportunity hotspots in Ag85C protein that would aid the structure-based design of novel chemical entities capable of resulting in potent antitubercular drugs.

ARTICLE HISTORY
Received 16 July 2020
Accepted 23 October 2020

KEYWORDS
Antigen 85; Adeniyi loops; drug-resistant TB; molecular dynamics simulations; β-isomer monocyclic enolphosphorus Cycliphostin (CyC8β); bioinformatics

1. Introduction

Tuberculosis (TB) is a lung infection caused by the pathogen known as *Mycobacterium tuberculosis* (*Mtb*) that has 'blown away' human generations. About 10 million new TB infections, which result in approximately 1.3 million deaths (Backus et al., 2014), are reported annually in humans (Meehan et al., 2019). Although there have been various strategies, including diagnosis, vaccine program and curative therapy channelled towards TB eradication, high morbidity and mortality rate of TB has continued for many decades. More challenging TB is the metamorphosising active form caused by the evolving *Mtb strains* due to mutation-mediated resistance such as the non-synonymous single-nucleotide polymorphisms (nsSNP) (Adewumi et al., 2020; Jiang et al., 2015) in Ag85 protein by insertions or deletions in the genes encoding from drug target/drug metabolic pathways. Besides, the induced-mutation resistance could also occur due to clonal expansion and independently acquired drug resistance (Iketleng, 2018).

Moreover, the slow growth rate of bacilli and the thick, greasy drug-impermeable nature of the cell wall contributes to the success of the *Mtb* drug resistance. The increasing rate of active TB (or extensively/multidrug) resistant TB is quite worrisome. Some of the frontline antitubercular drugs exhibiting resistance include isoniazid, rifampicin. Moreover, literature has shown that new drugs like bedaquiline exhibit frameshift and phenotypic drug resistance (Chawla et al., 2018; Veziris et al., 2017). Although some of these drugs are still therapeutically relevant, especially as combined anti-TB therapies, an associated severe problem is the many drug dose and dosage regimen, requiring two years treatment period, which results in an inherent difficulty in the patients' compliance (Nguyen, 2017).

The cell of mycobacteria contains different proteins that are responsible for the successes of the microorganisms in pathogenesis, survival and resistance to many drugs (Ahangar et al., 2018; Nguyen, 2017). Among these proteins, *Mtb* antigen (Ag) 85 protein complex (Ag85A, Ag85B and Ag85C) bind to the human fibronectin *via* ether heparin and

CONTACT Mahmoud E. S. Soliman soliman@ukzn.ac.za Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa