



**COMPUTER-AIDED APPROACHES IN DRUG DESIGN: THE
EXIGENT WAY FORWARD: DYNAMIC PERSPECTIVES INTO THE
MECHANISTIC ACTIVITIES OF SMALL MOLECULE INHIBITORS
TOWARD ANTIVIRAL, ANTITUBERCULAR AND ANTICANCER
THERAPEUTIC INTERVENTIONS**

CLEMENT AGONI

217036027

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

in

Health Sciences (Pharmaceutical Chemistry)

Supervisor

Prof. Mahmoud Soliman

2021

**COMPUTER-AIDED APPROACHES IN DRUG DESIGN: THE EXIGENT
WAY FORWARD: DYNAMIC PERSPECTIVES INTO THE
MECHANISTIC ACTIVITIES OF SMALL MOLECULE INHIBITORS
TOWARD ANTIVIRAL, ANTITUBERCULAR AND ANTICANCER
THERAPEUTIC INTERVENTIONS**

by

Clement Agoni

217036027

2021

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

in

Health Sciences (Pharmaceutical Chemistry)

This is the thesis in which the chapters are written as a set of discrete submitted manuscripts, with an overall introduction and final summary. In most cases, these chapters will have been published in reputable, internationally recognized, peer-reviewed journals.

This is to certify that the contents of this thesis are the original research work of Clement Agoni.

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

Supervisor: Signed:

Name: Prof. Mahmoud E. Soliman Date: 02/11/2020

ABSTRACT

The crucial role of CADD in the drug design process is now indisputable and has proven over the years that it can accelerate the discovery potential drug candidates while reducing the associated cost. Using knowledge and information about biological target or knowledge about a ligand with proven bioactivity, CADD, and its techniques can influence various drug discovery pipeline stages. The ability CADD approaches to elucidate drug-target interactions at the atomistic level allows for investigations of the mechanism of drugs' actions, revealing atomistic insights that influence drug design and improvement. CADD approaches also seek to augment traditional in vitro and in vivo experimental techniques and not replace them since CADD approaches can also allow modeling complex biological processes that hitherto seemed impossible to explore using experimental methods.

According to the World Health Organization (WHO), featuring prominently in the top ten causes of death are cancer, lower respiratory tract infection, tuberculosis (TB), and viral infections such as HIV/AIDS. Collectively, these diseases are of global health concerns, considering a large number of associated deaths yearly. Over the years, several therapeutic interventions have been employed to treat, manage, or cure these diseases, including chemotherapy, surgery, and radiotherapy. Of these options, small molecule inhibitors have constituted an integral component in chemotherapy, thereby undoubtedly playing an essential role in patient management.

Although significant success has been achieved using existing therapeutic approaches, the emergence of drug resistance and the challenges of associated adverse side effects has prompted the need for the drug design processes against these diseases to remain innovative, including combining existing drugs and establishing improved therapeutic options that could overcome resistance while maintaining minimal side effects to patients. Therefore, an exploration of drug-

target interactions towards unraveling mechanisms of actions as performed in the reports in this thesis are relevant since the molecular mechanism provided could form the basis for the design and identification of new therapeutic agents, improvement of the therapeutic activity of existing drugs, and also aid in the development of novel therapeutic strategies against these diseases of global health concern.

Therefore the studies in this thesis employed CADD approaches to investigate molecular mechanisms of actions of novel therapeutic strategies directed towards some crucial therapeutics implicated in viral infections, tuberculosis, and cancer. Therapeutic targets studied included; SARS-CoV-2 RNA dependent RNA polymerase (SARS-CoV-2 RdRp), Human Rhinovirus B14 (HRV-B14) and human *N*-myristoyltransferases in viral infections, Dihydrofolate reductase (DHFR) and Flavin-dependent thymidylate synthase (FDTS) in TB, human variants of TCR-CD1d, and Protein Tyrosine Phosphatase Receptor Zeta (PTPRZ) in cancer.

The studies in this thesis is divided into three domains and begins with a thorough review of the concept of druggability and drug-likeness since the crux of the subsequent reports revolved around therapeutic targets and their inhibitions by small molecule inhibitors. This review highlights the principles of druggability and drug-likeness while detailing the recent advancements in drug discovery. The review concludes by presenting the different computational, highlighting their reliability for predictive analysis.

In the first domain of the research, we sought to unravel the inhibitory mechanism of some small molecule inhibitors against some therapeutic targets in viral infections by explicitly focusing on the therapeutic targets; SARS-CoV-2 RdRp, HRV-B14, and *N*-myristoyltransferase.

Therapeutic targeting of SARS-CoV-2 RdRp has been extensively explored as a viable approach in the treatment of COVID-19. By examining the binding mechanism of Remdesivir, which

hitherto was unclear, this study sought to unravel the structural and conformational implications on SARS-CoV-2 RdRp and subsequently identify crucial pharmacophoric moieties of Remdesivir required for its inhibitory potency. Computational analysis showed that the modulatory activity of Remdesivir is characterized by an extensive array of high-affinity and consistent molecular interactions with specific active site residues that anchor Remdesivir within the binding pocket for efficient binding. Results also showed that Remdesivir binding induces minimal individual amino acid perturbations, subtly interferes with deviations of C- α atoms, and restricts the systematic transition of SARS-CoV-2 RdRp from the “buried” hydrophobic region to the “surface-exposed” hydrophilic region. Based on observed high-affinity interactions with SARS-CoV-2 RdRp, a pharmacophore model was generated, which showcased the crucial functional moieties of Remdesivir. The pharmacophore was subsequently employed for virtual screening to identify potential inhibitors of SARS-CoV-2 RdRp. The structural insights and the optimized pharmacophoric model provided would augment the design of improved analogs of Remdesivir that could expand treatment options for COVID-19.

The next study sought to explore the therapeutic targeting of human rhinoviruses (HRV) amidst challenges associated with the existence of a wide variety of HRV serotypes. By employing advanced computational techniques, the molecular mechanism of inhibition of a novel benzothiophene derivative that reportedly binds HRV-B14 was investigated. An analysis of the residue-residue interaction profile revealed of HRV upon the benzothiophene derivative binding revealed a distortion of the hitherto compacted and extensively networked HRV structure. This was evidenced by the fewer inter-residue hydrogen bonds, reduced van der Waals interactions, and increased residue flexibility. However, a decrease in the north-south wall's flexibility around the

canyon region also suggested that the benzothiophene derivative's binding impedes the “breathing motion” of HRV-B14; hence its inhibition.

The next study in the first domain of the research investigated the structural and molecular mechanisms of action associated with the dual inhibitory activity of IMP-1088. This novel compound reportedly inhibits human *N*-myristoyltransferase subtypes 1 and 2 towards common cold therapy. This is because it has emerged that the pharmacological inhibition of *N*-myristoyltransferase is an efficient non-cytotoxic strategy to completely thwart the replication process of rhinovirus toward common cold treatment. Using augmentative computational and nanosecond-based analyses, findings of the study revealed that the steady and consistent interactions of IMP-1088 with specific residues; Tyr296, Phe190, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182, and Asn246, shared within the binding pockets of both HNMT subtypes, in addition to peculiar structural changes account for its dual inhibitory potency. Findings thus unveiled atomistic and structural perspectives that could form the basis for designing novel dual-acting inhibitors of *N*-myristoyltransferase towards common cold therapy.

In the second domain of the research, the mechanism of action of some small molecule inhibitors against DHFR, FDTS, and *Mtb* ATP synthase in treating tuberculosis is extensively investigated and reportedly subsequently.

To begin with, the dual therapeutic targeting of crucial enzymes in the folate biosynthetic pathway was explored towards developing novel treatment methods for TB. Therefore, the study investigated the molecular mechanisms and structural dynamics associated with dual inhibitory activity of PAS-M against both DHFR and FDTS, which hitherto was unclear. MD simulations revealed that PAS-M binding towards DHFR and FDTS is characterized by a recurrence of strong conventional hydrogen bond interactions between a peculiar site residue the 2-amino-

decahydropteridin-4-ol group of PAS-M. Structural dynamics of the bound complexes of both enzymes revealed that, upon binding, PAS-M is anchored at the entrance of hydrophobic pockets by a strong hydrogen bond interaction while the rest of the structure gains access to deeper hydrophobic residues to engage in favorable interactions. Further analysis of atomistic changes of both enzymes showed increased C- α atom deviations and an increase C- α atoms radius of gyration consistent with structural disorientations. These conformational changes possibly interfered with the enzymes' biological functions and hence their inhibition as experimentally reported.

Additionally, in this domain, the therapeutic targeting of the ATP machinery of *Mtb* by Bedaquiline (BDQ) was explored towards unravelling the structures and atomistic perspectives that account for the ability of BDQ to selectively inhibits mycobacterial F1Fo-ATP synthase *via* its rotor c-ring. BDQ is shown to form strong interaction with Glu65B and Asp32B and, consequently, block these residues' role in proton binding and ion. BDQ binding was also revealed to impede the rotatory motion of the rotor c-ring by inducing a compact conformation on the ring with its bulky structure. Complementary binding of two molecules of BDQ to the rotor c-ring, proving that increasing the number of BDQ molecule enhances inhibitory potency.

The last study in this research domain investigated the impact of triple mutations (L59V, E61D, and I66M) on the binding of BDQ to *Mtb* F1F0 ATP-synthase. The study showed that the mutations significantly impacted the binding affinity of BDQ, evidenced by a decrease in the estimated binding free energy (ΔG). Likewise, the structural integrity and conformational architecture of F1F0 ATP-synthase was distorted due to the mutation, which could have interfered with the binding of BDQ.

The third domain of the research in this thesis investigated some small molecule inhibitors' inhibitory mechanism against some therapeutic targets in cancer, specifically PTPRZ and hTCR-

CD1d. Studies in the third domain of the research in the thesis began with the investigation of the inhibitory mechanism of NAZ2329, an allosteric inhibitor of PTPRZ, by specifically investigating its binding effect on the atomic flexibility of the WPD-loop. Having been established as a crucial determinant of the catalytic activity of PTPRZ, an implicated protein in glioblastoma cells, its successful therapeutic modulation could present a viable treatment option in glioblastoma. Structural insights from an MD simulation revealed that NAZ2329 binding induces an open conformation of the WPD-loop which subsequently prevents the participation of the catalytic aspartate of PTPRZ from participating in catalysis, hence inhibiting the activity of PTPRZ. A pharmacophore was also created based on high energy contributing residues which highlighted essential moieties of NAZ2329 and could be used in screening compound libraries for potential inhibitors of PTPRZ.

A second study in this domain sought to explore how structural modification could improve a therapeutic agent's potency from an atomistic perspective. This study was based on an earlier report in which the incorporation of a hydrocinnamoyl ester on C6'' and C4-OH truncation of the sphingoid base of KRN7000 generated a novel compound AH10-7 with high therapeutic potency and selectivity in human TCR-CD1d and subsequently results in the activation of invariant natural killer T cells (*i*NKT). The hydrocinnamoyl ester moiety was shown to engage in high-affinity interactions, possibly accounting for the selectivity and higher potency of AH10-7. Molecular and structural perspectives provided could aid in the design of novel α -GalCer derivatives for cancer immunotherapeutics.

Chapter 3 provides theoretical insights into the various molecular modeling tools and techniques employed to investigate the various conformational changes, structural conformations, and the

associated mechanism of inhibitions of the studied inhibitors towards viral, tuberculosis, and cancer therapy.

Chapter 4 provided sufficient details on druggability and drug-likeness principles and their recent advancements in the drug discovery field. The study also presents the different computational tools and their reliability of predictive analysis in the drug discovery domain. It thus provides a comprehensive guide for computational-oriented drug discovery research.

Chapter 5 provides an understanding of the binding mechanism of Remdesivir, providing structural and conformational implications on SARS-CoV-2 RdRp upon its binding and identifying its crucial pharmacophoric moieties.

Chapter 6 explains the mechanism of inhibition of a novel benzothiophene derivative, revealing its distortion of the native extensively networked and compact residue profile.

Chapter 7 unravels molecular and structural bases behind this dual inhibitory potential of the novel inhibitor IMP-1088 toward common cold therapy using augmentative computational and cheminformatics methods. The study also highlights the pharmacological propensities of IMP-1088.

Chapter 8 unravels the molecular mechanisms and structural dynamics of the dual inhibitory activity of PAS-M towards DHFR and FDTS.

Chapter 9 reports the structural dynamics and atomistic perspectives that account for the reported ability of BDQ to halt the ion shuttling ability of mycobacterial c-ring.

Chapter 10 presents the structural dynamics and conformational changes that occur on *Mtb* F1F0 ATP-synthase binding as a result of the triple mutations using molecular dynamics simulations, free energy binding, and residue interaction network (RIN) analyses.

Chapter 11 explored the impact of NAZ2329, a recently identified allosteric inhibitor of Protein Tyrosine Phosphatase Receptor Zeta (PTPRZ), on the atomic flexibility of the WPD-loop, an essential loop in the inhibition of PTPRZ. The study also presents the drug-likeness of NAZ2329 using in silico techniques and its general inhibitory mechanism.

Chapter 12 provides atomistic insights into the structural dynamics and selective mechanisms of AH10-7 for human TCR-CD1d towards activating *i*NKT cells.

The studies in this thesis collectively present a thorough and comprehensive in silico perspective that characterizes the pharmacological inhibition of some known therapeutic targets in viral infections, tuberculosis, and cancer. The augmentative integration of computational methods to provide structural insights could help design highly selective inhibitors of these therapeutic targets. Therefore, the findings presented are fundamental to the design and development of next-generation lead compounds with improved therapeutic activities and minimal toxicities.

DECLARATION 1 – PLAGIARISM

I, **Mr. Clement Agoni**, declare as follows:

That the work described in this thesis has not been submitted to UKZN or other tertiary institution for the purposes of obtaining an academic qualification, whether by myself or any other party.

This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

Their words have been re-written, but the general information attributed to them has been referenced.

Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.

This thesis does not contain text, graphics or tables copied from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references section. A detail contribution to publications that form part and/or include research presented in this thesis is stated (include submitted manuscripts).

Signed:

Date: 02/11/2020

DECLARATION 2 – LIST OF PUBLISHED ARTICLES

1. **Clement Agoni**, Fisayo A. Olotu, Pritika Ramharack and Mahmoud Soliman (2020), "Druggability and Drug-likeness Concepts in Drug Design- are Biomodelling and Predictive Tools having their say?"

Contribution:

Clement Agoni: Main contributor, contributed by literature reviews, writing of manuscripts

Dr. Pritika Ramaharack: Contributed to the manuscript writing and proof reading

Dr. Fisayo A. Olotu: Contributed to the manuscript writing and proof reading

Prof. Mahmoud E. Soliman: Supervisor

2. **Clement Agoni** and Mahmoud Soliman (2020). "The Binding of Remdesivir to SARS-CoV-2 RNA-Dependent Polymerase May Pave The Way Towards the Design of Potential Drugs for COVID-19 Treatment".

Contribution:

Clement Agoni: Main contributor to the project by performing all the experimental work, conducted literature reviews, and writing of manuscript

Prof. Mahmoud E. Soliman: Supervisor

3. **Clement Agoni**, Pritika Ramharack, Geraldene Munsamy and Mahmoud Soliman (2020), "Human Rhinovirus Inhibition through Capsid “Canyon” Perturbation: Structural Insights into the Role of a Novel Benzothiophene Derivative".

Contribution

Clement Agoni: Main contributor by performing all the experimental work, literature reviews, and writing of manuscripts

Dr. Pritika Ramaharack: Contributed to the manuscript writing and proof reading

Dr. Geraldene Munsamy Contributed to the manuscript writing

Prof. Mahmoud E. Soliman: Supervisor

4. **Clement Agoni**, Elliasu Y. Salifu and Mahmoud Soliman (2020). “Dual-Targeting of Human *N*-Myristoyltransferase Subtype 1/2 by IMP-1088 Halts Common Cold Pathogenesis: Atomistic insight”

Contribution

Clement Agoni: Main contributor by performing all the experimental work, literature reviews, and writing of manuscripts

Mr. Elliasu Y. Salifu: Contributed to the manuscript writing

Prof. Mahmoud E. Soliman: Supervisor

5. **Clement Agoni**, Pritika Ramharack, Elliasu Y. Salifu and Mahmoud Soliman (2020). “The Dual-Targeting Activity Of The Metabolite Substrate Of Para-Amino Salicylic Acid In The Mycobacterial Folate Pathway: Atomistic And Structural Perspectives”.

Contribution

Clement Agoni: Main contributor by performing all the experimental work, literature reviews, and writing of manuscripts

Dr. Pritika Ramharack: Contributed to the manuscript writing and proof reading

Mr. Elliasu Y. Salifu: Contributed to the manuscript writing

Prof. Mahmoud E. Soliman: Supervisor

6. Elliasu Y. Salifu, **Clement Agoni**, Fisayo A. Olotu, Yussif M. Dokurugu, Mahmoud E. S. Soliman (2019), “Halting ionic shuttle to disrupt the synthetic machinery – Structural and molecular insights into the inhibitory roles of Bedaquiline (BDQ) towards Mycobacterium tuberculosis ATP synthase in the treatment of tuberculosis”.

Contribution

Mr. Elliasu Y. Salifu: Main contributor by performing all the experimental work, literature reviews, and writing of manuscripts

Clement Agoni: Contributed to manuscript writing, editing and analysis of results

Dr. Fisayo Olotu: Contributed to manuscript proof reading

Dr. Yussif M. Dokurugu: Contributed to manuscript proof reading

Prof. Mahmoud E. Soliman: Supervisor

7. Elliasu Y. Salifu, **Clement Agoni**, Fisayo A. Olotu, Yussif M. Dokurugu, Mahmoud E. S. Soliman (2020), Triple *Mycobacterial* ATP-synthase mutations impede Bedaquiline binding: Atomistic and structural perspectives

Contribution:

Elliasu Y. Salifu: Author- Main contributor, performed literature reviews, system preparation, analyzed data, result interpretation, writing and compilation of manuscripts.

Clement Agoni: Contributed to manuscript writing, editing and analysis of results

Fisayo Olotu: Contributed to manuscript editing and proof reading

Dr. Yussif M. Dokurugu: Contributed to manuscript writing proof reading

Prof. Mahmoud E. Soliman: Supervisor

8. **Clement Agoni**, Pritika Ramharack and Mahmoud Soliman (2018), “Allosteric Inhibition Induces an Open WPD-Loop: A New Avenue Towards Glioblastoma Therapy”.

Contribution

Clement Agoni: Main contributor by performing all the experimental work, literature reviews, and writing of manuscripts

Dr. Pritika Ramaharack: Contributed to manuscript writing and proof reading

Prof. Mahmoud E. Soliman: Supervisor

9. Houda Washah, **Clement Agoni**, Fisayo A. Olotu, Geraldene Munsamy, and Mahmoud E. S. Soliman (2020), "Tweaking α -galactoceramides: Probing the dynamical mechanisms of improved recognition for invariant natural killer T-cell receptor in cancer immunotherapeutics", Current Pharmaceutical Biotechnology".

Contribution

Houda Washah: Contributed by performing all the experimental work, literature reviews, and writing of manuscripts

Clement Agoni: Contributed to analysis of results and writing of manuscripts

Dr. Fisayo Olotu: Contributed to manuscript proof reading

Geraldene Munsamy: Contributed to manuscript proof reading

Prof. Mahmoud E. Soliman: Supervisor

RESEARCH OUTPUT

1. **Clement Agoni**, Fisayo A. Olotu, Pritika Ramharack and Mahmoud Soliman (2020), "Druggability and Drug-likeness Concepts in Drug Design- are Biomodelling and Predictive Tools having their say?" *Journal of Molecular Modeling*, **26(6):120**
2. **Clement Agoni** and Mahmoud Soliman (2020). "The Binding of Remdesivir to SARS-CoV-2 RNA-Dependent Polymerase May Pave the Way Towards the Design of Potential Drugs for COVID-19 Treatment". *Current Pharmaceutical Biotechnology*, **(in press)**

APPENDIX A: Supplementary Documents

3. **Clement Agoni**, Pritika Ramharack, Geraldene Munsamy and Mahmoud Soliman (2020), "Human Rhinovirus Inhibition through Capsid "Canyon" Perturbation: Structural Insights into the Role of a Novel Benzothiophene Derivative". *Cell Biochemistry and Biophysics*, **78:3–13**.

APPENDIX B: Supplementary Documents

4. **Clement Agoni**, Elliasu Y. Salifu and Mahmoud Soliman (2020). "Dual-Targeting of Human *N*-Myristoyltransferase Subtype 1/2 by IMP-1088 Halts Common Cold Pathogenesis: Atomistic insight". *Future Virology* **(Submitted) Manuscript ID FVL-2020-0274**

APPENDIX C: Supplementary

5. **Clement Agoni**, Pritika Ramharack, Elliasu Y. Salifu and Mahmoud Soliman (2020). The Dual-Targeting Activity of The Metabolite Substrate of Para-Amino Salicylic Acid In The Mycobacterial Folate Pathway: Atomistic And Structural Perspectives. *The Protein Journal* **39:106–117**.

APPENDIX D: Supplementary Documents

6. Elliasu Y. Salifu, **Clement Agoni**, Fisayo A. Olotu, Yussif M. Dokurugu, Mahmoud E. S. Soliman (2019), “Halting ionic shuttle to disrupt the synthetic machinery – Structural and molecular insights into the inhibitory roles of Bedaquiline (BDQ) towards Mycobacterium tuberculosis ATP synthase in the treatment of tuberculosis” through an Individualistic Antifolate Approach. *Journal of Cellular Biochemistry*, **120(9):16108-16119**.

APPENDIX F: Supplementary Documents

7. Elliasu Y. Salifu, **Clement Agoni**, Fisayo A. Olotu, Yussif M. Dokurugu, Mahmoud E. S. Soliman (2020), Triple *Mycobacterial* ATP-synthase mutations impede Bedaquiline binding: Atomistic and structural perspectives. *Computational Biology and Chemistry*, **85:107204**.

APPENDIX G: Supplementary Documents

8. **Clement Agoni**, Pritika Ramharack and Mahmoud Soliman (2018), Allosteric Inhibition Induces an Open WPD-Loop: A New Avenue Towards Glioblastoma Therapy. *RSC Advances*, **8:40187-40197**.

APPENDIX H: Supplementary Documents

9. Houda Washah, **Clement Agoni**, Fisayo A. Olotu, Geraldene Munsamy, and Mahmoud E. S. Soliman (2020), "Tweaking α -galactoceramides: Probing the dynamical mechanisms of improved recognition for invariant natural killer T-cell receptor in cancer immunotherapeutics". *Current Pharmaceutical Biotechnology* (**in press**)

APPENDIX I: Supplementary Documents

OTHER RESEARCH OUTPUT

1. Kehinde F. Omolabi, **Clement Agoni**, Fisayo A. Olotu and Mahmoud E. S. Soliman (2020), "Molecular Basis of P131 Cryptosporidial-IMPDH Selectivity – A Structural, Dynamical and Mechanistic Stance", Cell Biochemistry and Biophysics, (*in print*)
2. Depika Dwarka, **Clement Agoni**, John Jason Mellem, Mahmoud E. Soliman and Himansu Baijjanath (2020), "Identification of potential SARS-CoV-2 inhibitors from South African medicinal plant extracts using molecular modelling approaches", South African Journal of Botany, 133:273-284.
3. Ransford Kumi Oduro, **Clement Agoni**, Abdul Rashid Issahaku, Olotu Fisayo and Mahmoud Soliman (2020) "'Polymorphism-Aided' Selective Targeting And Inhibition of Caspase-6 By a Novel Allosteric Inhibitor Towards Efficient Alzheimer's Disease Treatment" Therapy", Cell Biochemistry and Biophysics, 78(3):291-299.
4. Fisayo A. Olotu, **Clement Agoni**, Opeyemi Soremekun and Mahmoud E.S. Soliman (2020), "An update on the pharmacological usage of curcumin: has it failed in the drug discovery pipeline?", Cell Biochemistry and Biophysics, 78(3):267-289.
5. Iwuchukwu A. Emmanuel, **Clement Agoni**, Fisayo A. Olotu, Mahmoud E. S. Soliman (2020), "In Silico Repurposing of J147 For Neonatal Encephalopathy Treatment: Exploring Molecular Mechanisms of Mutant Mitochondrial ATP Synthase", Current Pharmaceutical Biotechnology (*in print*).
6. Fisayo A. Olotu, **Clement Agoni**, Opeyemi Soremekun and Mahmoud E.S. Soliman (2020), "The recent application of 3D-QSAR and docking studies to novel HIV-protease inhibitor drug discovery". Expert Opinion on drug discovery, 15(9):1095-1109.
7. John Oludele Olanlokun, Fisayo A. Olotu and Oyeboode Thomas Idowu, **Clement Agoni**,

- Moses Oluwole, Mahmoud Soliman and Olufunso Olabode Olorunsogu (2020), "In vitro, in silico studies of newly isolated tetrahydro-4-hydroxy-10-methoxy-6, 14-dimethyl-15-m-tolypentadec-13-enyl) pyran-2-one and isobutyryl acetate compounds from *Alstonia boonei* stem bark", *Journal of Molecular Structure*, 1216:128225.
8. Abdul Rashid Issahaku, **Clement Agoni**, Ransford Kumi, Olotu Fisayo and Mahmoud Soliman (2020), "Lipid-embedded molecular dynamics simulation model for exploring the reverse Prostaglandin D2 Agonism of CT-133 towards CRTh2 in the Treatment of Type-2 Inflammation Dependent Diseases", *Chemistry and Biodiversity*, 17(3):e1900548.
 9. Ransford Kumi Oduro, Opeyemi Soremekun Abdul Rashid Issahaku, **Clement Agoni**, Olotu Fisayo and Mahmoud Soliman (2020) "Exploring the Rings Potential of 2, 4-Diaminopyrimidine Derivatives Towards the Identification of Novel Caspase-1 Inhibitors in Alzheimer's Disease Therapy", *Journal of Molecular Modeling*, 26(4):68.
 10. **Clement Agoni**, Elliasu Y. Salifu, Geraldene Munsamy, Fisayo A. Olotu, Mahmoud Soliman (2019), "CF3-pyridinyl substitution on anti-malarial therapeutics: Probing differential ligand binding and dynamical inhibitory effects of a novel triazolopyrimidine-based inhibitor on *Plasmodium falciparum* Dihydroorotate dehydrogenase", *Chemistry and Biodiversity*, 16(12): e190036.
 11. Ransford Kumi Oduro, Abdul Rashid Issahaku, Opeyemi Soremekun, **Clement Agoni**, Olotu Fisayo and Mahmoud Soliman (2019), "From the Explored to the Unexplored: Computer-Tailored Drug Design Attempts in the Discovery of Selective Caspases Inhibitors", *Combinatorial Chemistry & High Throughput Screening*, 22(7):432-444.
 12. Xylia Q. Peters, Thembeke H. Malinga, **Clement Agoni**, Olotu Fisayo, Mahmoud Soliman

- (2019), "Zoning in on Tankyrases: A brief review on the past, present and prospective studies", *Anti-Cancer Agents in Medicinal Chemistry*, 19(16):1920-1934.
13. Iwuchukwu A. Emmanuel, Fisayo A. Olotu, **Clement Agoni**, Mahmoud E. S. Soliman (2019), "Broadening the horizon: Integrative pharmacophore-based and cheminformatics screening of novel chemical modulators of mitochondria ATP synthase towards interventive Alzheimer's disease therapy", *Medical Hypothesis*, 130:109277.
14. Opeyemi Soremekun, Fisayo A. Olotu, **Clement Agoni** and Mahmoud Soliman (2019), "Drug Promiscuity: Exploring the Polypharmacology Potential of 1, 3, 6-trisubstituted 1, 4-diazepane-7-ones as an Inhibitor of the 'god father' of Immune Checkpoint", *Computational Biology and Chemistry*, 80:433-440.
15. Elliasu Y. Salifu, **Clement Agoni**, Fisayo A. Olotu, Yussif M. Dokurugu, Mahmoud E. S. Soliman (2019), "Deciphering the canonical blockade of activated Hageman Factor (FXIIa) by Benzamidine in the coagulation cascade: A thorough dynamical perspective". *Chemical Biology and Drug Design*, 94:1905–1918.
16. Iwuchukwu A. Emmanuel, Fisayo A. Olotu, **Clement Agoni**, Mahmoud E. S. Soliman (2019), "Deciphering the "elixir of life": Dynamic perspectives into the allosteric modulation of mitochondrial ATP synthase by J147, a novel drug in the treatment of Alzheimer's disease". *Chemistry and Biodiversity*, 16(6):e1900085.
17. Abdul Rashid Issahaku, **Clement Agoni**, Opeyemi Soremekun, Patrick Appiah Kubi, Ransford Kumi, Olotu Fisayo and Mahmoud Soliman (2019), "Same Target, Different Therapeutic Outcome: The Case of CAY10471 and Fevipiprant on CRTh2 Receptor in Treatment of Allergic Rhinitis and Asthma", *Combinatorial Chemistry & High Throughput Screening*, 22(8):521-533.

18. Aimen K. Aljoundi, **Clement Agoni**, Olotu Fisayo, Mahmoud Soliman (2019), Turning to Computer-Aided Drug Design in the treatment of Diffuse Large B-cell Lymphoma: Has it been helpful? *Anti-Cancer Agents in Medicinal Chemistry*, 19(11):1325-1339.
19. Aljoundi AK, **Agoni C**, Olotu FA, Soliman ES (2019), 'Piperazing' the catalytic gatekeepers: Unraveling the pan-inhibition of SRC kinases; LYN, FYN, and BLK by Masitinib. *Future Medicinal Chemistry*, 11(18):2365-2380.
20. Opeyemi Soremekun, Olotu Fisayo, **Clement Agoni** and Mahmoud Soliman (2019), Recruiting monomer for dimer formation: Resolving the antagonistic mechanisms of novel immune checkpoint inhibitors against Programmed Death Ligand-1 in cancer immunotherapy. *Molecular Simulation*, 45(10):777-789.
21. Geraldene Munsamy, **Clement Agoni** and Mahmoud Soliman (2019) A Dual-Target of Plasmeprin IX and X: Unveiling the atomistic superiority of a core chemical scaffold in Malaria therapy, *Journal of Cellular Biochemistry*, 120(5):7876-7887.
22. Fisayo A. Olotu, **Clement Agoni**, Emmanuel A. Adeniji, Maryam Abdullahi and Mahmoud E.S. Soliman (2019) Probing gallate-mediated selectivity and high-affinity binding of Epigallocatechin gallate: A way-forward in the design of selective inhibitors for anti-apoptotic Bcl-2 proteins, *Applied Biochemistry and Biotechnology*, 187(3):1061-1080.
23. Fisayo A. Olotu, **Clement Agoni**, Emmanuel A. Adeniji, Maryam Abdullahi and Mahmoud E.S. Soliman (2019) Probing gallate-mediated selectivity and high-affinity binding of Epigallocatechin gallate: A way-forward in the design of selective inhibitors for anti-apoptotic Bcl-2 proteins, *Applied Biochemistry and Biotechnology*, 187(3):1061-1080.

24. Ali Rabbad, **C. Agoni**, A. Olotu and Soliman M. E (2019), Microbes, not humans: Exploring the molecular basis of Pseudouridimycin selectivity towards bacterial and not human RNA polymerase, *Biotechnology Letters*,41(1):115-128.
25. Maryam Lawal, Fisayo A. Olotu, **Clement Agoni** and Mahmoud Soliman (2018) Exploring the C-terminal tail dynamics: Structural and molecular perspectives into the therapeutic activities of novel CRMP-2 inhibitors, Naringenin and Naringenin-7-O-glucuronide, in the treatment of Alzheimer's disease, *Chemistry and Biodiversity*-15(12):e1800437.
26. Shama Khan, Imane Bjjj, Emmanuel A. Adeniji, Fisayo Olotu, **Clement Agoni** and Mahmoud E.S. Soliman (2018) Covalent Simulations of Covalent/Irreversible Enzyme Inhibition in Drug Discovery – A Reliable Technical Protocol, *Future Medicinal Chemistry* - 10(19):2265-2275.

SCIENTIFIC CONFERENCES ATTENDED

Antibiotic Stewardship and Conservation in Africa held at the University of KwaZulu-Natal, Durban, South Africa, 20-23 October 2019

Role: Participant

CHPC National Conference, 2-5 December 2018, Century City Hotel and Conference Centre, Cape Town, South Africa

Role: Poster presentation on “Allosteric inhibition induces an open WPD-loop: a new avenue towards glioblastoma therapy”

Signed:

Date: 02/11/2020

DEDICATION

This thesis is dedicated to my parents; Moses and Mary Agoni and my brothers; Isaac, Jacob, Abraham and Daniel Agoni for their love and support throughout my studies. They have been my pillar of strength during the challenging times of my studies.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and a gigantic thanks to my supervisor **Prof. Mahmoud E. Soliman** for being so supportive, listening, giving the best advice, guidance, patience and moral support during the course of my degree. He taught me so much, not only in the field pharmaceutical chemistry but in the journey of life with his periodic motivational speeches.

My thanks also go out to, Dr. Pritika Ramharack and Dr Fisayo A. Olotu. Your mentorship, guidance and technical support made me confident of myself as I journeyed through my doctoral studies. It was a privilege working with you.

A big thank you to the UKZN Molecular Bio-computation and drug design research group (2017/2020 group) for sharing their research knowledge with me.

You all are highly appreciated. It's been wonderful working with you all.

To the friends I made along this journey that have positively contributed to my progress at one stage or another, I am eternally grateful.

My profound gratitude goes to Centre for High Performance Computing, South Africa for their resources and technical support.

My appreciation also goes to UKZN College of Health Sciences for the financial support throughout the course of my study.

TABLE OF CONTENTS

ABSTRACT	i
DECLARATION 1 – PLAGIARISM	ix
DECLARATION 2 – LIST OF PUBLISHED ARTICLES	x
RESEARCH OUTPUT	xv
OTHER RESEARCH OUTPUT	xviii
DEDICATION	xxiii
ACKNOWLEDGEMENTS	xxiv
TABLE OF FIGURES	xxviii
LIST OF TABLES	xl
LIST OF ABBREVIATIONS	xli
LIST OF TOOLS, LIBRARIES AND WEB SERVERS	xliv
CHAPTER 1	1
1.0 Introduction	1
1.1 Background and rational of study	1
1.2 Aims and objectives of this study	3
1.3 Novelty and significance of study	11
REFERENCE	12
CHAPTER 2	14
2.0 Background on the Application of Computer Drug Design in Drug Discovery	14
2.1 Introduction to Computer-Aided Drug Design	14
2.2 Genesis and Chronological Evolution of Computer-Aided Drug Designs	21
2.3 Classification of Computer-Aided Drug Design	22
2.4 Challenges of the application of CADD in drug design development	26
2.6 Application of CADD in Developing Therapeutics for Viral Infections	27
2.7 Application of CADD in developing therapeutics for tuberculosis	33
2.8 Application of CADD in developing therapeutics for cancer	41
Reference	48
CHAPTER 3	78
3.0 Principles of Computational Chemistry and Molecular Modeling Techniques as Employed in Understanding the Interaction Dynamics of Biological Molecules with Therapeutic Agents	78
3.1 Introduction	78
3.2 Quantum Mechanics	80
3.3 Molecular Mechanics	87

3.4 Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM)	91
3.5 Molecular Dynamics Simulations	91
3.6 Force Fields	95
3.7 Post-Molecular Dynamics Analytical Methods	96
Reference	103
CHAPTER 4	116
Published Article	116
CHAPTER 5	151
Published Article	151
9.0 CONSENT FOR PUBLICATION	184
CHAPTER 6	196
Published Article	196
CHAPTER 7	223
Submitted Article	223
CHAPTER 8	255
Published Article	255
CHAPTER 9	282
Published Article	282
CHAPTER 10	310
Published Article	310
CHAPTER 11	340
Published Article	340
CHAPTER 12	371
Published Article	371
8.0 Consent for Publication	397
CHAPTER 13	405
13.0 Conclusion and Future Perspectives	405
13.1 Conclusions	405
13.2 Future Scope and beyond	410
APPENDIX	411
APPENDIX A	411
APPENDIX B	416
APPENDIX C	419
APPENDIX D	423
APPENDIX E	424

APPENDIX F 428
APPENDIX G 431

TABLE OF FIGURES

Figure 2.1: Schematic showing the different phases of the drug discovery process	16
Figure 2.2: In silico Computer-aided drug design	17
Figure 2.3: 2D structures of some small molecule inhibitors developed through CADD approaches	20
Figure 2.4: Schematic highlighting the classification of CADD.....	24
Figure 2.5: Linear structure of RNA Dependent RNA Polymerase (RdRp) showing the N-terminal domain and the main polymerase domain showing sub-domains (Finger, Palm, Thumb as well a ribbon diagram showing important domain in RdRp	30
Figure 2.6: Structure of icosahedral viral capsid demonstrating the unique VP1-VP3 domain orientations and the characteristic “canyon” drug-binding pocket	32
Figure 2.7: A) Human N-myristoyltransferase (NMT1) with Myristoyl-CoA and IMP-1088 inhibitor bound (PDB:5MU6). B) Human N-myristoyltransferase isoform 2 (NMT2) (PDB:4C2X).....	33
Figure 2.8: The Folate Biosynthetic pathway in <i>Mtb</i>	39
Figure 2.9: Structure of <i>Mtb</i> ATP synthase modelled from the structure of <i>E. Coli</i> ATP synthase. Structure highlights the various subunits	40
Figure 2.10: Structural and corresponding sequence representation of the unique regions of the PTPRZ protein; active site-red, allosteric site- green and WPD-loop- yellow	45
Figure 2.11: Mode of glycolipid presentation by CD1d to NKT cells	47

Figure 3.1: The scientific domains in which Applications of Quantum and Molecular Mechanics fit into.....	80
Figure 3.2: Schematic showing the Application of QM Methods to Biological Problems.CSP: Chemical Shift Perturbation.....	81
Figure 3.3: Schrödinger’s 1927 model of an atom.....	84
Figure 3.4: A graphical representation of a two-dimensional potential energy surface.....	87
Figure 3.5: Diagrammatical depiction of bonded and non-bonded interactions acting in molecular motion.....	89
Figure 3.6: Schematic description of basic MD algorithm where potential energy is E_{pot} ; simulation time = t , iteration time is given by dt ; x is atom coordinates; force component is F ; a is given by acceleration; atomic mass is m and v represents velocity.....	94
Figure 3.7: Flow chart showing steps involved in MD simulation of a drug-target complex.....	95
Figure 4.1: Methods for assessing the druggability of potential disease targets.....	126
Figure 5.1: Diagram of a Coronavirus virion structure.....	155
Figure 5.2: Comparative sequence alignment of the constructed homology model of SARS-CoV-2 RNA-dependent RNA polymerase and the accompanying template employed. Insert highlights the active site residue, VAL557 as previously reported	159
Figure 5.3: 3D representation of the SARS-CoV-2 RNA-dependent RNA polymerase model employed in this study, which was pruned to minimize computational cost. Also present in the structure are Zinc ions (purple). Interacting active site residues are also shown in black.....	162

Figure 5.4: A) Ramachandran plots for the built homology model SARS-CoV-2 RNA-dependent RNA polymerase created on the RAMPAGE online webserver. B) Overall model quality of the homology model of SARS-CoV-2 RNA-dependent RNA polymerase by Z-score of ProSA. C) Overall model quality of the homology model of SARS-CoV-2 RNA-dependent RNA polymerase by Z-score of ProSA.....166

Figure 5.5: Binding poses of Remdesivir in the active sites of SARS-CoV-2 RNA-dependent RNA polymerase and their accompanying docking score. Best binding pose of Remdesivir has a docking score of -5.9kcal/mol.....169

Figure 5.6: Comparative conformational analyses of the unbound (black) and Remdesivir-bound (red) SARS-CoV-2 RNA-dependent RNA polymerase model. [A] C- α RMSD plot showing the degree of atomistic deviations and instability. [B] C- α RMSF plot showing individual residue flexibility. Blue highlight shows the SARS-CoV-2 RNA-dependent RNA polymerase active site region and corresponding 3D representation is shown in the blue inset. Comparative radius of gyration plots of the unbound (black) and Remdesivir-bound (red) SARS-CoV-2 RNA-dependent RNA polymerase mode. [C] C- α ROG plot showing the degree of compactness of SARS-CoV-2 RNA-dependent RNA polymerase. [D] Comparative SASA plots of the unbound (black) Remdesivir-bound (red) SARS-CoV-2 RNA-dependent RNA polymerase model.....171

Figure 5.7: Molecular visualization of Remdesivir at the active sites (hydrophobic pockets) of SARS-CoV-2 RNA-dependent RNA polymerase. Inter-molecular interactions between Remdesivir and active site residues in SARS-CoV-2 RNA-dependent RNA polymerase at 20 ns, 40 ns and 60ns are shown in a, b and c respectively. A1, B1 and B1 and C1 shows a 3D

representation of the hydrophobicity of the Remdesivir bound at SARS-CoV-2 RNA-dependent active site.....175

Figure 5.8: Molecular visualization of Remdesivir at the active sites (hydrophobic pockets) of SARS-CoV-2 RNA-dependent RNA polymerase. Inter-molecular interactions between Remdesivir and active site residues in SARS-CoV-2 RNA-dependent RNA polymerase at 80 ns and 100ns are shown in a and b respectively. a1 and b1 shows a 3D representation of the hydrophobicity of the Remdesivir bound active site.....176-177

Figure 5.9: A) 3D homology model of model of the SARS-CoV-2 RNA dependent RNA polymerase complexed with Remdesivir. B) Per-residue energy contributions plot of the interacting residues at the SARS-CoV-2 RNA dependent RNA polymerase active site upon Remdesivir binding. C) The residue ligand interaction network illustrates stabilizing hydrophobic interactions pocketing Remdesivir at the active site from an averaged structure.....179

Figure 5.10: A) Specific Remdesivir moieties (the 3,4-dihydroxy-5-methyloxolane-2-carbonitrile moiety, pyrrolo[2,1-f][1,2,4]triazin-4-amine moiety and (1-methoxy-1-oxopropan-2-yl)(methyl)aminophosphonic acid) that interact with high energy contributing residues. B) 3-D pharmacophore model generated on ZINCPharmer (Green–hydrophobic/aromatic moiety, gold – hydrogen bond donor/acceptor)182

Figure 6.1: Structure of icosahedral viral capsid demonstrating the unique VP1-VP3 domain orientations and the characteristic “canyon” drug-binding pocket.....199

Figure 6.2: Root mean square fluctuations of HRV Capsid protein. Increased fluctuations are seen in the inhibitor bound conformations.....206

Figure 6.3: Residue Interaction Network of HRV-14 active site residues upon binding of compound 6g, Pleconaril and when unbound. The network shows an extensively tight and compact active site residue network in the unbound conformation which becomes distorted residue network with reduced interactions upon binding of both compound 6g and Pleconaril.....209

Figure 6.4: Plot of the estimated distances between the north and south wall as well as a visual representation of the flexibility of walls at the base of canyon region of HRV-B14 (orange), HRV-B14-Pleconaril (purple) and HRV-B14-compound 6g (pink).....212

Figure 6.5: Per-residue decomposition analyses revealing individual energy contributions of HRV-B14 active site residues to the binding and stability of compound 6g and Pleconaril. Active site residue that contributed the most to binding of both inhibitors was Tyr152.....215

Figure 7.1: Schematic showing of the crucial role of *N*-myristoyl transferase in the viral life cycle. This also highlights how infectious Rhinovirus polyproteins are generated in the cells of an infected host226

Figure 7.2: Showing the sequence alignment of binding sites residues of HsNMT1 and HsNMT2 and a 2D structure of compound IMP-1088, highlighting essential moieties. Sequence alignment reveals a similarity in binding site residue of both enzyme subtypes.....232

Figure 7.3: A) $C\alpha$ RMSD plot showing comparative stability and motions of IMP-1088 at the binding site HsNMT1 (blue) and HsNMT2 (red) over the 100ns simulation B) The differential positioning of IMP-1088 in the binding site of HsNMT2 and a structural comparison of the initial (yellow), optimized (purple) and final snapshot at 100 ns (blue) C) The differential positioning of

IMP-1088 in the binding site of HsNMT1 and a structural comparison of the initial (yellow), optimized (purple) and final snapshot at 100 ns.....233

Figure 7.4: Comparative time-based analysis of the interactions of IMP-1088 with the binding pocket of HsNMT1 and HsNMT2 with particular emphasis on interaction with Tyr296 chain. Shown in the table are individual residues engaged in various intermolecular interactions with IMP-1088. Residues highlighted in red are those that consistently interacted with both HsNMT1 and HsNMT2 throughout the simulation235

Figure 7.5: Per-residue energy plot of residues that consistently interacted with both HsNMT1 and HsNMT2 over the course of the 100ns MD simulation and their corresponding ligand-residue interaction plots highlighting specific interactions involved.....237

Figure 7.6: Characteristic secondary structure fluctuation among the unbound (blue) and IMP-1088-bound (red) HsNMT1/2. A) and C) RMSF plot showing regions with notable fluctuation which are indicated numerically. B) and D) Superimposed structures of the bound and apo conformation HsNMT1 and HsNMT2 showing regions with high fluctuations as indicated in 6A.....241

Figure 8.1: Schematic of the folate biosynthetic pathway.....258

Figure 8.2: Structural positioning and orientation of PAS-M at the hydrophobic grooves of DHFR depicting residues within the binding pockets at 10 ns [a], 50 ns [b] and 100 ns [c]. Hydrophobic surface representations of PAS-M-bound within the DHFR binding pockets (blue), highlighting hydrogen bond interacting residues at 10 ns [a1], 50 ns [b1] and 100 ns [c1]266

Figure 8.3: Hydrophobic surface representations of PAS-M-bound within the FDTS binding pockets (blue), highlighting hydrogen bond interacting residues at 10 ns [a], 50 ns [b] and 100 ns

[c]. Structural positioning and orientation of PAS-M at the hydrophobic grooves of FDTS depicting residues within the binding pockets at 10 ns [a1], 50 ns [b1] and 100 ns [c1]267

Figure 8.4: Per-residue decomposition plots showing individual energy contributions to the binding and stabilization of PAS-M at the hydrophobic grooves of DHFR[a] and FDTS [a1]. This reveals that energy contributions were highest in residues that formed hydrogen bond interactions with PAS-M. Corresponding inter-molecular interactions are shown in DHFR [b] and FDTS [b1].....269

Figure 8.5: Comparative C- α RMSD plots showing the degree of instability and disruption induced by PAS-M at the binding pocket of DHFR and FDTS. Also shown is the binding site superposition of unbound (violet) and PAS-M-bound (red) forms of the two enzymes.....271

Figure 8.6: A) Comparative compactness of PAS-M binding site residues of FDTS in the presence and absence of PAS-M. Binding site residues in the unbound FDTS were more compact relative to the PAS-M bound conformation with averagely lower RoG of 8.26 Å. B) Comparative compactness of PAS-M binding site residues of DHFR in the presence and absence of PAS-M. Binding site residues in the unbound DHFR were more compact relative to the PAS-M bound conformation with averagely lower RoG of 8.24 Å273

Figure 9.1: [A] Surface depiction of the mycobacterial c-ring F1F0-ATP synthase (green) bound by BDQ (grey). Also, active site and BDQ-interacting residues are depicted in yellow coloration and clearly shown in red inset. [B] 2D structure of BDQ showing constituent functional groups.....287

Figure 9.2: Conformational dynamics of F1F0 ATP synthase upon binding of BDQ. A) Plot of the RMSD showing a gradual increase in RMSD and subsequent steadying of the systems after about 20ns. B) RMSF showing the comparative fluctuations of Glu65B in the bound (BDQ) and the unbound (Apo) systems. Note: The scale used is from tleap numbering of residues, F1F0 of M. phlei has three chains of repeated residues as such the RMSF plot shows these chains in brackets and the position of the Glu65B in the two systems. C) Scatter plots showing fluctuations of the binding site residues among the bound and unbound systems. Indicated in green (unbound) and red (BDQ-bound) circles are the main residues (Glu65B) responsible for ion exchange in both systems D) Radius of gyration (Rg) of the binding site residues showing a low Rg due to the binding of BDQ resulting in a highly compact F1F0 structure.....294

Figure 9.3: Structural positioning and orientation of BDQ at the hydrophilic c rotor-stator region of F1F0 synthase BLK depicting steady, consistent and recurrent interaction of BDQ with the c-ring residues. Surface representations of BDQ-bound c-ring in the F1F0 synthase are shown in [. The degree of H-bonds ranging from H-bond donor (→ magenta) to H-bond acceptor (→ light green)297

Figure 9.4: Per-residue decomposition plots showing individual energy contributions to the binding and stabilization of BDQ at the c-ring of F1F0 synthase. Energy contributions were highest in conserved carboxylates (Glu65B, Asp32B) and Phe69B. Corresponding inter-molecular interactions between BDQ and the hydrophilic c rotor-stator region residues in F1F0 synthase is shown as well.....299

Figure 10.1: 3-D structure of Mtb F1F0 ATP synthase (light blue) bound by BDQ (grey) at the active site (yellow) and showing the three mutations 59(Leu → Val), 61(Glu → Asp) and 66(Ile → Met).....317

Figure 10.2: Per-residue decomposition plots showing individual energy contributions to the binding of BDQ with much emphasis on the residues in the wild ATP synthase-BDQ system and the corresponding mutants in the mutant ATP synthase-BDQ system. Corresponding intermolecular interactions between the wild and mutant systems is shown as well.....322

Figure 10.3: Residue Interaction network analysis of wild Mtb F1F0 ATP synthase in complex with BDQ. (A) Circled insert shows hotspot residues (Lue59A, Glu61B and Ile66A) which are subsequently mutated. (B) Network and topology representation highlighting both hotspot residues with the corresponding interacting neighboring residues. (C) Closer network view showing the residual interaction network and centrality of the hotspot residues.....325

Figure 10.4: Residue Interaction network analysis of mutant Mtb F1F0 ATP synthase in complex with BDQ. (A) Circled insert shows hotspot mutant residues (Val59A, Asp61B and Met66A). (B) Network and topology representation highlighting both hotspot residues with the corresponding interacting neighboring residues. (C) Closer network view showing the residual interaction network and centrality of the hotspot residues.....326

Figure 10.5: (A) Comparative hydrogen bond plots of within active site region of the wild and mutant bound Mtb ATP synthase. (B) 3D representation of the degree of hydrogen bond donors (magenta) and acceptors (light green) within the active of site of the wild bound Mtb ATP synthase. (C) 3D representation of the degree of hydrogen bond donors (magenta) and acceptors (light green)

within the active site of the mutant bound *Mtb* ATP synthase.....328

Figure 10.6: A) Comparative C- α RMSD plots wild and mutant BDQ bound *Mtb* ATP synthase. B) RMSF plots showing the degree of residue fluctuations that occurred on both wild and mutant *Mtb* ATP synthase over the simulation period. Inset shows 3D representation of some prominent residue fluctuation regions (region 37-67 and 160-175). C) C- α Rg plot showing distinct structural compactness and rigidity of wild and mutant bound *Mtb* ATP synthase.....332

Figure 11.1: Graphical representation of the WPD-loop (yellow) of PTPRZ. The active form of the protein is associated with a closed WPD-loop upon binding of the natural substrate, whereas, an open conformation of results in an inactive protein.....344

Figure 11.2: Structural and corresponding sequence representation of the unique regions of the PTPRZ protein; active site-red, allosteric site- green and WPD-loop- yellow (PDB Code: 5H08).....352

Figure 11.3: (A) Shows the comparative RMSF plot of the unbound (APO-red) and bound PTPRZ (complexed-blue), (B) Systematic NAZ2329 inhibitory activity in opening WPD-loop across the MD simulation period.....355

Figure 11.4: Cross-correlation matrices of the fluctuations of C- α atoms in (A) APO and (B) NAZ2329-complex.....357

Figure 11.5: PCA projection of the motion of Ca atoms constructed by plotting the first two principal components (PC1 and PC2) in the conformational space with APO (red) and NAZ2329-complex (black) respectively. PC1 and PC2, respectively, represent a covariance matrix after

elimination of eigenvectors (rotational movements). Each point between the single-directional motions represents a unique conformation during the simulation, whereby, similar structural conformations overlap in the graph.....358

Figure 11.6: Energy contributions of the interacting residues at the NAZ2329 active site. The residue ligand interaction network illustrates stabilizing hydrophobic interactions pocketing NAZ2329 at the active site. The highest energy contribution was two hydrogen bond interactions shared between ARG 237 with two separate oxygen moieties of NAZ2329.....362

Figure 11.7: (A) Common pharmacophoric features from alignment of the PTPRZ ligand/residue interaction plots, 3-D pharmacophore model generated on ZINCPharmer (Green-Hydrophobic/Aromatic moiety, gold - hydrogen bond donor/acceptor). (B) 2-D representation of the chemical features required for potential PTPRZ inhibitors.....364

Figure 12.1: 2D structures of AH10-7 (above) and KRN7000 (below) coupled with the 3D X-ray crystal structure of human *i*NKT-TCR (middle). Red inset showcases constituent hydrophobic pocket residues.....376

Figure 12.2: Comparative sequence analysis of AH10-7 and KRN7000-2C12 TCR-mCD1d complexes with corresponding highlights of interacting residues at the hydrophobic pockets. Green inserts indicate residues that directly interact with KRN7000 while red highlights indicate residues that directly interact with AH10-7.....380

Figure 12.3: Per-residue energy decomposition plot of uncommon binding site residues and their differential interactions with AH10-7 and KRN7000, with corresponding ligand interaction plots of average structures.....384

Figure 12.4: Complementary binding site-surface interactions of [A] C6” modified hydrocinnamoyl ester in AH10-7 and [B] “non-modified” KRN7000 galactosyl moiety. Conventional (OH--O) and non-conventional/carbon-hydrogen (CH--O) bonds are also indicated.....386-387

Figure 12.5: Structural superposition of AH10-7 and KRN7000 at the binding pocket of human *i*NKT-TCR to depict their analogous binding modes.....388

Figure 12.6: [A] Comparative time-based monitoring of complimentary interactions of “modified” and “non-modified” regions of AH10-7 and KRN7000, respectively at 10 ns and 40 ns. [B] Comparative time-based monitoring of complimentary interactions of “modified” and “non-modified” regions of AH10-7 and KRN7000, respectively at 70 ns and 100 ns.....392

Figure 12.7: A) Comparative SASA plots of residues that make up the binding pocket of both AH10-7 and KRN7000 as well as the unbound conformation. B) 3D surface representation of the higher SASA of the AH10-7 complex with hydrocinnamoyl ester region showcasing the highest SAS (deep blue). C) Comparative RMSF plots of binding site residues with the unbound *i*NKT-TCR showing the highest residue flexibility. D) 3D representation of superimposed X-ray crystal structures of unbound *i*NKT-TCR (black) and the AH10-7 (red) and KRN7000 (green) bound complexes.....395

LIST OF TABLES

Table 2.1: Highlights of drugs designed using CADD	19-20
Table 2.2: Highlights of some studies that have employed virtual screening in identifying anti-tubercular agents	35-36
Table 4.1: Table 4.1: In silico tools utilized in the prediction of molecule drug-likeness	133-138
Table 5.1: Comparative MM/GBSA-based binding free energy profile of compound Remdesivir.....	168
Table 5.2: Tabular presentation of the estimated energy contribution of individual active residues.....	180
Table 6.1: Measurement of the distance by which the north and south wall moves, measured in Angstroms. (Δa Change between maximum and minimum distance)	211
Table 6.2: MM/PBSA-based binding free energy profiles of compound 6g and Pleconaril.....	213
Table 7.1: MM/GBSA-based binding free energy profile of compound IMP-1088.....	238
Table 7.2: The Summarized ADMET Profile of compound IMP-1088.....	244
Table 8.1: MM/PBSA-based binding free energy profile of PAS-M.....	264
Table 9.1: MM/GBSA-based energy profiles of BDQ-singly and double-bound F1F0 ATP synthase.....	295
Table 10.1: Calculated final decomposition energies for all residues of interest.....	321
Table 10.2: Estimated Binding free energy profiles of BDQ to mutant and wild Mtb F1F0 ATP synthase.....	329
Table 11.1: The Swiss ADME Profile of NAZ2329.....	360
Table 11.2: MM/GBSA-based binding free energy profile of the simulated NAZ2329 complex.....	362
Table 12.1: MM/PBSA-based binding free energy profile of KRN7000 and AH10-7.....	382

LIST OF ABBREVIATIONS

2D	Two Dimension
3D	Three Dimension
AA	All Atom
ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity
ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
BDQ	Bedaquiline
CADD	Computer Aided Drug Design
CG	Coarse Grain
DCCM	Dynamic Cross Correlation Matrix
DDS	Dapsone
DNA	Deoxyribonucleic Acid
DHF	Dihydrofolate
DHFR	Dihydrofolate reductase
DHPS	dihydropteroate synthase
dUMP	2'-deoxyuridine-5'-monophosphate
dTMP	2'-deoxythymidine-5'-monophosphate
EM	cryo-Electron Microscopy
GAFF	General Amber force Field
FBDD	Fragment-Based drug design
FDTS	Flavin-dependent thymidylate synthase
HIV/AIDS	Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome
HNMT	Human <i>N</i> -myristoyltransferase
HRV-B14	Human Rhinovirus B14
HTS	High-Throughput Screening
IGF-1R	Insulin-like Growth Factor-1 Receptor
IKK- β	I-Kappa-B Kinase β
γ NKT	invariant Natural Killer T cells
LBDD	Ligand-Based Drug Design
MC	Monte Carlo
MD	Molecular Dynamics

MERS	Middle East Respiratory Syndrome
mTHF	Methylene tetrahydrofolate
MDR-TB	Multidrug-resistant Tuberculosis
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area
MM/PBSA	Molecular Mechanics/Poisson Boltzman Surface Area
MM	Molecular Mechanics
MFF	Molecular Force Fields
<i>Mtb</i>	<i>Mycobacterium Tuberculosis</i>
NADPH	Nicotinamide Adenine dinucleotide phosphate
NMR	Nuclear Magnetic Resonance
ONIOM	Our own N-layered Integrated Molecular Orbital and molecular mechanics
PAS	para-amino salicylic acid
PAS-M	Metabolite substrate of Para-amino Salicylic Acid
PTML	Perturbation-Theory Machine Learning model
PTP	Protein Tyrosine Phosphatase
PTK	Protein Tyrosine Kinases
PBDD	Pharmacophore-Based Drug Design
PDB	Protein Data Bank
PDBe	Protein Data Bank in Europe
PCA	Principal Component Analysis
PTPRZ	Protein Tyrosine Phosphatase Receptor Zeta
PUMA	p53 Upregulated Modulator of Apoptosis
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
QSAR	Quantitative Structure Activity Relationship
QSAR/QSPR	Quantitative Structure Activity Relationship/ Quantitative structure properties relations
RNA	Ribonucleic Acid
RIN	Residue Interaction Network
RR-TB	Rifampin Resistant tuberculosis
RPTP	Receptor-type Protein Tyrosine Phosphatases
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RoG	Radius of Gyration
SAR	Structure Activity Relationship

SASA	Solvent Accessible Solvent Area
SARSCoV-2 RdRP	SARS-CoV-2 RNA dependent RNA polymerase
SARSCoV	Severe Acute Respiratory Syndrome Coronavirus
SBDD	Structure-Based Drug Design
SHMT	Serine Hydroxymethyl Transferase
SMX	sulfamethoxazole
TB	Tuberculosis
TCR-CD1d	CD1d-restricted T cell Receptor (TCR)
THF	Tetrahydrofolate
TSase	Thymidylate synthase
UA	United Atom
WHO	World Health Organization

LIST OF TOOLS, LIBRARIES AND WEB SERVERS

AutoDockTools	http://autodock.scripps.edu/resources/adt
AutoDock Vina	http://vina.scripps.edu/
Avogadro	https://avogadro.cc/
MODELLER	https://salilab.org/modeller/
Open Babel	http://openbabel.org
PDB	http://www.rcsb.org/
SWISS-MODEL	https://swissmodel.expasy.org/
VMD	https://www.ks.uiuc.edu/Research/vmd/
ZINCpharma	http://zincpharmer.csb.pitt.edu/
Uniprot	https://www.uniprot.org/
PubChem	https://pubchem.ncbi.nlm.nih.gov/
SwissADME	http://www.swissadme.ch/
NCBI	https://www.ncbi.nlm.nih.gov/

CHAPTER 1

1.0 Introduction

1.1 Background and rational of study

The studies presented in this thesis are geared towards exploring drug-target interactions in antiviral, antitubercular, and anticancer therapeutic interventions using Computer-Aided Drug Design (CADD) approaches revealing structural and mechanistic events that could inform the drug design of novel and improved small molecule inhibitors. Drug discovery is multifaceted, comprising a diverse and multidisciplinary approach towards designing a therapeutically effective and commercially viable drug. Due to the complex nature of the drug discovery process, it is usually costly, risky, and generally time-consuming. On average, the development of one drug from discovery until it makes it to the market takes approximately 10-15 years, with an estimated average of \$1.3 billion [1]. As such, viable alternatives such as CADD have widely gained traction in the pharmaceutical industry with its ability to accelerate the drug discovery process while cutting down the cost involved [2–4]. Computer-aided approaches significantly optimize the lead discovery stage of drug development, which is known to majorly contribute to the overall drug discovery expenses due to various lead testing procedures [2–5]. By unraveling the molecular mechanism of drug molecules by exploring drug-target interactions, CADD approaches can also identify potential therapeutic agents against biological targets, thereby minimizing the search space to be investigated by experimental methods. Also, because drug-target interactions culminate into the eventual change in function/behavior of biological targets [6–9], the *in silico* elucidations of these interactions further establish CADD approaches current era of drug

discovery. CADD techniques also seek to augment traditional in vitro and in vivo experimental techniques and not to replace them since CADD approaches can also allow modeling of complex biological processes, which hitherto seemed impossible to explore using experimental methods.

According to the World Health Organization (WHO), featuring prominently in the top ten causes of death are cancer, lower respiratory tract infection, tuberculosis, and viral infections such as HIV/AIDS. Collectively, these diseases are of global health concerns, considering the large number of associated deaths yearly. Over the years, several therapeutic interventions have been employed to treat, manage, or cure these diseases, including; chemotherapy, surgery, and radiotherapy. Of these options, small molecule inhibitors have constituted an integral component in chemotherapy, thereby undoubtedly playing an essential role in patient management. Although significant success has been achieved using existing therapeutic approaches to the emergence of drug resistance, the drug design processes against these diseases remain innovative, including combining existing drugs and establishing improved therapeutic options that could overcome resistance. Therefore, an exploration of drug-target interactions as performed in the reports in this thesis is relevant since target interaction prediction plays a vital role in the drug discovery process, aiming to identify new drug compounds for biological targets.

In all, the structural and conformational elucidations unraveled this thesis coupled with investigated mechanisms of actions of the studied small molecule inhibitors culminates into vital chemical information that can contribute significantly towards the discovery of next-generation lead compounds with improved potency in the treatment of cancer, viral infections, and tuberculosis. Thus, this thesis investigates molecular and structural insights on some selected protein and enzyme targets implicated in the pathogenesis of the studied diseases and therefore are proven therapeutic targets for currently existing chemotherapies. These implicated biological

targets include; SARS-CoV-2 RNA dependent RNA polymerase (SARS-CoV-2 RdRp), Human Rhinovirus B14 (HRV-B14) and human *N*-myristoyltransferase (HNMT) in viral infections, Dihydrofolate reductase (DHFR) and Flavin-dependent thymidylate synthase (FDTS) in tuberculosis, as well as Protein Tyrosine Phosphatase Receptor Zeta (PTPRZ) and human variants of CD1d-restricted T Cell Receptor (TCR-CD1d) in cancer. Moreover, resulting insights are essential towards the structure-based design of highly specific small-molecule antagonists of these biological targets. Also reported herein are strategic optimization methods for enhancing some of the studied small molecule compounds' biological activity to minimize their toxic tendencies

1.2 Aims and objectives of this study

The primary intent of this study is to provide structural mechanistic actions and molecular insights on crucial therapeutic targets coupled with the *in silico* application of strategic drug design methods for optimizing the therapeutic activities of bioderived compounds and improving their pharmacokinetics in relation to their use in the treatment of viral infections, tuberculosis, and cancer. Target proteins investigated in this thesis include SARS-CoV-2 RdRp, HRV-B14, and human NMT in viral infections, DHFR and FDTS in tuberculosis, human variants of TCR-CD1d and PTPRZ in cancer, while the optimization methods for enhancing the biological activity of some of the studied small molecule compounds to minimize their toxic tendencies were also investigated.

The specific goals of this thesis are highlighted below:

1. To provide sufficient and current details on druggability and drug-likeness and principles their recent advancements in the drug discovery field while presenting the different computational tools and their reliability of predictive analysis in drug discovery:

- 1.1 To highlights the principles and methods that underline the assessment of the druggability of biological target
 - 1.2 To highlights and establish the reliability of available druggability predictive methods
 - 1.3 To outline methods for the identification of lead molecules by assessing drug-likeness
 - 1.4 To highlights physiochemical properties and pharmacokinetic properties in determining drug-likeness
 - 1.5 To highlights currently available *in silico* techniques of predicting drug-likeness while buttressing on their reliability in augmenting experimental methods
-
2. To provides an understanding of the binding mechanism of Remdesivir by unraveling the structural and conformational implications on SARS-CoV-2 RdRp upon Remdesivir binding and identify its crucial pharmacophoric moieties:
 - 2.1 To model the structure of SARS-CoV-2 RdRp through homology modeling in the absence of its X-ray crystal structure.
 - 2.2 To perform molecular dynamics simulations on the modelled structure of SARS-CoV-2 RdRp complexed with Remdesivir
 - 2.3 Investigate the conformational and structural changes on SARS-CoV-2

- RdRp upon Remdesivir binding relative to the unbound conformation
- 2.4 To measure the MM/PBSA-based free binding energy of Remdesivir
 - 2.5 To calculate the per-residue energy decomposition of individual amino acid residues that interact Remdesivir
 - 2.6 To calculate the per-residue energy decomposition of individual amino acid residues that interact Remdesivir
 - 2.7 To construct a pharmacophore that highlights the prominent interacting moieties of Remdesivir using ZINCPharma
 - 2.8 To screen the ZINC database for potential SARS-CoV-2 RdRp inhibitors using the generate pharmacophore
3. To provide the molecular and structural mechanism of action of a novel benzothiophene derivative shown to inhibit HRV viral replication:
- 3.1 To obtain the X-ray crystal structures of HRV-B14 capsid
 - 3.2 To model the structure of the novel benzothiophene derivative and subsequently subject it to molecular docking to generate complexes of the inhibitor and the HRV-B14 capsid.
 - 3.3 To perform molecular dynamics simulations on the generated HRV-B14 inhibitor complex and the unbound conformation of HRV-B14
 - 3.4 To explore the impact of the binding of the novel benzothiophene derivative on the structural dynamics of HRV-B14 upon binding to it “canyon region”

- 3.5 To explore the impact of the binding of the novel benzothiophene derivative on the residue interaction network of HRV-B14 upon binding to it “canyon region”
 - 3.6 To explore the impact of the binding of the novel benzothiophene derivative on the motion the North-South Wall of HRV-B14 upon binding to it “canyon region”
 - 3.7 To estimate the binding free energy profile of the the novel benzothiophene relative the known the known capsid inhibitor, Pleconaril
 - 3.8 To quantify the individual energy contributions of binding site residues towards the binding of the novel benzothiophene using per-residue energy decomposition approach
4. To unravels molecular and structural bases behind the dual inhibitory potential of the novel inhibitor IMP-1088 against human *N*-myristoyltransferase (HNMT) subtypes 1/2 toward common cold therapy using augmentative computational and cheminformatics methods. The study also highlights the pharmacological propensities of IMP-1088:
- 4.1 To perform a sequence of alignment of HNMT subtypes 1/2 to assess their sequence similarity
 - 4.2 To obtain the X-ray crystal structures of HNMT subtypes 1/2 and subject to molecular dynamics simulations complexed with IMP-1088

- 4.3 To quantify the individual energy contributions of binding site residues towards the binding of IMP-1088 in both subtypes 1/2 using per-residue energy decomposition approaches
 - 4.4 To estimate the relative conformational and structural changes on both HNMT subtypes upon binding of IMP-1088
 - 4.5 To perform *in silico* physicochemical and ADMET profiling of compound IMP-1088

5. To unravel the molecular mechanisms and structural dynamics that are associated with the dual inhibitory activity of bioactive metabolite of Para-amino Salicylic Acid (PAS-M) towards dihydrofolate reductase (DHFR) and flavin-dependent thymidylate synthase (FDTS):
 - 5.1 To highlight the crucial role of DHFR and FDTS as therapeutic targets in tuberculosis treatment
 - 5.2 To obtain the X-ray crystal structures of DHFR and FDTS and subject to molecular dynamics simulations complexed with PAS-M
 - 5.3 To measure the MM/PBSA free binding energy of PAS-M and per-residue interactions of interacting residues upon binding with both DHFR and FDTS.
 - 5.4 To assess the concomitant effects of PAS-M binding on the overall structural stability of DHFR and FDTS.

- 5.5 To reflect the dynamical interactions elicited by the PAS-M at the hydrophobic grooves (binding sites) of DHFR and FDTS and identify crucial residues involved in high-affinity binding.
 - 5.6 To apply appropriate metrics to evaluate the structural dynamics of the hydrophobic deep pockets of DHFR and FDTS upon binding of PAS-M
6. To describe the structural dynamics that explicate the experimentally reported antagonistic features of Bedaquiline (BDQ) in halting ion shuttling by the mycobacterial c-ring by binding *Mycobacterium tuberculosis* ATP synthase:
- 6.1 To obtain the X-ray crystal structure *Mycobacterium tuberculosis* ATP synthase complexed with BDQ and subject the complex to molecular dynamics simulations
 - 6.2 To investigate the dynamical and corresponding structural effects of the binding of BDQ on the primary structure of *Mtb* ATP synthase

- 6.3 To investigate how dual molecule or single molecule binding of BDQ on *Mtb* ATP synthase influence the inhibitory potency of BDQ.
 - 6.4 To perform a time-bound analysis of the interaction of BDQ at the c-ring F1F0 synthase
 - 6.5 To quantify the energetics of the individual residues to the total binding affinity of BDQ with particular emphasis on the electrostatic and van der Waal energy contributions using MM/PBSA calculations
7. To explore the impact of triple mutations on F1F0 structure and the consequential effect on the binding of BDQ:
- 7.1 Model the non-existent crystal structure of F1F0 subunit C of *Mtb* using online modelling tools such as SwissModel
 - 7.2 To generate a mutant *Mtb* F1F0 by substituting these three residues 59(Leu → Val), 61(Glu → Asp) and 66(Ile → Met) in its amino acid sequence.
 - 7.3 To perform a 150ns MD simulation on the mutant system and also a wild system with BDQ binding. In all two systems resulted including Mutant-BDQ and Wild-BDQ systems
 - 7.4 To assess the systems stability and other conformational dynamics such as flexibility and compactness.

- 7.5 To perform a Residue Interaction Network to assess the effect of mutation on the binding of BDQ
 - 7.6 To employ post-MD analysis such MM/GBSA based binding free energy calculation to estimate the binding affinity of BDQ with and without the triple mutation
 - 7.7 To conduct Hydrogen bond analysis on the simulated systems to ascertain the impact of the mutation of hydrogen formation
8. To explore the structural mechanism of inhibition of Protein Tyrosine Phosphatase Receptor Zeta (PTPRZ) by NAZ2329, a novel allosteric inhibitor:
 - 8.1 To highlight the role of PTPRZ as a therapeutic target in glioblastoma
 - 8.2 To prepare PTPRZ complexed with NAZ2329 and followed by molecular dynamics simulation studies to investigate the interaction dynamics over the simulation period
 - 8.3 To assess the impact of NAZ2329 binding on the structural stability of PTPRZ
 - 8.4 To monitor the structural and conformational dynamics of the WPD-loop upon the binding of order NAZ2329.
 - 8.5 To assess the physicochemical and drug-likeness of NAZ2329
9. To gain atomistic insights into the structural dynamics and selective mechanisms of AH10-7 against human variants of TCR-CD1d toward activating invariant natural killer T (*i*NKT):

- 9.1 To highlight the pharmacological ability of α -galactosylceramide derivatives as immunotherapeutic agents that in cancer treatment
- 9.2 To retrieve the X-ray crystal structure of 2C12 TCR-CD1d in complex with AH10-7 and KRN7000 and subject them to molecular dynamics simulations
- 9.3 To investigate the mechanistic selective binding affinity mediated by the incorporated hydrocinnamoyl ester on the “C6” of the sugar of AH10-7 using MM/PBSA calculations
- 9.4 To investigate the atomistic binding profiles of dually modified α -GalCer with individual hydrophobic pocket residues through per-residue energy decomposition
- 9.5 To explore how the dual sphinganine base and hydrocinnamoyl ester modification of α GalCer enhances binding interactions with TCR-CD1d
- 9.6 To examine the conformational changes on human TCR-CD1d upon binding of AH10-7 and KRN7000

1.3 Novelty and significance of study

Despite the availability of several therapeutic interventions against viral infections, tuberculosis, and cancer, these conditions persist, with these diseases featuring prominently in the top ten causes of death globally. Their persistence as global health concerns has been due to incidences of chemotherapeutic resistance and associated toxic effects over the years. Reports in this thesis implement several computational techniques towards augmenting experimental methods as an

avenue of potentially accelerating the drug discovery process of novel inhibitors that overcome challenges associated with existing conventional therapeutic remedies against these diseases. By unraveling the molecular mechanism of small molecule inhibitors against therapeutic targets investigated in this study, the employed CADD approaches reveal crucial atomistic and structural insights that could aid in designing small molecule inhibitors against the respective biological targets that possess improved therapeutic properties. This could also reduce the search space to be investigated in laboratory experiments while leading to the discovery of lasting therapeutic solutions since the biological targets investigated in this thesis are prominently implicated in the studied diseases' pathogenesis. With the use of CADD techniques, a comprehensive *in silico* perspective is offered to shed light on possible structural characteristics that allow for the inhibition of these implicated therapeutic targets while also revealing the amino acid residues implicated in the biological functions of the targets. Also, by defining these therapeutic targets' binding landscape, the reports in this thesis will present a prospective design of selective and unique inhibitors with critical pharmacophoric features that will aid in developing targeted and effective small molecule inhibitors towards overcoming chemotherapeutic resistance and minimizing toxicity effects.

Reference

1. Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA*. 2020;323:844–853. Available from: <https://doi.org/10.1001/jama.2020.1166>.
2. Usha T, Shanmugarajan D, Goyal AK, Kumar CS, Middha SK. Recent Updates on Computer-aided Drug Discovery: Time for a Paradigm Shift. *Curr Top Med Chem*. 2018;17:3296–

3307.

3. Prieto-Martínez FD, López-López E, Eurídice Juárez-Mercado K, Medina-Franco JL. Computational Drug Design Methods—Current and Future Perspectives. *Silico Drug Des.* 2019;3:19–44.
4. Siju E, Rajalakshmi G, Paulose A, Dhanya F, Hariraj N, Rahul K. CADD: pharmacological approaches in drug design and drug discovery. *World J Pharm Pharm Sci.* 2017;6:892–908.
5. Kalita JM. Advances In Computer Aided Drug Design. *Univ J Pharm Res.* 2015;1:17-22.
6. Settle W, Hegeman S, Featherstone RM. The Nature of Drug-Protein Interaction. In: *Concepts in Biochemical Pharmacology. Handbuch der experimentellen Pharmakologie/Handbook of Experimental Pharmacology.* Brodie BB, Gillette JR, Ackerman HS (Eds.), Springer, Berlin, Heidelberg. 1971.
7. Krasner J. Drug-Protein Interaction. *Pediatr Clin North Am.* 1972;19:51–63. Available from: <http://www.sciencedirect.com/science/article/pii/S0031395516326669>.
8. Sachdev K, Gupta MK. A comprehensive review of feature based methods for drug target interaction prediction. *J Biomed Inform.* 2019;93:103159. Available from: <http://www.sciencedirect.com/science/article/pii/S1532046419300772>.
9. Anusuya S, Keshewani M, Priya K, *et al.* Drug-Target Interactions: Prediction Methods and Applications. *Curr Protein Pept Sci.* 2018;19:537–561.

CHAPTER 2

2.0 Background on the Application of Computer Drug Design in Drug Discovery

2.1 Introduction to Computer-Aided Drug Design

Small molecule compounds play a crucial role in modern-day medicine due to their ability to either prevent, treat, cure, relief, or restore a diseased system to a non-diseased state. Over the years, development of small molecule compounds for therapeutic purposes have usually involved the conventional route of obtaining lead molecules and subsequently developing a scientifically structured process for finding small molecule compound that would possess essential biological properties through several trial-and-error processes. Identification of a single small molecule compound for patient use usually begins with an initial screening of over 10,000 candidates followed by hundreds of animal testing in preclinical stages and finally concluded by several clinical trial stages involving thousands of healthy volunteers and diseased individuals. However, the identification of the small molecule compounds is usually preceded by years of scientific study to determine the biochemistry behind a disease, for which pharmaceutical intervention is possible. These subsequently result in determining specific biological targets (receptors, enzymes, protein) whose functions must be modulated by the small molecule compounds to induce a therapeutic outcome [1].

With the advancement in technology, this process has gradually evolved into a more purposeful design, synthesis, and evaluation of the candidate small molecule compounds referred to as a “rational drug design process”, a process that thrives on the collaborative efforts of multiple disciplines such as process chemistry, animal pharmacology, drug metabolism, medicinal chemistry

and clinical research [2]. However, this conventional process is time and resource-consuming, with a recent study by Wouters et al. [3] revealing that an estimated median cost of \$985 million and an average cost of \$1.3 billion is required to get a small molecule compound from the laboratory to the market over a 7-to-12-year period. Aside from the huge cost and time-consuming nature of conventional drug discovery processes, studies have shown that only 1 out of 10 compounds that make it to clinical trials will successfully reach the market, suggesting a very low success rate of less than 0.001% [4,5]. These challenges have motivated the incorporation of new approaches into the drug discovery and development process, some of which aim to facilitate, expedite, and consequently to streamline the process.

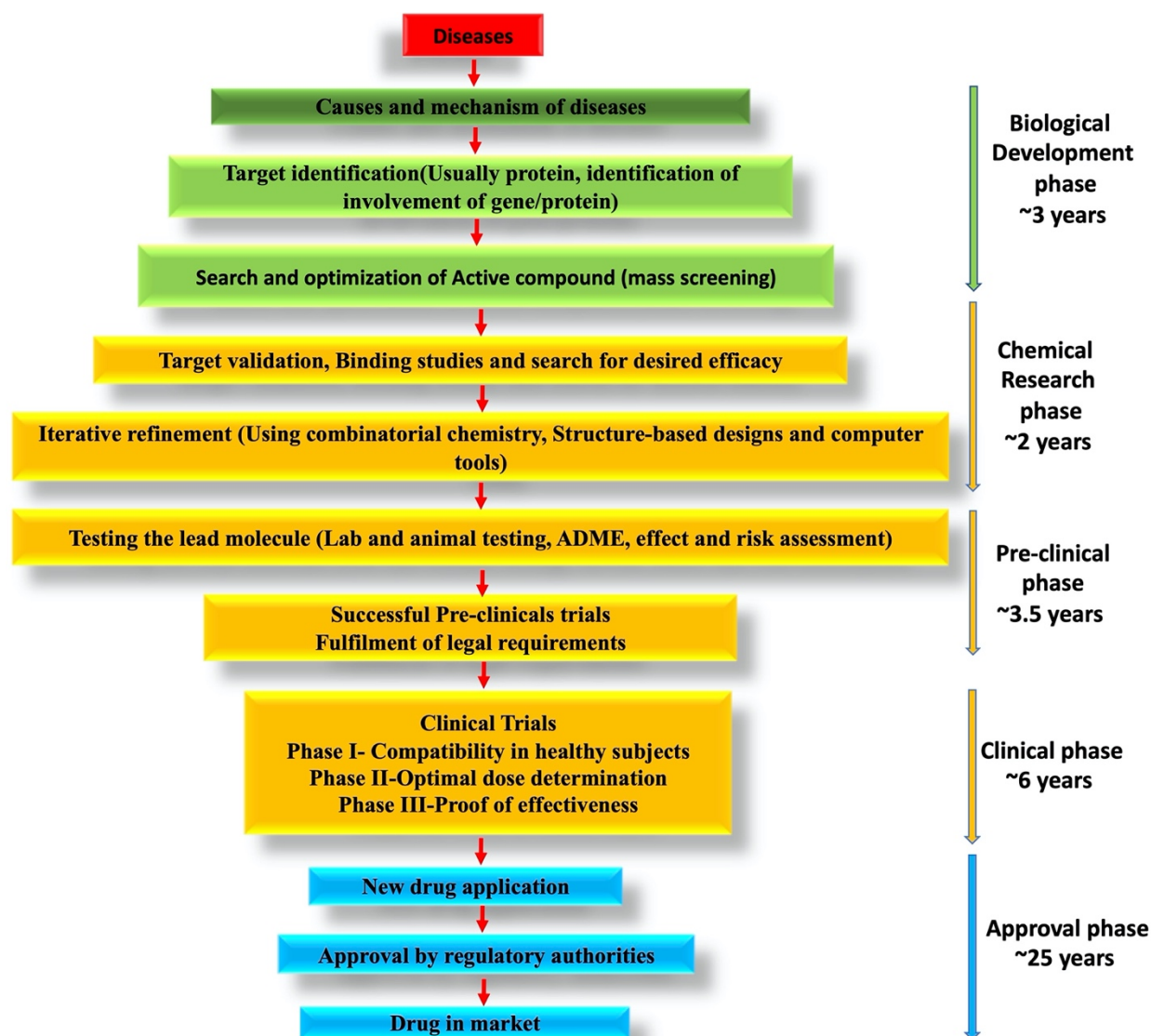


Figure 2.1: Schematic showing the different phases of the drug discovery process [1]

One such approach is Computer-Aided Drug Design, which has proven to save time, money, and resources [1,6–9]. Since its introduction, it is estimated that CADD currently accounts for about 10% of pharmaceutical research and development expenditure and continue to increase [4,10]. The incorporation of CADD in the drug discovery pipeline cuts across various stages of the process from lead discovery in its early years of application to its current application in target identification, target validation through to clinical trials through its ability to predict pharmacokinetic properties [4,7,11].

CADD's predictive power augments the drug design and development process by screening out only the most promising drug candidates from large compound libraries. Thus, CADD can prune compounds that could result in a “dead-end” due to their unfavorable pharmacological and pharmacokinetic properties, thereby accelerating getting drugs to the market. CADD also provides thorough atomistic binding features derived from the modeling of drug-target interactions [12]. These detailed atomistic scale insights are nearly impossible to assess in any other way yet are crucial in the design of new and improved small molecules compounds.

Once the possible binders are identified, combinatorial chemistry can be used to generate a series of derivatives. However, in the absence of a biological target, a QSAR pharmacophore can be created based on the structure and activity information of a ligand. The crucial pharmacophoric features can screen for new binders with similar features that bind to the same target. Further, properties such as absorption, distribution, metabolism, excretion, and toxicity can also be predicted using CADD tools and used to compare with bio-assay data. Compounds that make it through all preclinical assessments become eligible compounds for the following clinical trials.

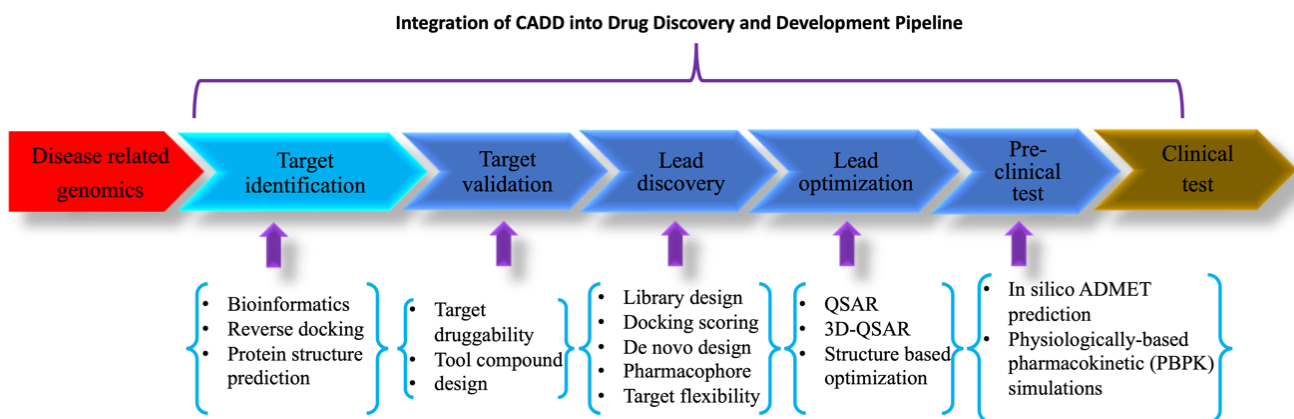


Figure 2.2: *In silico* Computer-aided drug design [13,14]

Over the years, CADD's application led to the successful discovery and design of some small-molecule compounds, notably the design of thymidylate synthase inhibitors as anticancer inhibitors by Appelt et al. in 1991[15]. Similarly, in 1994, Thaisrivongs (1994) reported HIV protease inhibitors that were also discovered through screening or via structural-based design based upon crystal structures of HIV protease complexes. The crystal complexes of HIV proteases allowed for a thorough evaluation of enzyme-inhibitor interaction and a practical effort toward the iterative cycle of structure-based design of inhibitors [16]. There have also been reports of the applications of CADD techniques in the successful discovery of neutrophil elastase inhibitors in 1998[17]. Other notable compounds discovered via CADD approaches include the carbonic anhydrase inhibitor discovered in 1989 by Baldwin et al. as an antiglaucoma agent inhibitor and a set of synthetic sweeteners that were also reported in 1990. Several of these CADD interventions have resulted in the discovery of many small inhibitors which have either been approved or in various stages of development as presented in table 2.1 amongst many others [18–20]. CADD has since become an essential part of the preliminary stages of the drug discovery pipeline of most renowned pharmaceutical companies, which has consequently expedited the drug development process in a more cost-efficient way towards minimizing failures in the final stage of clinical trials.

Table 2.1: Highlights of drugs designed using CADD

Drug	Year of Approval	Therapeutic action
Cimetidine	1979	H ₂ -receptor antagonist[21,22]
Captopril	1981	Antihypertensive [23]
Imatinib	1990	Tyrosine kinase inhibitor[24,25]
Saquinavir	1995	Human immunodeficiency Virus (HIV) inhibitor [26]
Dorzolamide	1995	Carbonic anhydrase inhibitor [27]
Indinavir	1996	Human immunodeficiency Virus (HIV) inhibitor [26]
Ritonavir	1996	Human immunodeficiency Virus (HIV) inhibitor[26,28]
Triofiban	1998	Fibrinogen antagonist[29]
Norfloxacin	1998	Inhibitor of bacterial DNA gyrase[30,31]
Oseltamivir	1999	Neuraminidase inhibitor[32]
Zanamivir	1999	Active against influenza A and B viruses.[33]
Gefitinib	2003	EGFR kinase inhibitor[34,35]
Fosamprenavir	2003	HIV protease inhibitor[36,37]
Atazanavir	2004	HIV protease Inhibitor[38,39]
Sorafenib	2005	VEGFR kinase inhibitor[40–42]
Erlotinib	2005	EGFR kinase inhibitor[43,44]
Tipranavir	2005	Nonpeptidic HIV-1 protease inhibitor[45]
Darunavir	2006	Nonpeptidic HIV-1 protease inhibitor[46,47]
Lapatinib	2007	EGFR / ERBB2 inhibitor[48,49]
Raltegravir	2007	Human immunodeficiency Virus (HIV) inhibitor[50]

Aliskiren	2007	Human renin inhibitor[51]
Boceprevir	2011	Hepatitis C virus (HCV) inhibitor[52]
Abiraterone	2011	Androgen synthesis inhibitor[53–55]
Crizotinib	2011	ALK inhibitor[56,57]
Rupintrivir	Phase II clinical trials	Human rhinovirus (HRV) 3C protease[58]
TMI-005	Phase II clinical trials	In Rheumatoid arthritis[59]
LY-517717	Phase II clinical trials	Serine protease Inhibitor[60]
Nolatrexed	Phase III clinical trials	In Liver cancer[61]
NVP-AUY922	Phase II clinical trials	Inhibitor for HSP90[62]
Oxymorphone	Phase III clinical trials	Peripheral opioid receptor antagonist [30,63,64]

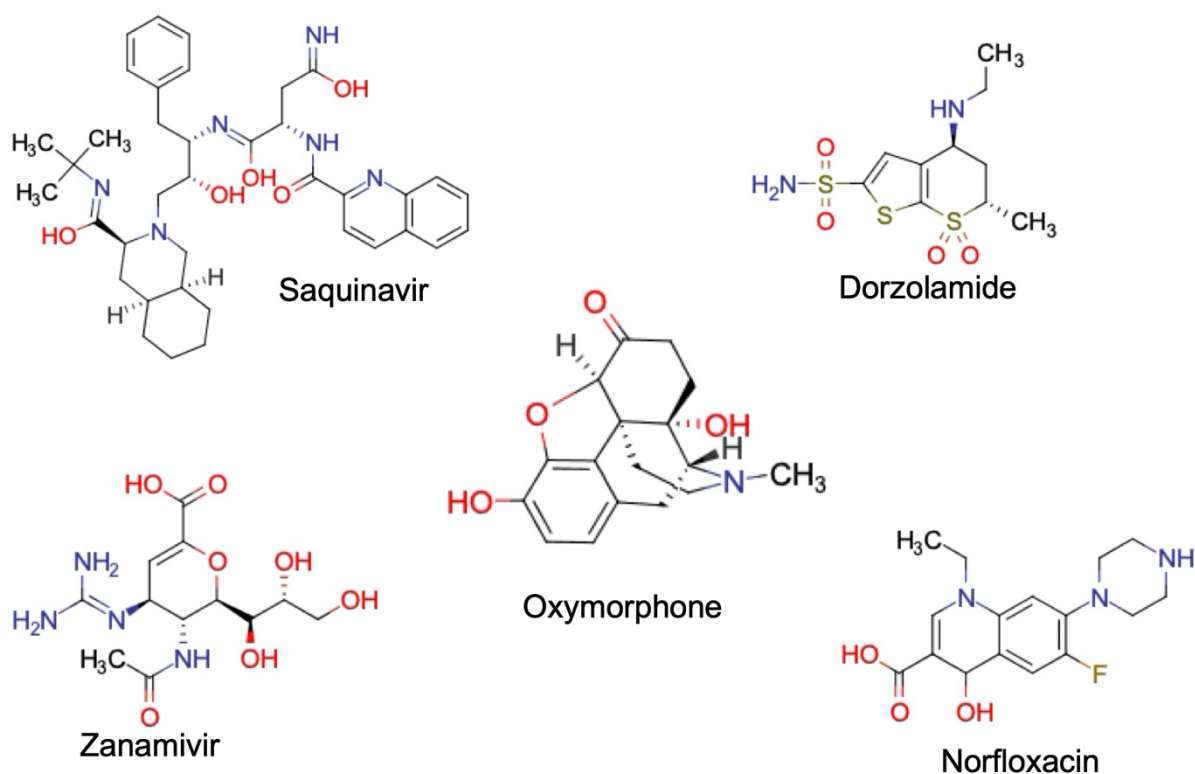


Figure 2.3: 2D structures of some small molecule inhibitors developed through CADD approaches

2.2 Genesis and Chronological Evolution of Computer-Aided Drug Designs

Although CADD is now very popular, with most globally renowned pharmaceutical companies extensively employing computational tools to design and discover therapeutic products for various life-threatening diseases [13], CADD has been around for several years. CADD became widely accepted as a concept at the beginning of the 20th century after the “lock and key” concept propounded by P. Ehrlich (1909), and E. Fischer (1894) had occupied a better of the 19th century. This was followed by subsequent incorporation of statistical techniques with biological activities of compounds to generate a new concept called Quantitative Structure-Activity Relationships (QSAR) a few decades later. [66] A publication titled “Next Industrial Revolution: Designing Drugs by Computer at Merck” in the Fortune magazine in 1981 heightened the popularity of CADD and its ensuing relevance in the drug discovery process [7,68]. This was in addition to the rising popularity of the combination of CADD with existing molecular biology concepts and its incorporation in the prediction of structures via Nuclear Magnetic Resonance and X-ray crystallography [7]. Several innovations that followed, including advancements in combinatorial chemistry, increased the number of compound databases covering large chemical spaces aided in expanding drug discovery and the eventual development of high-throughput screening (HTS) [69–71]. The advent of HTS expedited the drug discovery and development process by facilitating the rapid screening of large compound libraries for compounds that could elicit desirable therapeutic responses against new targets [72]. HTS eventually gained prominence in 1992 when about 40% of drug portfolios obtained their hits from HTS [72]. Regardless of the ability of HTS to facilitate the rapid screening of millions of compounds and its cost-effectiveness, the rates of obtaining successful hits are low, and identified hits might also not make it beyond preclinical and clinical investigations. The failure of hits was attributed to their poor pharmacokinetics and physicochemical properties such as adsorption, distribution, metabolism, excretion, and toxicology (ADME/T). Therefore, CADD presented a viable alternative that augmented

the existing methods by enhancing the success rates of hit compounds, making it as drugs while reducing the associated colossal cost of drug discovery and development. CADD thus improves the bioactive molecules, develops therapeutic alternatives, and understands biological events at the molecular level. The onset of the 21st witnessed a rise in CADD's prominence, with the advancement in computational power and the development of many models and new tools. With these improvements, CADD application in the pharmaceutical industry has taken center stage in the drug discovery process, particularly when combined with advanced forms of previous computational techniques such as HTS and X-ray crystallography. This has allowed for the application of CADD to find solutions to complex pharmaceutical solutions and discover numerous drugs in the treatment of diseases such as glaucoma, nonsmall-cell lung cancer, and acquired immunodeficiency syndrome (AIDS) [73–77].

2.3 Classification of Computer-Aided Drug Design

CADD has been classified mainly into two types, namely; Structure-based drug design (SBDD) and Ligand-based-drug design (LBDD). However, more recently, other subtypes have gained prominence, notably Pharmacophore-based drug design (PBDD) and Fragment-based drug design (FBDD), collectively referred to as hybrid-based methods.

2.3.1 Structured-Based Based Methods of Drug Design

Considered one of the most influential and innovative methods in drug design, structure-based drug design utilizes the three-dimensional (3D) structure of a biological target as the basis for drug design[9,69,71,78]. SBDD relies on understanding the mechanism of action of the bound ligands, the specific binding on the biological target, and thorough elucidation of the molecular dynamics

of the ligand-biological target complex [30,79–81]. As such, an existing X-ray crystal, Nuclear Magnetic Resonance (NMR) spectroscopy, or cryo-electron microscopy (EM) structures of the biological target complex is usually required to begin with [71,82,83]. In the absence of X-ray or NMR structures, molecular modelling techniques such as homology modelling is used to generate 3D structures of biological structures of interest [84–86]. These structures are retrieved from the known databases; Protein Data Bank (PDB) [87] and the Protein Data Bank in Europe (PDBe) [88] for further evaluation. To generate ligand-biological target complexes in the absence of deposited complexes, known computational techniques such as molecular docking are then employed.

Subsequently, molecular dynamics (MD) simulations are used to unravel the mechanism of action and interaction dynamics at an atomistic level [30,79–81]. Because the 3D structure of the biological target is the crux of SBDD, its identification of a valid biological target and understanding its structural dynamics is usually the first step in the entire SBDD process. SBDD is further sub-categorized into two approaches: de novo drug design and the ligand docking approach. In de novo drug design, large compound libraries are screened to identify new ligands that complementarily bind to a known biological target [84]. The recent advancement in technology has gradually seen a merger of both approaches [84].

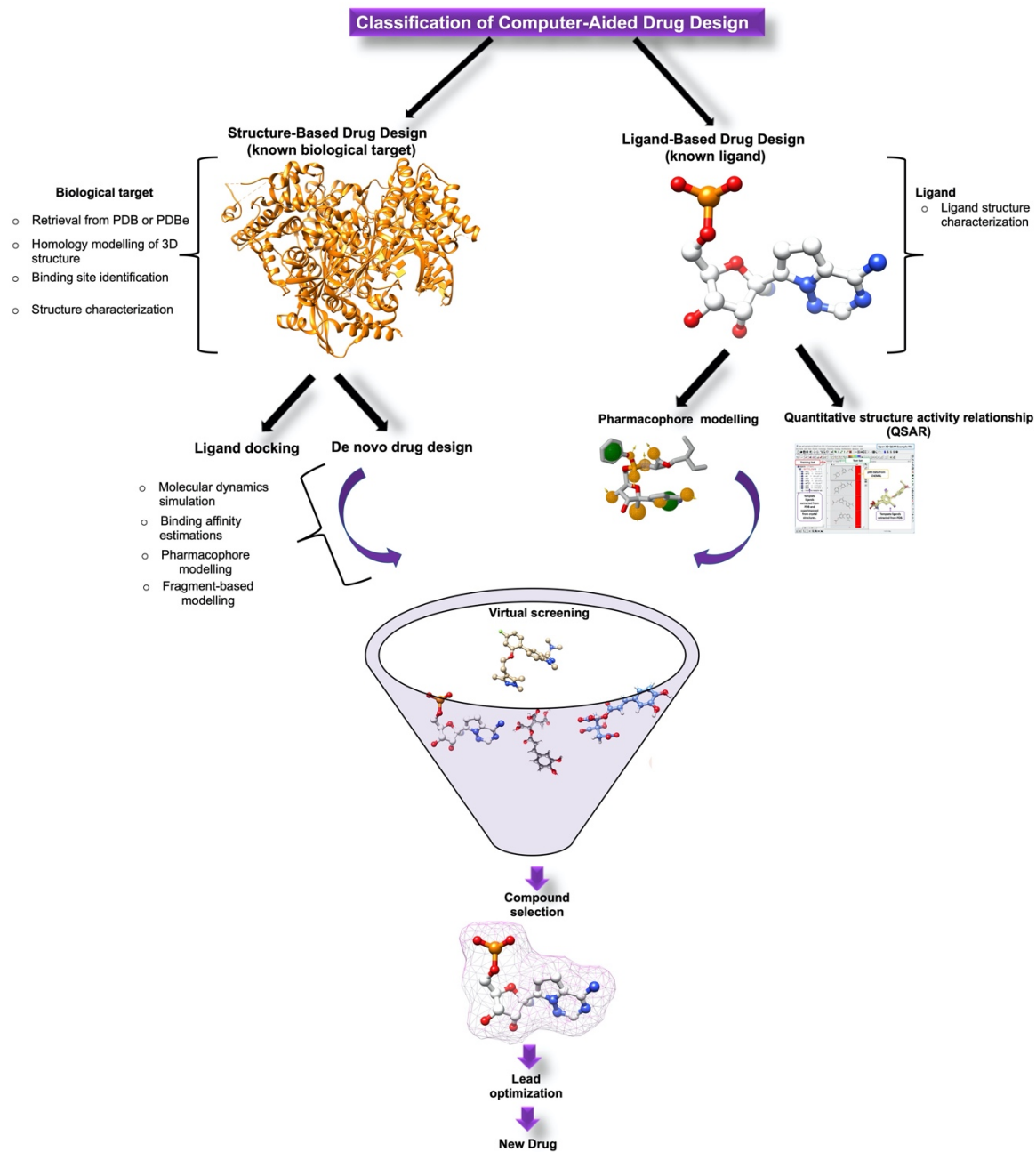


Figure 2.4: Schematic highlighting the classification of CADD

2.3.2 Ligand-Based Drug Design

This approach of drug design is employed when no information about the biological target is known. However, instead, enough information on the ligand that modulates that biological target's function is known [89,90]. Even though LBDD incorporates experimental approaches, its success rate is still lower than SBDD. In LBDD, notable computational techniques employed in quantitative structure-activity relationship/quantitative structure properties relations (QSAR/QSPR) are employed to select ligands with potential activity-based 2D/3D structural and physicochemical properties [9,84,90–92]. The selection of these molecular descriptors that inform the drug design process in LBDD is usually based on assuming that compounds that possess structural similarities could equally possess similar biological activity towards a given biological target [71,93]. The knowledge of QSAR/QSPR allows for identifying moieties on the ligands that could influence a particular therapeutic effect and inform any structural modification on the ligand to improve its therapeutic potency [9,84,94].

2.3.3 Hybrid Drug Design Methods

Interspersed between both SBDD and LBDD are methods that have gained prominence as CADD techniques, namely pharmacophore-based (PBDD) and fragment-based drug design. Hybrid methods also employ both SBDD and LBDD, thus in cases where there are a known biological target and a known ligand. In PDDD, the design of small molecules is based on the knowledge of chemical features of a known chemical compound required for the biological activity of that chemical compound [67,84,95–97]. These chemical features, referred to as pharmacophores, dictate the chemical compound's interaction with specific sites of a given biological target. PBDD has been widely explored because of its reliability over the years as a drug design method. Widely

explored methods under PBDD include; 3D-QSAR based pharmacophore modeling, common feature pharmacophore modeling, and 3D-pharmacophore modelling [84,98].

Fragment-based screening as an approach connects molecular fragments that have been shown to interact with specific sites on binding pockets to generate a small molecule compound that can be synthesized and be therapeutically useful [84,99]. The process of FBDD usually involves either a replacement of a chemical scaffold or a growing of new fragment for the purpose [84,99,100]. Several compounds designed *via* FBDD have made it through to clinical studies establishing its crucial role in the current day drug design and development [99,101].

2.4 Challenges of the application of CADD in drug design development

Although CADD is undoubtedly a substantial part of the drug design process, it has known challenges that impede its overall success. As a multidisciplinary specialty, CADD requires highly skilled professionals with intellectual proficiency in all relevant fields, a requirement that tends to be a significant challenge. Another considerable challenge of CADD is the accuracy and efficiency of computational techniques since, in scientific computing, several factors need to be considered, including; theoretical assumptions and algorithmic shortcuts, which could affect outcomes and predictions. This consequentially influences the quality and reliability of online computational resources and computational techniques such as virtual screening and ADMET predictions [102]. Other notable challenges include; deficiencies of CADD in chemo-genomics, the challenge of designing drugs with multi-targeting properties, improving the predictive capacity of toxicity models and side effects, and strengthening interaction with other disciplines to optimize the search for bioactive molecules for the treatment and/or prevention of diseases.

Although CADD has gained prominence, its application in the drug design and development process is still limited by many scientists' reluctance in integrating computational approaches with experimental techniques. Therefore, efforts should be made to ensure that CADD is successfully integrated with existing methods such as high throughput screening, combinatorial chemistry, and natural product-based drug discovery but not treated as a stand-alone technique of drug discovery [103–108].

2.6 Application of CADD in Developing Therapeutics for Viral Infections

Viral infections are the most notorious of all infectious diseases.

Viruses are obligate intracellular organisms that contain either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) genetic material enclosed within a capsid that causes numerous diseases in organisms. Several viruses have emerged in the last 40 years and have been the cause of significant disease outbreaks, notably; human immunodeficiency virus (HIV), influenza viruses, Hepatitis B virus, Ebola virus, Zika virus, severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome (MERS) and the most recently SARCoV-2 amongst many others [109–117]. As viruses continue to evolve with the emergence of drug-resistant strains, there is a need for new therapeutic agents to be designed as well. Available antivirals have targeted various viral proteins and enzymes while attempting to strengthen host immune responses. The earliest reported antiviral application in treatment was in the 1960s in which thiosemicarbazone was used to treat smallpox [118]. Many approved antivirals target viral replication; however, many of the structures required for the process, such as ribosomes, originate from humans, while some known antiviral have associated side effects. As such, the search for safer antiviral continues. Although the process of designing new antivirals may be expensive, tedious, and time-consuming,

technological advancement has allowed for the application of cheaper and less time-consuming approaches such as computer-aided techniques towards designing novel and more efficient antivirals. Computer-aided have been employed at various stages to discover and design various antivirals, with some even getting approved for clinical use, as presented in table 2.1, amongst many others [119,120].

With the aid of CADD techniques, the 3D structures of several structures in viruses have been determined via X-ray crystallography or homology modeling. Other molecular modeling techniques such as MD simulation and thermodynamics calculations have also allowed for the unraveling of atomistic insights and binding interaction dynamics that could help design novel antivirals with improved pharmacokinetics and physicochemical properties [118,121–123]. MD simulations have also helped identify hidden or allosteric binding pockets and augment existing antiviral drug discovery and development methods. To ensure efficient preclinical, experimental analyses of antiviral compounds in the ChEMBL dataset, Vázquez-Domínguez et al. proposed a Perturbation-Theory Machine Learning model (PTML) [124]. The PTML model allowed for simplifying the management of the data in ChEMBL by considering multiple features that combined with preclinical, experimental anti-retroviral tests. The application of CADD techniques such as virtual screening has also allowed for screening large compound libraries to identify potential ant-viral agents [125,126]. Some of these techniques, such as QSAR, also allow for the optimization of drugs against several viruses, which consequently optimize the metabolic profiles of the drugs [127,128].

2.6.1 Anti-Viral Targets Explored in This Thesis

This section will describe viral therapeutic targets that have been extensively investigated in relation to their interaction with specific small molecule inhibitors. The structure, functions, and implicated roles in viral disease pathogenesis will be discussed in detail.

2.6.1.1 SARCOV-2 RNA dependent RNA polymerase

One of the extensively investigated therapeutic targets in search of treatment methods against COVID-19 is the RNA-dependent RNA polymerase (RdRp) because of its crucial role in the replication and transcription of the SARSCOV-2 viral genome [129,130]. RdRp functions by controlling RNA strand elongation characterized by the addition of ~30,000 nucleotides. SARSCOV2 RdRp is the product of the cleavage of the polyproteins 1a and 1ab from ORF1a and ORF1ab [131]. Structurally, RdRp consists of three subunits a catalytic non-structural protein (nsp) 12 (398-919) and two nsp8 and nsp7 accessory subunits [132–134]. Nsp12 houses the polymerase domain in a cupped “right hand” shape with palm domains, a thumb, and fingers [131,135,136]. The majority of the catalytic motives (A-E) are located in the palm domain, known to be the most conserved. The remaining two domains (F and G) are located in the finger domains [136–138]. Several reports have explored the effectiveness of RdRp as a therapeutic target for the treatment of COVID-19 [139,140] due to their uniquely conserved core structure the similarities of the motifs essential for the catalytic activity of RdRp [141,142]. Its suitability as a therapeutic target is based on the fact that there is no homolog of RdRp in humans [130]. Targeting RdRp over

the years has involved nucleoside and nucleotide analog inhibitors, which antagonize the native nucleosides and nucleotides and thus block viral RNA synthesis while inhibitors others induce mutations in the viral genome [143,144]. Due to structural similarity in conserved amino acids between SARCOV-2 RdRp and other different positive-sense RNA viruses, several therapeutic agents targeting other RdRp have been repurposed as potential therapeutic agents in COVID-19 treatment [138,145]. Particularly at the Notable SARSCOV-2 RdRp inhibitors include at various advanced stages of clinical trials includes Remdesivir [146–149], Favipiravir [150–152], Ribavirin[153–155] and Galidesivir [130,156,157]

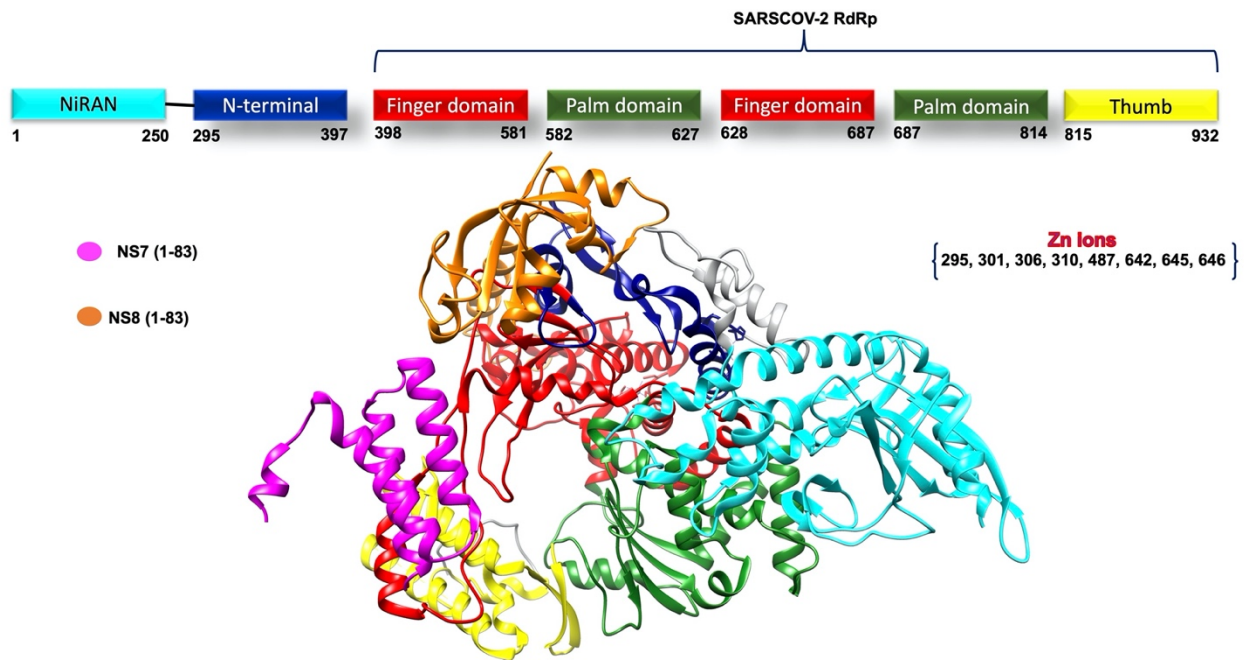


Figure 2.5: Linear structure of RNA Dependent RNA Polymerase (RdRp) showing the N-terminal domain and the main polymerase domain showing sub-domains (Finger, Palm, Thumb as well as a ribbon diagram showing important domain in RdRp.

2.6.1.2 Icosahedral Viral Capsid polyproteins of Human Rhinovirus B14 (HRV-B14)

Consisting of four viral capsid proteins, namely VP1, VP2, VP3, and VP4, which enclose the viral (+)-ssRNA strand, the icosahedral capsid polyprotein has been extensively explored as a therapeutic target for the treatment of HRV infectious [158,159]. Structurally, the capsid proteins' arrangements form a unique icosahedral symmetry, which houses 60 copies of each protein [160]. VP1, VP2, and VP3, which are the largest of the capsid proteins, make up the viral capsid polyprotein's external surface. At the same time, VP4 encompasses the polyprotein's inner surface, thus forming an interface between the viral (+)-ssRNA and the external capsid protein (VP1, VP2, and VP3). Peculiar narrow ditches around the 5-fold apex of the viral capsid called “canyons” are used for cell-receptor-binding, which results in the uncoating of the capsid for subsequent cell entry [161]. A hydrophobic groove within the “canyon” made up of residues; Leu25, Ala21, Leu106, Phe124, Tyr128, Iso130, Tyr152, Val176, Tyr197, Met221, Asp219A, and Ser172 has been targeted by many capsid inhibitors. Upon binding to the hydrophobic groove in the “canyon”, some of these capsid inhibitors stabilize the viral capsid dynamics and subsequently impede viral uncoating [162]. Notable viral capsid stabilizing inhibitors include; pleconaril (WIN-63843) and Vapendavir (BTA-798) [163–166].

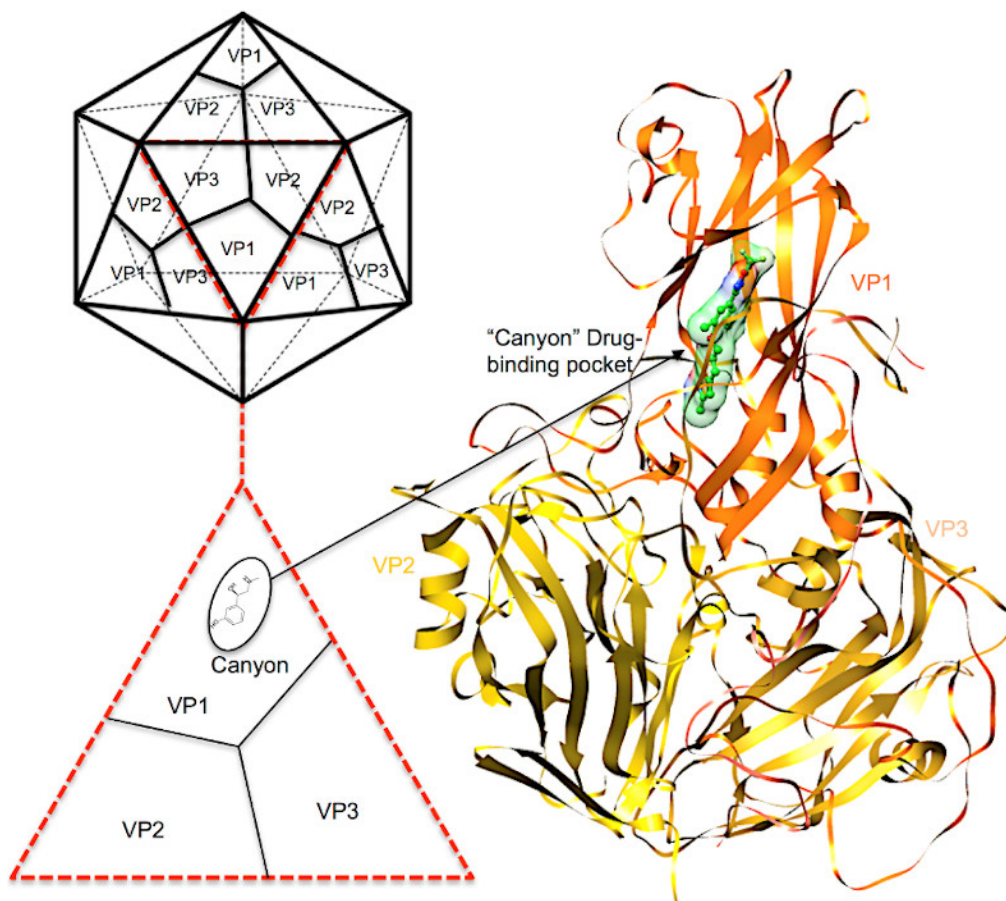


Figure 2.6: Structure of icosahedral viral capsid demonstrating the unique VP1-VP3 domain orientations and the characteristic “canyon” drug-binding pocket. [162]

2.6.1.3 Human *N*-myristoyltransferases

N-myristoyltransferase (NMT) is a widely conserved enzyme in eukaryotes that responsible for the post and co-translational modifications characterized by the transfer of myristate from myristoyl coenzyme A (Myr-CoA) to the N terminus of specific substrates [167,168]. NMTs belongs to the superfamily GCN5-related N-acetyltransferase (GNAT) [169]. In humans, NMT is expressed as two isoenzymes, namely; NMT1 and NMT2, and is implicated in several diseases, including viral infections such as common cold [170]. The implication of NMT in viral infection was evidenced by a mutagenesis study involving Poliovirus. It was suggested that the *N*-myristoylation of the

capsid protein VP0 is a crucial capsid assembly and viral infectivity [171–174]. This evidence presents human *N*-myristoyltransferase as a viable therapeutic target that is minimally affected by any potential mutation in the virus as well as an serotypic variation since the host *N*-myristoyltransferase is an invariant factor in the replication of the virus.

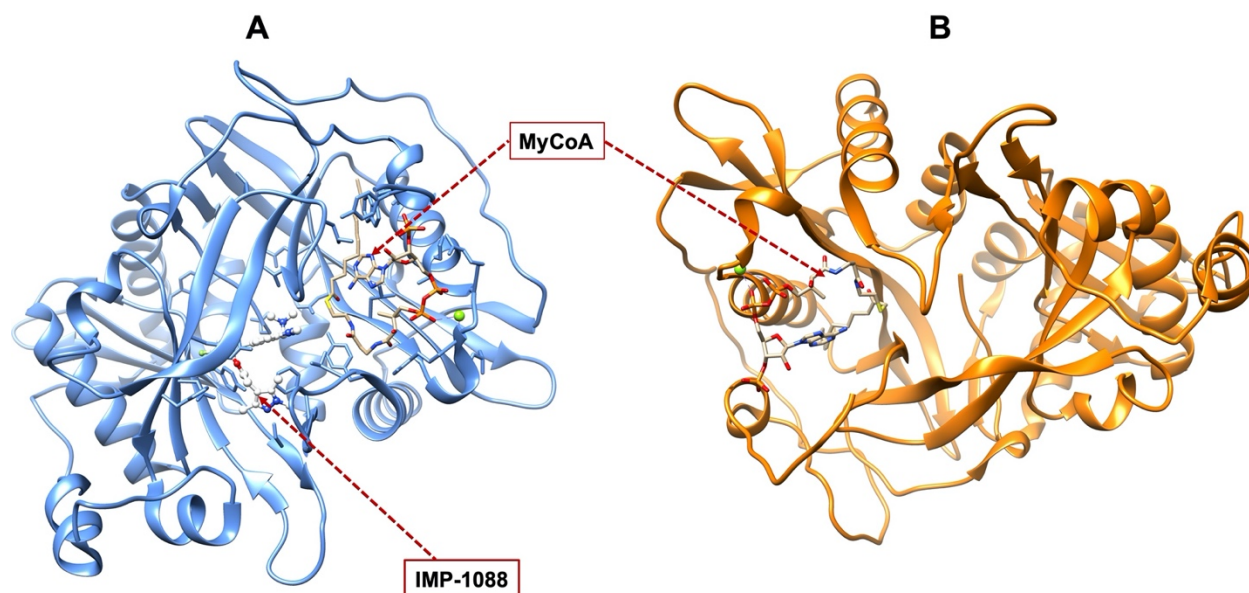


Figure 2.7: A) Human *N*-myristoyltransferase (NMT1) with Myristoyl-CoA and IMP-1088 inhibitor bound (PDB:5MU6). B) Human *N*-myristoyltransferase isoform 2 (NMT2) (PDB:4C2X)

2.7 Application of CADD in developing therapeutics for tuberculosis

Tuberculosis (TB) remains a major global health threat considering its persistence as one of the leading causes of death worldwide. TB is caused by the “tuberculosis complex,” namely; *Mycobacterium tuberculosis* (Mtb), *Mycobacterium bovis*, and *Mycobacterium Africanum* [175,176] of which Mtb is the most predominant causative agent amongst the complex. In 2018, about 10 million people reportedly contracted TB, of which over half a million were reported attributed to Rifampin resistant TB (RR-TB) An estimated total of 1.5 million reported died from

these infections [177]. Considering the center stage, the drug resistance has assumed in the number of deaths associated with TB, novel antitubercular agents with the improved mechanism of actions are of urgent need towards overcoming the burden of TB. The increase in biological data, changes in the number of identified biological targets, and advancement in HTS technologies and other technological tools have also resulted in a change in the drug discovery process for novel antitubercular agents. As such, CADD has gradually taken a center in the design of TB therapeutics. Nonetheless, the applications still have some challenges as elaborated in a report by Ekins et al. in 2011, which informed their recommendation that CADD techniques should be ideally incorporated with experimental procedures to ensure an acceleration of the TB drug discovery process [178]. Several *in silico* techniques such as virtual screening, molecular dynamics simulations, molecular docking, homology modeling, QSAR, and pharmacophore modeling, amongst many others, have been employed over the years towards designing novel TB therapeutics [179]. A report by Taira et al. employed virtual screening to identify compounds that could interact with the mycobacterial CmaA1 [180]. The virtual screening process was based on the structural similarity between the selected active compounds in the study. It resulted in the identification of two compounds that could potentially inhibit mycobacterial growth. In a more recent report by Kumar et al., virtual screening was combined with MD simulation and binding free energy calculations to identify potential inhibitors of the promising TB target, decaprenylphosphoryl- β -d-ribose 2'-epimerase (DprE1). Their report also identified four compounds (ChEMBL2441313, ChEMBL2338605, ChEMBL441373 and ChEMBL1607606) as potential inhibitors of DprE1. In another study, Billones et al. identified potential inhibitors of Mtb 1,d-transpeptidase 2 (LdtMt2), an enzyme responsible for the catalyzation of 3-to-3 cross-linking between amino acids in linear amino sugar chains needed to form the peptidoglycan layer of mycobacteria [181]. Billones et al.

again in 2017 also combined molecular modeling techniques and experimental methods to discover novel inhibitors of Mtb 7, 8-diaminopelargonic acid synthase (Mtb BioA) [182]. Using similarity search, QSAR models, and pharmacophore hypotheses, Mehra et al. also screened the ChemBridge database [183] for potential inhibitors of the GlmU protein that plays a key role in the biosynthesis of peptidoglycans and liposaccharides, molecules that promotes the growth of Mtb [184]. Among many others, as highlighted in table 2.2, these reports showcase some of the immense contributions of CADD in TB pharmaceutical research.

Table 2.2: Highlights of some studies that have employed virtual screening in identifying anti-tubercular agents [179]

System	Function	PDB	Anti-Mtb- Activity
Cyclopropane mycolic acid synthase 1[180]	Cell wall	1KPH	MIC50/5.1 μ M
l,d-transpeptidase 2[181]	Cell wall	3TUR	MIC94/25.0 μ M
GlmU protein[184]	Cell wall	3ST8	IC50/9.0 μ M
L-alanine dehydrogenase[185]	biosynthesis of l-alanine	2VHW	IC50/35.5 μ M
L-alanine dehydrogenase[186]	biosynthesis of l-alanine	4LMP	MIC/1.53 μ M
L-alanine dehydrogenase[187]	biosynthesis of l-alanine	2VOJ	MIC/11.81 μ M

Leucyl-tRNA synthetase[188,189]	Protein synthesis	2V0C	MIC/25 μ M
Haloalkane dehalogenase[190]	Unknown	2QVB	Kd/3.37 μ M
3-dehydroquinate dehydratase[191]	shikimate pathway	2Y71	MIC/6.25 μ g/mL
3-dehydroquinate dehydratase [192]	shikimate pathway	15 PDB structures	MIC/100 mg/ml
Dihydrofolate reductase[193]	folate pathway	Mtb: 1DF7; human:1OHJ	MIC/25 μ M
Salicylate synthase[194]	iron acquisition	3VEH	MIC99/156 μ M
Transcription factor IdeR[195]	iron acquisition contro	1U8R	MIC90/17.5 μ g/ml
7,8-Diaminopelargonic acid synthase[182]	biotin biosynthesis pathway	3TFU	MIC/25 μ M
7,8-Diaminopelargonic acid synthase[196]	biotin biosynthesis pathway	3TFU	MIC/7.86 μ M
Flavin-dependent oxidoreductase Melf[197]	needed to withstand ROS and RNS induced stress	2WGK	MIC/13.5 μ M
NAD⁺-dependent DNA ligase A[198]	DNA metabolism	1ZAU/1TAE	MIC50/15 μ M
DNA Gyrase[199]	DNA topology	4BAE	MIC/7.8 μ M
Flavin-dependent thymidylate synthase [200]	DNA metabolism	2AF6	MIC90/125 μ M
Flavin-dependent thymidylate synthase [201]	DNA metabolism	2AF6	IC29/100 μ Mb

2.7.1 Anti-tubercular targets explored in this thesis

This section will describe anti-tubercular therapeutic targets that have been extensively investigated in relation to their interaction with specific small molecule inhibitors. The structure, functions, and implicated roles in the pathogenesis will be discussed in detail.

2.7.1.1 Mycobacterial folate biosynthetic pathway

The folate biosynthetic pathway has been extensively explored as an avenue to the development of novel treatment measures for tuberculosis due to the many druggable targets available in the pathways [202]. Inhibitors that target the folate biosynthetic pathway, referred to as antifolate, impede the production of reduced folate cofactors via the inhibition of crucial enzymes in the pathway. To synthesize folates, the enzyme dihydrofolate reductase encoded (DHFR) by *folA* gene is known to catalyze the conversion of dihydrofolate (DHF) into tetrahydrofolate (THF). The process employs Nicotinamide adenine dinucleotide phosphate (NADPH) as an electron donor. The generated THF is subsequently converted to N5, N10 Methylene tetrahydrofolate, catalyzed by serine hydroxymethyltransferase (SHMT). The resultant N5, N10 Methylene tetrahydrofolate (mTHF), subsequently serves as a precursor for the synthesis of thymidylate in a processed catalyzed by the enzyme thymidylate synthase (TSase). In thymidine synthesis, TSase converts mTHF and 2'-deoxyuridine-5'-monophosphate (dUMP) to 2'-deoxythymidine-5'-monophosphate (dTMP) DHF, and the cycle repeats itself using DHFR. The mycobacterium then employs the generated dTMP for the synthesis of DNA. Due to this established vital role of DHFR in the

pathway and its powerful influence on mycobacterial DNA synthesis, it presents a viable therapeutic target in the design of new TB therapeutics [203,204].

Nonetheless, many of the known available *Mtb* DHFR inhibitors have not achieved therapeutic success due to their inability to permeate the cellular membranes or a generally poor potency against the enzyme [205]. Other notable enzymes in the pathway that have been explored for therapeutic purposes in TB include the enzymes flavin-dependent thymidylate synthase (FDTS) and the Rv267, a functional analog of DHFR [204,206]. Encoded by the gene, *thyX* FDTS reportedly facilitates dTMP production by catalyzing the conversion of DHF to mTHF, with several inhibitors having been designed to target it [207,208]. Therefore its inhibition will consequently reduce the production of mTHF.

Inhibitors of the folate pathways towards TB therapy over the years have included dapsone (DDS) and sulfamethoxazole (SMX), which inhibits dihydropteroate synthase (DHPS) [209,210], para-aminosalicylic acid (PAS) [211], a combination of trimethoprim and sulfamethoxazole as *Mtb* DHFR inhibitors [212] amongst many others. However, there is currently no approved treatment agents that target DHFR. Targeting the foliate pathways at various points using antifolates impedes the activated methyl cycle, which results in a decrease in the mycolic acids of the mycobacterium. Most recently, reports of multi-targeting agents of the cycles are gaining since these have been shown to possess higher therapeutic potency against the mycobacterium in relation to single targeting agents [204]

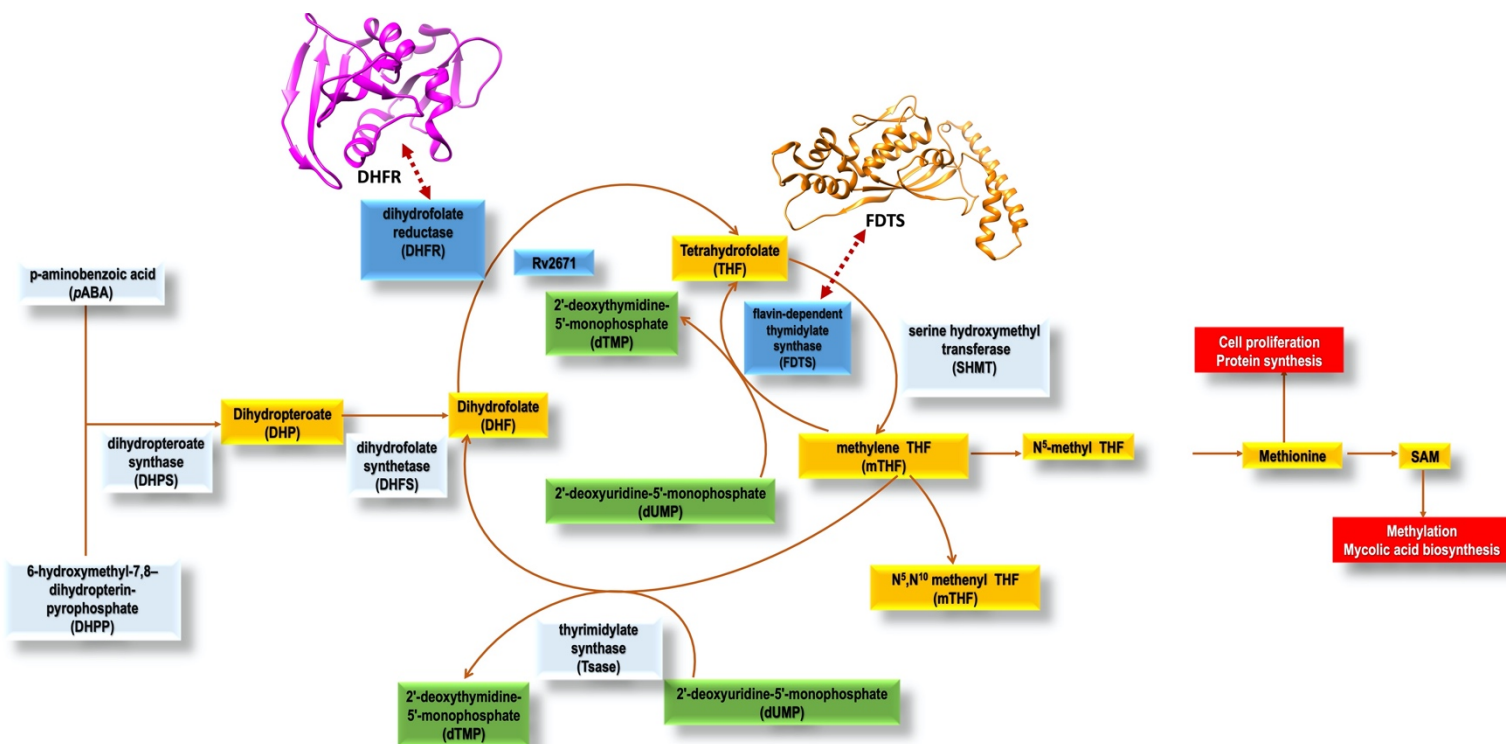


Figure 2.8: The Folate Biosynthetic pathway in *Mtb* [204]

2.7.1.2 C-rotor ring of *Mycobacterium tuberculosis* F1Fo-Adenosine Triphosphate (ATP) synthase

The c-rotor ring of *Mtb* F1F0 ATP synthase has been established as an essential structure in mycobacterial ATP synthesis. Hence, several inhibitors, notably Bedaquiline (BDQ) of the diarylquinoline class of compounds, have therapeutically targeted it towards TB therapy [213–219]. As a part of the membrane-embedded F1F0 ATP synthase, the c-rotor ring is harbored in the F0 domain. In mycobacterium, the F0 domain exists in identical forms of the c-subunit, assuming the shape of an hourglass-shaped cylinder and forms a central pore referred to as the c-ring [220,221]. As an integral of the F1F0 ATP synthase, the c subunit is made up of two transmembrane helices responsible for proton translocation. Both helices also house crucial

residues (D28, L59, E61, A63, I66) required for sensitivity to diarylquinolines like BDQ [214,222–224]. Structurally, F1F0 ATP synthase consists of two domains; the F1 and F0 domain. The F1 complex facilitates the transfer of H⁺ from the periplasmic space to the cytoplasmic space [225–227]. The F1 complex, which is also water-soluble, converts Adenosine Diphosphate (ADP) and inorganic phosphate (Pi) to ATP using H⁺-motive force [225,226,228] by a process referred to as the rotary mechanism [225,228,229]. As a crucial enzyme in the synthesis of ATP in the three catalytic active pockets of the F0 domain, F1Fo-ATP synthase facilitates the shuttling of ions across cellular membranes [228].

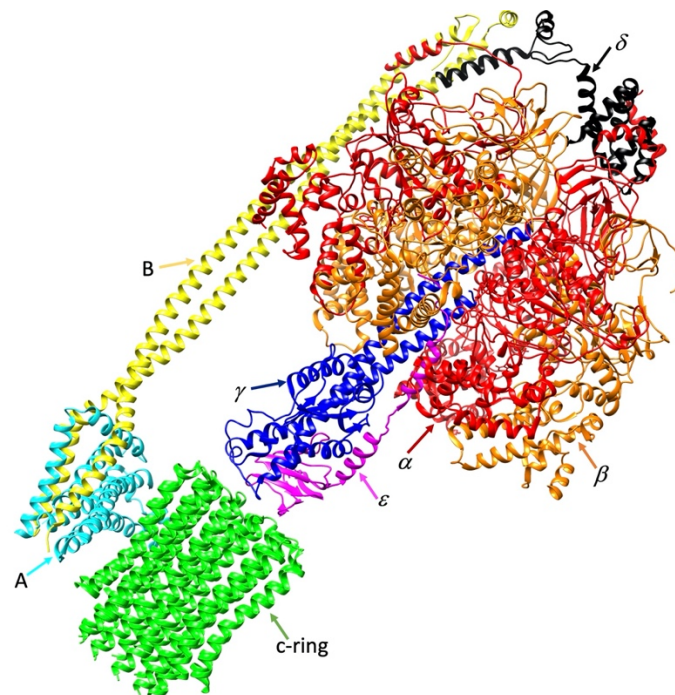


Figure 2.9: Structure of *Mtb* ATP synthase modelled from the structure of *E. Coli* ATP synthase. Structure highlights the various subunits

2.8 Application of CADD in developing therapeutics for cancer

As a disease, cancer has significantly contributed to the growing mortality worldwide, with more than half of cancer cases in low- and middle-income countries [230]. Treatment and management of cancer remain a challenge, with many available interventions been halted or not making it beyond the clinical phases of drug development. The challenges with existing cancer therapeutics are mainly due to cancerous cells' multifactorial characteristics, thus evading and developing drug resistance to cancer therapeutics [231,232]. Targeted therapies and precision medicine have gradually gained prominence as an advanced therapeutic intervention to address these challenges. However, disease relapse, adverse side effects, expensive treatment options, and drug resistance persist [233–238]. As such, the search for alternative cancer therapeutics with enhanced potency, cheap, minimal adverse effects, and overcome resistances remains unabated. Conventional drug discovery methods in cancer therapeutics are time-consuming and expensive. However, the advent of CADD has sort to ameliorate some of these challenges with the hope of developing novel cancer therapeutics with high selectivity and minimal toxicities [8,239]. Together with experimental methods, these computer-aided techniques have contributed to the identification and selection of crucial oncogenic targets coupled as well as the strategic development of highly selective inhibitory methods using low cost less toxic therapeutic molecules that have significantly contributed towards the discovery of next-generation lead compounds with great potency in cancer treatment. Several reports have reported the involvement of CADD in the development of cancer therapeutics. For instance, in the report by Mutata et al., structure-based pharmacophore modeling was employed to identify potential inhibitors of p53 upregulated modulator of apoptosis (PUMA), whose inhibition leads to a deficiency in apoptosis and thus minimizing the risk of cancer development as well as drug resistance. In a separate report by Liu et al., a combinatorial

computational approach was employed in discovering potential insulin-like growth factor-1 receptor (IGF-1R) inhibitors [240]. Considering the implication of IGF-1R in several tumors such as lung cancers, prostate, and breast cancers due to its crucial role in signaling various cancer hallmarks including cell growth, proliferation, and apoptosis, a successful inhibition could be a therapeutic breakthrough. In their report, virtual and pharmacophore modeling were employed to identify 15 compounds that exhibited potential inhibitory activity against IGF-1R. Another reported application of CADD in the discovery and design of anticancer inhibitors is a study by Chiang et al. in which ligand-based methods were used to identify tubulin inhibitors [241]. Cell cycle progression and cell division require tubulin polymerization. Therefore its inhibitions present a powerful avenue for cancer therapy. Another report by Liou in 2006 used a structure-activity relationship (SAR) to generate a model on a set of indole-derivatives to improve the potency and solubility of anticancer drug candidate BPR0L075. Computational techniques were also employed in a study by Noha et al. in 2011, where small molecules inhibitors of I-Kappa-B Kinase β were identified (IKK- β) [242]. The crucial role of IKK- β in the NF- κ B signaling pathway presents them as viable cancer targets. Employing ligand-based pharmacophore modeling, they identified novel inhibitors with an affinity toward IKK- β . These studies and many others establish CADD's application toward the design of anticancer agents in the last couple of years.

2.8.1 Anti-cancer targets explored in this thesis

This section will describe therapeutic targets in cancer that has been extensively investigated in relation to their interaction with specific small molecule inhibitors. The structure, functions, and implicated roles in tuberculosis disease pathogenesis will be discussed in detail.

2.8.1 Anti-cancer targets explored in this thesis

This section will describe cancer therapeutic targets that have been extensively investigated in relation to their interaction with specific small molecule inhibitors. The structure, functions and their implicated roles in tuberculosis disease pathogenesis will be discussed in detail.

2.8.1.1 Protein Tyrosine Phosphatase Zeta (PTPRZ)

Protein Tyrosine Phosphatase Receptor Zeta (PTPRZ) is a member of the R5 subfamily of receptor-type protein tyrosine phosphatases (RTPs), which also has Protein Tyrosine Phosphatase G (PTPRG) as the other member [243,244]. Although Protein Tyrosine Phosphatases (PTPs) have been implicated in many cellular processes and diseases, they are generally classified as "undruggable" due to the positively charged active pockets which are highly conserved [244,245]. PTPs are known tumor suppressors, unlike the protein tyrosine kinases (PTKs), whose phosphorylation regulates many cellular functions. Simultaneously, its dysregulation has also been linked to the cause of many cancers such as gliomas [246,247]. Structurally, PTPRZ and PTPRG are made of carbonic anhydrase (CAH)-like domain that occurs in extracellular, two tyrosine phosphatase domains which are intracellular, and a fibronectin type III-like domain [248]. Although PTPRZ has been extensively investigated as a viable therapeutic target for glioblastoma treatment due to its strong expression in gliomas, the pathophysiological implication of PTPRG is unclear even though it is significantly expressed in advanced astrocytomas [249–251]. By regulating the process of protein tyrosine phosphorylation, PTPRZ leads to the survival of glioblastoma cells, thus enhancing tumor growth [246]. Therefore, in glioblastoma cells, reports

have shown that the inhibition of PTPRZ interrupts cell proliferation, reduces the formations of tumors in vivo, and reduces the migration of malignant cells in vitro hence its extensive exploitation as a therapeutic target [251–254]. The structural architecture of the PTPRZ active site consists of 5 loops, namely; the P-loops (phosphate-binding loop), the WPD-loop (catalytic acid/base harboring), Q-loop (conserved glutamine containing), pTyr-loop (pTyr-recognition), and the E-loop (contains multiple conserved residues) [255,256]. By its flexible nature, the WPD loop encloses the catalytic site assuming the role of a gate. In the active form of PTPRZ, the flexible WPD-loop assumes a closed conformation, while an open conformation is assumed in an inactive form of PTPRZ[257]. When substrates bind to the flexible WPD-loop, it assumes a closed conformation, which allows the catalytic aspartate on the loop to partake in catalysis [255]. As such, a successful inhibition of the WPD-loop mobility could subsequently impede the process of catalytic activity of the PTPRZ [258,259]. Although several phosphotyrosine competitive inhibitors have been developed over the years, none have been able to achieve success in cancer therapy due to their inability to permeate cell wall or as a result of the highly conserved and positively charged catalytic pocket [260]

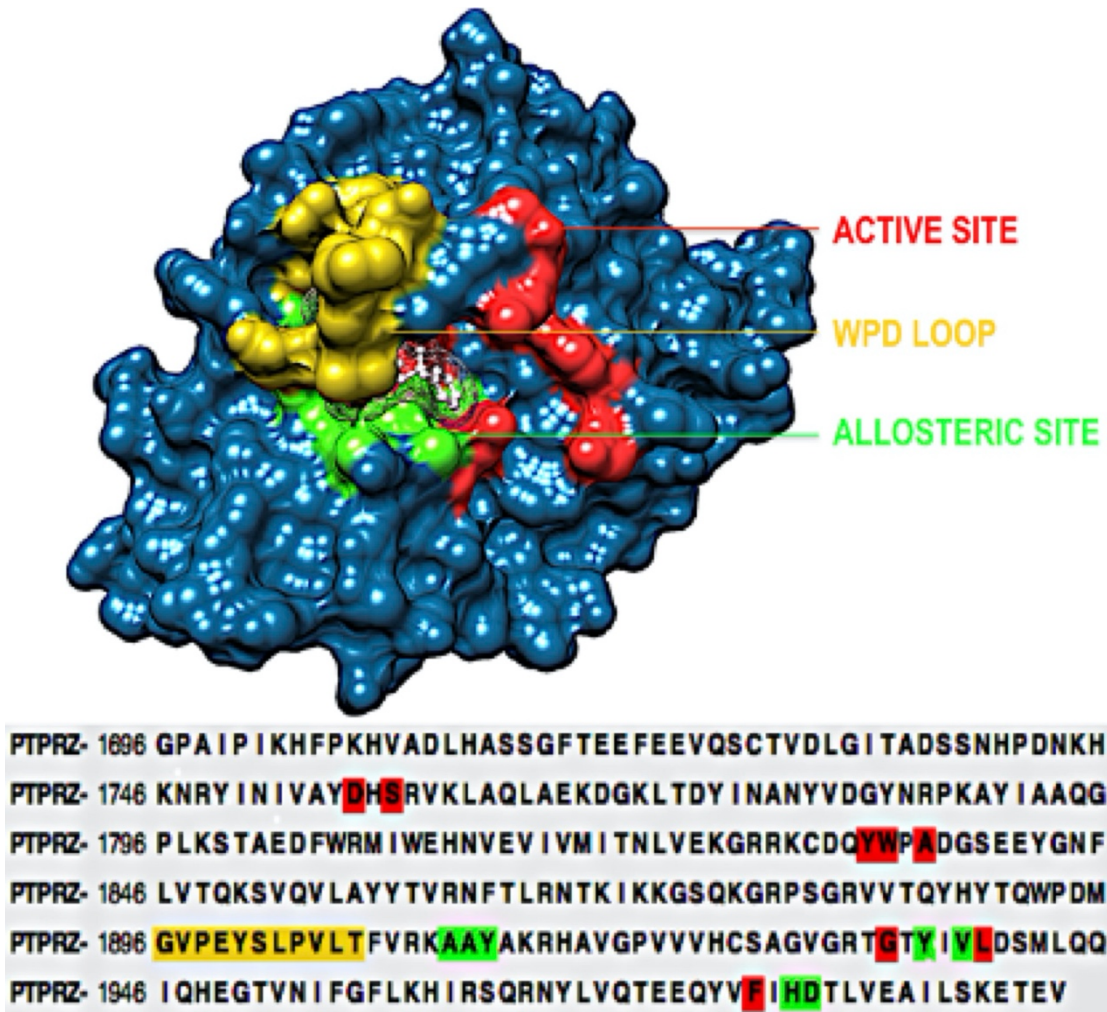


Figure 2.10: Structural and corresponding sequence representation of the unique regions of the PTPRZ protein; active site-red, allosteric site- green and WPD-loop- yellow[256].

2.8.1.2 Human CD1d-restricted Invariant natural killer T cells

Therapeutic targeting of human CD1d-restricted Invariant natural killer T (*i*NKT) cells has been proven by several reports to have potential therapeutic applications in supplementing immune responses against cancer and infections. *i*NKT cells form part of unconventional T cells involved in both innate and adaptive immune response [261,262]. A distinguishing characteristic of *i*NKT

cells, when compared to conventional T cells, is their ability to recognize lipid-based antigens rendered by MHC class I-like CD1d protein [263]. Also, a unique feature of *i*NKT cells in the presence of a semi-variant T cell receptor (TCR) found in both humans (V α 124J α 18) and mice (V α 14J α 18) which have been therapeutically exploited toward stimulating *i*NKT cells responses for anti-cancer purposes [264,265]. This has also expanded immunotherapeutic options as well as vaccine research [266–268]. *i*NKT cells exhibit selectivity towards CD1d-expressing thymocytes, unlike other natural killer cells that interact with class two MHC molecules [261,269,270]. The presence of TCR in *i*NKT cells also allows for their therapeutic targeting by prototype alpha-galactosylceramide inhibitors. *i*NKTs dominate in the adipose but occur in low frequency in the blood [271,272]. Activation of *i*NKTs results in an increased expression of NKTs, which leads to the stimulation of the conventional T cells through the secretion of cytokines and mature dendritic cells [273,274]. When triggered by antigens, *i*NKT s release cytokines (Th1, Th2, and Th17) within a space of some few hours together with the expression of CD40L, characterizing *i*NKTs role in immune response [269,275]. The antitumor immune response of *i*NKTs, on the other hand, is characterized by inducing the release of cytotoxic lymphocytes (CTLs), activation of other natural killer cells, and the modulation of the immunosuppressive tumor microenvironment, which altogether results in the killing of cancerous cells [266,276–278]. This has informed the extensive investigation of *i*NKT as immunotherapeutic targets [266,269,278]. Notable activators of *i*NKTs towards cancer therapy have included alpha-galactosylceramide (α -GalCer) and their resulting derivatives and synthetic forms; KRN7000 and the more recent AH10-7 [264,265,279–282].

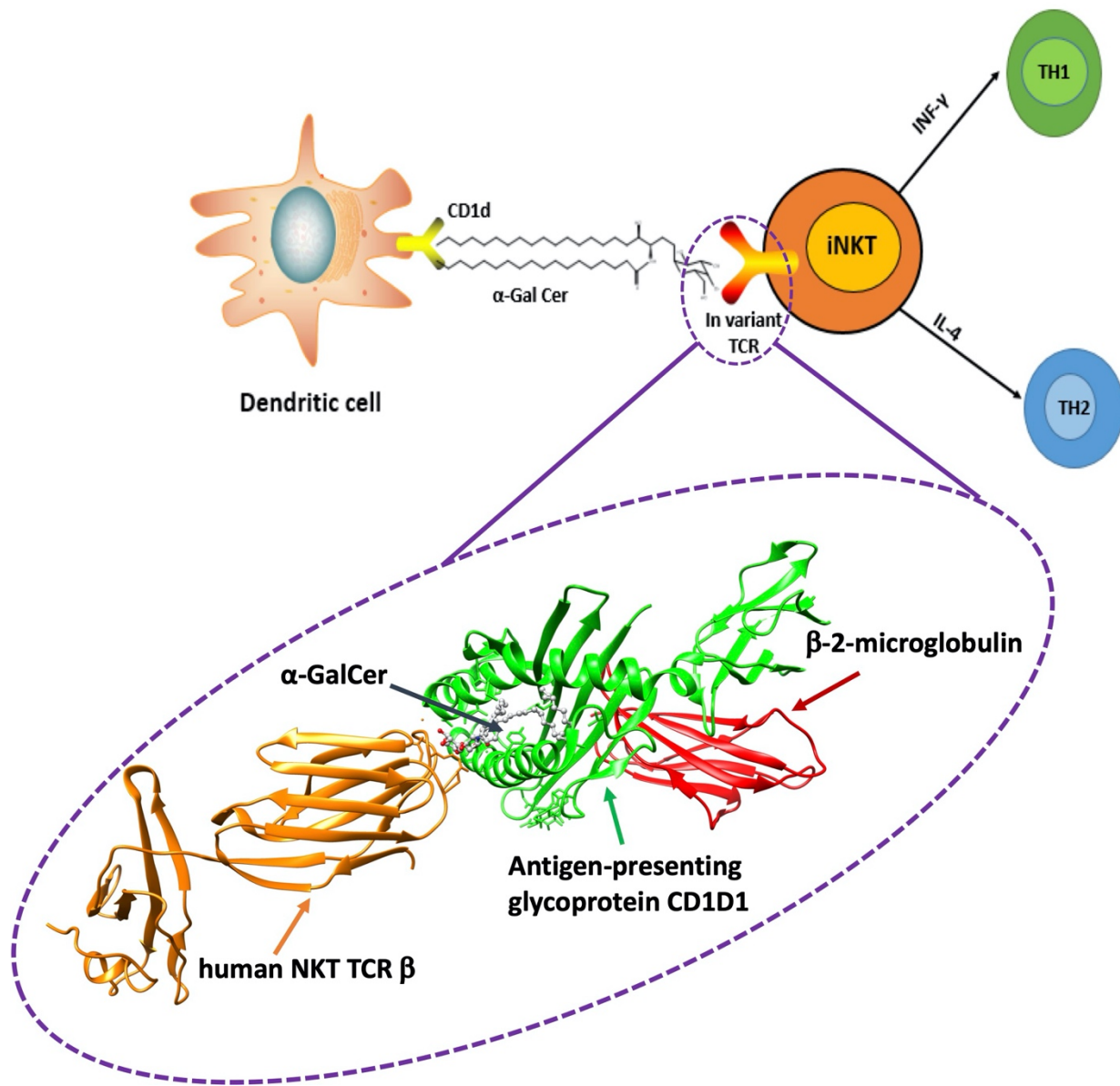


Figure 2.11: Mode of glycolipid presentation by CD1d to NKT cells

Reference

- [1] Baldi A. Computational approaches for drug design and discovery: An overview. *Syst Rev Pharm.* 2010;1:99–105.
- [2] Drews J. Drug discovery: a historical perspective. *Science.* 2000;287:1960–1964.
- [3] Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA.* 2020;323:844–853. Available from: <https://doi.org/10.1001/jama.2020.1166>.
- [4] Kapetanovic IM. CADD in silico chemico bio approach. *Chem Biol Interact.* 2009;171:165–176.
- [5] DiMasi J, Feldman L, Seckler A, et al. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther.* 2010;87:272–277.
- [6] Kumar N, Hendriks B, Janes V, et al. Applying computational modeling to drug discovery and development. *Drug Discov Today.* 2006;11:806–811.
- [7] Sliwoski G, Kothiwale S, Meiler J, et al. Computational methods in drug discovery. *Pharmacol Rev.* 2014;66:334–395.
- [8] Hung C-L, Chen C-C. Computational Approaches for Drug Discovery. *Drug Dev Res.* 2014;75:412–418.
- [9] Hassan Baig M, Ahmad K, Roy S, et al. Computer Aided Drug Design: Success and Limitations. *Curr Pharm Des.* 2016;22:572–581.
- [10] van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov.* 2003;2:192–204.
- [11] Bharath EN, Manjula SN, Vijaychand A. In silico drug design-tool for overcoming the innovation deficit in the drug discovery process. *Int J Pharm Pharm Sci.* 2011;3:8–12.

- [12] Saha D, Emran T Bin, Paul S. Bioinformatics : The effects on the cost of drug discovery. :44–50.
- [13] Yun T, Weiliang Z, Kaixian C, et al. New technologies in computer-aided drug design: toward target identification and new chemical entity discovery. *Drug Discov Today Technol.* 2006;3:307–313.
- [14] Pranita P, Madhavi M, Rishikesh V, et al. Computer-aided drug design: an innovative tool for modeling. *J Med Chem.* 2012;2:139–148.
- [15] Appelt K, Becquet R, Barlett C. Design of enzyme inhibitors using iterative protein crystallographic analysis. *J Med Chem.* 1991;34:1834–1925.
- [16] Thaisrivongs S. HIV protease inhibitors. *Annu Rep Med Chem.* 1994;29:133–144.
- [17] Hlasta D, Pagani E. Human leukocyte elastase inhibitors. *Annu Rep Med Chem.* 1998;29:195–204.
- [18] Rodrigues T, Schneider G. Flashback forward: Reactiondriven de novo design of bioactive compounds. *Synlett.* 2014;25:170–178.
- [19] Reynolds C. Impact of computational structure-based methods on drug discovery. *Curr Pharm Des.* 2014;20:3380–3386.
- [20] Shoichet B, MCGovern S, Wei B, et al. Lead discovery using molecular docking. *Curr Opin Chem Biol.* 2002;6:439–446.
- [21] Brimblecombe R, Duncan W, Durant G, et al. The pharmacology of cimetidine, a new histamine H₂-receptor antagonist. *Br J Pharmacol.* 1975;53:435–436.
- [22] Henn R, Isenberg J, Maxwell V, et al. Inhibition of gastric acid secretion by cimetidine in patients with duodenal ulcer. *N Engl J Med.* 1975;293:371–375.
- [23] Frohlich ED, Cooper RA, Lewis EJ. Review of the Overall Experience of Captopril in

Hypertension. Arch Intern Med. 1984;144:1441–1444. Available from: <https://doi.org/10.1001/archinte.1984.00350190137023>.

[24] Druker B, Tamura S, Buchdunger E, et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat Med. 1996;2:561–566.

[25] Buchdunger E, Zimmermann J, Mett H, et al. Inhibition of the Abl protein-tyrosine kinase in vitro and in vivo by a 2-phenylaminopyrimidine derivative. Cancer Res. 1996;56:100–104.

[26] Lv Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. HIV AIDS (Auckl). 2015;7:95–104. Available from: <https://pubmed.ncbi.nlm.nih.gov/25897264>.

[27] Hartenbaum D. The efficacy of dorzolamide, a topical carbonic anhydrase inhibitor, in combination with timolol in the treatment of patients with open-angle glaucoma and ocular hypertension. Clin Ther. 1996;18:460–465. Available from: [https://doi.org/10.1016/S0149-2918\(96\)80026-9](https://doi.org/10.1016/S0149-2918(96)80026-9).

[28] Kempf DJ, Marsh KC, Kumar G, et al. Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by coadministration with ritonavir. Antimicrob Agents Chemother. 1997;41:654–660.

[29] Jagroop IA, Mikhailidis DP. The effect of tirofiban on fibrinogen/agonist-induced platelet shape change and aggregation. Clin Appl Thromb Off J Int Acad Clin Appl Thromb. 2008;14:295–302.

[30] Prieto-Martínez FD, López-López E, Eurídice Juárez-Mercado K, et al. Computational Drug Design Methods—Current and Future Perspectives. Silico Drug Des. 2019;19–44.

[31] Dua K, Malipeddi VR, Madan J, et al. Norfloxacin and metronidazole topical formulations for effective treatment of bacterial infections and burn wounds. Interv Med Appl Sci. 2016;8:68–

76. Available from: <https://pubmed.ncbi.nlm.nih.gov/28386462>.

[32] Dobson J, Whitley RJ, Pocock S, et al. Oseltamivir treatment for influenza in adults: a meta-analysis of randomised controlled trials. *Lancet*. 2015;385:1729–1737. Available from: [https://doi.org/10.1016/S0140-6736\(14\)62449-1](https://doi.org/10.1016/S0140-6736(14)62449-1).

[33] Colman PM. Zanamivir: an influenza virus neuraminidase inhibitor. *Expert Rev Anti Infect Ther*. 2005;3:191–199. Available from: <https://doi.org/10.1586/14787210.3.2.191>.

[34] Sirotinak F, Zakowski M, Miller V, et al. Efficacy of cytotoxic agents against human tumor xenografts is markedly enhanced by coadministration of ZD1839 (Iressa), an inhibitor of EGFR tyrosine kinase. *Clin Cancer Res*. 2000;6:4885–4892.

[35] Baselga J, Averbuch S. ZD1839 ('Iressa') as an anticancer agent. *Drugs*. 2000;60:33–40.

[36] Falcoz C, Jenkins J, Bye C, et al. Pharmacokinetics of GW433908, a prodrug of amprenavir, in healthy male volunteers. *J Clin Pharmacol*. 2002;42:887–898.

[37] Shen C, Wang Y, Kovalevsky A, et al. Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *FEBS J*. 2010;277:3699–3714.

[38] Robinson B, Riccardi K, Gong Y, et al. BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob Agents Chemother*. 2000;44:2093–2099.

[39] Piliero P. Atazanavir: a novel HIV-1 protease inhibitor. *Expert Opin Investig Drugs*. 2002;11:1295–1301.

[40] Heim M, Sharifi M, Hilger R, et al. Antitumor effect and potentiation or reduction in cytotoxic drug activity in human colon carcinoma cells by the Raf kinase inhibitor (RKI) BAY 43-9006. *Int J Clin Pharmacol Ther*. 2003;41:616–617.

[41] Ahmad T, Eisen T. Kinase inhibition with BAY 43-9006 in renal cell carcinoma. *Clin*

Cancer Res. 2004;10:6388S-6392S.

[42] Wilhelm S, Carter C, Tang L, et al. BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF / MEK / ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res.* 2004;64:7099–7109.

[43] Ng S, Tsao M, Nicklee T, et al. Effects of the epidermal growth factor receptor inhibitor OSI-774, Tarceva, on downstream signaling pathways and apoptosis in human pancreatic adenocarcinoma. *Mol Cancer Ther.* 2002;777–783.

[44] Pollack V, Savage D, Baker D, et al. Inhibition of epidermal growth factor receptor-associated tyrosine phosphorylation in human carcinomas with CP-358,774: dynamics of receptor inhibition in situ and antitumor effects in athymic mice. *J Pharmacol Exp Ther.* 1999;291:739–748.

[45] Doyon L, Tremblay S, Bourgon L, et al. Selection and characterization of HIV-1 showing reduced susceptibility to the non-peptidic protease inhibitor tipranavir. *Antivir Res.* 2005;68:27–35.

[46] Tie Y, Boross P, Wang Y, et al. High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J Mol Biol.* 2004;338:341–352.

[47] Koh Y, Nakata H, Maeda K, et al. Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. *Antimicrob Agents Chemother.* 2003;47:3123–3129.

[48] Xia W, Liu L, Ho P, et al. Truncated ErbB2 receptor (p95ErbB2) is regulated by heregulin through heterodimer formation with ErbB3 yet remains sensitive to the dual EGFR / ErbB2 kinase

inhibitor GW572016. *Oncogene*. 2004;23:646–653.

[49] Wood E, Truesdale A, McDonald O, et al. A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res*. 2004;64:6652–6659.

[50] Hicks C, Gulick RM. Raltegravir: the first HIV type 1 integrase inhibitor. *Clin Infect Dis an Off Publ Infect Dis Soc Am*. 2009;48:931–939.

[51] Gradman AH, Schmieder RE, Lins RL, et al. Aliskiren, a novel orally effective renin inhibitor, provides dose-dependent antihypertensive efficacy and placebo-like tolerability in hypertensive patients. *Circulation*. 2005;111:1012–1018.

[52] Bacon B, Gordon S, Lawitz E et al. Boceprevir for previously treated chronic HCV genotype 1 infection. *N Engl J Med*. 2011;364.

[53] Jarman M, SE Barrie J, Llera M. The 16,17-double bond is needed for irreversible inhibition of human cytochrome p45017alpha by abiraterone (17- (3-pyridyl) androsta-5, 16-dien-3beta-ol) and related steroidal inhibitors. *J Med Chem*. 1998;41:5375–5381.

[54] O'Donnell A, Judson I, Dowsett M, et al. Hormonal impact of the 17alpha-hydroxylase / C (17,20) -lyase inhibitor abiraterone acetate (CB7630) in patients with prostate cancer. *Br J Cancer*. 2004;90:2317–2325.

[55] Jagusch C, Negri M, Hille U, et al. Synthesis, biological evaluation and molecular modeling studies of methyleneimidazole substituted biaryls as inhibitors of human 17alpha-hydroxylase-17,20-lyase (CYP17). Part I: Heterocyclic modifications of the core structure. *Bioorg Med Chem*. 2008;16:1992–2010.

[56] Rodig S, Shapiro G. Crizotinib, a small-molecule dual inhibitor of the c-Met and ALK receptor tyrosine kinases. 1477 - 1490. 2010;11:Curr Opin Investig Drugs.

- [57] Butrynski J, D'Adamo D, Hornick J, et al. Crizotinib in ALK-rearranged inflammatory myofibroblastic tumor. *N Engl J Med*. 2010;363:1727–1733.
- [58] Patick A, Binford S, Brothers M, et al. In vitro antiviral activity of AG7088, a potent inhibitor of human rhinovirus 3C protease. *Antimicrob Agents Chemother*. 1999;43:2444–50.
- [59] Shu C, Zhou H, Afsharvand M, et al. Pharmacokinetic-pharmacodynamic modeling of apratastat: a population-based approach. *J Clin Pharmacol*. 2011;51:472–481.
- [60] Agnelli G, Haas S, Ginsberg JS, et al. A phase II study of the oral factor Xa inhibitor LY517717 for the prevention of venous thromboembolism after hip or knee replacement. *J Thromb Haemost*. 2007;5:746–753.
- [61] Gish RG, Porta C, Lazar L, et al. Phase III Randomized Controlled Trial Comparing the Survival of Patients With Unresectable Hepatocellular Carcinoma Treated With Nilotrexed or Doxorubicin. *J Clin Oncol*. 2007;25:3069–3075. Available from: <https://doi.org/10.1200/JCO.2006.08.4046>.
- [62] Bendell JC, Bauer TM, Lamar R, et al. A Phase 2 Study of the Hsp90 Inhibitor AUY922 as Treatment for Patients with Refractory Gastrointestinal Stromal Tumors. *Cancer Invest*. 2016;34:265–270.
- [63] Sindt JE, Jenkinson RH. Nonintravenous opioids [Internet]. Second Edi. *Pharmacol. Physiol. Anesth. Found. Clin. Appl.* Elsevier Inc.; 2018. Available from: <https://doi.org/10.1016/B978-0-323-48110-6.00018-1>.
- [64] Hutchinson M, Menelaou A, Foster D et al. CYP2D6 and CYP3A4 involvement in the primary oxidative metabolism of hydrocodone by human liver microsomes. *Br J Pharmacol*. 2004;574:287–297.
- [65] Khosravi-Far R, Der C. The Ras signal transduction pathway. *Cancer Metastasis Rev*.

1994;13:67–89.

[66] Hansch C. Use of quantitative structure-activity relationships (QSAR) in drug design. *Pharm Chem J.* 1980;14:678–691.

[67] Valent P, Groner B, Schumacher U, et al. Paul Ehrlich (1854-1915) and His Contributions to the Foundation and Birth of Translational Medicine. *J Innate Immun.* 2016;8:111–120.

[68] Van Drie JH. Computer-aided drug design: the next 20 years. *J Comput Aided Mol Des.* 2007;21:591–601.

[69] Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: A critical review. *Curr Med Chem.* 2013;20:2839–2860.

[70] Jhoti H, Rees S, Solari R. High-throughput screening and structure-based approaches to hit discovery: Is there a clear winner? *Expert Opin Drug Discov.* 2013;8:1449–1453.

[71] Macalino SJY, Gosu V, Hong S, et al. Role of computer-aided drug design in modern drug discovery. *Arch Pharm Res.* 2015;38:1686–1701.

[72] Pereira DA, Williams JA. Origin and evolution of high throughput screening. *Br J Pharmacol.* 2007/07/02. 2007;152:53–61. Available from: <https://pubmed.ncbi.nlm.nih.gov/17603542>.

[73] Medina-Franco JL, Martínez-Mayorga K, Juárez-Gordiano C, et al. Pyridin-2(1H)-ones: A Promising Class of HIV-1 Non-nucleoside Reverse Transcriptase Inhibitors. *ChemMedChem.* 2007;2:1141–1147. Available from: <https://doi.org/10.1002/cmdc.200700054>.

[74] Talele TT, Rigby SAK and AC. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem.* 2010. p. 127–141. Available from: <http://www.eurekaselect.com/node/85602/article>.

[75] Prieto-Martínez F., Medina-Franco J. Charting the bromodomain BRD4: towards the

identification of novel inhibitors with molecular similarity and receptor mapping. *Lett Drug Des Discov.* 2018;15:1002–1011.

[76] Prieto-Martínez F, Medina-Franco J. Diseño de fármacos asistido por computadora: cuando la informática, la química y el arte se encuentran. *TIP Rev Espec En Ciencias Químico-Biológicas.* 2018;21:124–134.

[77] Talevi A. Computer-Aided Drug Design: An Overview. In: Gore M, Jagtap U, editors. *Methods Mol Biol.* New York, NY: Humana Press.; 2018. p. 1–9. Available from: https://doi.org/10.1007/978-1-4939-7756-7_1.

[78] Grinter S, Zou X. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules.* 2014;19:10150–10176.

[79] Sledz P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol.* 2018;93–102:48.

[80] Kalyaanamoorthy S, Chen YPP. Structure-based drug design to augment hit discovery. *Drug Discov. Today.* 2011. p. 831–839.

[81] Freddolino P, Harrison C, Liu Y, et al. Challenges in protein folding simulations: Timescale, representation, and analysis. *Nat Phys.* 2010;6:751–758.

[82] Scapin G. Structural biology and drug discovery. *Curr Pharm Des.* 2006;12:: 2087-2097.

[83] Renaud J-P, Egner U, Hillig RC. The Evolving Role of Structural Biology in Drug Discovery. *Struct. Biol. Drug Discov.* 2020. p. 1–22. Available from: <https://doi.org/10.1002/9781118681121.ch1>.

[84] Usha T, Shanmugarajan D, Goyal AK, et al. Recent Updates on Computer-aided Drug Discovery: Time for a Paradigm Shift. *Curr Top Med Chem.* 2018;17:3296–3307.

[85] Schmidt T, Bergner A, Schwede T. Modelling three-dimensional protein structures for

applications in drug design. *Drug Discov. Today*. 2014. p. 890–897.

[86] Munsamy G, Mahmoud ES. Homology Modeling in Drug Discovery-an Update on the Last Decade. *Lett. Drug Des. Discov.* 2017. p. 1099–1111. Available from: <http://www.eurekaselect.com/node/149119/article>.

[87] Berman HM, Battistuz T, Bhat TN, et al. The Protein Data Bank. *Biol Crystallogr.* 2002;58:899–907.

[88] Velankar S, van Ginkel G, Alhroub Y, et al. PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res.* 2016;44:D385-95.

[89] Choudhary L., Shukla A, Zade S, et al. (C.A.D.D.)-a new modern software-based approach in drug design and discovery. *Int J Pharm Chem.* 2011;1:10–12.

[90] Siju E, Rajalakshmi G, Paulose A, et al. CADD: pharmacological approaches in drug design and drug discovery. *World J Pharm Pharm Sci.* 2017;6:892–908.

[91] Perkins R, Fang H, Tong W, et al. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. *Environ Toxicol Chem.* 2003;22:1666–1679.

[92] Melo C, Braga R, Andrade C. 3D-QSAR approaches in drug design: Perspectives to generate reliable CoMFA models. *Curr Comput Aided Drug Des.* 2014;10:148–159.

[93] Prathipati P, Dixit A, Saxena A. Computer-aided drug design: Integration of structure-based and ligand-based approaches in drug design. *Curr Comput Aided Drug Des.* 2007;3:133–148.

[94] Esposito E, Hopfinger A, Madura J. Methods for applying the quantitative structure activity relationship paradigm. *Methods Mol Biol.* 2004;275:131–214.

[95] Ehrlich P. Ueber den jetzigen Stand der Chemotherapie. *Ber Dtsch Chem Ges.* 1909;42:17–47.

- [96] Vuorinen A, Schuster D. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods*. 2015;71:113–134.
- [97] Yadav D, Khan F, Negi A. Pharmacophore modeling, molecular docking, QSAR, and in silico ADMET studies of gallic acid derivatives for immunomodulatory activity. *J Mol Model*. 2012;18:2513–2525. Available from: 10.1007/s00894-011-1265-3.
- [98] Yadav D, Meena A, Srivastava A, et al. Development of QSAR model for immunomodulatory activity of natural coumarinolignoids. *Drug Des Dev Ther*. 2010;4:173–186.
- [99] Lu W, Zhang R, Jiang H, et al. Computer-Aided Drug Design in Epigenetics. *Front Chem*. 2018;6:57. Available from: <https://www.frontiersin.org/article/10.3389/fchem.2018.00057>.
- [100] Erlanson D, Fesik S, Hubbard R, et al. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov*. 2016;15:605–619. Available from: 10.1038/nrd.2016.109.
- [101] Erlanson DA, Fesik SW, Hubbard RE, et al. Twenty years on: The impact of fragments on drug discovery. *Nat. Rev. Drug Discov*. 2016. p. 605–619.
- [102] Yang H, Sun L, Li W, et al. In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Front Chem*. 2018;6:30.
- [103] Prieto-Martí'nez F, Medina-Franco J. Molecular docking: Current advances and challenges. *TIP Rev Espec en Ciencias Quim*. 2018;21:65–87.
- [104] Saldivar-Gonzalez F, Prieto-Martí'nez F, Medina-Franco J. Descubrimiento y desarrollo de fá'rmacos: un enfoque computacional. *Educ Quimica*. 2017;28:51–58.
- [105] Medina-Franco JL. Discovery and development of lead compounds from natural sources using computational approaches. In: Mukherjee P, editor. *Evidence-based Valid Herb Med*. Elsevier; 2015. p. 455–475.
- [106] Gonzalez-Medina M, Prieto-Martí'nez F, Naveja J, et al. Chemoinformatic expedition of

the chemical space of fungal products. *Future Med Chem.* 2016;6:1113–1126.

[107] Lopez-Vallejo F, Caulfield T, Martinez-Mayorga K, et al. Integrating virtual screening and combinatorial chemistry for accelerated drug discovery. *Comb Chem High Throughput Screen.* 2011;14:475–487.

[108] Moffat J, Vincent F, Lee JA, et al. Opportunities and challenges in phenotypic drugdiscovery: an industry perspective. *Nat Rev Drug Discov.* 2017;16:531–543.

[109] Herrington CS, Coates PJ, Duprex WP. Viruses and disease: emerging concepts for prevention, diagnosis and treatment. *J Pathol.* 2015;235:149–152. Available from: <https://doi.org/10.1002/path.4476>.

[110] Harapan H, Itoh N, Yufika A, et al. Coronavirus disease 2019 (COVID-19): A literature review. *J Infect Public Health.* 2020;13:667–673. Available from: <http://www.sciencedirect.com/science/article/pii/S1876034120304329>.

[111] Malvy D, McElroy AK, de Clerck H, et al. Ebola virus disease. *Lancet.* 2019;393:936–948. Available from: [https://doi.org/10.1016/S0140-6736\(18\)33132-5](https://doi.org/10.1016/S0140-6736(18)33132-5).

[112] Luo G (George), Gao S-J. Global health concerns stirred by emerging viral infections. *J Med Virol.* 2020;92:399–400. Available from: <https://doi.org/10.1002/jmv.25683>.

[113] Parvez MK, Parveen S. Evolution and Emergence of Pathogenic Viruses: Past, Present, and Future. *Intervirol.* 2017;60:1–7. Available from: <https://www.karger.com/DOI/10.1159/000478729>.

[114] Arora HS. A to Z of Zika Virus: A Comprehensive Review for Clinicians. *Glob Pediatr Heal.* 2020;7.

[115] Madihi S, Syed H, Lazar F, et al. A Systematic Review of the Current Hepatitis B Viral Infection and Hepatocellular Carcinoma Situation in Mediterranean Countries. Komatsu H, editor.

Biomed Res Int. 2020;2020:7027169. Available from: <https://doi.org/10.1155/2020/7027169>.

[116] Abdelrahman Z, Li M, Wang X. Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses. *Front Immunol*. 2020;11:2309. Available from: <https://www.frontiersin.org/article/10.3389/fimmu.2020.552909>.

[117] German Advisory Committee Blood (Arbeitskreis Blut) S ‘Assessment of PT by B. Human Immunodeficiency Virus (HIV). *Transfus Med Hemother*. 2016/05/09. 2016;43:203–222. Available from: <https://pubmed.ncbi.nlm.nih.gov/27403093>.

[118] Bean B. Antiviral therapy: current concepts and practices. *Clin Microbiol Rev*. 1992;5:146–182.

[119] Mallipeddi PL, Kumar G, White SW, et al. Recent advances in computer-aided drug design as applied to anti-influenza drug discovery. *Curr Top Med Chem*. 2014;14:1875–1889.

[120] Ganesan A, Barakat K. Applications of computer-aided approaches in the development of hepatitis C antiviral agents. *Expert Opin Drug Discov*. 2017;12:407–425. Available from: <https://doi.org/10.1080/17460441.2017.1291628>.

[121] Santos LH, Ferreira RS, Caffarena ER. Computational drug design strategies applied to the modelling of human immunodeficiency virus-1 reverse transcriptase inhibitors. *Mem Inst Oswaldo Cruz*. 2015;110:847–864. Available from: <https://pubmed.ncbi.nlm.nih.gov/26560977>.

[122] Mottin M, Borba JVVB, Melo-Filho CC, et al. Computational drug discovery for the Zika virus. *Brazilian J Pharm Sci*. 2018;54.

[123] Peele KA, Potla Durthi C, Srihansa T, et al. Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: A computational study. *Informatics Med unlocked*. 2020/05/11. 2020;19:100345. Available from: <https://pubmed.ncbi.nlm.nih.gov/32395606>.

[124] Vásquez-Domínguez E, Armijos-Jaramillo V, Tejera E, et al. Multioutput Perturbation-

Theory Machine Learning (PTML) Model of ChEMBL Data for Antiretroviral Compounds. *Mol Pharm.* 2019;16:4200–4212.

[125] Kirchmair J, Distinto S, Liedl KR, et al. Development of anti-viral agents using molecular modeling and virtual screening techniques. *Infect Disord Drug Targets.* 2011;11:64–93.

[126] Islam R, Parves MR, Paul AS, et al. A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *J Biomol Struct Dyn.* 2020;0:1–12. Available from: <https://doi.org/10.1080/07391102.2020.1761883>.

[127] Vora J, Patel S, Sinha S, et al. Molecular docking, QSAR and ADMET based mining of natural compounds against prime targets of HIV. *J Biomol Struct Dyn.* 2019;37:131–146.

[128] Yuan H, Parrill A. QSAR development to describe HIV-1 integrase inhibition. *J Mol Struct Theochem.* 2000;529:273–282.

[129] Ahlquist P. RNA-Dependent RNA Polymerases, Viruses, and RNA Silencing. *Science* (80-). 2002;296:1270.

[130] Zhu W, Chen CZ, Gorshkov K, et al. RNA-Dependent RNA Polymerase as a Target for COVID-19 Drug Discovery. *SLAS Discov.* 2020;

[131] Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun.* 2019;10:2342. Available from: <https://doi.org/10.1038/s41467-019-10280-3>.

[132] Ahn D-G, Choi J-K, Taylor DR, et al. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Arch Virol.* 2012;157:2095–2104. Available from: <https://doi.org/10.1007/s00705-012-1404-x>.

[133] Posthuma CC, te Velthuis AJW, Snijder EJ. Nidovirus RNA polymerases: Complex

enzymes handling exceptional RNA genomes. *Virus Res.* 2017;234:58–73. Available from: <http://www.sciencedirect.com/science/article/pii/S0168170216308103>.

[134] Subissi L, Posthuma CC, Collet A, et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci.* 2014;111:E3900 LP-E3909. Available from: <http://www.pnas.org/content/111/37/E3900.abstract>.

[135] McDonald SM. RNA synthetic mechanisms employed by diverse families of RNA viruses. *WIREs RNA.* 2013;4:351–367. Available from: <https://doi.org/10.1002/wrna.1164>.

[136] Gorbalenya AE, Pringle F, Zeddani J-L et al. The Palm Subdomain-Based Active Site Is Internally Permuted in Viral RNA-Dependent RNA Polymerases of an Ancient Lineage. *J Molec Biol.* 2002;324.

[137] Venkataraman S, Prasad, B. V. L. S., Selvarajan R. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses.* 2018;10:76.

[138] Gao Y, Yan L, Huang Y, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science.* 2020;7498:1–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32277040>.

[139] Aanouz I, Belhassan A, El-Khatibi K, et al. Moroccan Medicinal plants as inhibitors against SARS-CoV-2 main protease: Computational investigations. *J Biomol Struct Dyn.* 2020;1–9.

[140] Boopathi S, Poma AB, Kolandaivel P. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J Biomol Struct Dyn.* 2020;1–10.

[141] Shu B, Gong P. Structural basis of viral RNA-dependent RNA polymerase catalysis and

- translocation. *Proc Natl Acad Sci.* 2016;113:E4005 LP-E4014. Available from: <http://www.pnas.org/content/113/28/E4005.abstract>.
- [142] Yin W, Mao C, Luan X, et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science.* 2020;1504:eabc1560.
- [143] Campagnola G, Gong P, Peersen O. High-Throughput Screening Identification of Poliovirus RNA-Dependent RNA Polymerase Inhibitors. *Antivir Res.* 2011;91:241–251.
- [144] Li G, De Clercq E. Therapeutic Options for the 2019 Novel Coronavirus (2019-nCoV). *Nat Rev Drug Discov.* 2020;19:149–150.
- [145] Wang Q, Wu J, Wang H, et al. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell.* 2020;182:417-428.e13. Available from: <http://www.sciencedirect.com/science/article/pii/S0092867420306292>.
- [146] Gordon C, Tchesnokov E, Feng J, et al. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J Biol Chem.* 2020;295:4773–4779.
- [147] Eastman RT, Roth JS, Brimacombe KR, et al. Remdesivir: A Review of Its Discovery and Development Leading to Emergency Use Authorization for Treatment of COVID-19. *ACS Cent Sci.* 2020/05/04. 2020;6:672–683. Available from: <https://pubmed.ncbi.nlm.nih.gov/32483554>.
- [148] Murphy B, Perron M, Murakami E et al. The Nucleoside Analog GS-441524 Strongly Inhibits Feline Infectious Peritonitis (FIP) Virus in Tissue Culture and Experimental Cat Infection Studies. *Vet Microbiol.* 2018;219:226–233.
- [149] Choy K-T, Wong A-L, Kaewpreedee P et al. Remdesivir, Lopinavir, Emetine, and Homoharringtonine Inhibit SARS-CoV-2 Replication In Vitro. *Antivir Res.* 2020;178:104786.
- [150] Furuta Y, Komeno T, Nakamura T. Favipiravir (T-705), a broad spectrum inhibitor of viral

RNA polymerase. *Proc Jpn Acad Ser B Phys Biol Sci.* 2017;93:449–463. Available from: <https://pubmed.ncbi.nlm.nih.gov/28769016>.

[151] Furuta Y, Takahashi K, Kuno-Maekawa M et al. Mechanism of Action of T-705 against Influenza Virus. *Antimicrob Agents Chemother.* 2005;49:981.

[152] Furuta Y, Takahashi K, Fukuda Y, et al. In Vitro and In Vivo Activities of AntiInfluenza Virus Compound T-705. *Antimicrob Agents Chemother.* 2002;977–981.

[153] Graci J, Cameron C. Mechanisms of Action of Ribavirin against Distinct Viruses. *Rev Med Virol.* 2006;16:37–48.

[154] Khalili JS, Zhu H, Mak NSA, et al. Novel coronavirus treatment with ribavirin: Groundwork for an evaluation concerning COVID-19. *J Med Virol.* 2020/04/10. 2020;92:740–746. Available from: <https://pubmed.ncbi.nlm.nih.gov/32227493>.

[155] Tong S, Su Y, Yu Y, et al. Ribavirin therapy for severe COVID-19: a retrospective cohort study. *Int J Antimicrob Agents.* 2020;56:106114. Available from: <http://www.sciencedirect.com/science/article/pii/S0924857920302971>.

[156] Warren T, Wells J, Panchal R et al. Protection against Filovirus Diseases by a Novel Broad-Spectrum Nucleoside Analogue BCX4430. *Nature.* 2014;508:402–405.

[157] Zhang W, Stephen P, Thériault J, et al. Novel coronavirus polymerase and nucleotidyl-transferase structures: potential to target new outbreaks. *J Phys Chem Lett.* 2020;11:4430-4435.

[158] Rossmann M, Arnold E, Erickson J, et al. Structure of a Human Common Cold Virus and Functional Relationship to Other Picornaviruses. *Nature.* 1985;317:145–153.

[159] Kim J, Jung YK, Kim C, et al. A Novel Series of Highly Potent Small Molecule Inhibitors of Rhinovirus Replication. *J Med Chem.* 2017;60:5472–5492. Available from: <https://doi.org/10.1021/acs.jmedchem.7b00175>.

- [160] Panjwani A, Strauss M, Chou J et al. Capsid Protein VP4 of Human Rhinovirus Induces Membrane Permeability by the Formation of a Size-Selective Multimeric Pore. *PLoS Pathog.* 2014;10:1–12.
- [161] Reisdorph N, Thomas JJ, Katpally U, et al. Human Rhinovirus Capsid Dynamics is Controlled by Canyon Flexibility. 2003. 314:34–44.
- [162] Agoni C, Munsamy G, Ramhrack P, et al. Human Rhinovirus Inhibition Through Capsid “Canyon” Perturbation: Structural Insights into The Role of a Novel Benzothiophene Derivative. *Cell Biochem Biophys.* 2020;78:3–13.
- [163] De Palma A, Vliegen I, De Clercq E, et al. Selective Inhibitors of Picornavirus Replication. *Med Res Rev.* 2008;28:823–884.
- [164] Diana G, Rudewicz P, Pevear D, et al. Picornavirus Inhibitors: Trifluoromethyl Substitution Provides a Global Protective Effect against Hepatic Metabolism. *J Med Chem.* 1995;38:1355–1371.
- [165] Feil S, Hamilton S, Krippner G, et al. An Orally Available 3-Ethoxybenzoxazole Capsid Binder with Clinical Activity against Human Rhinovirus. *ACS Med Chem Lett.* 2012;3:303–307.
- [166] Shih S-R, Chen S-J, Hakimelahi GH, et al. Selective Human Enterovirus and Rhinovirus Inhibitors: An Overview of Capsid-Binding and Protease-Inhibiting Molecules. *Med Res Rev.* 2004;24:449–474.
- [167] Wright MH, Heal WP, Mann DJ, et al. Protein myristoylation in health and disease. *J Chem Biol.* 2010;3:19–35.
- [168] Boutin JA. Myristoylation. *Cell Signal.* 1997;9:15–35.
- [169] Salah Ud-Din AI, Tikhomirova A, Roujeinikova A. Structure and Functional Diversity of GCN5-Related N-Acetyltransferases (GNAT). *Int. J. Mol. Sci.* . 2016.

- [170] Thinon E, Serwa RA, Broncel M, et al. Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat Commun.* 2014;5:4919. Available from: <https://doi.org/10.1038/ncomms5919>.
- [171] Marc D, Masson G, Girard M, et al. Lack of myristoylation of poliovirus capsid polypeptide VP0 prevents the formation of virions or results in the assembly of noninfectious virus particles. *J Virol.* 1990;
- [172] Marc D, Drugeon G, Haenni A-L, et al. Role of myristoylation of poliovirus capsid protein VP4 as determined by site-directed mutagenesis of its N-terminal sequence. *EMBO J.* 1989;
- [173] Marc D, Girard M, Van Der Werf S. A Gly1 to Ala substitution in poliovirus capsid protein VP0 blocks its myristoylation and prevents viral assembly. *J Gen Virol.* 1991;
- [174] Moscufo N, Simons J, Chow M. Myristoylation is important at multiple stages in poliovirus assembly. *J Virol.* 1991;
- [175] Wolinsky E. Tuberculosis. In: Wyngaarden J, Smith Jr., L, Bennett J, editors. *Cecil Textb Med.* 19th ed. Philadelphia, PA: W.B. Saunders Company; 1992. p. 1733–1742.
- [176] Sensi P, Grass I. Antimycobacterial agents. In: Burger A, Wolff M, editors. *Burger's Med Chem Drug Discov.* fifth. New York, NY: John Wiley & Sons; 1996. p. 575–635.
- [177] World Health Organization W. *Global tuberculosis report.* Geneva; 2019.
- [178] Ekins S, Freundlich JS, Choi I, et al. Computational databases, pathway and cheminformatics tools for tuberculosis drug discovery. *Trends Microbiol.* 2011;19:65–74.
- [179] Aleksandrov A, Myllykallio H. Advances and challenges in drug design against tuberculosis: application of in silico approaches. *Expert Opin Drug Discov.* 2019;14:35–46. Available from: <https://doi.org/10.1080/17460441.2019.1550482>.

- [180] Taira J, Ito T, Nakatani H et al. In silico structure-based drug screening of novel antimycobacterial pharmacophores by DOCK-GOLD tandem screening. *Int J Mycobacteriol.* 2017;6:142–148.
- [181] Billones J, Carrillo M, Organo V et al. Toward antituberculosis drugs: in silico screening of synthetic compounds against *Mycobacterium tuberculosis* D-transpeptidase 2. *Drug Des Devel Ther.* 2016;10:1147–1157.
- [182] Billones JB, O MC, Sy JBA, et al. In silico discovery and in vitro activity of inhibitors against *Mycobacterium tuberculosis* 7,8-diaminopelargonic acid synthase (Mtb BioA). *2017;563–574.*
- [183] ChemBridge database. 2018.
- [184] Mehra R, Rani C, Mahajan P et al. Computationally guided identification of novel *Mycobacterium tuberculosis* GlmU inhibitory leads, their optimization, and in vitro validation. *ACS Comb Sci.* 2016;18:100–116.
- [185] Saxena S, Devi P, Soni V et al. Identification of novel inhibitors against *Mycobacterium tuberculosis* L-alanine dehydrogenase (MTB-AlaDH) through structure-based virtual screening. *J MolGraph Model.* 2014;47:37–43.
- [186] Saxena S, Samala G, Sridevi J et al. Design and development of novel *Mycobacterium tuberculosis* L-alanine dehydrogenase inhibitors. *Eur J Med Chem.* 2015;92:401–414.
- [187] Reshma R, Saxena S, Bobesh K et al. Design and development of new class of *Mycobacterium tuberculosis* L-alanine dehydrogenase inhibitors. *Bioorg Med Chem.* 2016;24:4499–4508.
- [188] Gudzera O, Golub A, Bdzhola V et al. Discovery of potent anti-tuberculosis agents targeting leucyl-tRNA synthetase. *Bioorg Med ChemMed Chem.* 2016;24:1023–1031.

- [189] Gudzera O, Golub A, Bdzhola, VG et al. Identification of *Mycobacterium tuberculosis* leucyl-tRNA synthetase (LeuRS) inhibitors among the derivatives of 5-phenylamino-2H-b[1,2,4]triazin-3-one. *J Enzym Inhib Med Chem*. 2016;31:201–207.
- [190] Buryska T, Daniel L, Kunka A et al. Discovery of novel haloalkane dehalogenase inhibitors. *Appl Env Microbiol*. 2016;82:1958–1965.
- [191] Petersen G, Saxena S, Renuka J et al. Structure-based virtual screening as a tool for the identification of novel inhibitors against *Mycobacterium tuberculosis* 3-dehydroquinate dehydratase. *J Mol Graph Model*. 2015;60:124–131.
- [192] Lone M, Athar M, Gupta V et al. Prioritization of natural compounds against *Mycobacterium tuberculosis* 3-dehydroquinate dehydratase: a combined in-silico and in-vitro study. *Biochem Biophys Res Commun*. 2017;491:1105–1111.
- [193] Sharma K, Tanwar O, Sharma S et al. Structural comparison of Mtb-DHFR and h-DHFR for design, synthesis and evaluation of selective non-pteridine analogues as antitubercular agents. *Bioorg Chem*. 2018;80:319–333.
- [194] Chiarelli L, Mori M, Barlocco D et al. Discovery and development of novel salicylate synthase (MbtI) furanic inhibitors as antitubercular agents. *Eur J Med Chem*. 2018;155:754–763.
- [195] Rohilla A, Khare G, Tyagi A. Virtual screening, pharmacophore development and structure based similarity search to identify inhibitors against IdeR, a transcription factor of *Mycobacterium tuberculosis*. *Sci Rep*. 2017;7:4653.
- [196] Singh S, Khare G, Bahal R et al. Identification of *Mycobacterium tuberculosis* BioA inhibitors by using structure-based virtual screening. *Drug Des Devel Ther*. 2018;12:1065–1079.
- [197] Dharra R, Talwar S, Singh Y et al. Rational design of drug-like compounds targeting *Mycobacterium marinum* MefF protein. *PLoS One*. 2017;12:e0183060.

- [198] Korycka-Machala M, Nowosielski M, Kuron A et al. Naphthalimides selectively inhibit the activity of bacterial, replicative DNA ligases and display bactericidal effects against tubercle bacilli. *Molecules*. 2017;22:1.
- [199] Jeankumar V, Reshma R, Vats R et al. Engineering another class of anti-tubercular lead: hit to lead optimization of an intriguing class of gyrase ATPase inhibitors. *Eur J Med Chem*. 2016;122:216–231.
- [200] Djaout K, Singh V, Boum Y et al. Predictive modeling targets thymidylate synthase ThyX in *Mycobacterium tuberculosis*. *Sci Rep*. 2016;6:27792.
- [201] Luciani R, Saxena P, Surade S et al. Virtual screening and X-ray crystallography identify non-substrate analog inhibitors of flavin-dependent thymidylate synthase. *J Med Chem*. 2016;59:9269–9275.
- [202] Scocchera E, Wright DL. The Antifolates BT - Antibacterials: Volume II. In: Fisher JF, Mobashery S, Miller MJ, editors. Cham: Springer International Publishing; 2018. p. 123–149. Available from: https://doi.org/10.1007/7355_2017_16.
- [203] Visentin M, Zhao R, Goldman ID. The Antifolates. *Hematol Oncol Clin North Am*. 2012;26:629–648.
- [204] Hajian B, Scocchera E, Shoen C, et al. Drugging the Folate Pathway in *Mycobacterium tuberculosis*: The Role of Multi-targeting Agents. *Cell Chem Biol*. 2019;26:781-791.e6. Available from: <http://www.sciencedirect.com/science/article/pii/S2451945619300686>.
- [205] Nixon M, Saionz K, Koo M, et al. Folate pathway disruption leads to critical disruption of methionine derivatives in *Mycobacterium tuberculosis*.
- [206] Kumar A, Guardia A, Colmenarejo G, et al. A focused screen identifies antifolates with activity on *Mycobacterium tuberculosis*. *ACS Infect Dis*. 2015;1:604–614.

- [207] Zheng J, Rubin E, Bifani P, et al. para-Aminosalicylic acid is a prodrug targeting dihydrofolate reductase in Mycobacterium tuberculosis. *J Biol Chem*. 2013;288:23447–23456.
- [208] Choi M, Karunaratne K, Kohen A. Flavin-dependent thymidylate synthase as a new antibiotic target. *Molecules*. 2016;21.
- [209] Barr J. A short history of dapsone, or an alternative model of drug development. *Pharm Hist*. 2011;53:123–147.
- [210] Wallace Jr. R, Nash D, Steele L, et al. Susceptibility testing of slowly growing mycobacteria by a microdilution MIC method with 7H9 broth. *J Clin Microbiol*. 1986;24:976–981.
- [211] Chakraborty S, Gruber T, Barry 3rd C, et al. Para-aminosalicylic acid acts as an alternative substrate of folate metabolism in Mycobacterium tuberculosis. *Science* (80-). 2013;339:88–91.
- [212] Vilchèze C, Jacobs Jr., W. The combination of sulfamethoxazole, trimethoprim, and isoniazid or rifampin is bactericidal and prevents the emergence of drug resistance in Mycobacterium tuberculosis. *Agents ChemotherAntimicrob*. 2012;56:5142–5148.
- [213] Koul A, Dendouga N, Vergauwen K, et al. Diarylquinolines target subunit c of mycobacterial ATP synthase. :323–324.
- [214] Andries K, Verhasselt P, Guillemont J, et al. A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science*. 2005;307:223–227.
- [215] Koul A, Vranckx L, Dendouga N, et al. Diarylquinolines Are Bactericidal for Dormant Mycobacteria as a Result of Disturbed ATP Homeostasis. 2008;
- [216] Haagsma AC, Podasca I, Koul A, et al. Probing the interaction of the diarylquinoline TMC207 with its target mycobacterial ATP synthase. *PLoS One*. 2011;6.
- [217] Haagsma AC, Abdillahi-ibrahim R, Wagner MJ, et al. Selectivity of TMC207 towards

- Mycobacterial ATP Synthase Compared with That towards the Eukaryotic Homologue □.
2009;53:1290–1292.
- [218] Cohen J. Infectious disease. Approval of novel TB drug celebrated—with restraint. *Science*. 2013;339:130.
- [219] D. J. Tuberculosis Success. *Nat Rev Drug Discov*. 2013;12:175–176.
- [220] Preiss L, Langer JD, Yildiz Ö, et al. Structure of the mycobacterial ATP synthase Forotor ring in complex with the anti-TB drug bedaquiline. *Sci Adv*. 2015;1:1–8.
- [221] Meier T, Polzer P, Diederichs K, et al. Structure of the rotor ring of F-Type Na⁺-ATPase from *Ilyobacter tararicus*. *Science*. 2005;308:659-662.
- [222] Petrella S, Cambau E, Chauffour A, et al. Genetic Basis for Natural and Acquired Resistance to the Diarylquinoline R207910 in Mycobacteria. 2006;50:2853–2856.
- [223] Huitric E, Verhasselt P, Andries K, et al. In Vitro Antimycobacterial Spectrum of a Diarylquinoline ATP Synthase Inhibitor. 2007;51:4202–4204.
- [224] Segala E, Sougakoff W, Jarlier V, et al. Synthase : New Insights into the Binding of Synthase C-Ring Structure New Mutations in the Mycobacterial ATP Synthase : New Insights into the Binding of the Diarylquinoline TMC207 to the ATP Synthase. 2012;
- [225] Junge W, Sielaff H, Engelbrecht S. Torque generation and elastic power transmission in the rotary F O F 1-ATPase. *Nature*. 2009;459:364–370.
- [226] Noji H, Yasuda R, Yoshida M, et al. Direct observation of the rotation of F1-ATPase. *Nature*. 1997. p. 299–302.
- [227] Walker J. The ATP synthase: the understood, the uncertain and the unknown. *Biochem Soc Trans*. 2013;41:1–16.
- [228] Walker JE. The ATP synthase: the understood, the uncertain and the unknown. *Biochem*

Soc Trans. 2013;41:1–16.

[229] Boyer PD. the Atp Synthase—a Splendid Molecular Machine. *Annu Rev Biochem.* 1997;66:717–749.

[230] Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, Fr Int Agency Res Cancer; 2013 Available from <http://globocan.iarc.fr>, acedido em 28/04/2016. 2012;

[231] Holohan C, Van Schaeybroeck S, Longley DB, et al. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer.* 2013;13:714–726. Available from: <https://doi.org/10.1038/nrc3599>.

[232] Housman G, Byler S, Heerboth S, et al. Drug resistance in cancer: An overview. *Cancers (Basel).* 2014.

[233] Bode AM, Dong Z. Precision oncology- the future of personalized cancer medicine? *npj Precis Oncol.* 2017;1:2.

[234] Shin SH, Bode AM, Dong Z. Addressing the challenges of applying precision oncology. *npj Precis Oncol.* 2017;1:28.

[235] Palumbo MO, Kavan P, Miller WH, et al. Systemic cancer therapy: Achievements and challenges that lie ahead. *Front. Pharmacol.* 2013.

[236] Block KI, Gyllenhaal C, Lowe L, et al. Designing a broad-spectrum integrative approach for cancer prevention and treatment. *Semin. Cancer Biol.* 2015. p. S276–S304.

[237] Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. *Science.* 2013. p. 1546–1558.

[238] Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. *CA Cancer J Clin.* 2016;66:75–88.

[239] Prajapat P, Agarwal S, Talesara G, et al. Significance of Computer Aided Drug Design and

3D QSAR in Modern Drug Discovery. *J Med Org Chem J Med Org Chem*. 2017;11.

[240] Liu X, Xie H, Luo C, et al. Discovery and SAR of thiazolidine-2,4-dione analogues as insulin-like growth factor-1 receptor (IGF-1R) inhibitors via hierarchical virtual screening. *J Med Chem*. 2010;53:2661–2665.

[241] Chiang Y, Kuo C, Wu Y, et al. Generation of ligand-based pharmacophore model and virtual screening for identification of novel tubulin inhibitors with potent anticancer activity. *J Med Chem*. 2009;52:4221–4233.

[242] Noha S, Atanasov A, Schuster D, et al. Discovery of a novel IKK- β inhibitor by ligand-based virtual screening techniques. *Bioorg Med Chem Lett*. 2011;21:577–583.

[243] He R, Yu Z, Zhang R, et al. Protein tyrosine phosphatases as potential therapeutic targets. *Acta Pharmacol Sin*. 2014;35:1227–1246.

[244] Fujikawa A, Sugawara H, Tanaka T, et al. Targeting PTPRZ inhibits stem cell-like properties and tumorigenicity in glioblastoma cells. *Sci Rep*. 2017;7:1–17.

[245] Lazo J, Sharlow E. Drugging Undruggable Molecular Cancer Targets. *Annu Rev Pharmacol Toxicol*. 2016;56:23–40.

[246] Navis AC, Van Den Eijnden M, Schepens JTG, et al. Protein tyrosine phosphatases in glioma biology. *Acta Neuropathol*. 2010;119:157–175.

[247] Barr A. Protein tyrosine phosphatases as drug targets: strategies and challenges of inhibitor development. *Futur Med Chem*. 2010;2:1563–1576.

[248] Andersen JN et al. Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Mol Cell Biol*. 2001;21:7117–7136.

[249] Vezzalini M et al. Expression of transmembrane protein tyrosine phosphatase gamma (PTPgamma) in normal and neoplastic human tissues. *Histopathology*. 2007;50:615–628.

- [250] Müller S, Kunkel P, Lamszus K, et al. A role for receptor tyrosine phosphatase ζ in glioma cell migration. *Oncogene*. 2003;22:6661–6668. Available from: <https://doi.org/10.1038/sj.onc.1206763>.
- [251] Ulbricht U, Brockmann MA, Aigner A, et al. Expression and function of the receptor protein tyrosine phosphatase zeta and its ligand pleiotrophin in human astrocytomas. *J Neuropathol Exp Neurol*. 2003;62:1265–1275.
- [252] Fujikawa A, Nagahira A, Sugawara H, et al. Small-molecule inhibition of PTPRZ reduces tumor growth in a rat model of glioblastoma. *Sci Rep*. 2016;6:1–16.
- [253] Foehr ED, Lorente G, Kuo J, et al. Targeting of the receptor protein tyrosine phosphatase β with a monoclonal antibody delays tumor growth in a glioblastoma model. *Cancer Res*. 2006;66:2271–2278.
- [254] Wang V, Davis DA, Veeranna RP, et al. Characterization of the activation of protein tyrosine phosphatase, receptor-type, Z polypeptide 1 (PTPRZ1) by hypoxia inducible factor-2 alpha. *PLoS One*. 2010;5.
- [255] Tautz L, Critton DA, Grotegut S. Protein tyrosine phosphatases: structure, function, and implication in human disease. *Methods Mol Biol*. 2013;1053:179–221. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23860656>.
- [256] Agoni C, Ramharack P, Soliman MES. Allosteric inhibition induces an open WPD-loop: a new avenue towards glioblastoma therapy. *RSC Adv*. 2018;8:40187–40197. Available from: <http://xlink.rsc.org/?DOI=C8RA08427K>.
- [257] Barr AJ, Ugochukwu E, Lee WH, et al. Large-Scale Structural Analysis of the Classical Human Protein Tyrosine Phosphatome. *Cell*. 2009;136:352–363.
- [258] Hoff RH, Hengge AC, Wu L, et al. Effects on general acid catalysis from mutations of the

invariant tryptophan and arginine residues in the protein tyrosine phosphatase from *Yersinia*. *Biochemistry*. 2000;39:46–54.

[259] Keng YF, Wu L, Zhang ZY. Probing the function of the conserved tryptophan in the flexible loop of the *Yersinia* protein-tyrosine phosphatase. *Eur J Biochem*. 1999;259:809–814.

[260] Combs AP, Yue EW, Bower M, et al. Structure-based design and discovery of protein tyrosine phosphatase inhibitors incorporating novel isothiazolidinone heterocyclic phosphotyrosine mimetics. *J Med Chem*. 2005;48:6544–6548.

[261] Brennan PJ, Brigl M, Brenner MB. Invariant natural killer T cells: an innate activation scheme linked to diverse effector functions. *Nat Rev Immunol*. 2013;13:101–17.

[262] Bendelac A, Savage PB, Teyton L. The Biology of NKT Cells. *Annu Rev Immunol*. 2007;25:297–336.

[263] Rossjohn J, Pellicci DG, Patel O, et al. Recognition of CD1d-restricted antigens by natural killer T cells. *Nat Rev Immunol*. 2012;12:845–857.

[264] Kawano T, Cui J, Koezuka Y, et al. CD1d-restricted and TCR-mediated activation of α 14 NKT cells by glycosylceramides. *Science*. 1997;278:1626–1629.

[265] Morita M, Motoki K, Akimoto K, et al. Structure-activity relationship of α -galactosylceramides against B16-bearing mice. *J Med Chem*. 1995;38:2176–2187.

[266] Bedard M, Salio M, Cerundolo V. Harnessing the power of invariant natural killer T cells in cancer immunotherapy. *Front Immunol*. 2017;8.

[267] Cerundolo V, Silk J, Masri S, et al. Harnessing invariant NKT cells in vaccination strategies. *Nat Rev Immunol*. 2009;9:28–38.

[268] Cerundolo V, Salio M. Harnessing NKT cells for therapeutic applications. *Curr Top Microbiol Immunol*. 2007;314:325–340.

- [269] Lam PY, Nissen MD, Mattarollo SR. Invariant natural killer T cells in immune regulation of blood cancers: Harnessing their potential in immunotherapies. *Front Immunol.* 2017;8:1–12.
- [270] Griewank K, Borowski C, Rietdijk S, et al. Homotypic interactions mediated by Slamf1 and Slamf6 receptors control NKT cell lineage development. *Immunity.* 2007;27:751–62.
- [271] Lynch L, Michelet X, Zhang S, et al. Regulatory iNKT cells lack expression of the transcription factor PLZF and control the homeostasis of T(reg) cells and macrophages in adipose tissue. *Nat Immunol.* 2015;16:85–96.
- [272] Lynch L, Nowak M, Varghese B, et al. Adipose tissue invariant NKT cells protect against diet-induced obesity and metabolic disorder through regulatory cytokine production. *Immunity.* 2012;37:574–87.
- [273] Nieuwenhuis EE, Matsumoto T, Exley M, et al. CD1d-dependent macrophage-mediated clearance of *Pseudomonas aeruginosa* from lung. *Nat Med.* 2002;8:588–93.
- [274] Schmiege J, Yang G, Franck RW, et al. Glycolipid presentation to natural killer T cells differs in an organ-dependent fashion. *Proc Natl Acad Sci U S A.* 2005;102:1127–32.
- [275] Monteiro M, Graca L. iNKT Cells: Innate lymphocytes with a diverse response. *Crit Rev Immuno.* 2014;34:81–90.
- [276] Yamashita K, Arimoto A, Nishi M, et al. Application of iNKT Cell-targeted Active Immunotherapy in Cancer Treatment. *Anticancer Res.* 2018;38:4233–4239. Available from: <http://ar.iijournals.org/lookup/doi/10.21873/anticanres.12719>.
- [277] Crowe NY, Smyth MJ, Godfrey DI. A critical role for natural killer T cells in immunosurveillance of methylcholanthrene-induced sarcomas. *J Exp Med.* 2002;196:119–27.
- [278] Cerundolo V, Salio M. Harnessing NKT cells for therapeutic applications. *Curr Top Microbiol Immunol.* 2007;314:325–340.

- [279] Natori T, Morita M, Akimoto K, et al. Agelasphins, novel antitumor and immunostimulatory cerebrosides from the marine sponge agelas-mauritanus. *Tetrahedron*. 1994;50:2771–84.
- [280] Kobayashi E, Motoki K, Uchida T, et al. KRN7000, a novel immunomodulator, and its antitumor activities. *Oncol Res*. 1995;7:529–534.
- [281] Kronenberg M, Gapin L. The unconventional lifestyle of NKT cells. *Nat Rev Immunol*. 2002;2:557–68.
- [282] Chennamadhavuni D, Saavedra-Avila NA, Carreño LJ, et al. Dual Modifications of α -Galactosylceramide Synergize to Promote Activation of Human Invariant Natural Killer T Cells and Stimulate Anti-tumor Immunity. *Cell Chem Biol*. 2018;25:P571-584.E8.

CHAPTER 3

3.0 Principles of Computational Chemistry and Molecular Modeling Techniques as Employed in Understanding the Interaction Dynamics of Biological Molecules with Therapeutic Agents

3.1 Introduction

As a fast-emerging field, computational chemistry entails the modeling and computational simulation of systems such as polymers, drugs, organic and inorganic molecules, and biomolecules, amongst many others [1,2]. Initially introduced in the 1960s, computational chemistry has gained immense prominence in the last several decades due to the remarkable improvement in computer hardware and the continuous development of advanced computational software [3–5]. Technological advancement has substantially increased computing power since faster and more efficient computational algorithms are being developed. Computational chemistry has simplified the hitherto complex biological and chemical problems by integrating physics, chemistry, biology, and mathematics. As applied in this research, computational chemistry has enabled the investigation of the structural mechanisms of actions and inhibitory activities associated with the binding of some small molecule inhibitors against some therapeutic targets against viruses, Mycobacterium tuberculosis, and cancer. Experimental techniques are associated with the high cost and time-consuming, based on proven theories, computational chemistry, and molecular modeling, which can address these challenges using software and computational tools based on empirical information. To ensure the reliability of molecular modeling procedures, models prepared must adhere to standard computational techniques and consider all the factors that ensure that the simulated models assume similarity to natural systems. Molecular modeling

methods incorporate the appropriate application of computational chemistry and molecular modeling, which in turn have been developed into several techniques such as molecular dynamics (MD) simulations, free binding energy analysis, molecular docking, 3D-QSAR, cheminformatics, and drug repurposing techniques, among several others as applied in the studies of this thesis. This chapter principally discusses the computational theories governing many of these computational techniques as employed in this thesis. The unique approaches that allow for the description of the interaction between atoms and molecules in the application of computational models to study chemical and biological systems include; quantum mechanics (QM) [6–13], molecular mechanics (MM)[14–16], hybrid QM/MM[17–20] and the ‘Our own N-layered Integrated Molecular Orbital and molecular mechanics’ (ONIOM) [21,22]. Also described are molecular force fields (MFF). These approaches enable a versatile and apt description of chemical, biological systems so long as the systems' parameters are impeccably attuned and balanced among each other [4]. Collectively, these approaches have allowed for the biological and chemical challenges, including but not limited to systems biology [23], enzyme catalysis [24], bioenergy transduction [25] biomimetics, and ligand design [26]. The subsequent sessions describe these approaches as applied to achieve the objectives outlined in this thesis. The ensuing sessions also describe other computational methods as employed in this thesis.

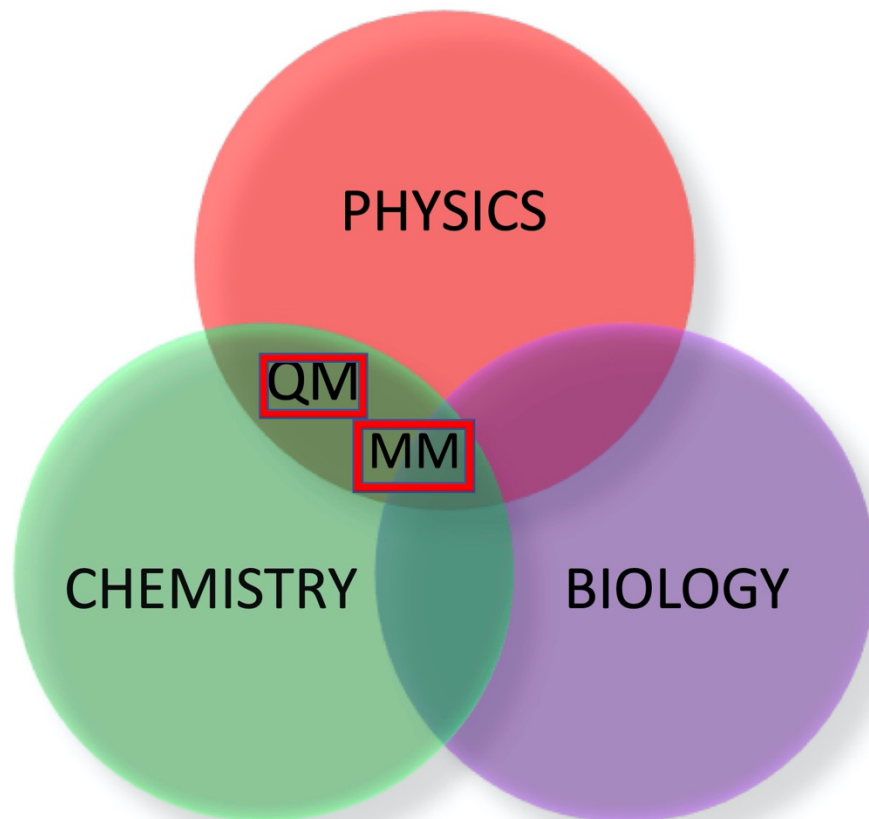


Figure 3.1: The scientific domains in which Applications of Quantum and Molecular Mechanics fit into

3.2 Quantum Mechanics

Quantum mechanisms describe biological and chemical systems by partitioning the atoms that make up the system into two components: the nuclei and the electrons that enclose the nuclei. A characteristic feature of QM description is that it considers all shifts in electron density that emanates from charge transfers, polarization, and many-body contributions. Therefore, as applied in molecular modeling, QM allows for the investigation of the electronic behavior of molecules and spatial interactions. The application of QM in addressing biological challenges has been widely explored in areas such as structured based drug design, protein structure and solvation, and refinement of biological structures by X-ray spectroscopy and NMR [27]. As employed in this

thesis, QM methods also allow for the unraveling of the fundamental physical basis of the catalysis of enzymes, a phenomenon that inspired the Noble prize winning work in chemistry by Karplus et al. in 2013 [28–30]. QM approaches do not require empirical parameters; as such, their application is not restricted to specific molecule class. Comparatively, although QM has these advantages, it demands increased computational effort and resources relative to MM. QM employs Schrödinger’s wave function to measure the system’s energetics, whereas electron mapping is performed using the electron density approach [31,32]. On the other hand, the electron density in QM is also estimated using the Born-Oppenheimer approximation theory [33]. The Schrödinger’s wave function and the Born-Oppenheimer approximation theory are discussed in detail in the following sections. The model proposed by Schrödinger’s wave function forms the basis for the current-day concept of an atom.

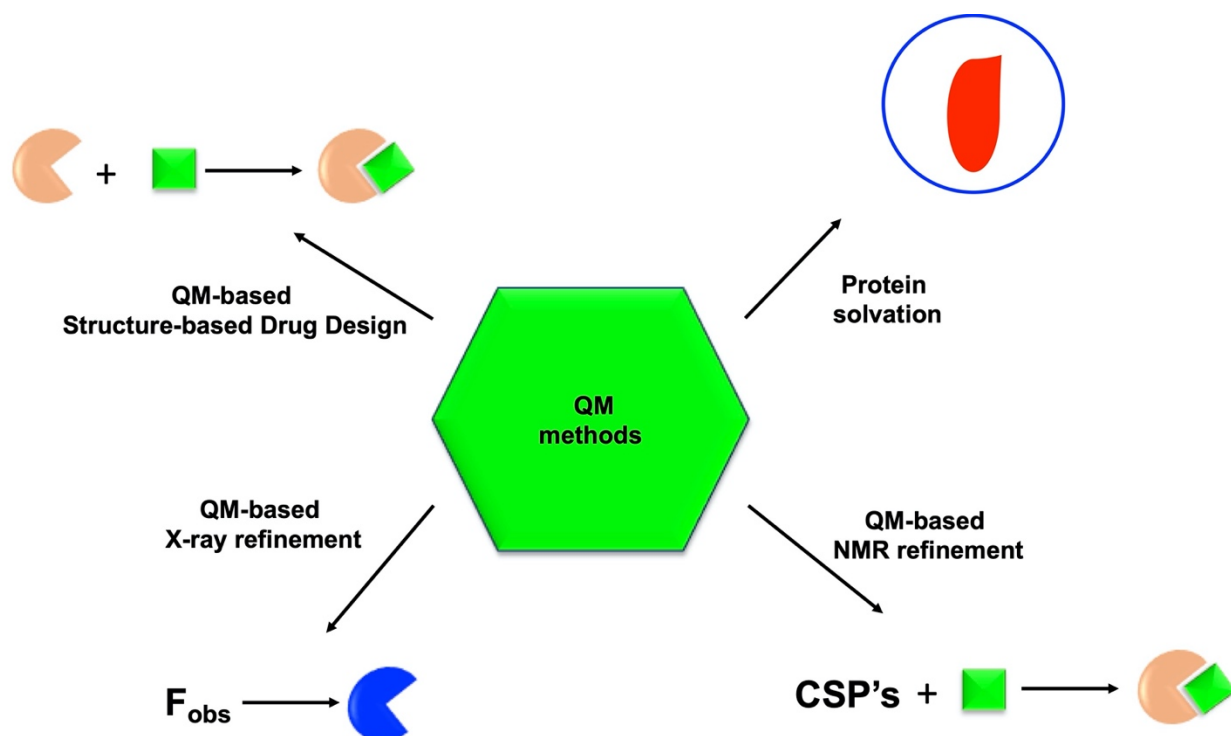


Figure 3.2: Schematic showing the Application of QM Methods to Biological Problems[27]
CSP: Chemical Shift Perturbation

3.2.1 The Schrödinger's wave function

Inspired by De Broglie's doctoral thesis in 1925, Schrödinger initiated a study to obtain a wave equation applied to the hydrogen atom. Schrödinger added his ideas, such as his interest in establishing stationary waves in three dimensions, analogous to sound waves in cavities [34]. In wave mechanics, the Schrödinger equation is used as a wave function for a particle and enables creating a complete model for an atom [34]. As an improvement in De Broglie's work, the Schrödinger wave equation addresses De Broglie's theory's limitations, such as its inability to cater to accelerated motions and rotations [34]. Schrödinger's wave equation also attempted to reconcile the theoretical assumptions propounded by Heisenberg and those postulated by De Broglie while also showcasing a 3D orbital model for each energy sub-level described and thus assumed that electrons could be described as matter waves [34]. This eventually allowed for the molecular geometry of numerous chemical substances to be determined [34]. The Schrödinger's wave equation also addresses significant challenges associated with Bohr's model, which treated electrons as particles that existed in precisely-defined orbits. The basic form of Schrödinger's equation is described as follows;

$$\mathbf{H}\psi = \mathbf{E}\psi \quad (\text{Eq. 3.2.1})$$

The equation's wave function is denoted by Ψ , whereas the H and E represent the Hamiltonian operator and electron binding energy, respectively. The wave function cannot predict the exact location of electrons, although it can describe the probability of the electrons' locations. For the Schrödinger equation to present a physically relevant model, the wave function must be normalized, continuous, single-valued, and anti-symmetric. The molecular Hamiltonian operator

(H) is obtained by a sum of the potential energy (V) and the kinetic energy (T) and mathematically described as follows;

$$\mathbf{H} = \mathbf{T} + \mathbf{V} \quad (\text{Eq. 3.2.2})$$

However, the Hamiltonian operator constitutes all kinetic and potential energy operators for all the electrons and nuclei in a molecule when relativistic effects are neglected, and particles are described as end masses. Subsequently, the advanced form of the equation is represented as follows;

$$\mathbf{H} = \left[-\frac{\hbar^2}{8\pi^2} \sum_i \frac{1}{m_j} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + \sum_i \sum_j \left\langle \frac{eie_i}{r_{ij}} \right\rangle \right] \quad (\text{Eq. 3.2.3})$$

Where i and j represent the electrons, m_j represent mass of electron j while r_{ij} being the distance between i and j electrons.

Due to the highly complex nature of the Schrödinger's wave equation and its associated limitations in solving for molecular systems which usually consists of thousands of atoms, the Born-Oppenheimer approximation theory as described in the ensuing session has been shown to compensate for this limitation in QM [35–37]

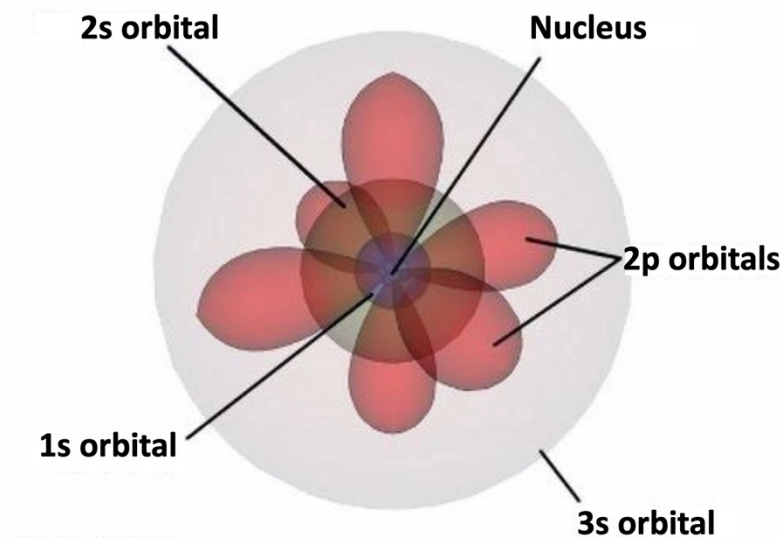


Figure 3.3: Schrödinger's 1927 model of an atom

3.2.2 The Born-Oppenheimer Approximation

As one of the underlying fundamental concepts in quantum mechanics that describe the quantum states of molecules, the Born-Oppenheimer approximation allows for the separation of nuclei and electrons motions [38,39]. In doing so, the Born-Oppenheimer approximation can address the significant limitation of Schrödinger's wave equation. Propounded by Max Born and J. Robert Oppenheimer in 1927, the approximation thus describes the wave function's uncoupling to that of the electrons [18,27,40–42]. Electrons are considered to be lighter in weight than nuclei; thus, they have increased velocity and move instantaneously relative to the movement of nuclei [43]. The difference in weight between the electrons and nuclei is accounted for by the Born-Oppenheimer approximation. Therefore, the electrons can almost instantaneously respond to nuclei displacement [44–47]. Therefore the distribution of electrons within molecules is described by nuclei location [44,45]. The Born-Oppenheimer approximation makes it possible for the Schrödinger equation to be solved for the kinetic energy of the electrons alone since the nuclei's kinetic energy will remain

constant. The difference in velocities of the nuclei and electrons allows for the Born-Oppenheimer approximation to be applied, minimizing the complexity of the wave function of the Hamiltonian equation [40]. The approximation of the Hamiltonian is shown below:

$$\mathbf{H} = -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 - \frac{\hbar^2}{2} \sum_A \frac{1}{M_A} \nabla_A^2 - \sum_A \sum_i \frac{Z_A e^2}{r_{Ai}} + \sum_i \sum_{j>i} \frac{e^2}{r_{ij}} + \sum_A \sum_{B>A} \frac{Z_A Z_B e^2}{R_{AB}} \quad (\text{Eq. 3.2.4})$$

Where A and B denote the nuclei, while i and j represent the electrons. M_A represent the nucleus A mass while m_e denote an electron mass. R_{AB} represents a nuclei A and B inter-distance with r_{ij} being the distance between i and j electrons. Z_A is nucleus A charge while r_{Ai} represents nucleus A and electron i inter-distance.

The complexity of the Hamiltonian equation's wave function is minimized due to the variance in the velocities of the nuclei and the electrons when the Born-Oppenheimer is applied [40]. When simplified as a result of the instantaneous response of electrons to nuclear motions, the wave function is subsequently expressed as follows;

$$\psi(\mathbf{r}_{elec}) = \psi(\mathbf{r}_{elec})(\mathbf{r}_{nucl}) \quad (\text{Eq. 3.2.5})$$

Which is subsequently changed to

$$\mathbf{H}_{EN} \psi(\mathbf{r}_{el}) = E_{EN} \psi(\mathbf{r}_{el}) \quad (\text{Eq. 3.2.6})$$

\mathbf{H}_{EN} symbolizes a difference between terms based on their dependence on fixed nuclear positions (\mathbf{V}_{NN}) or their dependence on the non-fixed electron positions (\mathbf{H}_{el}). An energy term E_{EN} is integrated, which is derived from two sources; one being the fluctuating electron co-ordinates \mathbf{E} and fixed nuclear co-ordinates \mathbf{N} :

$$\mathbf{H}_{el} + V_{NN}) \psi(\mathbf{r}_{el}) = E_{EN} \psi(\mathbf{r}_{el}) \quad (\text{Eq. 3.2.7})$$

The application of Born-Oppenheimer approximation to electrons in the ground states is more accurate than when applied to electrons in the excited states [31,43]. Upon solving the equation, it enables the construction of a molecular potential energy curve and a potential energy surface of a polyatomic molecule [43] while also permitting the identification and assessment of the equilibrated conformations given molecule with the lowest point on the surface [43,46].

3.2.3 Potential Energy Surface Function

A potential energy surface is an efficient mathematical function that gives a molecule's energy as a function of its geometry [31]. Originating from the Born-Oppenheimer approximation, the potential energy surface functions play a crucial role in modeling molecular systems, describing and simulating molecular systems by presenting a basis for comprehending the processes associated with nuclear motions in molecules [48]. A full dimensional and accurate potential energy surface provides insights into spectroscopy, structure, and reactivity of molecules by probing further than the usual stationary points and barriers and solving the time-dependent Schrödinger equation [49–51]. In a Potential energy surface, the regions dominated by van der Waals interactions are vital for low-temperature phenomena and molecular stacking, which are essential in understanding DNA and RNA molecules. On the other hand, most stable structures on potential energy surfaces are covalent or ionic bond related. As displayed in the figure below, high potential energy regions displayed specifies high-energy nuclear arrangements or molecular conformations, whereas low energy regions signify low nuclear energy conformation and can be applied in computational chemistry for this purpose [31,52].

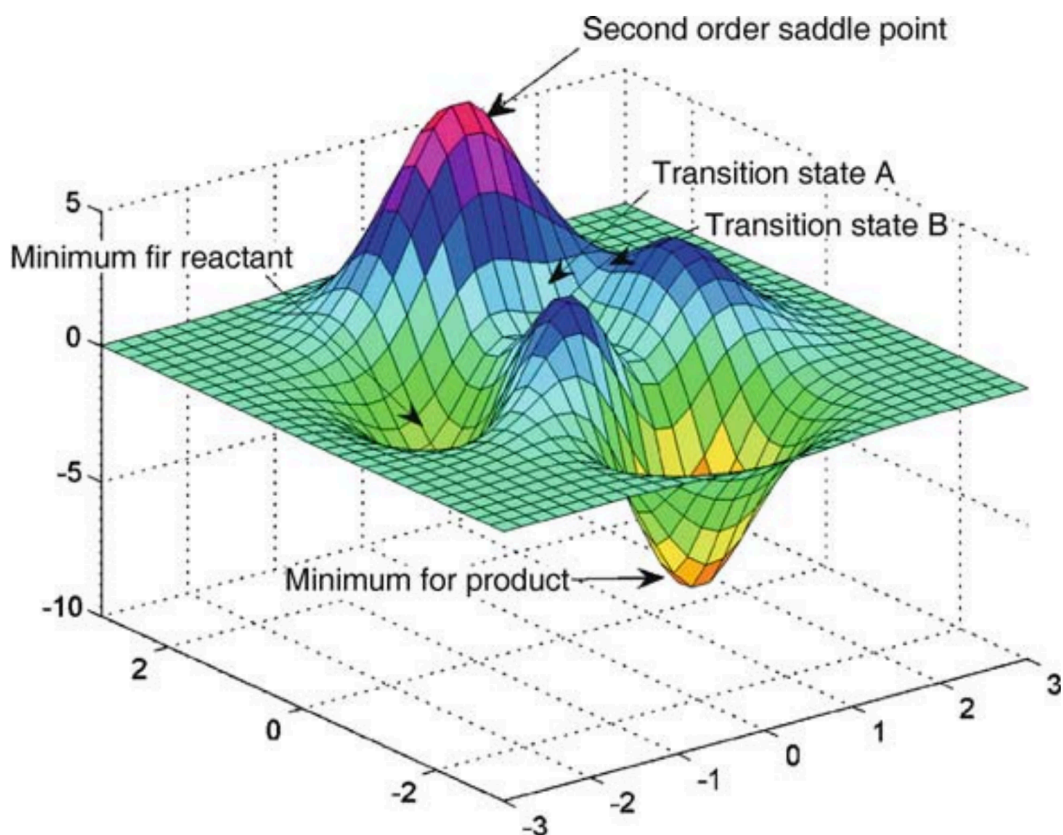


Figure 3.4: A graphical representation of a two-dimensional potential energy surface

3.3 Molecular Mechanics

Molecular mechanics (MM) describes bonds, valence angles, and Coulomb and Lennard-Jones interactions to account for charge-charge and non-bonded contributions by incorporating classical type models [1,53]. In doing so, MM methods can predict the energy of a molecular system as a function of their conformation [1]. Relative to QM methods, MM force fields dominate as the preferred method choice for protein simulations towards understanding conformational dynamics of these protein structures because potential energy functions in MM methods provide a relatively

more accurate representation of dispersion interactions while adhering to the Newtonian laws of classical mechanisms [54,55]. MM has been employed to study a wide range of systems, including hydrocarbons, nucleic acids, membrane fragments, and material assemblies, which consist of thousands of atoms. The application of MM in more complex phenomena such as polarization and many-body contributions has also gained attention in recent years [4,56–58]. MM has also allowed for the study of areas (solid-state interfaces, semi-conducting systems, metals, alloys, etc.) in which the application of simplistic force field methods was deemed undependable [4]. s corresponding to the minimum-energy geometry – or more precisely, to the various probable potential energy surface minima [59,60]. Nonetheless, MM is appropriate for bond-breaking reactions, just like most ab initio methods [61]. MM or the force-field method is mathematically described as follows;

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} \quad (\text{Eq. 3.3.1})$$

Where,

E_{tot} = total energy, E_{str} = bond-stretching energy,

E_{bend} = angle-bending energy, E_{tors} = torsional energy,

E_{vdw} = van der Waals energy E_{elec} = electrostatic energy

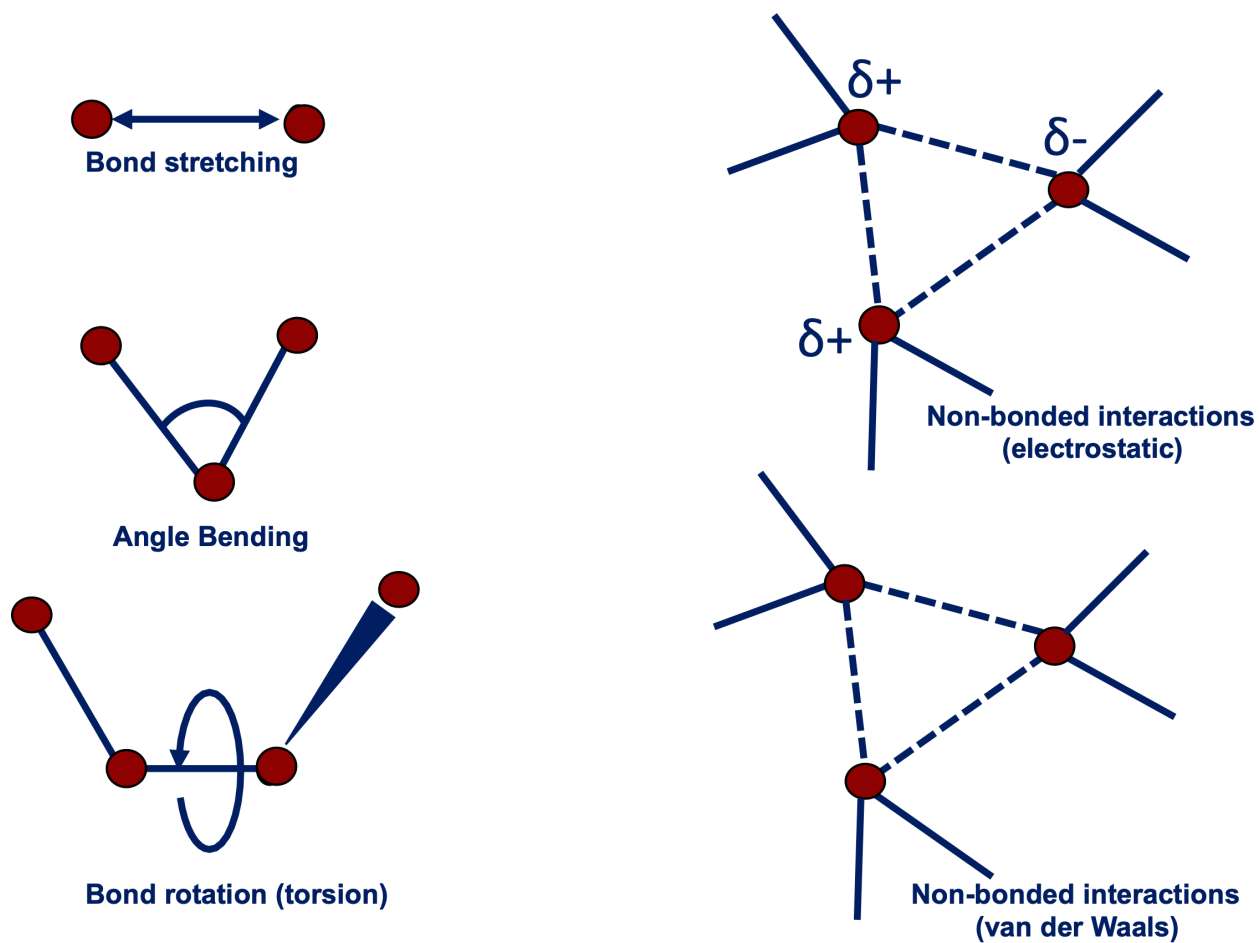


Figure 3.5: Diagrammatical depiction of bonded and non-bonded interactions acting in molecular motion [62].

Also, MM accounts for energy contributions from hydrogen bonding and stretch-bend coupling interactions. In methodical force fields, all constituent atoms are considered building blocks, whereas electrons are not described as separate particles. As such, rather than implementing the Schrödinger equation, more descriptions about the bonds are provided. This accounts for the ball-and-spring model earlier mentioned with different atomic sizes and bond lengths while the atoms themselves are treated as spheres. This model was therefore termed “atom types,” which is dependent on the atomic number and the interconnecting chemical bonds [31,63]. From Eq. 3.8;

Bond stretching, which occurs between directly bonded atoms, is mathematically represented as:

$$E_{str} = \sum K_r (\mathbf{r} - \mathbf{r}_0)^2 \quad (\text{Eq. 3.3.2})$$

Angle bending, which occurs between atoms that are surrounded by the same central atoms is mathematically expressed as:

$$E_{\theta} = \sum K_{\theta} (\theta - \theta_0)^2 \quad (\text{Eq. 3.3.3})$$

Also, bond torsion is expressed as

$$E_{\phi} = \sum K_{\phi} [1 + \cos(n\phi - \phi_0)] \quad (\text{Eq. 3.3.4})$$

In all, non-bonded interactions, which constitute van der Waals and electrostatic forces, we have

$$E_{nb} = \left[\sum \sum \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \right] + \left[\sum \sum \left(\frac{q_i q_j}{D r_{ij}} \right) \right] \quad (\text{Eq. 3.3.5})$$

Where, K_r , K_{θ} and K_{ϕ} depict the force constants for bond, angle and dihedral angles while r_0 , θ_0 and ϕ_0 represent the equilibrium distance, angle and phase angle. Moreover, the parameter for distance is r_{ij} while A_{ij} and B_{ij} depict the parameters for van der Waals forces. The molecular dielectric constant is represented by D while the point charges are q_i and q_j .

3.4 Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM)

Pioneered in the 1970s by Martin Karplus, Michael Levitt, and Arieh Warshel, QM/MM attempts to harness the advantages of both QM and MM [4,17,18,64,65]. In studying a molecular system with hybrid QM/MM approaches, the QM component is employed to treat the most significant part of the system, whereas the classical MM component is subsequently employed to model the rest of the system [4]. The crux of QM/MM as a method centers around the electrostatic interaction between QM and MM subsystems. As an advantage over individual QM and MM, hybrid QM/MM methods give more accurate quantum calculations and are associated with lesser computational cost [66,67]. The technique of QM/MM encompasses two steps-wise domains, a QM, and an MM phase [130]. In the QM phase, the reactive domain, which includes ligands and amino acids, is treated with QM techniques, which defines molecules' electronic structures [129]. This MM phase defines interactions between atoms by incorporating a force field, acting as a potential energy function [128,131]. Amidst the notable advantages of the hybrid QM/MM method over QM and MM, it may not be appropriate for every structure-based drug design study; hence further exploration of the technique may be required [129]. In drug discovery research, the application of QM/MM encompasses areas such as simulation of chemical reactions such as the catalytic mechanism of enzymes, the investigation of electronic properties, and calculation of spectra in single simulation towards understanding the mechanism of action of some drugs [68–70].

3.5 Molecular Dynamics Simulations

MD simulation is a widely used computational technique based on classical Newtonian mechanics that employs MM force fields to provide essential atomistic insights on biomolecular structures by simulating particulate motion as a function of time [71–74]. MD simulation was primarily employed in achieving the objectives of the various reports in this thesis by providing an atomistic time-scale insight into the conformational and structural dynamics of the studied biological targets with their respective bound small molecule inhibitors. Based on varying accuracy requirements, all forms of MD simulations including; coarse-grained (CG), united-atom (UA), and all-atom (AA), including explicit/implicit solvent models which are known to permit simulations of temporal and spatial scales, have been shown to play a vital role in drug discovery in recent years [75–77]. The application of MD simulation in drug discovery has included; exploring the mechanism of actions of therapeutic agents, identifying potential binding sites, and calculating binding free energies of therapeutic agents, amongst many others [77–79]. Originating in the late 1950s, MD simulations have gained center stage in drug design in recent years, evolving to include the simulation of complex macromolecules such as DNA and RNA even though it began with its application in the simulation of simple gasses [80,81]. In the late 1970s, the first protein MD simulation was also reported [82]. The modeling of complex systems in MD simulations are carried out at atomic levels. At the same time, the equation of motions is mathematically solved to depict the motions and dynamics of biomolecules undergoing simulation, which allows their kinetic and thermodynamic properties to be derived [71,83]. MD simulation techniques are divided into two, namely classical and Monte Carlo (MC) MD simulation. In a Classical MD simulation, atoms and molecules' trajectories are generated by mathematically solving Newton's equation of motion for a system of interacting particles, in which molecular mechanics defines forces between the particles and potential energy force fields as presented below [145]. Trajectories specify the

particle position and velocity, which varies with time [146]. In performing a Classical MD, the positions and velocities of particles, the force field, and boundary conditions must be initially considered.

$$\mathbf{F}_i = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} \quad (\text{Eq. 3.5.1})$$

From the equation, $\mathbf{r}_i(t) = (x_i(t), y_i(t), z_i(t))$ represents the (i) vector of the particulate position with a force, F_i , acting on the (i) particle at time t and particle mass m_i

On the other hand, MC methods seek to create an ensemble of representative configurations under specific thermodynamics conditions for a complex macromolecular system [84]. The preference of Classical MD to Monte Carlo is attributed to its ability to permit time-dependent response and rheological properties, amongst other dynamical properties [140]. Moreover, in classical MD, Newton's equation of motions is integrated into computational algorithms (highly-advanced physical and mathematical algorithms), which in turn enables near-accurate real-time mechanistic and conformational definitions of occurring chemical reactions [85,86].

The basic algorithm of MD simulations is shown below;

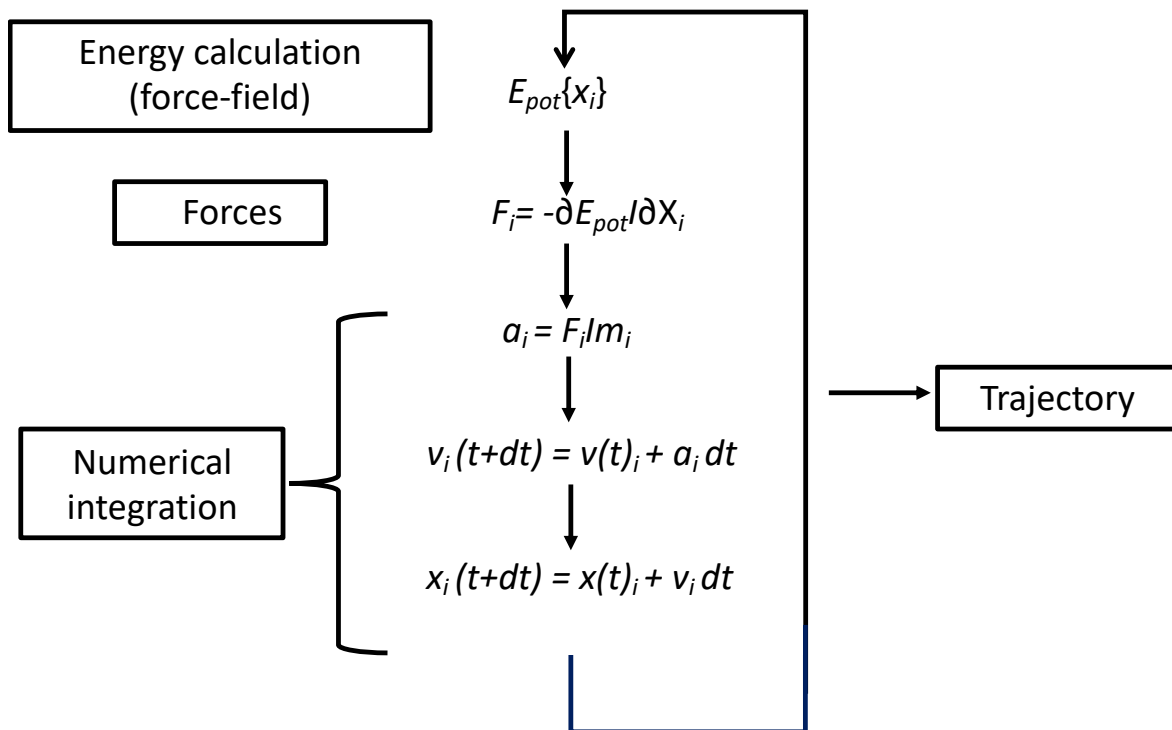


Figure 3.6: Schematic description of basic MD algorithm where potential energy is E_{pot} ; simulation time = t , iteration time is given by dt ; x is atom coordinates; force component is F ; a is given by acceleration; atomic mass is m and v represents velocity (Image adapted from [80]).

The trajectories generated from the simulated systems' time-scale atomistic motions are further analyzed to assess dynamical occurrences in simulated systems at the atomic level [87]. A summary of the steps followed in performing MD simulations for studies in this thesis is presented in the flowchart below.

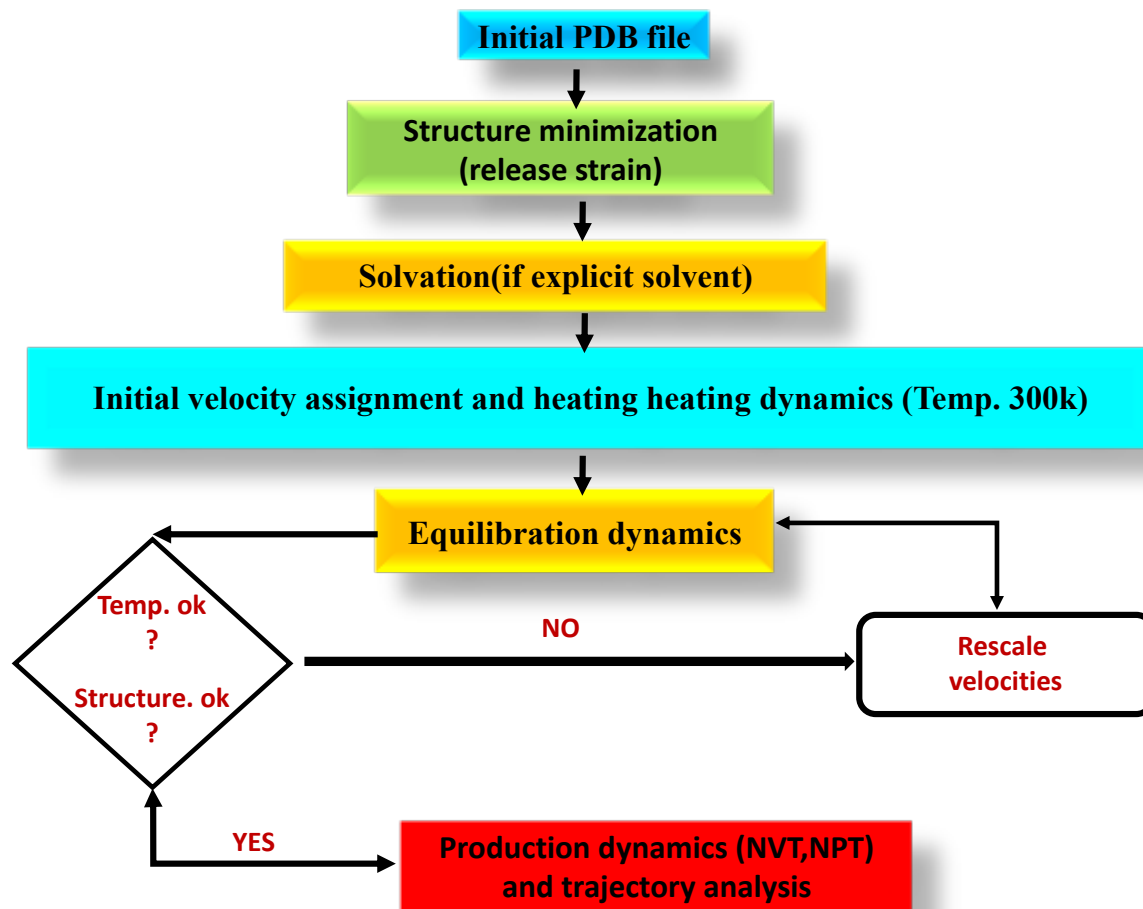


Figure 3.7: Flow chart showing steps involved in MD simulation of a drug-target complex[88]

3.6 Force Fields

Force fields refer to a set of potential energy functions through which force can be obtained and can be used to describe the energies associated with conformations of molecules [89]. Mathematically, force fields consist of set parameters that describe the correlative function between a system's energy and the coordinate of its constituents. In MD simulations performed for all investigations in this thesis, the AMBER force field [83] was employed for protein parameterization, while the General Amber Force Field (GAFF) was used for the parameterization

of bound inhibitors. However, several other force field has been widely used in performing MD simulations for various biomolecular systems, notably; CHARMM [90], GROMOS [91], NAMD [87] OPLS-AA [87] and ENCAD [49,92] amongst many others. Most force fields applied in MD simulation of proteins are harmonic motions and a description of bond lengths and angles [89]. The main variations between the different force field results come from the varied approaches to obtain the individual parameters. However, they mainly yield similar findings which corroborate with experimental results [75].

3.7 Post-Molecular Dynamics Analytical Methods

Molecular dynamics simulations generate trajectories that encode the events that have occurred across the biomolecular system over a specified time-frame, either through a visual or graphical representation. These visual or graphical occurrences along the simulation are essential in unraveling appropriate events or insights, which could explain the structural activities or attributes of biological and non-biological molecules. These trajectories can also typify sequential snapshots described by velocity vectors and positional coordinates, while the time-scale evolution of the systems in phase space is also captured [71,93]. Several techniques and calculations have been established to analyze these trajectories depending on the type of MD simulation study though specific quantitative analyses are necessary to assist any visual systemization. In these studies, post analysis was required to ascertain the system's conformational stability and energetics, the dynamics of the biomolecular system, and the ligands binding landscape and thermodynamic plasticity of energies with the combined trajectories.

3.7.1 System Stability, Convergence and Structural Deviations

A system's convergence describes its dynamics based on typical angular vibrations of atomic bonds during a folding or unfolding event until it attains a state of equilibrium. Convergence measures the reproducibility and accuracy of a trajectory and unveiling details on whether the system has reached an energetically stable conformation [38]. For MD trajectories to be accurate, it is essential to consider the equilibrium points, conformational plateaus, and ultimate energetics. Metrics that are used to predict the stability of a system coupled with corresponding structural events include the C α root mean square deviation (RMSD), the C α root mean square fluctuation (RMSF), the radius of gyration (RoG), Surface accessible surface area (SASA) and Dynamic Cross-Correlation Matrix (DCCM).

3.7.1.1 Root Mean Square Deviation (RMSD)

The C α RMSD is used to predict a biomolecular system's stability by estimating the corresponding deviations and motions that occur across the protein's backbone atoms compared to their starting structure over the simulation time. In other terms, the spatial alterations that occur between two static structures of a steady trajectory can be used to evaluate system divergence. Mathematically, RMSD is expressed as:

$$RMSD = \left(\frac{\sum_N (R_1 - R_1^0)^2}{N} \right)^{\frac{1}{2}} \quad (\text{Eq. 3.7.1})$$

The total number of atoms in the system is N. The vector position of the C α atom of particle i is R_i, which is predicted from the starting conformation (O) after alignment with the ultimate conformation with the aid of the least square fitting. Therefore, the RMSD is suitable to calculate the conformational alterations that occur within ligands, receptors, or complex systems during a simulation period and can be obtained by estimating the frame number average in each resulting trajectory [94,95]. In most conformational studies, an increase in atomistic motions or deviations depicts an unstable structure, while a decrease in these deviations is an indication of atomistic stability.

3.7.1.2 Root Mean Square Fluctuation (RMSF)

While the RMSD focuses on the whole macromolecular structure, each protein residue's corresponding motions can be estimated using the RMSF metrics. RMSF measures fluctuations of discrete amino acid residues relative to their average positions during the simulation. This, therefore, quantifies the relative motions of constituent residues of a protein and forms an important determinant in predicting the flexibility of such proteins since a high residual fluctuation can correlate with an increase in flexibility while low fluctuation in residual motions can indicate a structurally rigid protein [96,97]. Mathematically, RMSF is expressed as:

$$sRMSF_i = \frac{(RMSF_i - \overline{RMSF})}{\sigma(RMSF)} \quad (\text{Eq. 3.7.2})$$

Where the RMSF of the ith residue, is RMSF_i, from which the average RMSF is subtracted and further divided with the RMSF's standard deviation, resulting in the standardized RMSF.

3.7.1.3 Radius of Gyration (RoG)

The RoG of a protein can be defined as the root mean square atomistic distance from the center of gravity and is essential for predicting structural compactness in a protein [98,99]. Several studies have been used to reflect the RoG metrics' accuracy in determining protein compactness [100,101].

The RoG of a system is therefore mathematically represented as:

$$r^2_{gyr} = \frac{(\sum_{i=1}^n w_i (r_i - r^-)^2)}{\sum_{i=1}^n w_i} \quad (\text{Eq. 3.7.3})$$

Where the position of the i th atom is denoted as r_i and the center weight of atom i is shown as r .

3.7.1.4 Solvent Accessible Surface Area

SASA has been used over the years to depict the transitional occurrence of protein atoms between the hydrophilic and hydrophobic regions. In other words, the mechanistic motions of constituent atoms from the initially exposed outer environment to the hydrophobic core region of the protein can be estimated. Since this transitional event can likely correlate with massive structural changes in proteins relative to the initial structure, this parameter has been previously adopted to predict protein folding and unfolding events [102–104].

3.7.1.5 Dynamic Cross Correlation Matrix

Dynamic cross-correlation matrix (DCCM) is a computational technique used to quantify correlation coefficients of motions between atoms within a simulated system [110]. DCCM presents a 3D matrix that allows for a graphical depiction of time-correlated insights on the amino acid residues that make up the protein structure that undergoes MD simulation. The cross-correlation coefficient varies from -1 (completely anticorrelated motion) to +1 (completely correlated motion). The formula used to describe dynamic cross-correlation is presented as follows:

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle)^{\frac{1}{2}}} \quad (\text{Eq. 3.7.4})$$

i and j represents the i^{th} and j^{th} residue respectively. The displacement vectors that correspond to the i^{th} and j^{th} residue respectively is represented as Δr_i or Δr_j .

3.7.2 Free binding energy predictions

Assessing binding energy is vital in protein-ligand interaction systems and other aspects of computational drug design discovery, such as protein-protein system behaviors [105,106]. Computer-based free binding energies of macromolecules and other varying molecular systems are calculated using either the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) or Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) techniques. Both computational tools have been applied widely in protein-ligand interactions and are known for accurate results. Both methods were also used to calculate the energy contributions of individual residues using the per-residue energy decomposition approach after each MD

simulation. The binding free energy (ΔG_{bind}) was determined by the equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} + G_{\text{ligand}} \quad (1) \quad (\text{Eq. 3.7.5})$$

$$\Delta G_{\text{bind}} = \Delta G_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S, \quad (2) \quad (\text{Eq. 3.7.6})$$

where ΔG_{bind} is considered to be the summation of the gas phase and solvation energy terms less the entropy ($T\Delta S$) term

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} \quad (3) \quad (\text{Eq. 3.7.7})$$

ΔE_{gas} denotes the sum of AMBER force field internal energy terms ΔE_{int} (bond, angle and torsion), the covalent van der Waals (ΔE_{vdw}) and the non-bonded electrostatic energy component (ΔE_{elec}).

The solvation energy can be estimated from the following equation:

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{non-polar}} \quad (4) \quad (\text{Eq. 3.7.8})$$

$$G_{\text{non-polar}} = \gamma \text{SASA} + b \quad (5) \quad (\text{Eq. 3.7.9})$$

G_{GB} is the polar solvation contribution and the non-polar solvation contribution is represented as $G_{\text{non-polar}}$. Both of which are determined from SASA, derived from a 1.4 Å water probe radius.

3.7.3 Residue Interaction Network

Residue interaction network (RIN) is a computational tool used to explore macromolecular systems' structural complexity. RIN makes it possible to easily visualize the kinds of interaction

between the residues that make up a protein and the interaction of these residues and a drug molecule [107,108]. The type of interactions exhibited by amino acids gives insights into the functional dynamics of a protein and may also account for the stronger binding of a drug [109]. In the studies reported herein, snapshots were taken during the molecular dynamic simulations at different time intervals and visualized on Discovery studio, a molecular visualization software.

3.7.4 Principal Component Analysis

As a covariance-matrix-based method, Principal Component Analysis (PCA) is a covariance-matrix-based approach that can be used to elucidate the vital dynamics of a simulated system on a low-dimensional free energy landscape [111]. PCA also allows for an elucidation of any displacement of atoms and any associated structural dynamics that may have occurred over a simulation period with a protein structure [112]. PCA symbolizes a linear transformation that diagonalizes the covariance matrix of a simulated system and thus deletes any instantaneous linear correlations in the simulated system [111,113]. The fluctuations in a simulated system can be described by the first few principal components of a simulated system through a decreasingly ordering of the eigenvalues of the transformation [111,114–116]. PCA as employed in this study involves the calculation of the first two principal components (PC1 and PC2) and covariance matrices of the C- α atoms of the simulated structure. A 2x2 covariance matrix is subsequently constructed from generated MD trajectories using Cartesian coordinates of C- α atoms.

Reference§§

- [1] Lewars EG. Computational Chemistry. 2011. Available from: <http://link.springer.com/10.1007/978-90-481-3862-3>.
- [2] Chaudhari R, Tan Z, Huang B, et al. Computational polypharmacology: a new paradigm for drug discovery. *Expert Opin Drug Discov.* 2017;12:279–291.
- [3] Moore G. Cramming more components onto integrated circuits. *Electronics.* 1965;38:114–117.
- [4] Hofer TS, de Visser SP. Editorial: Quantum Mechanical/Molecular Mechanical Approaches for the Investigation of Chemical Systems – Recent Developments and Advanced Applications. *Front Chem.* 2018;6:357. Available from: <https://www.frontiersin.org/article/10.3389/fchem.2018.00357>.
- [5] Whitworth I. Review of microprocessor architecture. *Microprocess Microsy.* 1979;3:21–28.
- [6] Parr RG, Yang W. *Density-Functional Theory of Atoms and Molecules.* Oxford: University Press; 1994.
- [7] Szabo A, Ostlund NS. *Modern Quantum Chemistry.* New York, NY: Dover Publications; 1996.
- [8] Náráy-Szabó G, Oláh J, Krámos B. Quantum mechanical modeling: a tool for the understanding of enzyme reactions. *Biomolecules.* 2013;3:662–702.
- [9] Helgaker T, Jørgensen P, Olsen J. *Molecular Electronic Structure Theory.* Chichester: Wiley; 2000.

- [10] Koch W, Holthausen M. A Chemist's Guide to Density Functional Theory. 2nd Editio. Weinheim: Wiley-VCH; 2002.
- [11] Cook DB. Handbook of Computational Chemistry. New York, NY: Dover Publications; 2005.
- [12] Sholl D, Steckel J. Density Functional Theory-A Practical Introduction. Hoboken, NY: Wiley; 2009.
- [13] Drummond B. Understanding quantum mechanics: A review and synthesis in precise language. *Open Phys.* 2019;17:390–437.
- [14] Leach AR. Molecular Modelling. 2nd Edn. Harlow: Prentice-Hall; 2001.
- [15] Jensen F. Introduction to Computational Chemistry. 2nd Editio. Chichester: John Wiley & Sons Ltd.; 2006.
- [16] Ramachandran KI, Deepa G, Namboori K. Computational Chemistry and Molecular Modeling: Principles and Applications. Berlin: Springer; 2008.
- [17] Zheng M, Waller MP. Adaptive quantum mechanics/molecular mechanics methods. *WIREs Comput Mol Sci.* 2016;6:369–385.
- [18] Pezeshki S, Lin H. Recent progress in adaptive-partitioning QM/MM methods for Born-Oppenheimer molecular dynamics. In: Rivail J-L, Ruiz-Lopez M, editors. Quantum Model Complex Mol Syst Challenges Sdvances Comput Chem Phys. Springer, Cham; 2015;93–113.
- [19] Metz S, Kästner J, Sokol AA, et al. ChemShell-a modular software package for QM/MM simulations. *WIREs Comput Mol Sci.* 2014;4:101.
- [20] Senn HM, Thiel W. QM/MM Methods for Biomolecular Systems. *Angew Chemie Int Ed.*

2009;48:1198–1229.

[21] Chung LW, Sameera WMC, Ramozzi R, et al. The ONIOM Method and Its Applications. *Chem Rev.* 2015;115:5678–5796. Available from: <https://doi.org/10.1021/cr5004419>.

[22] Chung LW, Hirao H, Li X, et al. The ONIOM method: its foundation and applications to metalloenzymes and photobiology. *WIREs Comput Mol Sci.* 2012;2:327–350. Available from: <https://doi.org/10.1002/wcms.85>.

[23] Brunk E, Neri M, Tavernelli I, et al. Integrating computational methods to retrofit enzymes to synthetic pathways. *Biotech Bioeng.* 2012;109:572–582.

[24] van der Kamp MW, Mulholland AJ. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry.* 2013;52:2708–2728. Available from: <https://doi.org/10.1021/bi400215w>.

[25] Hayashi S, Ueno H, Shaikh AR, et al. Molecular Mechanism of ATP Hydrolysis in F1-ATPase Revealed by Molecular Simulations and Single-Molecule Observations. *J Am Chem Soc.* 2012;134:8447–8454. Available from: <https://doi.org/10.1021/ja211027m>.

[26] Bozkurt E, Ashari N, Browning N, et al. Lessons from Nature: Computational Design of Biomimetic Compounds and Processes. *Chimia (Aarau).* 2014;68:642–647.

[27] Merz KM, Jr. Using quantum mechanical approaches to study biological systems. *Acc Chem Res.* 2014;47:2804–2811.

[28] Warshel A, Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol.* 1976;103:227–249. Available from: <http://www.sciencedirect.com/science/article/pii/0022283676903119>.

- [29] Cui Q. Perspective: Quantum mechanical methods in biochemistry and biophysics. *J Chem Phys.* 2016;145:140901. Available from: <https://doi.org/10.1063/1.4964410>.
- [30] Smith JC, Roux B. Eppur Si Muove! The 2013 Nobel Prize in Chemistry. *Structure.* 2013;21:2102–2105. Available from: <http://www.sciencedirect.com/science/article/pii/S0969212613004395>.
- [31] Jensen F. Introduction to Computational Chemistry [Internet]. *Angew. Chemie Int. Ed.* 2007. Available from: https://books.google.com/books/about/Introduction_to_Computational_Chemistry.html?id=RDI G48UcZfYC&pgis=1.
- [32] Trabesinger A. The short version. *Nat Phys.* 2009;5:383. Available from: <https://doi.org/10.1038/nphys1279>.
- [33] Shen L, Wu J, Yang W. Multiscale Quantum Mechanics/Molecular Mechanics Simulations with Neural Networks. *J Chem Theory Comput.* 2016;12:4934–4946. Available from: <https://doi.org/10.1021/acs.jctc.6b00663>.
- [34] Levada CL, Maceti H, Lautenschleguer IJ. Review of the Schrödinger Wave Equation. *IOSR J Appl Chem.* 2018;11:1–7.
- [35] Nakatsuji H. Scaled Schrodinger equation and the exact wave function. *Phys Rev Lett.* 2004;93:pp.30403–1.
- [36] Barde NP et al. Deriving time dependent Schrödinger equation from Wave-Mechanics, Schrödinger time independent equation, classical and Hamilton-Jacobi equations. *Leonardo Electron J Pract Technol.* 2015;14:pp.31–48.

- [37] Bahrami M et al. The Schrodinger-Newton equation and its foundations. *New J Phys.* 2014;pp.1–17.
- [38] Bechstedt F. Many-Body Approach to Electronic Excitations. 2014;181:3–12. Available from: <http://link.springer.com/book/10.1007/978-3-662-44593-8>.
- [39] Akbas H, Turgut OT. Born-Oppenheimer approximation for a singular system. *J Math Phys.* 2018;59. Available from: <http://dx.doi.org/10.1063/1.5021364>.
- [40] Huang XL, Yi XX. Born-Oppenheimer approximation in open systems. *Phys Rev A.* 2009;80:7. Available from: <http://arxiv.org/abs/0905.2741>.
- [41] Liehr AD. On the use of the Born-Oppenheimer approximation in molecular problems. *Ann Phys (N Y).* 1957;1:221–232. Available from: <http://www.sciencedirect.com/science/article/pii/000349165790009X>.
- [42] Born M, Oppenheimer JR. Born-Oppenheimer approximation. *Ann Phys.* 1927;84:457.
- [43] Jecko T. On the mathematical treatment of the born-Oppenheimer approximation. *J Math Phys.* 2014;55.
- [44] Liehr AD. On the use of the Born-Oppenheimer approximation in molecular problems. *Ann Phys (N Y).* 1957;1:pp.221–232. Available from: <http://www.sciencedirect.com/science/article/pii/000349165790009X>.
- [45] Ochkur VI. The Born-Oppenheimer method in the theory of atomic collisions. *Sov Phys JETP.* 1965;18:p.503.
- [46] Matsika S. The Born-Oppenheimer approximation. *J Chem Phys.* 2010;133:224103. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21241076>.

- [47] Born M, Oppenheimer R. Quantum theory of the molecules. *Ann d Phys.* 1927;457–484.
- [48] Peyerimhoff S. J. N. Murrell, S. Carter, S. C. Farantos, P. Huxley, and A. J. C. Varandas: *Molecular Potential Energy Functions*, Verlag John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore 1984. 197 Seiten, Preis: £ 19.95. *Berichte der Bunsengesellschaft für Phys. Chemie.* John Wiley & Sons, Ltd; 1985. Available from: <https://doi.org/10.1002/bbpc.19850891027>.
- [49] Levitt M, Hirshberg M, Sharon R, et al. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comput Phys Commun.* 1995;91:215–231.
- [50] Woolley R. Quantum chemistry beyond the Born-Oppenheimer approximation. *J Mol Struct THEOCHEM.* 1991;230:17–46.
- [51] Ashworth SH. *Molecular Quantum Mechanics*, 5th edn, by Peter Atkins and Ronald Friedman. *Contemp Phys.* 2012;53:372–373. Available from: <https://doi.org/10.1080/00107514.2012.678277>.
- [52] Jakobsson E. *Computational Biochemistry and Biophysics* Edited by Oren M. Becker (Tel Aviv University), Alexander D. MacKerell, Jr. (University of Maryland), Benoît Roux (Cornell University), and Masa-katsu Watanabe (Wavefunction, Inc.). Marcel Dekker: New York and Basel. 2001. xii + 512 pp. \$195.00. ISBN: 0-8247-0455-X. *J Am Chem Soc.* 2001;123:12745. Available from: <https://doi.org/10.1021/ja015256z>.
- [53] Hehre WJ. *A Guide to Molecular Mechanics and Quantum Chemical Calculations.*
- [54] Hobza P, Sponer J. Structure, energetics, and dynamics of the nucleic Acid base pairs:

nonempirical ab initio calculations. *Chem Rev.* 1999;99:3247–3276.

[55] Vanommeslaeghe K, Guvench O, MacKerell AD, et al. Molecular mechanics. *Curr Pharm Des.* 2014;20:3281–3292.

[56] Lemkul JA, Huang J, Roux B, et al. An empirical polarizable force field based on the classical drude oscillator model: development history and recent applications. *Chem Rev.* 2016;116:4983–5013.

[57] Baker CM. Polarizable force fields for molecular dynamics simulations of biomolecules. *WIREs Comp Mol Sci.* 2015;5:241–254.

[58] Stone AJ. *The Theory of Intermolecular Forces.* Oxford, UK: Oxford University Press; 1995.

[59] Boeyens* J. Molecular mechanics: theoretical basis, rules, scope and limits. *Coord Chem Rev.* 2001;212:3–10.

[60] Zimmer M. Are classical molecular mechanics calculations still useful in bioinorganic simulations? *Coord Chem Rev.* 2009;253:817–826.

[61] Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew Chem Int.* 2009;1198–1229.

[62] Molecular Mechanics. *Indipedia India's Wikipedia OSDD.*

[63] Lewars EG. *Computational chemistry: Introduction to the theory and applications of molecular and quantum mechanics: Third Edition 2016.* *Comput. Chem. Introd. to Theory Appl. Mol. Quantum Mech. Third Ed.* 2016. 2016.

[64] Aqvist J, Warshel A. Simulation of enzyme reactions using valence bond force fields and

other hybrid quantum/classical approaches. *Chem Rev.* 1993;93:2523.

[65] Warshel A. Molecular dynamics simulations of biological reactions. *Acc Chem Res.* 2002;35:385.

[66] Sauer J, Sierka M. Combining quantum mechanics and interatomic potential functions in ab initio studies of extended systems. *J Comput Chem.* 2000;21:1470–1493.

[67] Honarparvar B, Govender T, Maguire GEM, et al. Integrated approach to structure-based enzymatic drug design: Molecular modeling, spectroscopy, and experimental bioactivity. *Chem. Rev.* 2014. p. 493–537.

[68] Fan J, Lin L, Wang C-K. Excited state properties of non-doped thermally activated delayed fluorescence emitters with aggregation-induced emission: a QM/MM study. *J Mater Chem C.* 2017;5:8390–8399. Available from: <http://dx.doi.org/10.1039/C7TC02541F>.

[69] Morzan UN, Alonso de Armiño DJ, Foglia NO, et al. Spectroscopy in Complex Environments from QM–MM Simulations. *Chem Rev.* 2018;118:4071–4113. Available from: <https://doi.org/10.1021/acs.chemrev.8b00026>.

[70] Jindal G, Warshel A. Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region. *J Phys Chem B.* 2016;120:9913–9921. Available from: <https://doi.org/10.1021/acs.jpcc.6b07203>.

[71] Heinecke A, Eckhardt W, Horsch M, et al. Molecular dynamics simulation. *SpringerBriefs Comput Sci.* 2015.

[72] Buchete N-V, Hummer G. Peptide folding kinetics from replica exchange molecular dynamics. *Phys Rev E.* 2008;77:30902. Available from:

<https://link.aps.org/doi/10.1103/PhysRevE.77.030902>.

[73] Takada S, Kanada R, Tan C, et al. Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Acc Chem Res.* 2015;48:3026–3035. Available from: <https://doi.org/10.1021/acs.accounts.5b00338>.

[74] Mortier J, Rakers C, Bermudez M, et al. The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. *Drug Discov Today.* 2015;20:686–702. Available from: <http://www.sciencedirect.com/science/article/pii/S1359644615000215>.

[75] Durrant J, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol.* 2011;9:71.

[76] Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. *J Comput Aided Mol Des.* 2012;26:15–26. Available from: <https://doi.org/10.1007/s10822-011-9517-y>.

[77] Lin X, Li X, Lin X. A review on applications of computational methods in drug screening and design. *Molecules.* 2020;25:1–17.

[78] Wang Y, Lupala CS, Lin* HL and X. Identification of Drug Binding Sites and Action Mechanisms with Molecular Dynamics Simulations. *Curr. Top. Med. Chem.* 2018. p. 2268–2277. Available from: <http://www.eurekaselect.com/node/168292/article>.

[79] Hou T, Wang J, Li Y, et al. Assessing the Performance of the MM_PBSA and MM_GBSA Methods. 1. The Accuracy.pdf. 2011;69–82.

[80] Hospital A, Goñi JR, Orozco M, et al. Molecular dynamics simulations: Advances and

applications. *Adv. Appl. Bioinforma. Chem.* 2015.

[81] Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. *Neuron.* 2018;99:1129–1143. Available from: <https://doi.org/10.1016/j.neuron.2018.08.011>.

[82] McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature.* 1977;267:585–590. Available from: <http://www.nature.com/doi/10.1038/267585a0>.

[83] Case DA, Cheatham TE, Darden T, et al. The Amber biomolecular simulation programs. *J. Comput. Chem.* 2005. p. 1668–1688.

[84] Fichthorn KA, Weinberg WH. Theoretical foundations of dynamical Monte Carlo simulations. *J Chem Phys.* 1991;95:1090–1096.

[85] Vitalini F, Mey ASJS, Noé F, et al. Dynamic properties of force fields. *J Chem Phys.* 2015;142:084101.

[86] Monticelli L, Tieleman DP. Force fields for classical molecular dynamics. *Methods Mol Biol.* 2013;924:197-213.

[87] Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005;26:1781–1802.

[88] Patodia S. Molecular Dynamics Simulation of Proteins: A Brief Overview. *J Phys Chem Biophys.* 2014;4:4–7.

[89] González MA. Force fields and molecular dynamics simulations. *Collect SFN.* 2011;12:169–200.

[90] Brooks BR, Brooks CL, Mackerell AD, et al. CHARMM: The biomolecular simulation program. *J Comput Chem.* 2009;30:1545–1614.

- [91] Christen M, Hünenberger PH, Bakowies D, et al. The GROMOS software for biomolecular simulation: GR[OMOS05. *J. Comput. Chem.* 2005;26:1719–1751.
- [92] Monticelli L, National F. Force fields for classical molecular dynamics. *Methods Mol Biol [Internet]*. 2013;924:197–213. Available from: <http://link.springer.com/10.1007/978-1-62703-017-5>.
- [93] Likhachev I V., Balabaev NK, Galzitskaya O V. Available Instruments for Analyzing Molecular Dynamics Trajectories. *Open Biochem J.* 2016; 10:1–11.
- [94] Forster MJ. Molecular modelling in structural biology. *Micron.* 2002; 33:365-84.
- [95] Aloy P, Russell RB. Structural systems biology: Modelling protein interactions. *Nat. Rev. Mol. Cell Biol.* 2006;7:188-97.
- [96] Karshikoff A, Nilsson L, Ladenstein R. Rigidity versus flexibility: The dilemma of understanding protein thermal stability. *FEBS J.* 2015;282:3899–3917.
- [97] Bornot A, Etchebest C, De Brevern AG. Predicting protein flexibility through the prediction of local structures. *Proteins Struct Funct Bioinforma.* 2011;79:839–852.
- [98] Lobanov M.Y, Bogatyreva, N.S., and Galziskaya O. Radius of gyration as an indicator of protein structure compactness. *Mol Biol.* 2008;42:623–628.
- [99] Salleh AB, Rahim ASMA, Rahman RNZRA, et al. The Role of Arg157Ser in Improving the Compactness and Stability of ARM Lipase. *J Comput Sci Syst Biol.* 2012;05:38–46.
- [100] Abdullahi M, Olotu FA, Soliman ME. Allosteric inhibition abrogates dysregulated LFA-1 activation: Structural insight into mechanisms of diminished immunologic disease. *Comput Biol Chem.* 2018;73:49–56.

- [101] Huynh T, Smith JC, Sanson A. Protein unfolding transitions in an intrinsically unstable annexin domain: Molecular dynamics simulation and comparison with nuclear magnetic resonance data. *Biophys J*. 2002;83:681–698.
- [102] Lins L, Thomas A, Brasseur R. Analysis of accessible surface of residues in proteins. *Protein Sci*. 2003;12:1406–1417.
- [103] Malleshappa Gowder S, Chatterjee J, Chaudhuri T, et al. Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. *ScientificWorldJournal*. 2014;2014:971258.
- [104] Lins L, Thomas A, and R. B. Analysis of accessible surface of residues in proteins. *Protein Sci*. 2003;12:1406–1417.
- [105] Kalra P, Reddy TV, Jayaram B. Free Energy Component Analysis for Drug Design: A Case Study of HIV-1 Protease–Inhibitor Binding. *J Med Chem*. 2001;44:4325–4338. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-0035818886&partnerID=tZOtx3y1%5Cnhttp://pubs.acs.org/doi/abs/10.1021/jm010175z>.
- [106] Mhlongo NN, Soliman MES. Single H5N1 influenza A neuraminidase mutation develops resistance to oseltamivir due to distorted conformational and drug binding landscape: multiple molecular dynamics analyses. *RSC Adv*. 2015;5:10849–10861. Available from: <http://xlink.rsc.org/?DOI=C4RA13494J>.
- [107] Moonsamy S, Bhakat S, Walker RC, et al. Single Active Site Mutation Causes Serious Resistance of HIV Reverse Transcriptase to Lamivudine: Insight from Multiple Molecular Dynamics Simulations. *Cell Biochem Biophys*. 2016;74:35–48.
- [108] Ndagi U, Mhlongo NN, Soliman ME. The impact of Thr91 mutation on c-Src resistance to

UM-164: molecular dynamics study revealed a new opportunity for drug design. *Mol BioSyst.* 2017;13:1157–1171. Available from: <http://xlink.rsc.org/?DOI=C6MB00848H>.

[109] Amitai G, Shemesh A, Sitbon E, et al. Network analysis of protein structures identifies functional residues. *J Mol Biol.* 2004;344:1135–1146.

[110] McCammon J. Protein dynamics. *Reports Prog Phys.* 1984;47:1.

[111] Sittel F, Jain A, Stock G. Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. *J Chem Phys.* 2014;141:014111. Available from: <http://dx.doi.org/10.1063/1.4885338>.

[112] Martinez AM, Kak AC. PCA versus LDA. *Trans Pattern Anal Mach Intell.* 2001;23:228–233.

[113] Jolliffe TI. *Principal Component Analysis.* Springer, New York; 2002.

[114] Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Proteins Struct Funct Bioinforma.* 1993;17:412–425. Available from: <https://doi.org/10.1002/prot.340170408>.

[115] Mu Y, Nguyen PH, Stock G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins.* 2005;58:45–52.

[116] Maisuradze GG, Liwo A, Scheraga HA. How adequate are one- and two-dimensional free energy landscapes for protein folding dynamics? *Phys Rev Lett.* 2009;102:238102. Available from: <http://europepmc.org/abstract/MED/19658975>.

CHAPTER 4

Published Article

Druggability and Drug-likeness Concepts in Drug Design- are Biomodelling and Predictive Tools having their say?

Clement Agoni ^a Fisayo A. Olotu ^a, Pritika Ramharack ^a, Mahmoud Soliman ^{a*}

^a Molecular Bio-Computation & Drug Design Lab, School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4000, South Africa.

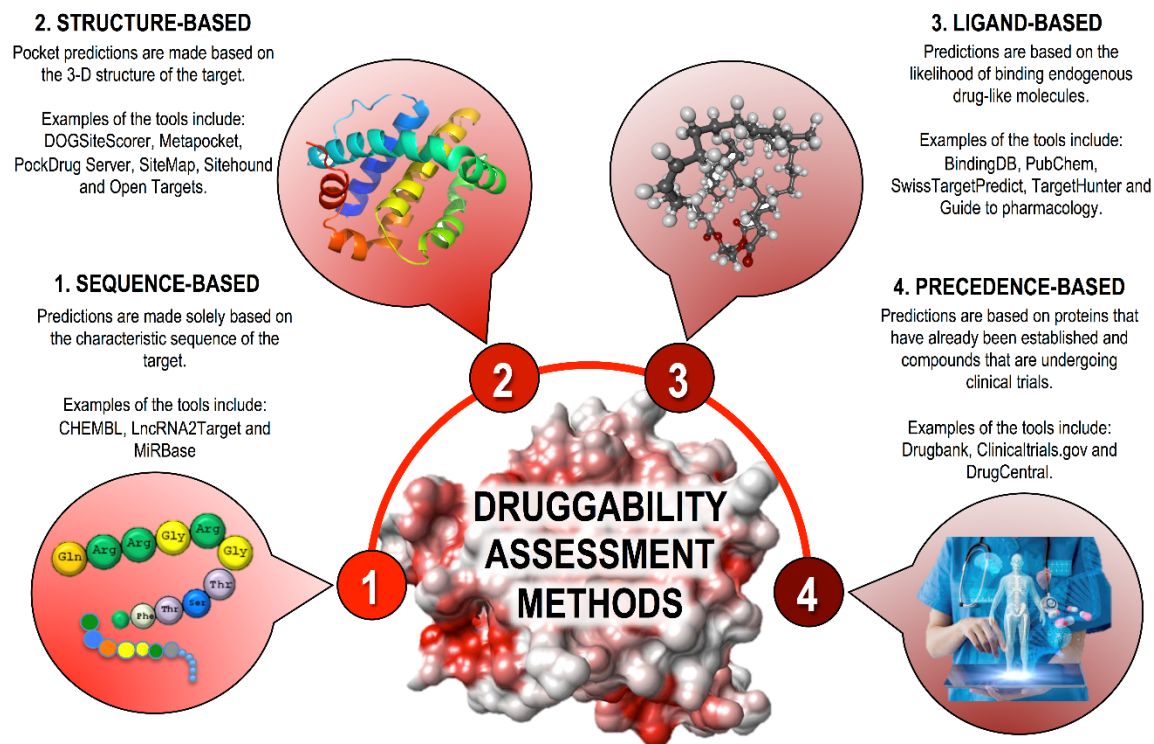

* Corresponding author: Mahmoud E. Soliman, email: soliman@ukzn.ac.za

Telephone: +27 31 260 8048, Fax: +27 31 260 7872

<http://soliman.ukzn.ac.za>

GRAPHICAL ABSTRACT

INCREASING RELIABILITY OF PREDICTIONS



Abstract

The drug discovery process typically involves target identification and design of suitable drug molecules against these targets. Despite decades of experimental investigations in the drug discovery domain, about 96% overall failure rate has been recorded in drug development due to the “undruggability” of various identified disease targets, in addition to other challenges. Likewise, the high attrition rate of drug candidates in the drug discovery process has also become an enormous challenge for the pharmaceutical industry. To alleviate this negative outlook, new trends in drug discovery have emerged. By drifting away from experimental research methods, computational tools and big data are becoming valuable in the prediction of biological target druggability and the drug-likeness of potential therapeutic agents. These tools have proven to be useful in saving time and reducing research costs. As with any emerging technique, however, controversial opinions have been presented regarding the validation of predictive computational tools. To address the challenges associated with these varying opinions, this review attempts to highlight the principles of druggability and drug-likeness and their recent advancements in the drug discovery field. Herein, we present the different computational tools and their reliability of predictive analysis in the drug discovery domain. We believe that this report would serve as a comprehensive guide toward computational-oriented drug discovery research.

Keyword: druggability, drug-likeness, computer-aided drug design, drug discovery

1.0 Introduction

The design and development of several therapeutic agents centered on their ability to interact with crucial biomolecules involved in disease progression. These biological molecules majorly include proteins and enzymes that drive pathogenic pathways and could represent therapeutic avenues for halting undesirable pathogenic development such as in cancer, TB, and neurodegenerative diseases, some of which have been termed ‘undruggable’ (1–4). Of the 10% of the human genome representing druggable targets for therapeutic agents, only half of those are relevant to disease (5–7). Over the years a significant advancement has been in the experimental and computational techniques employed in the drug discovery process, nonetheless, a remarkable failure of about 96% overall failure rate been recorded in diverse drug discovery projects due to the ‘undruggability’ of various identified protein targets, amidst other challenges (8–10). The drug discovery process is usually a time-consuming process and substantial financial investment, as such, a 96% failure of the process due to target undruggability and other challenges necessitates an understanding of challenges associated with target’s undruggability.

As essential as lead identification and lead optimization techniques are in drug discovery, the concept of protein druggability remains vital, usually descriptive of members of expressed protein cohort (proteome) that can selectively be bound and modulated by high-affinity ligands with innate drug-likeness (11,12). Studies have shown that among the disparate protein families that constitute the human proteome, about 3,000 have been targeted by small synthetic/organic drug molecules with regards to therapeutic intervention in diseases (13–16). As such, there remains a large pool of other proteins that belong to families that are not targetable or ‘uninhibitable’ by small drug molecules.

The concept of druggability has also expanded to encompass the ‘druggable genome,’ which describes encoding genes for proteins that are not only physiologically relevant but whose activity can also be regulated by high-affinity ligands (11,14,16). It is also important to mention that druggability as a term is more suitable to describe protein targets than DNA and RNA sequences, even though they may also serve as active sites for drug molecules (12,17).

An integral component of drug development in recent years has, therefore, been target validation. Existing conventional processes of target validation techniques are known to investigate the interplay between alteration in protein biological functionality and corresponding therapeutic impacts. Thus, druggability incorporates a structure-based approach to evaluate tendencies that small drug-like molecules possess innate potencies to bind and modulate protein activities (11,13). Consequently, several approaches have emerged over the years to investigate the druggability of potential protein targets as succinctly discussed by Egner and Hillig (18).

Approaches employed to determine the druggability of a protein have been categorized based on 1) those that employ empirical analyses of established associations between drugs and proteins 2) structural analyses of target proteins (13,19–23). Moreover, when the protein in question is a member of a well-studied family with closely related homologs from which drug association details can be retrieved, the former method is more suitably applied (15,16). On the other hand, where there is no prior knowledge about the potential target or in cases where the target protein belongs to families for which no drug candidates are known, the latter method that employs structure-based predictive models is most appropriate (13).

A notable and pioneering contribution to the development of the structural predictive model was carried out by Hajduk et al., wherein they measured druggability using hit rates of NMR-based fragment screening for a diverse set of 23 proteins. This model incorporates a simple regression analysis that defined descriptors of the binding pocket such as the surface area, roughness, polar/apolar contact area, and the number of charged residues (21,24). Their analysis generated a predictive ‘druggability score,’ which is defines the logarithm of the predicted rate of NMR hit. Over time, a variety of methods have been developed to improve the prediction of protein druggability (11,13,19,23,25–27), some of which form an integral part of this review.

Drug-likeness is a significant consideration in the early stages of the drug discovery process when selecting compounds as it helps to optimize pharmacokinetics and pharmaceutical attributes. These include chemical stability, solubility, bioavailability, and distribution profile (12). Drug-like compounds also describe chemical molecules that possess functional groups and or have physical properties consistent with the majority of known drugs and hence can be inferred as compounds that might be active biologically or might show therapeutic potential (28,29). Thus these compounds fall within critical physicochemical thresholds such as molecular mass, hydrophobicity, and polarity when compared to the narrow range distribution of approved drugs (30,31). Regardless, drug-like compounds hold no apparent structural similarity to any approved drug. Although high rates of attrition have shown in drug discovery, the preferential selection of compounds with innate drug-likeness increases the likelihood of surviving this occurrence (32).

A critical understanding of the concepts of target druggability coupled with the drug-likeness of a chemical molecule could, therefore, further potentiate the success rate of novel chemical molecules towards becoming viable therapeutic options in disease intervention.

Therefore, in this review, we discuss the underlying principles of both druggability and drug-likeness as it relates to the drug discovery process. Also, we list the various methods that have been employed over the years to predict, determine, and achieve the druggability of a novel protein target as previously studied. We extend this methodical review to cover evolutionary advances made in determining the drug-likeness of a compound coupled with relevant experimental and computational techniques aimed at enhancing the drug-likeness of a non-drug hit compound.

2.0 Assessing Druggability of Biological targets

Advancements in the computational domain have paved a path toward algorithms that assess the druggability of a target and store any information in large virtual libraries. We now have access to a plethora of data, including protein and gene expression profiles, protein-protein interactions, and gene regulatory networks, thus making it possible to build combined predictive models. Initial algorithms used to assess druggability were designed to reproduce nuclear magnetic resonance (NMR) fragment screening hit rates (21,33,34). This approach, however, leads to proprietary data that was expensive to obtain and difficult to access publicly. Another method of assessing druggability was reported by Cheng *et al.* in which the maximum achievable affinity for a binding site based on the structure of the pocket was estimated (17). Likewise, a separate report by Sheridan *et al.* also described an equation that could approximate the “bindability” of a given binding pocket on a protein, based on similar functions and some particular descriptors as methods for assessing druggability(19). Experimental druggability assessments involved soaking protein crystals in various solvents and identifying binding sites based on X-ray diffraction data. Although this

method was the most reliable and validated technique, the crystallized protein was required (34,35).

Various computational tools that have been developed over the year to assess the druggability of a target based on the availability of a protein's 3D structure. Some of these tools are based on algorithms that consider structural features of a protein, including surface polarity, surface hydrophobicity, and pocket dimensions. The algorithms used are roughly classified into geometric and energetic approaches (36,37). In the geometric approach, the shape of the protein is investigated to identify any voids between the protein's atoms. The energetic approach uses interactive molecular energies, which are mapped onto the 3D structure of the protein through molecular dynamics simulations. The binding pockets are then identified based on the interaction energy profiles that are generated (34,37). The downfall of the approaches mentioned above is that a 3D structure of the protein is needed, the process may be slow, and the predictive pockets cannot be validated until an experimental analysis is performed. Because not all druggable targets are proteins, there are a few databases that identify other druggable targets, notably Therapeutic Target database (38), LncRNA2Target (39), and MiRBase (40).

Four primary methods are currently used for assessing druggability; these include the precedence-based, ligand-based, structure-based, and sequence-based methods, as highlighted in figure 4.1 (11). Although each method comes with its benefits, numerous caveats need to be considered and accounted for when identifying an ideal target. One of the most important considerations that are frequently overlooked is that compounds do not always target a protein. Other existing biological targets include; nucleic acids, other small molecules, protein-protein interactions and nucleic acid complexes such as ribosomes (35).

2.1 Sequenced-based druggability assessment methods

The basic unit of proteins, amino acids, which determine the function of a protein, has also been employed in predicting the druggability of protein drug targets. As such, sequence-based druggability assessment methods analyze sequence alignments of protein targets, identifying possible conserved regions that may be deemed predictive binding pockets(34). Assessment of potential target sequence also provides insights into essential physicochemical properties like solvent accessibility, hydrophobicity, charge, and polarity, amongst other properties(41). Sequence-based methods have also been widely employed in Machine-learning algorithms in assessing protein target druggability(41). Even though a knowledge of the target sequence could aid in the prediction of functional domains on the target, the target sequence alone is insufficient in providing an overall structural and functional insight of a protein target. Thus, the knowledge of target sequence alone is unable to provide details on the accessibility of these domains to either gene expression level, a drug molecule, or its relevance in the interactome(41).

2.2 Structure-based druggability assessment

Structure-based druggability assessment methods rely on a knowledge of the binding pocket located on the biological target to which a ligand can bind (34). Thus, for a small drug-like molecule to bind a protein, it requires a binding site that contains complementary characteristics, including an appropriate size that can accommodate a molecule and a deep hydrophobic pocket that can bury the drug within the protein. Large exposed polar binding sites are considered less druggable than smaller hydrophobic pockets (34,42). The structure-based assessment methods generally involve an initial identification of possible binding pockets followed by sorting the identified binding pockets based on the physicochemical properties of the pocket. Sorted binding pockets are then compared to a reference set of known biological targets to assess their degrees of

difficulty or druggability (43). Numerous software and online platforms have been developed to aid in the accurate identification of binding pockets, notably; DOGSiteScorer (26), Metapocket (44), PockDrug Server (45), SiteMap (23), Sitehound (46) and Open Targets (47). These online platforms provide essential binding pocket information, which could in determining the druggability of given biological targets.

2.3 Ligand-based druggability assessment

In ligand-based druggability assessment methods, the druggability of a therapeutic target is based on a knowledge of endogenous compounds that could bind to the target and are yet to reach clinical trials (34). Several databases available online contains sufficient information about several ligands that could be used in predicting the druggability of a given target. Notable examples of these online platforms include BindingDB (48), PubChem (49), SwissTargetPredict (50), TargetHunter (51), Guide to pharmacology (52), National Centre for Biotechnology Information(NCBI) (53) and Protein data bank (54).

2.4 Precedence-based druggability assessment

In employing precedence-based methods for assessing druggability, knowledge of established proteins or drug targets is required. Precedence-based methods also rely on a knowledge of chemical molecules that are already undergoing clinical trials (55). This method demonstrates the highest level of druggability as the given target has already been proven to be safe. However, precedence-based methods of assessing druggability are still not a guarantee of success for different and or product profiles (42,56). Various databases contain information about compounds, targets, and published works that could be employed in precedence-based druggability

assessments. Notable examples of some of these freely available databases include DrugCentral (57), Clinicaltrials.gov (58), and Drugbank (59). As of February 2020, the latest release of DrugBank(2020-01-03) alone contained about 13,491 drug entries, including 2,641 approved small molecule drugs, 1,364 approved biologics (proteins, peptides, vaccines, and allergenics), 130 nutraceuticals and over 6,347 experimental (discovery-phase) drugs.

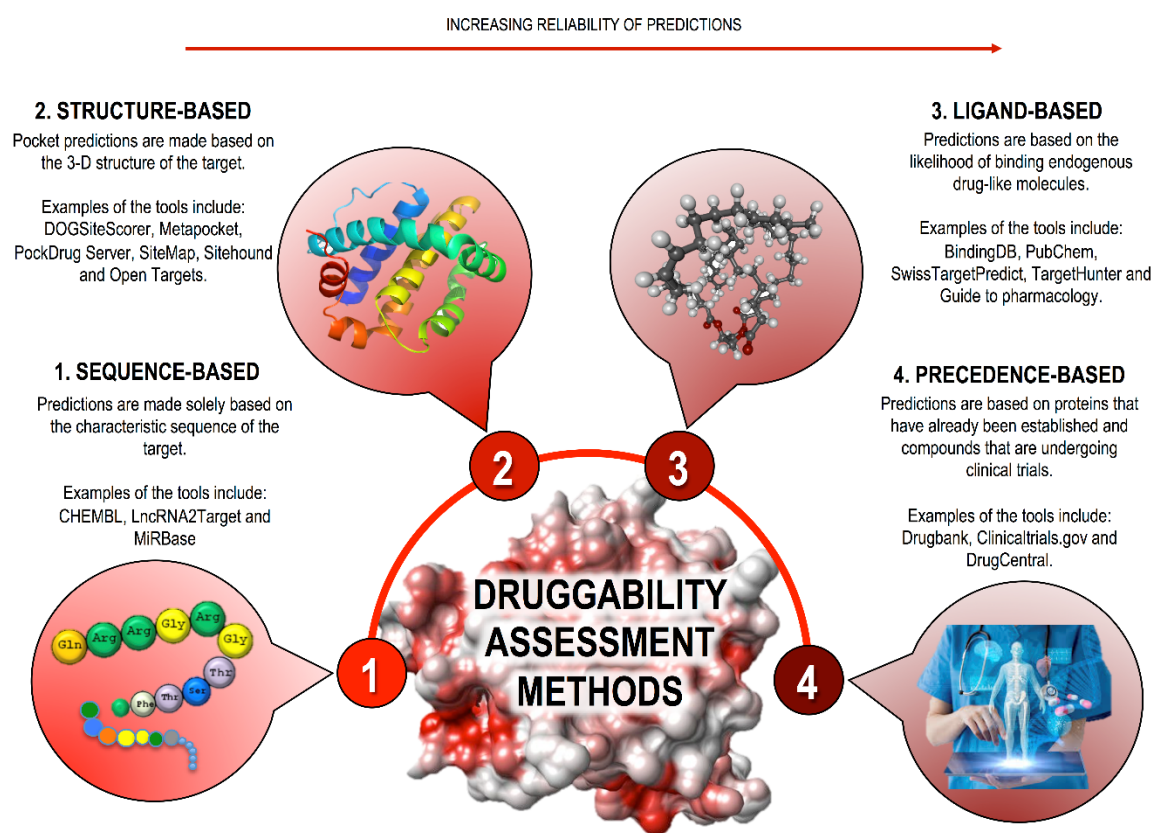


Figure 4.1: Methods for assessing the druggability of potential disease targets.

3.0 Are Predictive Druggability Assessment Methods Reliable?

While investigations of most disease targets have potential inhibitors, there is minimal experimental data associated with the research. This observation has led to almost all scientific domains to turn to machine learning and predictive computational tools. Not only do these methods save laborious hours in the wet-laboratory setting, but it also saves elaborate amounts of money that would have been spent on unnecessarily large stocks of reagents. Therefore, the use of predictive tools without validation proves to be controversial. Thus, the reliability of the results is questioned at every stage.

Druggability assessment methods utilize a wide range of algorithms, as discussed. It may be noted that according to the choice of predictors and parameters, bias results may be obtained. A geometry-based method may find the most suitable pocket based solely on the orientation and shape of the protein, whereas the energy-based method may find greater molecular forces in a different pocket of the same protein. It is, therefore, essential to choose the appropriate predictive software when identifying potentially druggable pockets of a protein since the principle and level of accuracy of each predictive model vary. For instance, in binding pocket prediction, several improved approaches have been developed over the years, since many of the existing prediction provide varying results. One such improved binding pocket methods is the metapocket method (44), which combined four different methods namely, SURFNET(60), PASS(61), LIGSITE(cs)(62) and Q-SiteFinder(63) towards improving its binding pocket prediction success rate. Therefore, although most of the various algorithms that have been employed by many existing druggability prediction models have been validated against different training sets, a comparative study of ten representative prediction methods suggest that, of the top three

performing methods, 30% of clinically known binding sites were not detected with existing prediction tool (64). Another shortfall of druggable models is that they do not adequately account for dynamic protein-ligand binding interactions (35). Thus, the relevance of binding site target interactions further indicates that although binding pocket prediction methods are useful in identifying putative druggable sites, they have limitations, and are thus not allowing for validation analysis. Therefore, the further integration of druggability information with other parameters, including the lead compound's drug-likeness, would provide a more effective and holistic mechanism of disease management.

4.0 Identification of lead molecules by assessing drug-likeness

The concept of drug-likeness has advanced extensively over the years, taking into consideration the structural, physicochemical, biochemical, pharmacokinetic (PK), and toxicity characteristics of a compound, thus a drug-like compound possesses sufficiently acceptable ADME properties, as well as sufficiently acceptable toxicity properties (65,66). Understanding these properties have become an integral part of drug discovery and has enabled the accurate selections of hits that are suitable starting points for the identification of new clinical candidates. These properties also allow a streamlining of drug discovery efforts towards the identification of a compound that possesses a higher chance of pharmacokinetic success and overall safety.

The time-consuming nature and enormous resources required in performing *in vivo* studies to determine the drug-likeness of a compound have presented computational techniques as a viable alternative. Christopher Lipinski and coworkers reported the earliest report on a decisive criteria for identifying drug-like compounds at Pfizer in the mid-1990s (67). Since then, several reports

have attempted to describe the drug-likeness of compounds, including the report by Sadowski *et al.* (68) in 1998, in which, for the first time, a scoring function was employed in predicting the drug-likeness of the compounds using neural networks (NN). Subsequently, many other reports have also employed neural networks in prediction of drug-likeness (69–71). Other computer-based approaches and molecular descriptors that have also been employed over the years in the prediction of drug-likeness include, recursive partitioning (RP) (72,73), support vector machine (SVM) (69,74) and genetic algorithm (GA) (75,76), amongst others. Over the years, these techniques have gained prominence in their ability to predict drug-likeness of compounds from non-drug compounds satisfactorily. This eventual incorporation of *in silico* techniques in predicting drug-likeness of compounds has allowed for the analysis of large compound libraries to identify compounds that could be subsequently prioritized for synthesis, *in vitro* and *in vivo*, and for thorough experimental evaluation. *In silico* approaches have thus minimized time and resources that would have been wasted in assessing non-drug molecules. Several molecular descriptors have been applied in profiling chemical compounds towards their drug-likeness. A report by Feher and Schmidt succinctly highlights several molecular descriptors that can be used in characterizing chemical compounds as employed in their study (77). Nonetheless, these are usually of little practical help to medicinal chemist since the assessment of drug-likeness with these molecular descriptors are usually based on any type of countable feature or computable property instead of profiling based on biologically relevant parameters.

4.1 Physicochemical properties and Pharmacokinetic properties in determining drug-likeness

Assessing the pharmacokinetic and physicochemical properties of compounds in determining their drug-likeness occurs in the early stages of the drug discovery process. This process takes into

consideration structural features and physicochemical properties that complement a given pharmacophore without affecting it. These pharmacokinetic properties inform the behavior of compounds in the body with regards to absorption, distribution, and excretion. Early pharmacokinetics and physicochemical assessment relied on molecular properties such as polarity (polar surface area (PSA) and topological surface area), number of aromatic rings, and number of heavy atoms, lipophilicity, molecular mass, number of chiral centers, number of hydrogen bond donors and acceptors and number of rotatable bonds. However, recently these have been expanded to include complex properties such as metabolic stability, permeability, transporters, solubility, physiologically based pharmacokinetics, and pharmacokinetic/pharmacodynamic modeling (78–80). A combination of all these complex properties allows for the optimization and identification of drug-like molecules while taking into consideration safety, selectivity, efficacy, and pharmacokinetics (81). Based on these molecular properties several guidelines for predicting drug-likeness of chemical molecules have been proposed, notably, the Lipinski's rule of five (67), Veber rules(82), Waring drug candidate permeability (83) and Golden triangle rules(84) amongst numerous other existing guidelines for the prediction of drug-likeness.

4.1.1 Lipinski's rule of five (RO5)

As one of the earliest guidelines proposed for predicting the drug-likeness of compounds, the RO5 emerged from the deposition criteria where analysis of drug-likeness is based on lipophilicity, molecular weight, and counts of hydrogen bond acceptors and hydrogen bond donors (85,86). According Lipinski, compounds with sufficiently acceptable ADME/T properties (thus a molecular weight not greater than 500 Da, log P not greater than 5, hydrogen-bond donor not more than 5 and hydrogen bond acceptor (nitrogen and oxygen atoms) not more than 10) to survive

through the Phase I clinical trials, were considered drug-like. As a guideline, the Lipinski's RO5 is conceptually simple and straightforward to implement, hence its widespread adoption. However, it does not apply to natural products or substrates of biological transporters (87). Based on Lipinski's assumptions, chemical molecules that violate these rules are less likely to possess many desirable characteristics of drugs. Nonetheless, there are plenty of examples available for RO5 violation amongst the existing drugs, notably antibiotics, antifungals, vitamins, and cardiac glycosides, but yet these are orally bioavailable(88). These classes of compounds are orally bioavailable because they possess groups that act as substrates for transporters(88). In RO5, the consideration of the number of bonds is due to their direct correlation with an increase in aqueous solubility and a decrease in lipid bilayer permeability. On the other hand, molecular weight directly correlates with the size of the compounds, since large molecular weighted compound could hinder the diffusion of molecules across lipid bilayer membranes as well as reduce the molecule solubility. High lipophilicity ($\log P > 5$) often contributes to high metabolic turnover, aqueous solubility, and reduced intestinal absorption. However, high lipophilicity of compounds is also associated with an increased risk of promiscuity and toxicity since these compounds tend to bind to hydrophobic biological targets other than the desired target.

4.1.2 Golden triangle rules

The golden triangle as criteria for selecting drug-like compounds takes into consideration the metabolic stability and permeability of the compound (84). In the report by Johnson *et al.* structural properties that were assessed included molecular weight and lipophilicity ($\log D_{7.4}$) since these parameters affected almost all other drug-likeness properties, could represent other molecular properties such as rotatable bonds, hydrogen bond acceptors and donors, molecular volume,

heteroatoms and PSA (84,89). Lipophilicity and molecular weight are also good predictors of the survival of compounds in development and are also compatible with chemistry tools ligand efficiency (LE) and ligand-lipophilicity efficiency (LLE) (90).

Drug-like compounds under the golden triangle rule, therefore, possess high membrane permeability and exhibits favorable metabolic stability if they have a triangle baseline: molecular weight of 200 and lipophilicity of $-2 < \log D_{7.4} < 5$ as well as a triangle apex: molecular weight of 450 and $\log D_{7.4} = 1.0-2.0$.

4.1.3 Waring drug candidate permeability guideline

The membrane permeability of small molecule compounds is a crucial metric in the drug design pipeline that is essential in the overall potency and *in vivo* efficacy of the compounds since membrane permeability ensures the compounds reach their intended target. As such, using membrane permeability as a parameter, the drug-likeness of compounds could be predicted. Using permeability as a parameter, Waring *et al.* (83), propounded a guideline for the prediction of highly permeable drug-like compounds with particular consideration of the apparent permeability (Papp) of the compounds. In their report, the Papp of the studied compounds was set within a range of $< 100 \text{ nm/s}$ (low) and $> 100 \text{ nm/s}$ (high). In generating the guideline, the set range of Papp was statistically contrasted with other molecular property predictors such as total hydrogen bonding, hydrogen bond acceptors and hydrogen-bond donors, molecular weight, rotatable bonds, logD, PSA and logP. Their report concluded that close to property limits, the probability of obtaining high permeability compounds was lower. Their report also showed that the permeability of the compounds varies over different ranges of molecular weight.

5.0 *In silico* Techniques in predicting drug-likeness

The advent of computer-aided drug design has influenced a paradigm shift towards incorporating computational techniques in identifying drug-like compounds while eliminating compounds that may be unsuccessful in the drug discovery process. Over the years, experiment filters have been at the forefront in the drug discovery process as the reliable methods are isolating drug-like compounds. However, these conventional approaches are time demanding and resource-intensive (28,91). As such, computational approaches have gained prominence in the last two decades as reliable approaches in the prediction of drug-like compounds with much ease. Computational techniques have typically been involved the filtering out of non-drug compounds as well as compounds regarded as unfit for screening purposes by taking into consideration fundamental physiochemical and pharmacokinetic properties (29,92,93). In recent decades, several computational software and online platforms have been developed in predicting essential molecular properties of compounds towards the identification of drug molecules with promising potential of achieving success in the drug discovery process, as highlighted in Table 4.1.

Table 4.1: *In silico* tools utilized in the prediction of molecule drug-likeness (94)

<i>In silico</i> tool	Function
QikProp(https://www.schrodinger.com/qikprop) Distributed by Schrodinger	Predicts and evaluates the ADME properties of thousands of compounds within a short time

<p>VolSurf+ (https://www.moldiscovery.com/software/volsurf/) Distributed by Molecular Discovery Ltd. (moldiscovery.com)</p>	<p>A computational tool that calculates and predicts essential ADME properties of compounds using predictive 3D molecular interaction energy grid maps.</p>
<p>MedChem Studio (https://www.simulations-plus.com/software/admetpredictor/medchem-studio/) Distributed by Simulation Plus, Inc</p>	<p>A platform commonly used by medicinal chemists and computational chemists for identifying lead compounds. It could also be employed for the optimization of the identified lead compounds as well as the prediction of ADME properties. MedChem studio is a complete computer-aided drug design package that can also be used for ligand-based drug design and the categorization of libraries of compounds.</p>
<p>ADMET Predictor (https://www.simulations-plus.com/software/admetpredictor/) Distributed by Simulations Plus, Inc</p>	<p>A computational software that allows for the prediction of essential physicochemical ADMET properties as well crucial CYP metabolism kinetics of molecular structures. ADMET Predictor can also be used to generate predictive models of entirely new properties from the user's library of structures using the inbuilt ADMET Modeler module</p>
<p>ADMEWORKS ModelBuilder (https://www.fqs.pl/en/products) Distributed by Fujitsu</p>	<p>ADMEWORKS ModelBuilder generates QSAR/QSPR models that can be employed in predicting physicochemical and pharmacokinetic properties of compounds.</p>
<p>ADMEWORKS Predictor (https://www.fqs.pl/en/products) Distributed by Fujitsu</p>	<p>An <i>in silico</i> tool for evaluating and predicting essential pharmacological properties of compounds by employing QSAR based virtual screening</p>

<p>IMPACT-F (http://www.pharmainformatic.com/html/impact-f.html) Developed by PharmaInformatic, Germany</p>	<p>A reliable tool made up of QSAR models for evaluating the human oral bioavailability of drug candidates.</p>
<p>ADMET Modeler (http://www.simulations-plus.com/software/admetpredictor/admet-modeler/) Distributed by Simulation Plus, Inc</p>	<p>A plugin of the ADMET Predictor that addresses the tedious process of building reliable QSAR/QSPR models from experimentally derived data sets by automating the entire process.</p>
<p>ACD/PhysChem Suite (https://www.acdlabs.com/products/percepta/) Distributed by ACD/Labs</p>	<p>Employs fragment-based models to rapidly predict various physicochemical and pharmacokinetic properties of compounds such as lipophilicity (logD, logP), aqueous solubility, pKa values amongst other molecular descriptors</p>
<p>ACD/ADME Suite (https://www.acdlabs.com/products/percepta/index.php) Distributed by ACD/Labs</p>	<p>A computational tool that predicts the ADME properties of compounds by estimating various molecular descriptors such as oral bioavailability, blood-brain barrier permeability, absorption amongst other</p>
<p>ACD/DMSO (https://www.acdlabs.com/products/percepta/index.php) Distributed by ACD/Labs</p>	<p>Predicts solubility in DMSO solution.</p>
<p>Discovery Studio ADMET Software (https://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/qsar-admet-and-predictive-toxicology.html) Distributed by Accelrys</p>	<p>Predicts the ADMET properties for compounds</p>
<p>MedChem Designer (http://www.simulations-plus.com/software/medchem-designer/) Distributed by Simulation Plus, Inc</p>	<p>ADMET properties predicting tools that incorporate chemical structure drawing features.</p>

Filter-it (http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html) Distributed by Silicos-it	Based on pre-programmed molecular descriptors, the Filter-it tool screens out compounds with undesirable properties from a set of compounds.
PreADME (https://preadmet.bmdrc.kr/) hosted by Yonsei University, Seoul, Republic of Korea.	An <i>in silico</i> tool for predicting the drug-likeness of compounds as well as the calculation of molecular descriptors.
PrologP/PprologD (https://www.compuDrug.com/?q=node/42) Distributed by CompuDrug	Employs both neural network methods and linear methods towards the prediction of the lipophilicity (logP/logD) of compounds.
Web Services for predicting drug-likeness	
Chemicalize (https://chemicalize.com/). Provided by ChemAxon	Online service for the prediction of physicochemical and pharmacokinetic properties such as logP, molecular weight, hydrogen bond donors, and acceptors. It also calculates molecular properties such as tautomers and PSA.
SwissADME (http://www.swissadme.ch/). Developed and maintained by the Molecular Modeling Group of the SIB Swiss Institute of Bioinformatics	A robust and reliable online service for predicting drug-likeness of compounds as well as their pharmacokinetic and physicochemical properties based on reliable inhouse methods notably; the Bioavailability rader, BOILED-Egg, synthetic accessibility score and, iLOGP.
AquaSol(http://cdb.ics.uci.edu/cgibin/tools/AquaSolWeb.py). Provided by the University of California, Irvine	Employs UG-RNN ensembles to predicts aqueous solubility of compounds.

<p>Molinfo (http://cdb.ics.uci.edu/cgi-bin/tools/MolInfoWeb.py). Provided by the University of California, Irvine</p>	<p>An online platform for predicting various essential molecular properties of compounds that are of interest to medicinal chemists.</p>
<p>DrugMint (http://crdd.osdd.net/oscadd/drugmint/) Maintained by Indraprastha Institute of Information Technology</p>	<p>Web server predicting the drug-likeness of compounds.</p>
<p>DrugLogit(http://hermes.chem.ut.ee/~alfx/druglogit.html) Maintained by the Institute of Chemistry, University of Tartu, Estonia</p>	<p>An online that predicts the chances of chemical structural being categorized as a drug or a non-drug compound Web service to predict the probability of a compound. DrugLogit also performs the classification of compounds based on the disease category.</p>
<p>AdmetSAR(http://lmm.d.ecust.edu.cn/admetsar1/) Maintained by Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology</p>	<p>A web server that predicts ADMET properties of compounds. A peculiar feature of this server is its ability to predict about 50 ADMET endpoints using a cheminformatics-based tool called ADMET-Simulator.</p>
<p>PkCSM (http://biosig.unimelb.edu.au/pkcsm/) Developed by the University of Cambridge, UK.</p>	<p>This online service employs graph-based signatures to calculate the pharmacokinetic properties of compounds.</p>
<p>MODEL-Molecular Descriptor Lab (http://jing.cz3.nus.edu.sg/cgi%E2%80%90bin/model/model.cgi) Maintained by the University of Singapore.</p>	<p>Based on 3D structures, this online service calculates the physiochemical and structural properties of compounds.</p>
<p>OSIRIS Property Explorer (https://www.organic-chemistry.org/prog/peo/) Maintained by the Virtual Computational Chemistry Laboratory</p>	<p>This web service allows for the calculation of the druglike properties of drawn structures as well as the prediction of toxicity profiles. It forms part of the</p>

	Actelion's in-house substance registration system.
Property calculator (https://mcule.com/apps/property-calculator/) Provided by mcule	Generate the physicochemical properties of chemical structures.
PreADMET (https://preadmet.bmdrc.kr/) Maintained by the Bioinformatics and molecular design research center of Yonsei University, Korea	An online web service that calculates the ADME properties of compounds. It also generates a drug-like compound library using computational methods.
Databases	
ADME-AP (http://bidd.nus.edu.sg/group/admeap/admeap.asp) Maintained by the Dept. Computational Science. NUS	ADME-AP is a database that houses a vast library of data about drug targets such as tissue distributions, functions, substrates, and known ADME linked targets. It simplifies the search for drug ADME related proteins.
The ADME database (https://www.fujitsu.com/jp/group/kyushu/en/solutions/industry/lifescience/admedatabase/) Distributed by Aureus	An extensive database of structurally different compounds coupled with known ADME features.

Nonetheless, the challenge with many of such tools is the unavailability of validating sets of compounds, thus some compounds are likely not to fall under any of the criteria being used by the computational tool. Also, many of these computational tools identify drug-like compounds based on features of existing drugs, hence drug-like compounds with new properties could easily be mistaken as non-drug compounds (91,95). Therefore, these computational tools play an essential role in the early stages of the drug discovery process, where they aid in the prioritization compound for further experimental exploration and other techniques such as high throughput screening.

6.0 Conclusion

In this review, the fundamental principles and the central concepts of druggability and drug-likeness have been briefly communicated. There are vast amounts of information and opinions in the assessment and analysis of both the drug-likeness of a molecule and the identification of its biological target. In this study, however, computational approaches were categorized and briefly explained. This provided a deeper appreciation of how online predictive tools and techniques streamline the drug discovery process by not wasting a copious number of years with “trial and error” experimental investigations. Although these “click of a button” procedures may shorten the rational drug discovery process, it is always important to keep in mind that predictive data is only ever “potential” information, until validated with experimental findings. We deduce that the synergistic capabilities of predictive and experimental investigations may lead to a greater understanding of drug-resistance in diseases, thereby allowing for more rapid and efficient treatment regimes.

7.0 Acknowledgement

The authors appreciate the College of Health Sciences, University of KwaZulu-Natal for financial and infrastructural support while we also thank the Center for High-Performance Computing (CHPC, www.chpc.ac.za) Cape-Town, South Africa for providing computational resources.

8.0 Conflict of Interest

Authors declare no conflict of interest.

9.0 References

1. Dang C V., Reddy EP, Shokat KM, Soucek L. Drugging the “undruggable” cancer targets. *Nature Reviews Cancer*. 2017; 17(8):502-508.
2. Galdeano C. Drugging the undruggable: targeting challenging E3 ligases for personalized medicine. *Future Med Chem*. 2017; 9(4):347-350.
3. Zhang ZY. Drugging the undruggable: Therapeutic potential of targeting protein tyrosine phosphatases. *Acc Chem Res*. 2017; 50(1):122-129.
4. Machado D, Girardini M, Viveiros M, Pieroni M. Challenging the drug-likeness dogma for new drug discovery in Tuberculosis. *Frontiers in Microbiology*. 2018; 9:1367.
5. Sakharkar M, Sakharkar K. Targetability of human disease genes. *Curr Drug Discov Tech*. 2007;4:48–58.
6. Taboureau O, Nielsen S, Audouze K. ChemProt: a disease chemical biology database. *Nucleic Acids Res*. 2011;39:D367–D372.
7. Dixon S, Stockwell B. Identifying druggable disease-modifying gene products. *Curr Opin Chem Biol*. 2009;13:549–555.
8. Scannell J, Blanckley A, Boldon H et al. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov*. 2012;11:191–200.
9. Hingorani A, Kuan V, Finan C et al. Improving the odds of drug development success through human genomics: modelling study. *Sci Rep*. 2019;9:18911.
10. Hay M, Thomas D, Craighead J, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol*. 2014;32:40–51.

11. Schmidtke P, Barril X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J Med Chem*. 2010;53(15):5858–67.
12. Vistoli G, Pedretti A, Testa B. Assessing drug-likeness - what are we missing? *Drug Discovery Today*. 2008; 13(7-8):285-94.
13. Kozakov D, Hall DR, Napoleon RL, Yueh C, Whitty A, Vajda S. New Frontiers in Druggability. *Journal of Medicinal Chemistry*. 2015; 58, 23, 9063-9088.
14. Oprea TI, Bologna CG, Brunak S, Campbell A, Gan GN, Gaulton A, et al. Unexplored therapeutic opportunities in the human genome. *Nature Reviews Drug Discovery*. 2018; 7, pages317–332.
15. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med*. 2017; 9(383). pii: eaag1166.
16. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002; 1(9):727-30.
17. Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, et al. Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol*. 2007; 25(1):71-5.
18. Egner U, Hillig RC. A structural biology view of target drugability. *Expert Opin Drug Discov*. 2008; 3(4):391-401.
19. Sheridan RP, Maiorov VN, Holloway MK, Cornell WD, Gao YD. Drug-like density: A method of quantifying the “bindability” of a protein target based on a very large set of pockets and drug-like ligands from the protein data bank. *J Chem Inf Model*. 2010;50(11):2029-40.

20. Rathi PC, Ludlow RF, Hall RJ, Murray CW, Mortenson PN, Verdonk ML. Predicting “Hot” and “Warm” Spots for Fragment Binding. *J Med Chem*. 2017; 60(9):4036-4046.
21. Hajduk PJ, Huth JR, Fesik SW. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem*. 2005; 48(7):2518-25.
22. Seco J, Luque FJ, Barril X. Binding site detection and druggability index from first principles. *J Med Chem*. 2009;52(8):2363-71.
23. Halgren TA. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model*. 2009;49(2):377–89.
24. Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. *Drug Discovery Today*. 2005; 10(23-24):1675-82.
25. Krasowski A, Muthas D, Sarkar A, Schmitt S, Brenk R. DrugPred: A structure-based approach to predict protein druggability developed using an extensive nonredundant data set. *J Chem Inf Model*. 2011; 51(11):2829-42.
26. Volkamer A, Kuhn D, Rippmann F, Rarey M. Dogsitescorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*. 2012;28(15):2074–5.
27. Perola E, Herman L, Weiss J. Development of a rule-based method for the assessment of protein druggability. *J Chem Inf Model*. 2012; 2(4):1027-38.
28. Walters W, Ajay A, Murcko M. Recognizing molecules with drug-like properties. *Curr Opin Chem Biol*. 1999;3(4):384–7.
29. Walters W, Stahl M, Murcko M. Virtual screening – An overview. *Drug Discov Today*.

1998;3:160–178.

30. Bickerton GR, Paolini G V., Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. *Nat Chem.* 2012; 4:90–98

31. Oprea TI. Property distribution of drug-related chemical databases. *J Comput Aided Mol Des.* 2000; 14(3):251-64.

32. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov.* 2007; 6(11):881-90.

33. Katsila T, Spyroulias GA, Patrinos GP, Matsoukas MT. Computational approaches in target identification and drug discovery. *Comput Struct Biotechnol J.* 2016;14(1):177–84.

34. Barril X. Druggability predictions: Methods, limitations, and applications. *Wiley Interdiscip Rev Comput Mol Sci.* 2013;3(4):327–38.

35. Cheng T, Hao M, Takeda T, Bryant SH, Wang Y. Large-Scale Prediction of Drug-Target Interaction: a Data-Centric Review. *AAPS J.* 2017;19(5):1264–75.

36. Neal K. B, Mahmoud E. S. Can We Rely on Computational Predictions To Correctly Identify Ligand Binding Sites on Novel Protein Drug Targets? Assessment of Binding Site Prediction Methods and a Protocol for Validation of Predicted Binding Sites. *Cell Biochem Biophys.* 2017;75(1):15–23.

37. Sun T, Lai L, Pei J. Analysis of protein features and machine learning algorithms for prediction of druggable proteins. *Quant Biol.* 2018;6(4):334–43.

38. Li YH, Yu CY, Li XX, Zhang P, Tang J, Yang Q, et al. Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic*

Acids Res. 2018;46(D1):D1121–7.

39. Jiang Q, Wang J, Wu X, Ma R, Zhang T, Jin S, et al. LncRNA2Target: A database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res.* 2015;43(D1):D193–6.

40. Griffiths-Jones S. MiRBase: MicroRNA sequences and annotation. *Curr Protoc Bioinforma.* 2010;34(29):1291–12910.

41. Kandoi G, Acencio ML, Lemke N. Prediction of druggable proteins using machine learning and systems biology: A mini-review. *Front Physiol.* 2015;6(9):54–65.

42. G. Wyatt P, H. Gilbert I, D. Read K, H. Fairlamb A. Target Validation: Linking Target and Chemical Properties to Desired Product Profile. *Curr Top Med Chem.* 2011;11(10):1275–83.

43. Fauman E, Rai B, Huang E. Structure-based druggability assessment—identifying suitable targets for small molecule therapeutics. *Curr Opin Chem Biol.* 2011;15(4):463–468.

44. Huang B. MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. *Omi A J Integr Biol* [Internet]. 2009;13(4):325–30. Available from: <http://www.liebertonline.com/doi/abs/10.1089/omi.2009.0045>

45. Hussein HA, Borrel A, Geneix C, Petitjean M, Regad L, Camproux AC. PockDrug-Server: A new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Res.* 2015;43(W1):W436–42.

46. Hernandez M, Ghersi D, Sanchez R. SITEHOUND-web: A server for ligand binding site identification in protein structures. *Nucleic Acids Res.* 2009;37:413–6.

47. Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, et al. Open

Targets: A platform for therapeutic target identification and Validation. *Nucleic Acids Res.* 2017;45(D1):D985–94.

48. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 2007;35(D4):198–201.

49. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* 2019;47(D1):D1102–9.

50. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. SwissTargetPrediction: A web server for target prediction of bioactive small molecules. *Nucleic Acids Res.* 2014;42(W1):32–8.

51. Wang L, Ma C, Wipf P, Liu H, Su W, Xie X-Q. TargetHunter: An In Silico Target Identification Tool for Predicting Therapeutic Potential of Small Organic Molecules Based on Chemogenomic Database. *AAPS J.* 2013;15(2):395–406.

52. Harding SD, Sharman JL, Faccenda E, Southan C, Pawson AJ, Ireland S, et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. *Nucleic Acids Res.* 2017;46(D1):D1091–106.

53. Bethesda (MD): National Library of Medicine (US). National Center for Biotechnology Information (NCBI). Accessed [11/03/2018]. 1988.

54. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. 2000;28(1):235–42.

55. Loving KA, Lin A, Cheng AC. Structure-Based Druggability Assessment of the

Mammalian Structural Proteome with Inclusion of Light Protein Flexibility. *PLoS Comput Biol.* 2014;10(7):321–9.

56. Edfeldt F, Folmer R, Breeze A. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov Today.* 2011;16(7–8):284–7.

57. Ursu O, Holmes J, Bologna CG, Yang JJ, Mathias SL, Stathias V, et al. DrugCentral 2018: An update. *Nucleic Acids Res.* 2019;47(D1):D963–70.

58. Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials.gov in research. *Bmj.* 2018;361(4):1452–9.

59. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46(D1):D1074–82.

60. Laskowski R. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph.* 1995;13:323–30.

61. Brady G, Stouten P. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des.* 2000;14:383–401.

62. Huang B, Schroeder M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct Biol.* 2006;6:19.

63. Laurie A, Jackson R. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics.* 2005;21(9):1908–16.

64. Leis S, Schneider S, Zacharias M. In Silico Prediction of Binding Sites on Proteins. *Curr Med Chem.* 2010;17(15):1550–1562.

65. Kerns E, Di L. *Drug-like Properties: Concepts, Structure Design and Methods.* Burlington:

Academic Press; 2008; 552

66. Schneider G. Prediction of Drug-Like Properties. In: Madame Curie Bioscience Database. Austin (TX): Landes Bioscience;
67. Lipinski C, Dominy B, Feeney P. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 1997;23:3–25.
68. Sadowski J, Kubinyi H. A scoring scheme for discriminating between drugs and nondrugs. *J Med Chem.* 1998;41:3325–3329.
69. Byvatov E, Fechner U, Sadowski J, Schneider G. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci.* 2003;43:1882–9.
70. Takaoka Y, Endo Y, Yamanobe S, Kakinuma H, Okubo T, Shimazaki Y, et al. Development of a method for evaluating drug-likeness and ease of synthesis using a data set in which compounds are assigned scores based on chemists' intuition. *J Chem Inf Comput Sci.* 2003;43:1269–75.
71. Ajay A, Walters W, Murcko M. Can we learn to distinguish between “drug-like” and “nondrug-like” molecules? *J Med Chem.* 1998;41:3314–3324.
72. Wagener M, van Geerestein V. Potential drugs and nondrugs: Prediction and identification of important structural features. *J Chem Inf Comput Sci.* 2000;40:280–92.
73. Schneider N, Jaeckels C, Andres C, Hutter M. Gradual in silico filtering for druglike substances. *J Chem Inf Model.* 2008;48:613–28.

74. Zernov V, Balakin K, Ivaschenko A, Savchuk N, Pletnev I. Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions. *J Chem Inf Model.* 43(2003):2048–56.
75. Gillet V, Khatib W, Willett P, Fleming P, Green D. Combinatorial library design using a multiobjective genetic algorithm. *J Chem Inf Comput Sci.* 2002;42:375–85.
76. Gillet V, Willett P, Bradshaw J. Identification of biological activity profiles using substructural analysis and genetic algorithms. *J Chem Inf Comput Sci.* 1998;38:165–79.
77. Feher M, Schmidt J. Property distribution: differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci.* 2003;43:218–227.
78. Rowland M, Peck C, Tucker G. Physiologically-based pharmacokinetics in drug development and regulatory science. *Annu Rev Pharmacol Toxicol.* 2011;51:45–73.
79. Gabrielsson J, Green A. Quantitative pharmacology or pharmacokinetic pharmacodynamic integration should be a vital component in integrative pharmacology. *J Pharmacol Exp Ther.* 2009;331:767–774.
80. Wager T, Hou X, Villalobos A. Moving beyond rules: The development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem Neurosci.* 2010;1:435–449.
81. Li D, Kerns E, Carter G. Drug-like property concepts in pharmaceutical design. *Curr Pharm Des.* 2009;15:2184–2194.
82. Veber D, Johnson S, Cheng H, Smith B, Ward K, Kopple K. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem.* 2002;45:2615–23.

83. Waring M. Defining optimum lipophilicity and molecular weight ranges for drug candidates—molecular weight dependent lower logD limits based on permeability. *Bioorg Med Chem Lett.* 2009;19:2844–2851.
84. Johnson T, Dress K, Edwards M. Using the golden triangle to optimize clearance and oral absorption. *Bioorg Med Chem Lett.* 2009;19:5560–5564.
85. Lipinski C. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods.* 2000;44(1):235–49.
86. Lipinski CA. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. *Adv Drug Deliv Rev.* 2016;101:34–41.
87. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development setting. *Adv Drug Deliv Rev.* 2012;64:4–17.
88. Kadam R, Roy N. Recent trends in drug-likeness prediction: A comprehensive review of In silico methods. *Indian J Pharm Sci.* 2007;69(5):609–15.
89. Bhal S, Kassam K, Peirson I, Pearl G. The rule of five revisited: applying log D in place of log P in drug-likeness filters. *Mol Pharm.* 2007;4:556–560.
90. Murphy RB, Philipp DM, Friesner RA. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J Comput Chem.* 2000 Dec;21(16):1442–57.
91. Clark D, Pickett S. Computational methods for the prediction of “drug-likeness.” *Drug*

Discov Today. 2000;5(2):49–58.

92. Lewis R, Mason J, McLay I. Similarity measures for rational set selection and analysis of combinatorial libraries: The diverse property-derived (DPD) approach. *J Chem Inf Comput Sci.* 1997;37:599–614.
93. Rishton G. Reactive compounds and in vitro false positives in HTS. *Drug Discov Today.* 1997;2:382–384.
94. Bioinformatics SI of. Click2Drug. 2013. p. 1–10.
95. Polinsky A. Combichem and chemoinformatics. *curr opin Drug Discov Devel.* 1999;2:197–203.

CHAPTER 5

Published Article

The Binding of Remdesivir to SARS-CoV-2 RNA-Dependent RNA polymerase May Pave The Way Towards the Design of Potential Drugs for COVID-19 Treatment

Clement Agoni^a, and Mahmoud E. S. Soliman^{a*}

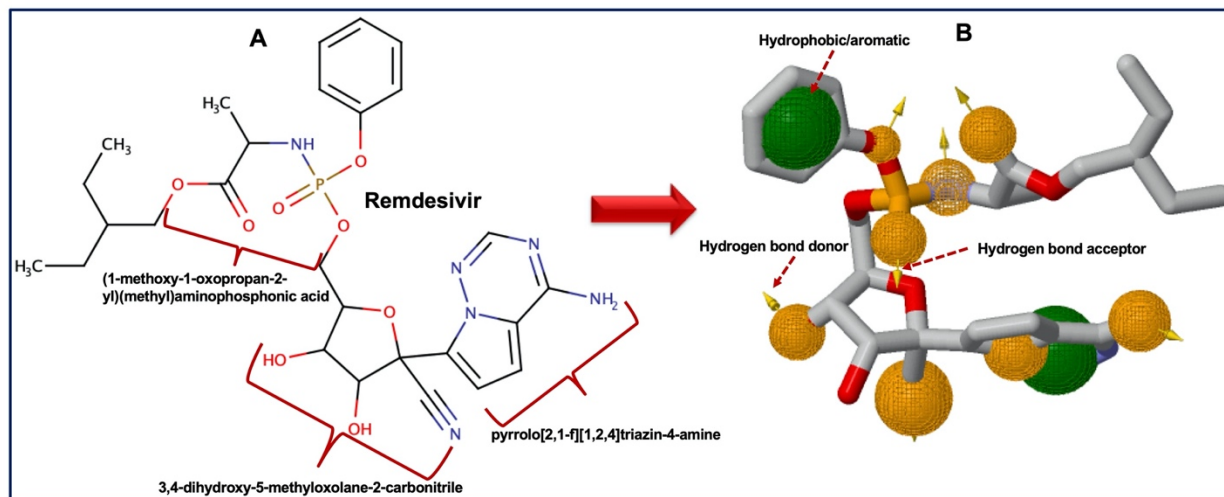
^aMolecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University
of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 8048, Fax: +27 (0) 31 260 7872

GRAPHICAL ABSTRACT



ABSTRACT

Aim: We seek to provide an understanding of the binding mechanism of Remdesivir unravel the structural and conformational implications on SARS-CoV-2 virus RNA-dependent RNA polymerase upon its binding and identify its crucial pharmacophoric moieties.

Background: The coronavirus disease of 2019 (COVID-19) pandemic has infected over a million people, with over 65,000 deaths as of the first quarter of 2020. The current limitation of effective treatment options with no approved vaccine or targeted therapeutics for the treatment of COVID-19 has posed serious global health threats. This has necessitated several drug and vaccine development efforts across the globe. To date, the farthest in the drug development pipeline so far is Remdesivir.

Objectives: We perform molecular dynamics simulation, quantify the energy contributions of binding site residues using per-residue energy decomposition calculations, and subsequently generate a pharmacophore model for the identification of potential SARS-CoV-2 virus RNA-dependent RNA polymerase inhibitors.

Methods: Integrative molecular dynamics simulations and thermodynamic calculations coupled with advanced post-molecular dynamics analysis techniques were employed.

Results: Our analysis showed that the modulatory activity of Remdesivir is characterized by an extensive array of high-affinity and consistent molecular interactions with specific active site residues that anchor Remdesivir within the binding pocket for efficient binding. These residues are ASP452, THR456, ARG555, THR556, VAL557, ARG624, THR680, SER681, and SER682. Results also showed that Remdesivir binding, induces minimal individual amino acid perturbations, subtly interferes with deviations of C- α atoms and restricts the systematic transition of SARS-CoV-2 RNA-dependent RNA polymerase from the “buried” hydrophobic region to the “surface-exposed” hydrophilic region. We also mapped a pharmacophore model based on observed high-affinity interactions with SARS-CoV-2 virus RNA-dependent RNA polymerase, which showcased the crucial functional moieties of Remdesivir and was subsequently employed for virtual screening.

Conclusion: The structural insights and the optimized pharmacophoric model provided would augment the design of improved analogs of Remdesivir that could expand treatment options for COVID-19

Keywords: COVID-19; SARS-CoV-2 RNA-dependent RNA polymerase; Remdesivir, Homology modelling; per-residue energy decomposition; Pharmacophore; molecular dynamic simulations

1.0 INTRODUCTION

The current global pandemic, Coronavirus disease (COVID-19), is caused by coronavirus SARS-CoV-2, a severe acute respiratory syndrome virus [1]. Even though the origin of the disease is still not clear yet [2], the COVID-19 pandemic manifested first in the city of Wuhan in China in December of 2019[1,3–5]. As of the end of July 2020, COVID-19 has caused about 666,285 deaths, with about 17,042,890 confirmed infected cases globally [6].

SARS-CoV-2 is a member of an enveloped positive-sense RNA family of viruses that are about 60nm-140nm in diameter. They possess spike-like projections on their outer surfaces, as shown in figure 5.1, that gives them a crown-shape appearance under an electron microscope [7]. The first severe acute respiratory syndrome caused by coronaviruses, SARS-CoV, occurred in 2003 [8–10]. SARS-CoV affected five different, with an 11% lethal rate and a global total of 8422 cases [8,9,11]. The Middle East respiratory syndrome coronavirus (MERS-CoV) diseases that also emerged in 2012, affecting 2494 individuals with a 34% lethal rate was also caused by coronaviruses[12,13]. Although SARS-CoV-2 is less lethal compared to MERS-CoV and SARS-CoV, it however, has a faster rate of human to human transmission.

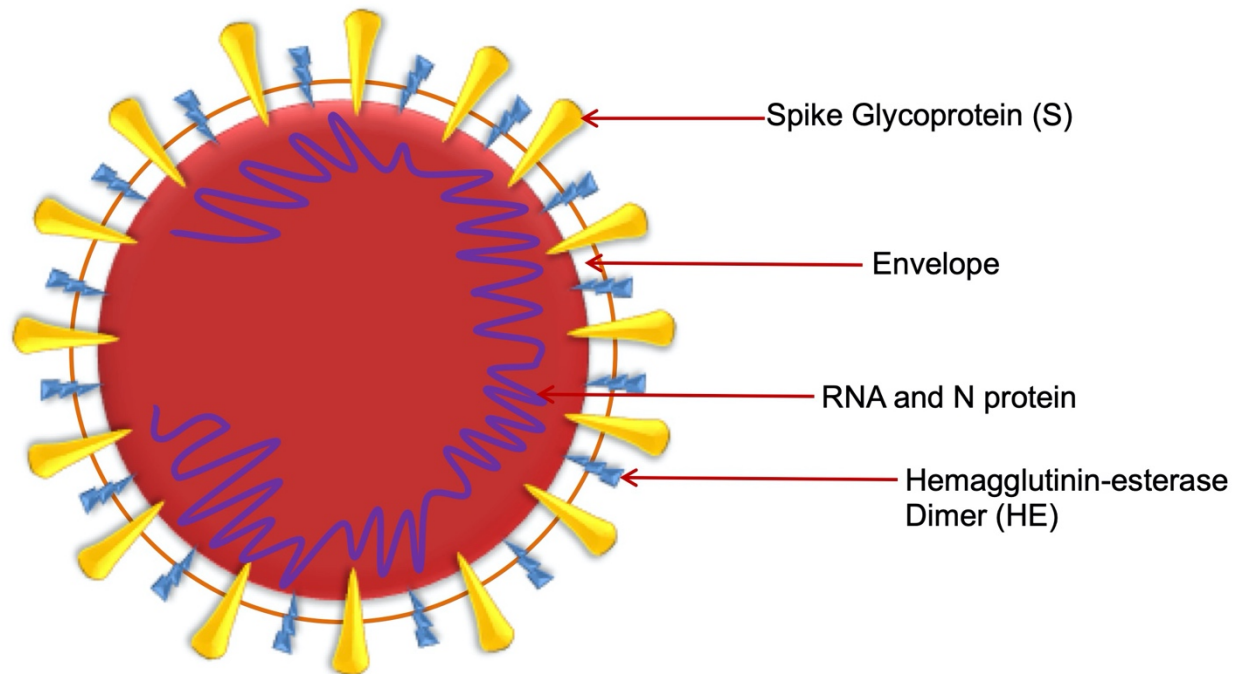


Figure 5.1: Diagram of a Coronavirus virion structure

Transmission of COVID-19 amongst humans is through the inhalation of large droplets released through coughing and sneezing of an infected individual [3,14,15]. Infection of the virus can also occur through the touching of contaminated surfaces and then touching one's eyes, nose, and face[16]. This is because the virus has been shown to survive on surfaces for days if the atmosphere is favourable [14,16]. Transmission *via* aerosolization/feco oral route has also been postulated[17]. Reports have shown that infected individuals can continue to transmit the disease as the symptoms persist, and even after the individual has clinically recovered [3,15]. Transmission can also occur in asymptomatic individuals before the onset of disease symptoms [14,18]. Symptomatic individuals usually experience fever, dyspnea, respiratory symptoms, and shortness of breath [14,19,20]. Severe acute respiratory syndrome, kidney failure, pneumonia, and even death may occur in individuals with severe forms of the disease [19]. Although COVIDS-19 affects all ages

with an overall fatality rate of 3%, fatality is higher in the elderly population and individuals with underlying co-morbidities [6,14].

There is currently no approved vaccine or targeted therapeutics for human coronavirus treatment, particularly COVID-19 [21]. Thus there is a limitation of effective treatment options. As such, all available treatment protocols are mainly supportive [22–24]. However, several drug development efforts are underway. The World Health Organisation is also currently focusing on four therapies which, include; Remdesivir, Chloroquine, and Hydroxychloroquine, a combination of Lopinavir and ritonavir with the possible addition of interferon-beta [25–27].

As one of the small molecule inhibitors being considered for COVID-19 therapy, Remdesivir is probably the most advanced of all the compounds in clinical trials [28]. It is an adenosine analog with broad-spectrum antiviral activity in animal models, both *in vitro* and *in vivo*, which failed as an Ebola drug [29]. Remdesivir has also been shown to inhibit coronaviruses that cause SARS and MERS [30–32]. It impedes viral replication by inhibiting the RNA-dependent RNA polymerase, thereby causing premature termination of RNA transcription [30,33].

Although Remdesivir has been shown to display potent *in vitro* activity against SARS-CoV-2 RNA-dependent RNA polymerase with an EC₅₀ at 48 hours of 0.77 μM in Vero E6 cells, and hence its advancement in clinical trials, the conformational, and structural changes associated with SARS-CoV-2 RNA-dependent RNA polymerase upon Remdesivir binding remains unexplored. Conformational changes of enzymes are the hallmarks of their dynamics that have a correlation with the overall functions of these enzymes [34]. Conformation changes may be induced by the binding of ligands or other biological molecules that can modulate the function of the enzyme [35]. In this case, the binding of Remdesivir would have induced significant conformations. Therefore

to unravel these conformational changes, we used molecular dynamics simulation, a formidable tool that allows for an atomistic time-resolved investigation of the properties of enzyme structures [36,37]. Based on the structural dynamics and associated interaction profile, we also construct a pharmacophore model that can be used in the design of novel inhibitors with improved therapeutic activity against all other coronaviruses. A pharmacophore model elucidates and identifies the common chemical moieties of ligands which, are crucial in eliciting a biological function [38–44]. Therefore the model that will be generated in this study would reveal for the first time, the crucial moieties of Remdesivir required for its inhibition of SARS-CoV-2 RNA-dependent RNA polymerase using a per-residue-based energy decomposition technique. This model could subsequently be employed in virtual screening towards the discovery of novel SARS-CoV-2 RNA-dependent RNA polymerase inhibitors with improved pharmacokinetics and pharmacodynamics properties [45] and could possibly overcome any potential drawbacks limiting the efficacy of existing inhibitors such as cross-resistance [46].

2.0 COMPUTATIONAL METHODOLOGY

2.1 System preparation

2.1.1 Homology modelling model validation of SARS-CoV-2 RNA-dependent RNA polymerase

The unavailability of an experimentally resolved X-ray crystal structure of SARS-CoV-2 RNA dependent RNA polymerase at the time of the study prompted a construction of a putative model. We, therefore, retrieved the amino acid sequence of SARS-CoV-2 RNA-dependent RNA polymerase from the NCBI database with reference sequence YP_009725307.1. Using the online

server, SWISS model, [47,48], we performed structural modelling of the 3D structure of SARS-CoV-2 RNA-dependent RNA polymerase. The homology model of SARS-CoV-2 RNA-dependent RNA polymerase constructed in this study was based on the experimentally resolved X-ray crystal structure of chain A of SARS-CoV RNA-dependent RNA polymerase structure (PDB ID: 6NUR) [49]. The chain A of SARS-CoV RNA-dependent RNA polymerase possessed a sequence similarity of 96.35%, with SARS-CoV-2 RNA dependent RNA polymerase and hence was the most suitable template to be used in the construction of our model. To ensure the reliability of assumptions derived from the structural investigations of modelled structure, we performed a thorough model validation using the online platforms RAMPAGE [50] ProSA [51] and VERIFY-3D [52]. We visually analysed the structural and similarities of the constructed model and the template employed by superimposing the two structures on UCSF Chimera [53]. This also allowed for the identification of the putative Remdesivir binding site as shown in figure 5.2.

RMSD: ca									
model.pdb, chain A	117	QRLTKYTMAD	LVYALRHFDE	GNCDTLKEIL	VTYNCCDDDY	FNKKDWYDFV			
6nur.1.A.pdb.gz, chain A	119	QRLTKYTMAD	LVYALRHFDE	GNCDTLKEIL	VTYNCCDDDY	FNKKDWYDFV			
RMSD: ca									
model.pdb, chain A	167	ENPDILRVYA	NLGERVRQAL	LKTVQFCDAM	RNAGIVGVLT	LDNQDLNGNW			
6nur.1.A.pdb.gz, chain A	169	ENPDILRVYA	NLGERVRQSL	LKTVQFCDAM	RDAGIVGVLT	LDNQDLNGNW			
RMSD: ca									
model.pdb, chain A	217	YDFGDFIQTT	PGSGVPVVDV	YYSLLMPILT	LTRALTAESH	VDTDLTKPYI			
6nur.1.A.pdb.gz, chain A	219	YDFGDFVQVA	PGCGVPIVDS	YYSLLMPILT	LTRALAAESH	MADDLAKPLI			
RMSD: ca									
model.pdb, chain A	267	KWDLKLYDFT	EERLKLFDRI	FKYWDQTYHP	NCVNCLDDRC	ILHCANFNVL			
6nur.1.A.pdb.gz, chain A	269	KWDLKLYDFT	EERLCLFDRI	FKYWDQTYHP	NCINCLDDRC	ILHCANFNVL			
RMSD: ca									
model.pdb, chain A	317	FSTVFPPTSF	GPLVRKIFVD	GVPFVSTGY	HFRELGVVHN	QDVNLHSSRL			
6nur.1.A.pdb.gz, chain A	319	FSTVFPPTSF	GPLVRKIFVD	GVPFVSTGY	HFRELGVVHN	QDVNLHSSRL			
RMSD: ca									
model.pdb, chain A	367	SFKELLVYAA	DPAMHAASGN	LLLDKRTTCF	SVAALTNNVA	FQTVKPGNFM			
6nur.1.A.pdb.gz, chain A	369	SFKELLVYAA	DPAMHAASGN	LLLDKRTTCF	SVAALTNNVA	FQTVKPGNFM			
RMSD: ca									
model.pdb, chain A	417	KDFYDFAVSK	GFFKEGSSVE	LKHFFFAQDG	NAAISDYDYY	RYNLPTMCDI			
6nur.1.A.pdb.gz, chain A	419	KDFYDFAVSK	GFFKEGSSVE	LKHFFFAQDG	NAAISDYDYY	RYNLPTMCDI			
RMSD: ca									
model.pdb, chain A	467	RQLLFVVEVV	DKYFDCYDGG	CINANQVIVN	NLDKSAGFPF	NKWGKARLYY			
6nur.1.A.pdb.gz, chain A	469	RQLLFVVEVV	DKYFDCYDGG	CINANQVIVN	NLDKSAGFPF	NKWGKARLYY			
RMSD: ca									
model.pdb, chain A	517	DSMSYEDQDA	LFAYTKRNV	PTITQMNLY	AISAKNRART	VAGVVICSTM			
6nur.1.A.pdb.gz, chain A	519	DSMSYEDQDA	LFAYTKRNV	PTITQMNLY	AISAKNRART	VAGVVICSTM			
RMSD: ca									
model.pdb, chain A	567	TNRQFHQKLL	KSIAATRGT	VVIGTSKIFYG	GWHNMLKTVY	SDVENPHLMG			
6nur.1.A.pdb.gz, chain A	569	TNRQFHQKLL	KSIAATRGT	VVIGTSKIFYG	GWHNMLKTVY	SDVETPHLMG			
RMSD: ca									
model.pdb, chain A	617	WDYPKCDRAM	PNMLRIMASL	VLARKHTTCC	SLSHRFYRLA	NECAQVLSSE			
6nur.1.A.pdb.gz, chain A	619	WDYPKCDRAM	PNMLRIMASL	VLARKHNTCC	NLSHRFYRLA	NECAQVLSSE			
RMSD: ca									
model.pdb, chain A	667	VMCGGSLYVK	PGGTSSGDAT	TAYANSVFNI	CQAVTANVNA	LLSTDGNKIA			
6nur.1.A.pdb.gz, chain A	669	VMCGGSLYVK	PGGTSSGDAT	TAYANSVFNI	CQAVTANVNA	LLSTDGNKIA			
RMSD: ca									
model.pdb, chain A	717	DKYVRNLQHR	LYECLYRNRD	VDTDFVNEFY	AYLRKHFSMM	ILSDDAVVCF			
6nur.1.A.pdb.gz, chain A	719	DKYVRNLQHR	LYECLYRNRD	VDHEFVDEFY	AYLRKHFSMM	ILSDDAVVCF			
RMSD: ca									
model.pdb, chain A	767	NSTYASQGLV	ASIKNFKSVL	YYQNNVFMSE	AKCWTEETDLT	KGPHEFCSQH			
6nur.1.A.pdb.gz, chain A	769	NSNYAAQGLV	ASIKNFKAVL	YYQNNVFMSE	AKCWTEETDLT	KGPHEFCSQH			
RMSD: ca									
model.pdb, chain A	817	TMLVKQGDDY	VYLPYPDPSR	ILGAGCFVDD	IVKTDGTLMI	ERFVSLAIDA			
6nur.1.A.pdb.gz, chain A	819	TMLVKQGDDY	VYLPYPDPSR	ILGAGCFVDD	IVKTDGTLMI	ERFVSLAIDA			
RMSD: ca									
model.pdb, chain A	867	YPLTKHPNQE	YADVFLYLYQ	YIRKLHDEL	GHMLDMYSVM	LTNDNTRSRYW			
6nur.1.A.pdb.gz, chain A	869	YPLTKHPNQE	YADVFLYLYQ	YIRKLHDELM	LTNDNTRSRYW			
RMSD: ca									
model.pdb, chain A	917	EPE							
6nur.1.A.pdb.gz, chain A	909	EPE							

Active site



Figure 5.2: Comparative sequence alignment of the constructed homology model of SARS-CoV-2 RNA-dependent RNA polymerase and the accompanying template employed. Insert highlights the active site residue, VAL557 as previously reported

2.1.2 Remdesivir acquisition and preparation

In preparing for molecular docking, Remdesivir was modelled using the MarvinSketch software (39). The modelled structures were then taken through energy minimization using Avogadro 1.2.0 software (40). Avogadro is integrated with UFF forcefield, and this force field was used to optimize the molecular geometries of the compounds and a steepest descent algorithm for structural minimization. The pdbqt formats of all the compounds were then generated using AutoDock Vina (41). In preparation for molecular docking, hydrogen ions removed, and Gasteiger charges added using the Graphical User Interface (GUI) of UCSF Chimera [54]. The structure was subsequently saved in mol2 format.

2.2 Molecular docking and active site identification

According to recent reports by Lung *et al.*, the active site of SARS-CoV RNA-dependent RNA polymerase encompasses VAL557 and its surrounding residues [49,55]. As such, the active site was mapped by considering residues that were within a 5Å radius of VAL557. A grid box that encompassed a 5Å radius around VAL557 was used to map out the corresponding coordinates with the AutoDock Vina [56] incorporated in UCSF Chimera [53]. The defined coordinates of the Remdesivir binding site region on the modelled structure were obtained from the grid box coordinates $x = 144.452$, $y = 148.961$, $z = 163.495$ for the centre and coordinates $x=10.9051$, $y=18.7302$, $z= 15.2392$ for the size of the grid box. In preparing for molecular docking, the pdbqt formats of Remdesivir and the SARS-CoV-2 RNA dependent RNA polymerase model were generated using AutoDock Vina [56]. Molecular docking was subsequently performed using

AutoDock Vina. In preparing for molecular docking, the SARS-CoV-2 RNA dependent RNA polymerase structure was prepared using the Graphical User Interface (GUI) of UCSF Chimera [54], which involved the addition of hydrogen atoms. The output of the docking was viewed on UCSF Chimera using the integrated ViewDock module after docked complexes were saved for further analysis. Residues that interacted with Remdesivir in the best binding pose of the docking results are shown in figure 5.3. The docked complexes were then prepared for MD simulation following in-house pre-MD preparation protocols [57,58]. Two systems were then set up for the MD simulation, an apo conformation of SARS-CoV-2 RNA dependent RNA polymerase and the Remdesivir bound conformation.

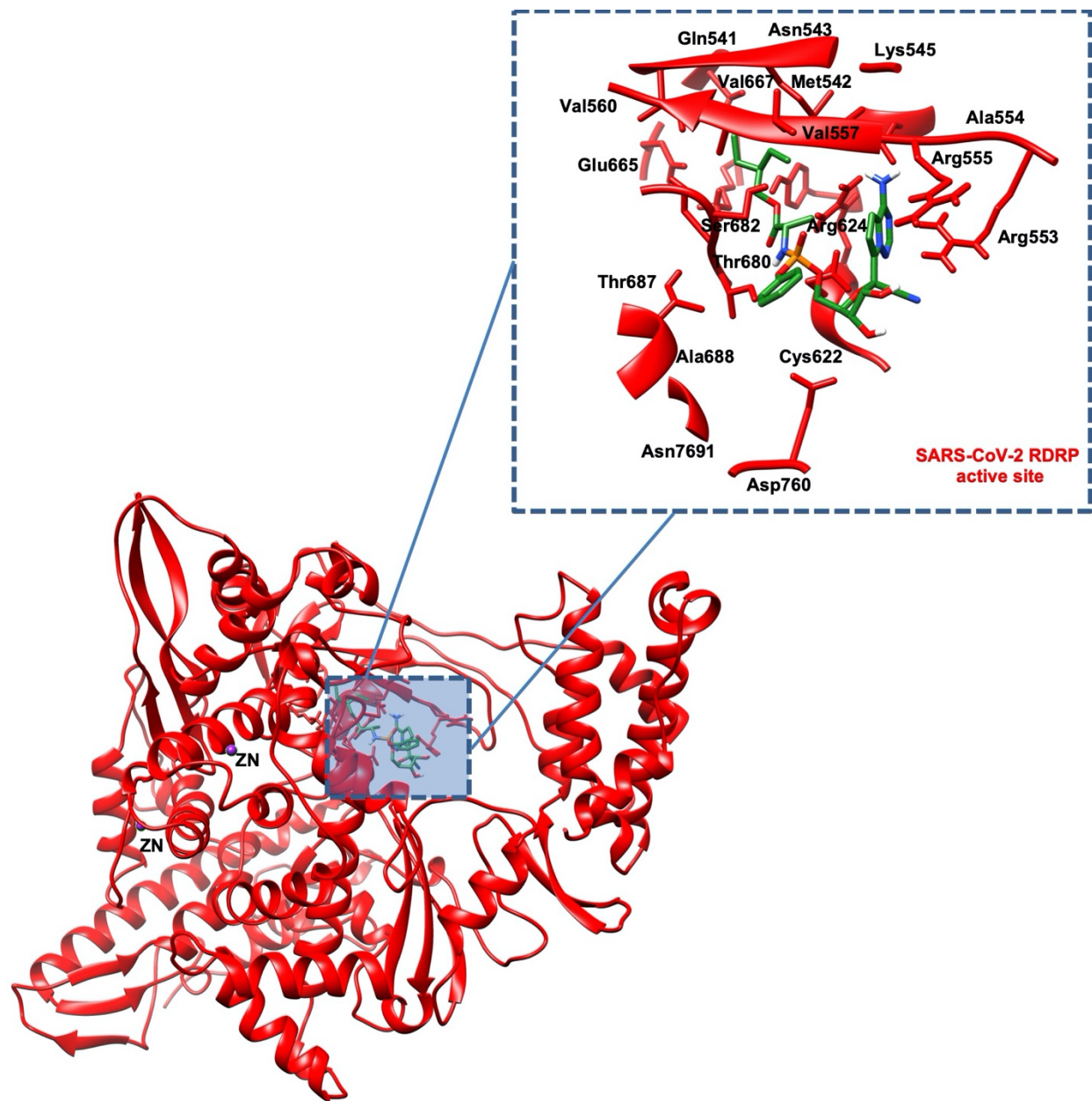


Figure 5.3: 3D representation of the SARS-CoV-2 RNA-dependent RNA polymerase model employed in this study, which was pruned to minimize computational cost. Also present in the structure are Zinc ions (purple). Interacting active site residues are also shown in **black**

2.3 Molecular dynamics (MD) simulations

To reveal the conformational changes that occurred on SARS-CoV-2 RNA-dependent RNA polymerase upon Remdesivir binding, we performed an atomic-scale MD simulation using the AMBER 18 GPU with an integrated PMEMD module [59,60]. This method is reliable and has been employed extensively in several previous studies [57,61–64]. Prior to molecular dynamics (MD) simulations, additional co-crystallized molecules such as crystal water, which were not applicable to this study, were removed from the enzyme structure using the graphical user interface of UCSF Chimera. In preparing Remdesivir for the MD simulation, the ANTECHAMBER module was employed for its parameterization, in which atomic partial charges (AM1BCC) gaff, using the bcc charge scheme were created [65]. On the other hand, the studied SARS-CoV-2 RNA-dependent RNA polymerase model was also parameterized using the FF14SB AMBER force field [66]. Protonation of histidine residues was performed using the pdb4amber script at a constant pH (cpH) to ensure the compatibility of the prepared enzyme model with the LEAP module. Subsequently, hydrogen ions were added, followed by a neutralization of the entire prepared system using the LEAP module. Neutralization was carried out by the addition of either of the counter ions, Na⁺ or Cl⁻. Afterward, topology and coordinate files for Remdesivir, SARS-CoV-2 RNA-dependent RNA polymerase, and their resultant complex were generated. Using a TIP3P orthorhombic box size of 12Å, water molecules were then added to explicitly solvate the entire system [67]. The two prepared systems (apo and complex) were subjected to an initial 2000 minimization steps at a restraint potential of 500kcal/mol. This was subsequently followed by a 1000 steps minimization of steepest descent with no restraint. The systems were gradually heated from 0K to 300K for 50ps. After heating, a 500ps equilibration was performed at a constant pressure of 1bar. The pressure was maintained constant using Berendsen barostat [68]. Using the

SHAKE algorithm [69] to constrict all atomic hydrogen bonds, a 100ns MD simulation was performed on all systems. The MD production run was performed using a 1fs time step. Coordinates for generated trajectories were saved at 1ps interval and were further analysed using the PTRJ and CPPTRAJ modules of AMBER 18 [70]. Graphical plots for analysis of the generated trajectories created with the Microcal Origin analytical software [71].

2.4 Per-residue energy decomposition calculations

The quantification of the energy contributions of individual binding site residues were performed using per-residue decomposition analyses technique. The Per-residue decomposition analysis is incorporated in the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) [72–75] method. This allows for the identification of crucial residues required for the efficient binding and stabilization of Remdesivir to the SARS-CoV-2 RNA-dependent RNA polymerase.

2.5 Pharmacophore model creation.

Following the MD simulation of Remdesivir complexed with RNA-dependent RNA polymerase of SARS-CoV-2, the online platform, ZincPharmer [76], was used to construct a pharmacophore model. The pharmacophore model was constructed using the pharmacophoric moieties that exhibited prominent interaction with the highest energy contributing amino acids, as identified in the per-residue energy decomposition analysis.

3.0 RESULTS AND DISCUSSION

3.1 Validation of modelled 3D structure of SARS-CoV-2 RNA-dependent polymerase

The homology model of the SARS-CoV-2 RNA-dependent RNA polymerase investigated in this study was built using the X-ray crystal structure of SARS-CoV RNA-dependent RNA polymerase [49] due to its highest sequence identity of 96.35%. To guarantee the reliability of the atomistic

predictions postulated in this report based on the modelled structure, we validated our model using the online server, RAMPAGE. From RAMPAGE, it was revealed that 98.1% (786 amino acids) of the amino acid residues that constitute our model were within the favoured region, as shown in the Ramachandran plot on figure 5.4A. Also, only 15 amino acid residues (1.9%) of our model were in the allowed region, while no residue was found in the outlier region. The high percentage of residues in the favoured and allowed regions suggested a reliable accuracy of the dihedral and backbones of our constructed homology model. To further ensure the reliability of our postulations from the model, the structure was also validated the PROSA online web server with the resultant Z-score plots presented in figure 5.4B. Validation of the PROSA web server allows for an evaluation of a homology model with regards to its acceptability within experimental X-ray and NMR limits. As shown in the Z-core plot, our modelled SARS-CoV-2 RNA-dependent RNA polymerase structured exhibited a Z-score of -13.23 , consistent with a relatively good quality homology model.

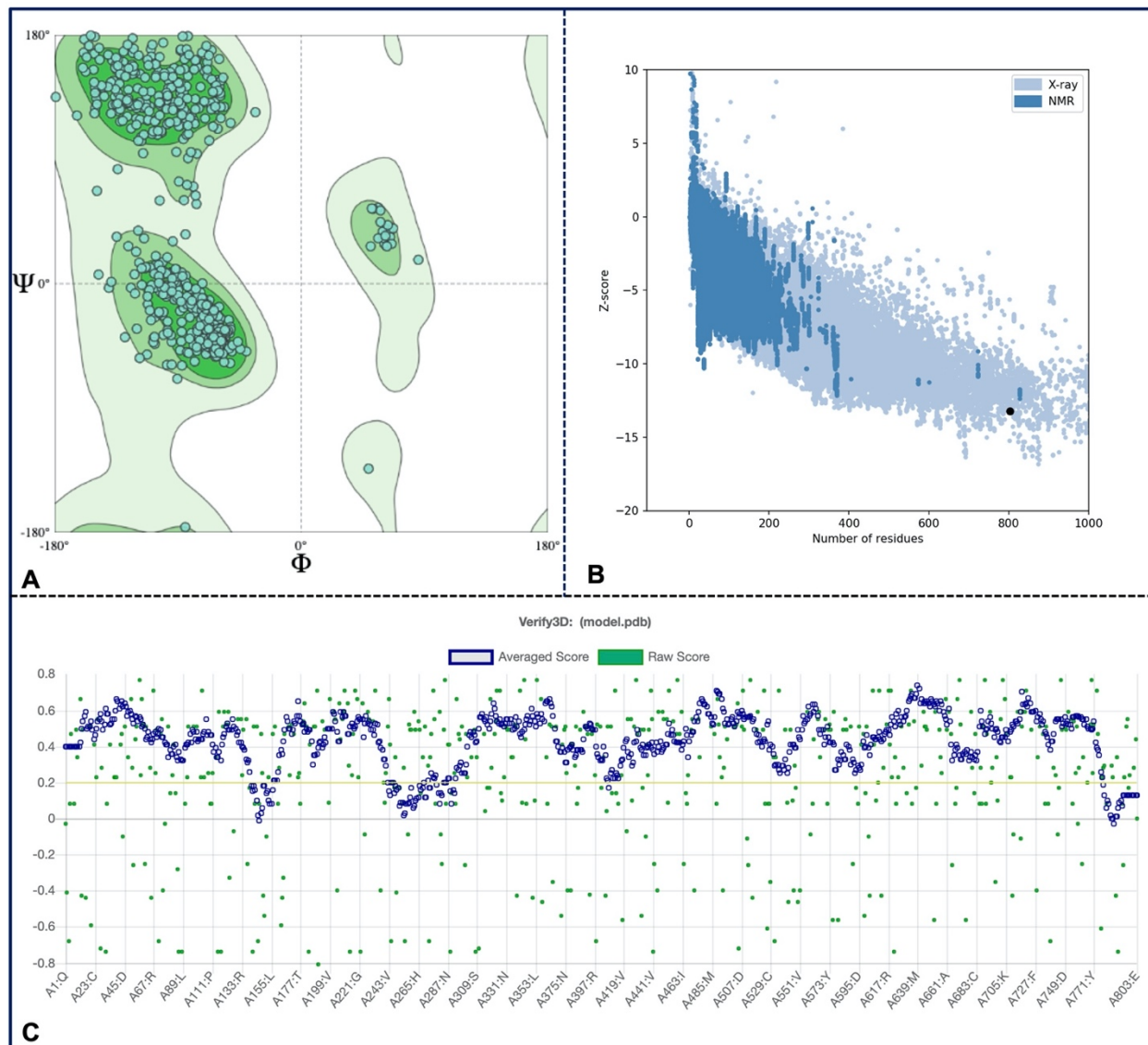


Figure 5.4: **A)** Ramachandran plots for the built homology model SARS-CoV-2 RNA-dependent RNA polymerase created on the RAMPAGE online webserver. **B)** Overall model quality of the homology model of SARS-CoV-2 RNA-dependent RNA polymerase by Z-score of ProSA. **C)** Overall model quality of the homology model of SARS-CoV-2 RNA-dependent RNA polymerase by Z-score of ProSA

Finally, the quality of our model was also accessed using the online server, Verify_3D, a platform that employs empirical and energetic techniques to ascertain model quality. Verify_3D also assesses individual residue quality, as presented in figure 11.4C. As shown, our modelled SARS-

CoV-2 RNA-dependent RNA polymerase recorded a Verify-3D score of 89.91%, with an average 3D-1D score ≥ 0.2 score.

3.2 Validation of model with experimentally resolved X-ray crystal structure of SARS-CoV-2 virus RNA-dependent RNA polymerase

Upon completion of this study, the X-ray crystal structure of SARS-CoV-2 virus RNA-dependent RNA polymerase bound with Remdesivir was released on the 8th of April 2020. However, a structural comparison and subsequent sequence alignment of our homology model and the experimentally resolved structure revealed a 90.78% sequence identity between both structures, as presented in supplementary figure S1. Particularly, the active site region of both structures exhibited a 100% sequence identity. The only notable difference is the incorporation of two catalytic magnesium ions in the experimental structure. Therefore, in light of the results, authors do not see any significant deviation from our results on the modelled structure. To further validate the reliability of the structural analysis revealed in this study from the modelled complex, we estimated the binding free of Remdesivir in both the modelled structure and the recently reported experimental complex by Yin *et al.*[77]. As shown in table 5.1, Remdesivir exhibited a binding free energy of -13.28kcal/mol and -10.16kcal/mol in the experimental structure and the homology model respectively. With a minimal binding free energy difference of about -3kcal/mol between both structures as observed, it further suggests that there would be minimal difference in the binding site interaction dynamics of Remdesivir within both structures as well. Therefore, the structural elucidations from the homology model that are provided in this report would be similar to what would occur in the experimentally resolved structure, hence reliable.

Table 5.1:Comparative MM/GBSA-based binding free energy profile of compound Remdesivir

Complexes	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
Modelled complex	-13.95±0.05	-69.79±0.53	-30.26±0.51	20.10±0.43	-10.16±0.11
Experimental complexed	-28.01± 0.24	-31.74 ±0.56	-59.77±0.13	46.49±0.47	-13.28±0.25

ΔE_{ele} = electrostatic energy; ΔE_{vdW} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG = gas phase free energy.

3.3 Insights from binding mode exploration of Remdesivir within SARS-CoV-2 RNA-dependent RNA polymerase active site toward COVID-19 therapy

Molecular docking was used to predict the best binding pose of Remdesivir within the active site region of modelled of SARS-CoV-2 RNA-dependent RNA polymerase. As presented in figure 5.5, several binding poses of Remdesivir were postulated with the corresponding docking score provided. The best binding pose, which is consistent with the most negative docking, was identified with a docking score of 5.9kcal/mol. Complexes generated from molecular docking, however, only mimic the static “key and lock” mechanism of ligand-receptor binding [78] and do not provide insights on the accurate dynamics of ligand-target interactions since both are flexible in reality. As such, assessment of the conformational changes of SARS-CoV-2 RNA-dependent RNA polymerase after molecular docking would not be truly representative of what would happen in reality upon Remdesivir binding. We, therefore, proceeded with molecular dynamics simulation of the best docked complex.

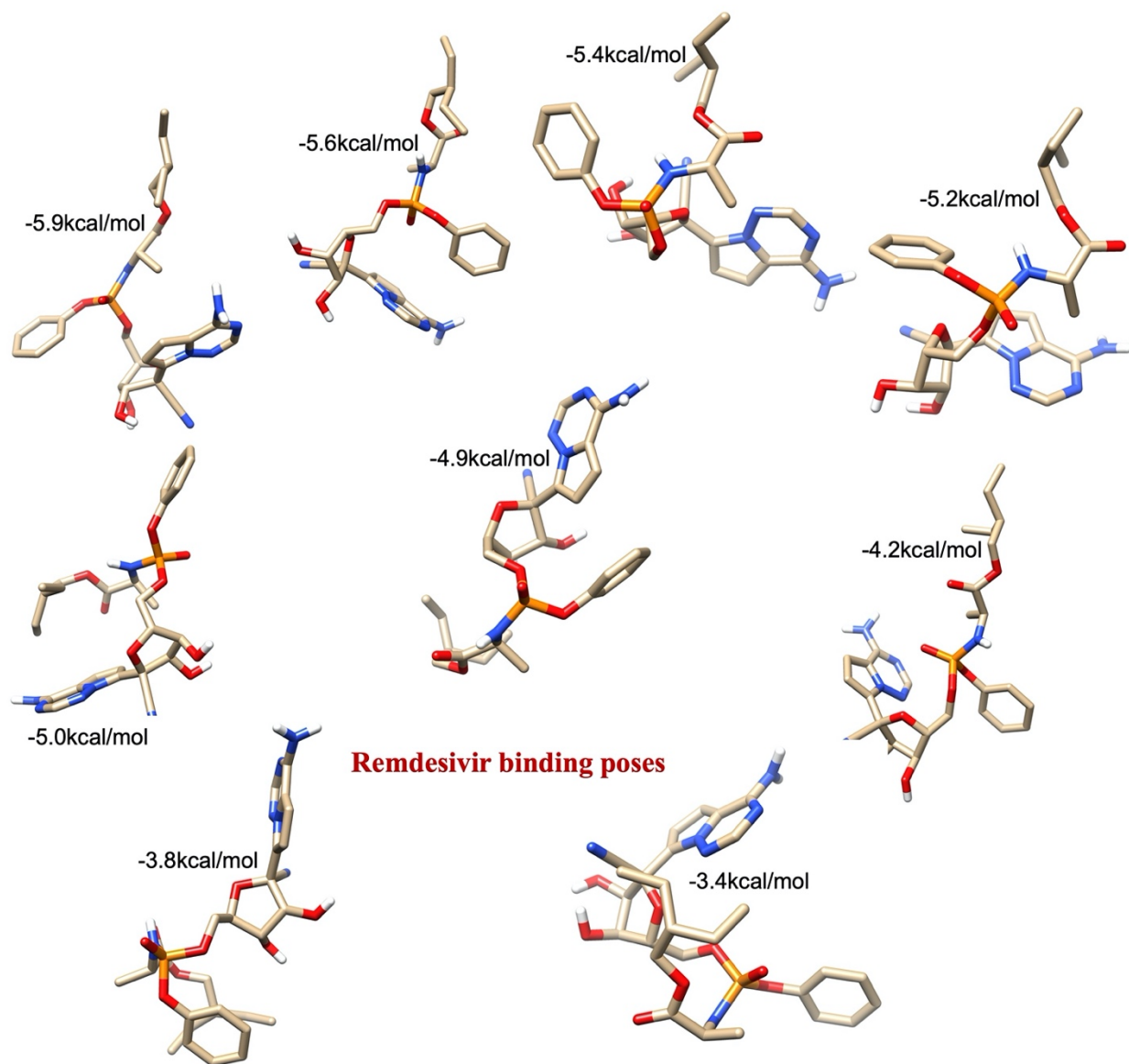


Figure 5.5: Binding poses of Remdesivir in the active sites of SARS-CoV-2 RNA-dependent RNA polymerase and their accompanying docking score. Best binding pose of Remdesivir has a docking score of -5.9kcal/mol

3.4 Structural And Conformational Dynamics Of SARS-Cov-2 RNA-Dependent RNA Polymerase Upon Remdesivir Binding

Conformational changes of enzymes are the hallmarks of their dynamics that have a correlation with the overall functions of these enzymes [34]. In many cases, such conformation changes may be induced by the binding of ligands or other biological molecules that are able to modulate the function of the enzyme [35]. In this case, the binding of Remdesivir would have induced significant conformations. We, therefore, proceeded to investigate the structural and conformational dynamics of SARS-CoV-2 RNA-dependent RNA polymerase as relates to Remdesivir binding using MD simulation. This reliable computational technique allows the study of time-resolved properties of protein or enzyme structures in atomic details [36]. Post-MD simulation metrics such as C- α root mean square fluctuation (RMSF), C- α radius of gyration (RoG), and Solvent Accessible Surface Area (SASA) was thus employed assess the conformational dynamics of SARS-CoV-2 RNA-dependent RNA polymerase.

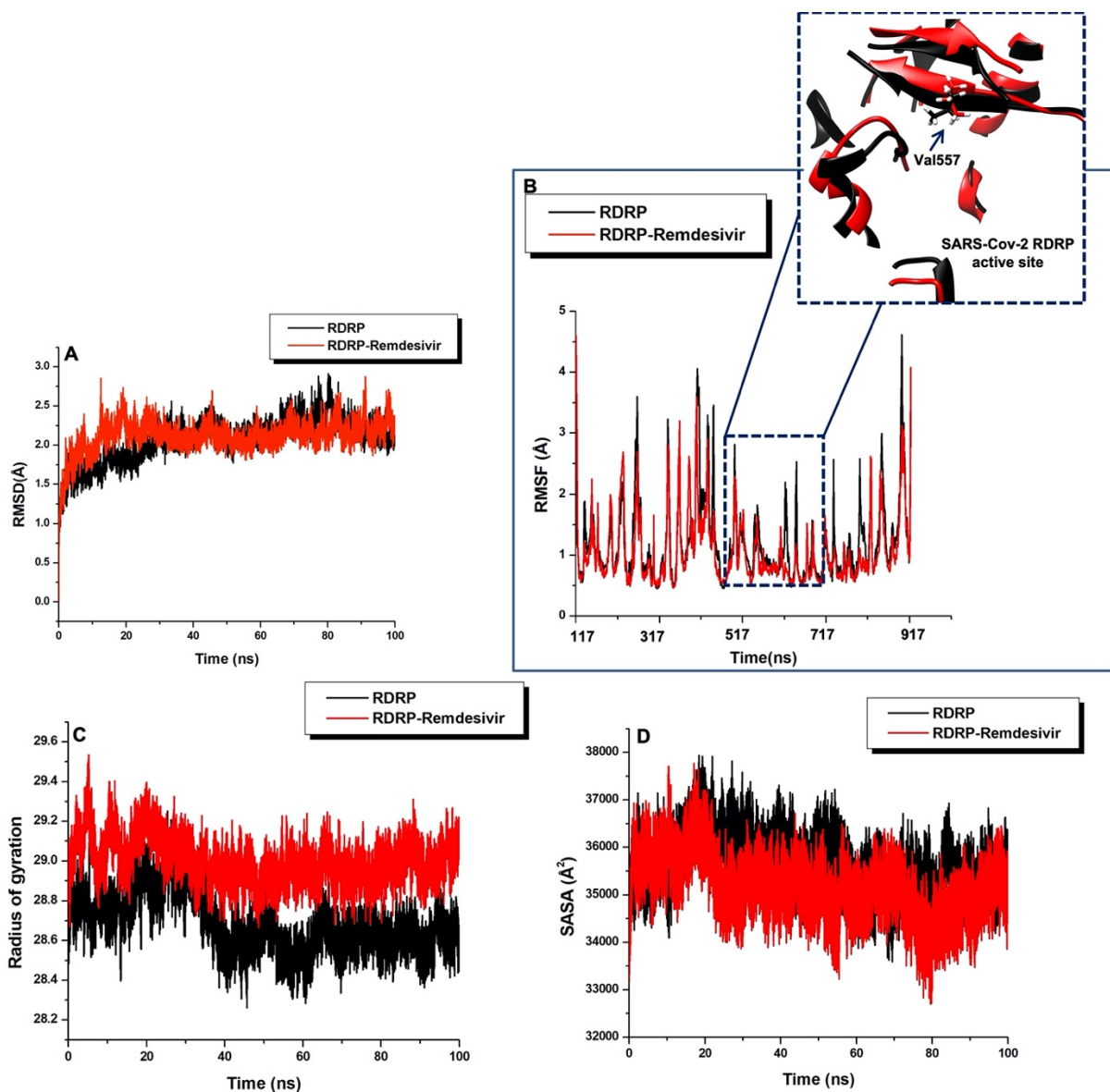


Figure 5.6: Comparative conformational analyses of the unbound (**black**) and Remdesivir-bound (**red**) SARS-CoV-2 RNA-dependent RNA polymerase model. **[A]** C- α RMSD plot showing the degree of atomistic deviations and instability. **[B]** C- α RMSF plot showing individual residue flexibility. **Blue highlight** show the SARS-CoV-2 RNA-dependent RNA polymerase active site region and corresponding 3D representation is shown in the **blue inset**. Comparative radius of gyration plots of the unbound (**black**) and Remdesivir-bound (**red**) SARS-CoV-2 RNA-dependent RNA polymerase mode. **[C]** C- α ROG plot showing the degree of compactness of SARS-CoV-2 RNA-dependent RNA polymerase. **[D]** Comparative SASA plots of the unbound (**black**) Remdesivir-bound (**red**) SARS-CoV-2 RNA-dependent RNA polymerase model

C- α RMSF can be employed to describe per-residue fluctuations across the enzyme's secondary structure, while C- α RoG can be used to predict structural compactness, perturbations, and atomistic mobility [79–83]. C- α RMSF calculations for the unbound and Remdesivir bound SARS-CoV-2 RNA-dependent RNA polymerase are presented as line plots in figure 5.6B, which revealed that the binding of Remdesivir generally induced minimal fluctuations of individual residues in the overall structure when compared to unbound conformation while showed relatively higher residual fluctuations. On the average, unbound SARS-CoV-2 RNA-dependent RNA polymerase had RMSF values of 1.20Å, while Remdesivir-bound conformation had an average RMSF of 1.06Å. Presumably, the minimal perturbation in the structure SARS-CoV-2 RNA-dependent RNA polymerase upon Remdesivir binding could highlight a possible mechanism by which Remdesivir exhibits its therapeutic modulatory activity against SARS-CoV-2 RNA-dependent RNA polymerase. Specific peak fluctuations regions of interest occurred between residues “517-717”, which encompassed most of the active site residues. Thus, although there was minimal flexibility of residues in the entire structure of SARS-CoV-2 RNA-dependent RNA polymerase upon Remdesivir binding, there was an apparent disruption of the active site residues, as shown. This could be attributed to the direct strong molecular interactions elicited upon Remdesivir binding. These presumptions were further verified by calculating the C- α RMSD, which is able to depict the stability of SARS-CoV-2 RNA-dependent RNA polymerase across the MD simulation. As shown in figure 5.6A, C- α atomistic deviations of both simulated models were almost the same, with a difference of 0.7Å. This suggested that the binding of Remdesivir barely interfered with deviations of C- α atoms of SARS-CoV-2 RNA-dependent RNA polymerase, corroborating with the minimal residue fluctuations observed. The mean RMSD values estimated for the Remdesivir bound and the unbound SARS-CoV-2 RNA-dependent RNA polymerase were

2.14Å and 2.07Å, respectively. Furthermore, the conformational dynamics of SARS-CoV-2 RNA-dependent RNA polymerase were assessed from the MD simulation by estimating the C- α RoG, which provides insights on enzyme structure compactness. As shown in figure 5.6C, the Remdesivir-bound SARS-CoV-2 RNA-dependent RNA polymerase exhibited an average RoG of 29.00Å, while the unbound conformation exhibited an average RoG 28.68Å. This suggested that Remdesivir binding induced a less compact structure consistent with structural unfolding. Nonetheless, the difference in comparison with the unbound is 0.32Å, further suggesting the structural unfolding with regards to Remdesivir binding is minimal. Estimations of the SASA provided more insights into the mechanistic perturbations of SARS-CoV-2 RNA-dependent RNA polymerase from the hydrophobic region (buried) to the hydrophilic phase (surface-exposed) according to previous studies [80,84–86]. As shown in the SASA plot on figure 5.6D, residues of the Remdesivir-bound SARS-CoV-2 RNA-dependent RNA polymerase appeared to be less surface-exposed when compared to the unbound conformation with an average SASA value of 35164.07Å². The unbound SARS-CoV-2 RNA-dependent RNA polymerase exhibited a higher average SASA of 35641.23Å². This could then imply that Remdesivir binding restricts the systematic transition of SARS-CoV-2 RNA-dependent RNA polymerase the hydrophobic region (buried) to the hydrophilic phase (surface-exposed)

3.5 Extensive Remdesivir-active site modulation favour SARS-CoV-2 RNA-dependent RNA polymerase inhibition

To understand the basis for the observed conformational and structural changes that occurred on SARS-CoV-2 RNA-dependent RNA polymerase upon Remdesivir binding, an investigation of the enzyme's fundamental interaction profile, particularly at the active site, is essential. Based on

previous studies, it has been established that the active site of SARS-CoV, a homolog with 96.35% sequence similarity, is located at VAL557 [49,55]. Sequence alignment of both SARS-CoV and SARS-CoV-2 RNA-dependent RNA polymerase, as presented in figure 5.2, showed VAL557 was conserved in both enzymes. This conserved pattern of VAL557 on both structures suggests it could be the active site of SARS-CoV-2 RNA-dependent RNA polymerase, whose therapeutic modulation, in addition to other binding pocket residues, could interfere with the overall activity of RNA-dependent polymerase coronaviruses. Therefore to ascertain the role of VAL557 and accompanying binding pocket residues in Remdesivir binding, we performed a molecular visualization of the interactions it elicited with Remdesivir upon binding, over a 100ns MD simulation using representative snapshots, 20 ns, 40 ns, 60 ns, 80 ns, 100 ns. This allowed for visualization of specific interactions that the active site residues engaged in with Remdesivir, which have also influenced its stability and binding affinity. The nature of these interactions could further establish or discredit the role of VAL557 in therapeutic inhibition of coronavirus RNA-dependent RNA polymerase by Redemsivir. Also, the exploration of the molecular interactions of other residues that were within 5Å radius from VAL557 could highlight other crucial residues that influence the therapeutic efficiency of Remdesivir against SARS-CoV-2 RNA-dependent RNA polymerase as experimentally predicted.

occurred at 20 ns were maintained. Again VAL557 retains in carbon-hydrogen bond interaction with Remdesivir, while ARG624 is also shown to engage in a π -anion interaction. At 60 ns, as shown in figure 5.7C, even as many interactions were again maintained, the number of hydrophobic interactions increased relative to interactions at 40 ns. However, the salt bridges elicited between ASP623 and Remdesivir were lost, with a conventional hydrogen bond interaction forming instead.

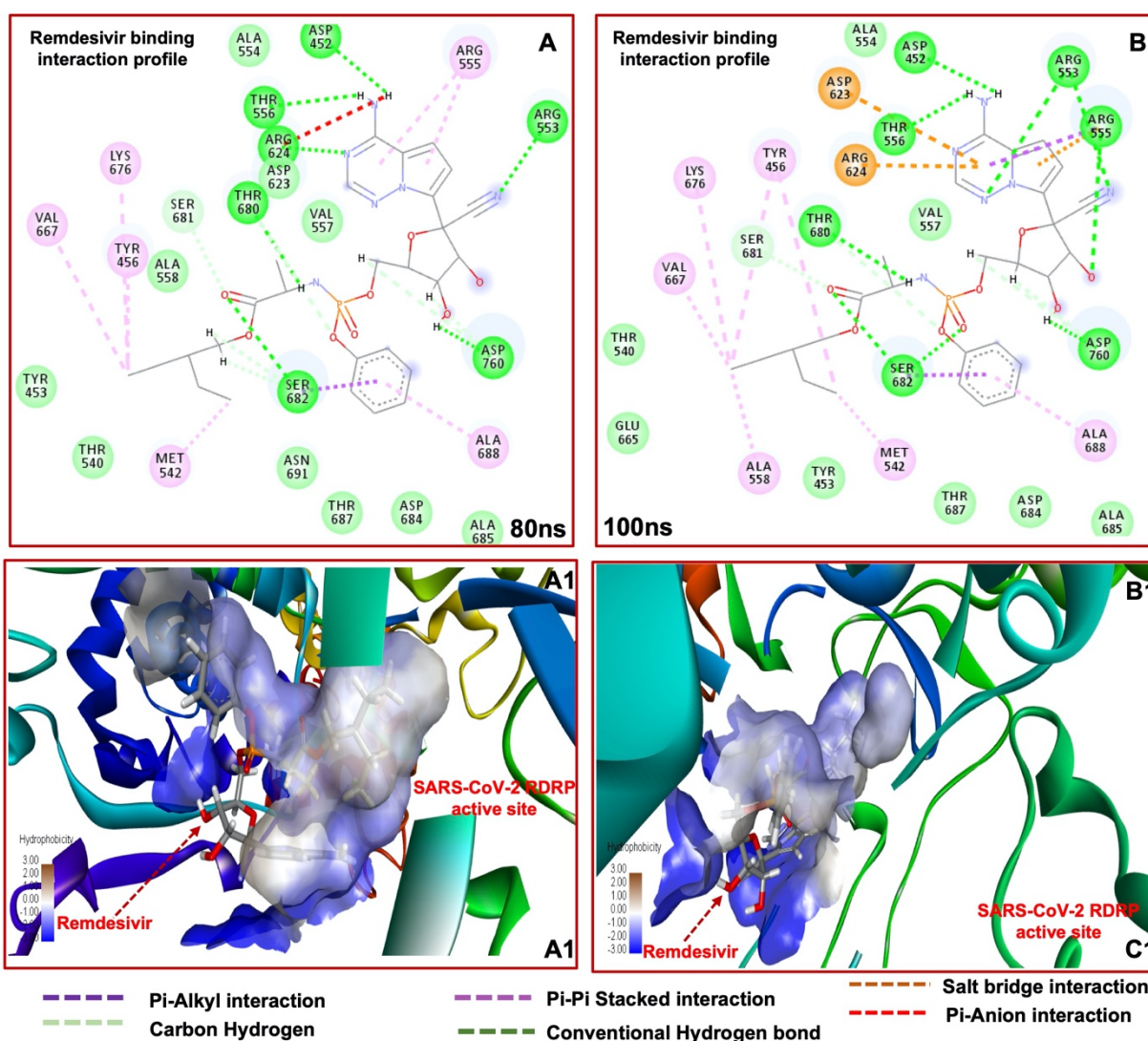


Figure 5.8: Molecular visualization of Remdesivir at the active sites (hydrophobic pockets) of SARS-CoV-2 RNA-dependent RNA polymerase. Inter-molecular interactions between

Remdesivir and active site residues in SARS-CoV-2 RNA-dependent RNA polymerase at 80 ns and 100ns are shown in **a** and **b** respectively. **a1** and **b1** shows a 3D representation of the hydrophobicity of the Remdesivir bound active site.

Towards the end of the simulation at 80 ns, as shown in figure 5.8A, the number of hydrophobic interactions with Remdesivir increased with many interactions observed from the beginning of the simulation again maintained. In addition to the conventional hydrogen bond interactions, other notable interactions included π - π stacked interactions with residues such as LYS676, VAL667, MET542, ALA688, ARG555, and TYR456. At the end of the MD simulation at 100 ns, as shown in figure 5.8B, the extensive pool of strong molecular interactions observed at the beginning of the simulations was sustained. Also, the two salt bridge interactions occurring between Remdesivir and ASP623, as well as ARG624, were again observed. From the observed interaction dynamics of Remdesivir in the SARS-CoV-2 RNA dependent RNA polymerase active site, it could be deduced specific residues interacted consistently, steadily, and intermittently across the 100ns MD simulation time. The consistency in engaging in strong interactions with Remdesivir implies that these interactions are essential to its therapeutic function against SARS-CoV-2 RNA dependent RNA polymerase. Notable interactions elicited included strong ionic and hydrogen bond formation and salt bridges. Thus, these interactions could have favoured the high-affinity binding and stability of Remdesivir in the active site of SARS-CoV-2 RNA dependent RNA polymerase and could also suggest the possible structural mechanism of its therapeutic efficacy. Also, the observed increase in interacting residues as the simulation proceeded further suggests that Remdesivir assumed various binding orientations that allowed it to interact with more hydrophobic residues towards achieving stability and *via* favourable molecular interactions as observed. Altogether, the observed interaction dynamics of Remdesivir could relatively provide necessary details and insights into the molecular basis and mechanisms of its high-affinity binding and therapeutic

modulation of SARS-CoV-2 RNA dependent RNA polymerase towards as considerations as therapeutic for COVID-19.

3.6 Quantitative insights into the contributions of active site residues towards Remdesivir binding

The critical roles of the putative active site residues of SARS-CoV-2 RNA dependent RNA polymerase were further revealed via per-residue decomposition analyses, which provided energy contributions of individual residues that interacted with Remdesivir at over the 100ns MD simulation duration. Overall, energies were decomposed into van der Waals and electrostatic energy, polar solvation energy, non-polar solvation energy, including their total energy contributions. As shown in figure 5.9 and table 5.2, many of the active site residues contributed significantly to the total binding of Remdesivir.

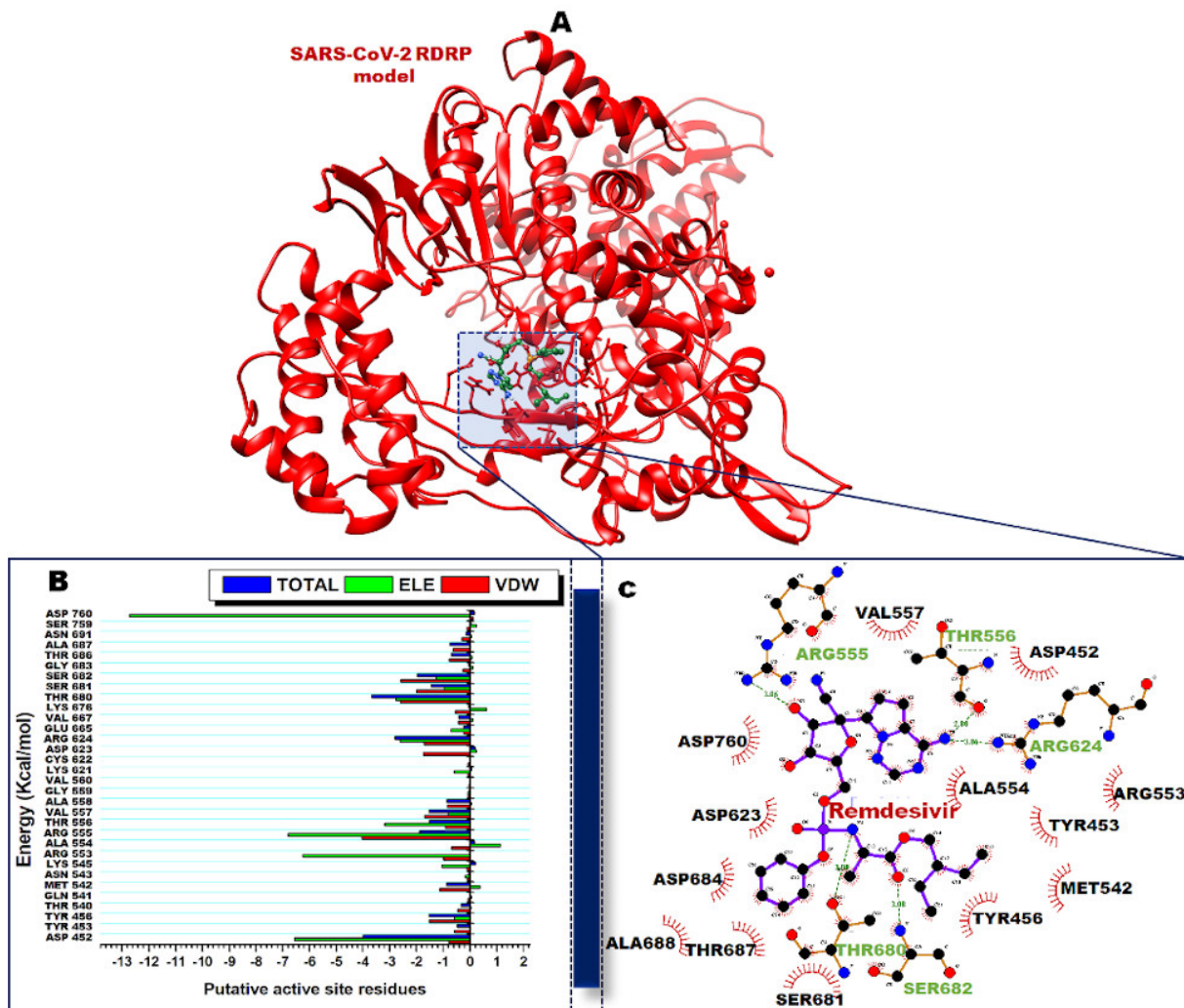


Figure 5.9. A) 3D homology model of model of the SARS-CoV-2 RNA dependent RNA polymerase complexed with Remdesivir. B) Per-residue energy contributions plot of the interacting residues at the SARS-CoV-2 RNA dependent RNA polymerase active site upon Remdesivir binding. C) The residue ligand interaction network illustrates stabilizing hydrophobic interactions pocketing Remdesivir at the active site from an averaged structure.

Table 5.2: Tabular presentation of the estimated energy contribution of individual active residues

Residue	Van der Waals (kcal/mol)	Electrostatic (kcal/mol)	Polar solvation (kcal/mol)	Non-Polar solvation (kcal/mol)	Total (kcal/mol)
ASP 452	-0.78±0.58	-6.56±1.19	3.37±1.21	-0.01±0.01	-3.99±1.06
TYR 453	-0.57±0.19	0.01±0.04	0.09± 0.05	-0.01±0.01	-0.47±0.17
TYR 456	-1.51±0.39	-0.57±0.26	0.72±0.18	-0.16 ±0.08	-1.53±0.47
THR 540	-0.46±0.11	0.03±0.04	0.12± 0.06	-0.02±0.01	-0.34±0.11
GLN 541	-0.09±0.01	0.05±0.04	0.05±0.05	0.00±0.00	0.01±0.03
MET 542	-1.12±0.27	0.38±0.08	-0.04±0.09	-0.08± 0.01	-0.87±0.29
ASN 543	-0.11±0.01	-0.19±0.10	0.24± 0.10	0.000±0.00	-0.06±0.05
LYS 545	-0.14±0.06	-1.04±0.73	1.38±0.82	-0.00 ±0.00	0.20±0.10
ARG 553	-0.99±0.39	-6.25±1.89	7.41±2.03	-0.22±0.07	-0.04±0.39
ALA 554	-0.69±0.16	1.13±0.32	-0.24±0.27	-0.03 ±0.02	0.16±0.30
ARG 555	-4.04±0.70	-6.79±2.14	9.33±2.01	-0.38±0.06	-1.88±0.77
THR 556	-0.92±0.71	-3.18±0.78	2.71± 0.30	-0.14 ±0.02	-1.52±0.56
VAL 557	-1.68±0.25	-0.79±0.21	1.05±0.25	-0.11±0.04	-1.53±0.27
ALA 558	-0.84±0.27	0.05±0.01	-0.01±0.09	-0.05±0.01	-0.85±0.29
GLY 559	-0.08±0.16	-0.08±0.03	0.12±0.03	0.00 ±0.00	-0.04±0.03
VAL 560	-0.05±0.01	0.01±0.01	0.04±0.02	0.00±0.00	0.001±0.2
LYS 621	-0.08±0.02	-0.59±0.34	0.72±0.38	-0.00 ±0.01	0.06±0.08
CYS 622	-0.08±0.02	-0.01±0.05	0.08±0.06	0.00±0.00	-0.02±0.03
ASP 623	-1.74±0.30	0.23±0.95	1.93 ±1.09	-0.23±0.03	0.19±0.69
ARG 624	-1.71±0.59	-2.60±1.61	1.72±1.13	-0.20±0.03	-2.80±0.99
GLU 665	-0.25±0.07	-0.70±0.31	0.69±0.39	-0.00±0.00	-0.21±0.21
VAL 667	-0.44±0.10	0.10±0.02	-0.06±0.03	-0.02±0.01	-0.41±0.11
LYS 676	-0.54±0.13	0.61±0.42	-0.12±0.45	-0.02±0.01	-0.06±0.39
THR 680	-2.60±0.46	-2.77±1.11	1.88±0.51	-0.19±0.03	-3.67±0.88
SER 681	-1.99±0.43	-0.95±0.38	1.58±0.30	-0.07±0.02	-1.44±0.65
SER 682	-2.58±0.50	-1.25±0.56	2.21±0.60	-0.35±0.04	-1.97±0.71
GLY 683	-0.28±0.06	0.10±0.05	0.12±0.06	-0.00±0.00	-0.05±0.06
THR 686	-0.77±0.29	0.08±0.26	0.07±0.22	-0.05±0.04	-0.68±0.42
ALA 687	-0.62±0.29	0.02±0.11	-0.01±0.10	-0.14±0.06	-0.75±0.38
ASN 691	-0.31±0.20	0.06±0.25	0.14±0.29	-0.03±0.04	-0.15±0.26
SER 759	-0.13±0.05	0.24±0.18	-0.08±0.22	-0.00±0.01	0.02±0.08
ASP 670	0.11±0.90	-12.71±2.02	13.13±1.63	-0.37±0.03	0.16±0.80

Also, many of these high energy contributing residues are amongst residues that consistently interacted with Remdesivir. The high energies estimated are also consistent with the strong molecular interactions elicited upon Remdesivir binding, as highlighted over the simulation period. Many of these residues contributed total energies <-1 kcal/mol establishing their essentiality in overall Remdesivir binding and stability [87]. The highest energy contributing residues with total energies <-1 kcal/mol included, ASP452(-3.99kcal/mol), THR456(-1.53kcal), ARG555(-1.88kcal/mol), THR556(-1.52kcal/mol), VAL557(-1.53kcal/mol), ARG624(-2.80kcal), THR680(-3.67kcal/mol), SER681(-1.44kcal/mol) and SER682 (-1.97kcal/mol). The identification of these residues could aid in the determinations of crucial pharmacophoric moieties of Remdesivir, which could be used in the design of improved Remdesivir analogs.

3.7 Per-Residue based pharmacophore model generation reveals essential moieties of Remdesivir

Previously, our research group has extensively published on methods that improve pharmacophore-based modelling. The earliest of such methods was the target-bound pharmacophore modelling approach, which highlighted the importance of incorporating the conformation of bound ligands in the pharmacophore model generation [88]. In improving this method, we outlined another pharmacophore workflow whereby the pharmacophore model was constructed based on high energy-contributing active amino acid residues to overall ligand binding [43,89,90]. High energy contributing residues are identified from per-residue energy decomposition analysis integrated with the MM/GBSA thermodynamics calculation. Identified high energy contributing residues are then mapped with corresponding interacting moieties on the

ligand. In this report, we followed the same protocol to generate an informative structural ensemble of steric and electronic features that may be in the supramolecular interaction of Remdesivir with SARS-CoV-2 RNA dependent RNA polymerase in figure 10B.

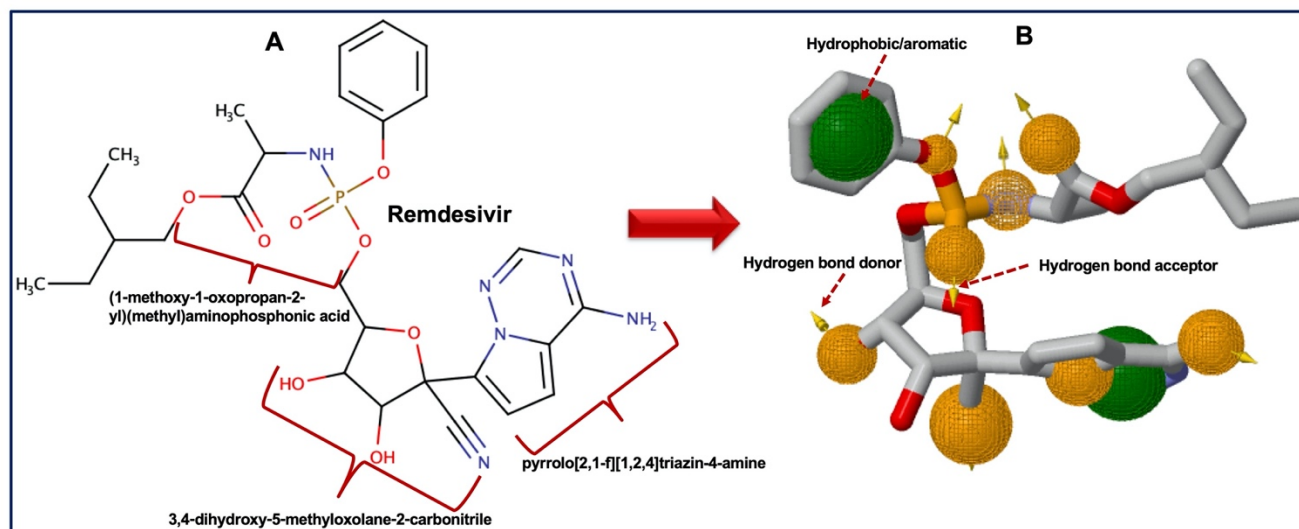


Figure 5.10: A) Specific Remdesivir moieties (the 3,4-dihydroxy-5-methyloxolane-2-carbonitrile moiety, pyrrolo[2,1-f][1,2,4]triazin-4-amine moiety and (1-methoxy-1-oxopropan-2-yl)(methyl)aminophosphonic acid) that interact with high energy contributing residues. B) 3-D pharmacophore model generated on ZINCPharmer (Green–hydrophobic/aromatic moiety, gold – hydrogen bond donor/acceptor).

As observed in the ligand interaction plots many of the high energy contributing residues, notably, ASP452, THR456, ARG555, THR556, VAL557, ARG624, THR680, SER681 and, SER682 interacted with specific moieties on Remdesivir by forming strong interactions with could have favoured its stability within the active site of SARS-CoV-2 RNA dependent RNA polymerase. Specific moieties that these residues interacted with include; the 3,4-dihydroxy-5-methyloxolane-2-carbonitrile moiety, pyrrolo[2,1-f][1,2,4]triazin-4-amine moiety and the (1-methoxy-1-oxopropan-2-yl)(methyl)aminophosphonic acid on Remdesivir as highlighted in figure 5.10A. The chosen essential pharmacophoric features of Remdesivir, as identified in the report, include aromatic/hydrophobic rings, hydrogen bond donors, and hydrogen bond acceptors, as shown in

figure 5.10A. To validate the pharmacophore model generated in this report, we screened the ZINC database for potential SARS-CoV-2 RNA dependent RNA polymerase inhibitors and identified 11 hits compounds as presented in TableS1. A good pharmacophore model should successfully differentiate active compounds from inactive ones. As such, with this extensively explored technique used in constructing our pharmacophore model, we believe the model could be used in screening large compound libraries for hit identification. The pharmacophore model also presents a useful chemical scaffold that could serve as a starting point for the discovery of improved therapeutic analogs of Remdesivir that target SARS-CoV-2 RNA dependent RNA polymerase. Nonetheless, as an inconclusive predictive model, a further validation of the pharmacophore model is required to guarantee its reliability.

4.0 CONCLUSION

In this study, we employed advanced computational techniques to understand the binding mechanism as well as structural and conformational implications that occur upon binding of Remdesivir to SARS-CoV-2 virus RNA-dependent RNA polymerase. Using a validated homology model of SARS-CoV-2 virus RNA-dependent RNA polymerase, our results revealed that Remdesivir binding is mediated by specific residues (ASP452, THR456, ARG555, THR556, VAL557, ARG624, THR680, SER681, and SER682) which engage in an extensive array of high-affinity interactive with essential Remdesivir moieties. Conformational changes observed upon Remdesivir included minimal individual amino acid perturbations, subtle interference in C- α atoms deviations, and a restriction in the systematic transition of SARS-CoV-2 RNA-dependent RNA polymerase from the “buried” hydrophobic region to the “surface-exposed” hydrophilic region. High energy contributing residues identified from per-residue energy decomposition

analysis were subsequently used in the construction of a pharmacophore model. This pharmacophore highlighted the essential moieties on Remdesivir required for SARS-CoV-2 RNA-dependent RNA polymerase inhibition while presenting pharmacophoric features that could aid in the design of improved analogs of Remdesivir to augment the current treatment options for COVID-19. As an ongoing study, a future exploration of pharmacophore model generation towards hit identification against SARS-CoV-2 RNA-dependent RNA polymerase would involve the assembling of multiple inhibitors to constitute training sets that could in turn be employed in building a more robust pharmacophore model *via* feature mapping.

5.0 ACKNOWLEDGEMENTS

The authors acknowledge the School of Health Science, University of KwaZulu-Natal, Westville campus for financial assistance, and The Centre of High Performance Computing (CHPC, www.chpc.ac.za), Cape Town, Republic of South Africa, for computational resources.

6.0 CONFLICT OF INTEREST

Authors declare no conflict of interests.

7.0 ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable

8.0 HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used in the course of this study.

9.0 CONSENT FOR PUBLICATION

Not applicable

10. REFERENCE

- [1] Wang, C.; Horby, P.; Hayden, F.; Gao, G. A Novel Coronavirus Outbreak of Global Health Concern. *Lancet*, **2020**, *395*, 470–473.
- [2] Forster, P.; Forster, L.; Renfrew, C.; Forster, M. Phylogenetic Network Analysis of SARS-CoV-2 Genomes. *Proc. Natl. Acad. Sci.*, **2020**, *117*, 9241 LP – 9243.
- [3] Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K.S.M.; Lau, E.H.Y.; Wong, J.Y.; Xing, X.; Xiang, N.; Wu, Y.; Li, C.; Chen, Q.; Li, D.; Liu, T.; Zhao, J.; Liu, M.; Tu, W.; Chen, C.; Jin, L.; Yang, R.; Wang, Q.; Zhou, S.; Wang, R.; Liu, H.; Luo, Y.; Liu, Y.; Shao, G.; Li, H.; Tao, Z.; Yang, Y.; Deng, Z.; Liu, B.; Ma, Z.; Zhang, Y.; Shi, G.; Lam, T.T.Y.; Wu, J.T.; Gao, G.F.; Cowling, B.J.; Yang, B.; Leung, G.M.; Feng, Z. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N. Engl. J. Med.*, **2020**, 1199–1207.
- [4] Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M.-L.; Zhang, Y.-L.; Dai, F.-H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E.C.; Zhang, Y.-Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature*, **2020**, *579*, 265–269.
- [5] Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; Chen, H.-D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R.-D.; Liu, M.-Q.; Chen, Y.; Shen, X.-R.; Wang, X.; Zheng, X.-S.; Zhao, K.; Chen, Q.-J.; Deng, F.; Liu, L.-L.; Yan, B.; Zhan, F.-X.; Wang, Y.-Y.; Xiao, G.-F.; Shi, Z.-L. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature*, **2020**, *579*, 270–273.
- [6] Outbreak, C. Coronavirus Outbreak <https://www.worldometers.info/coronavirus/> (accessed July 29, 2020).

- [7] Romero, J.R. *Enteroviruses*; Richman, D.; Whitley, R.; Hayden, F., Eds.; 4th ed.; ASM Press: Washington, **2016**.
- [8] Cheng, V.; Lau, S.; Woo, P.; Yuen, K. Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clin Microbiol Rev*, **2007**, *20*, 660–694.
- [9] Lee, N.; Hui, D.; Wu, A.; Chan, P.; Cameron, P.; Joynt, G. et al. A Major Outbreak of Severe Acute Respiratory Syndrome in Hong Kong. *N Engl J Med*, **2003**, *348*, 1986–1994.
- [10] Loon, S.; Lun, K. SARS: A Timely Reminder. *Br J Ophthalmol*, **2013**, *97*, 1217–1218.
- [11] Chan-Yeung, M.; Xu, R. SARS: Epidemiology. *Respirology*, **2003**, *8*, S9–14.
- [12] World Health Organization, (WHO). Middle East Respiratory Syndrome Coronavirus <https://www.who.int/emergencies/mers-cov/en/> (accessed July 29, 2020).
- [13] Chafekar, A.; Fielding, B. MERS-CoV: Understanding the Latest Human Coronavirus Threat. *Viruses*, **2018**, *10*, 93.
- [14] Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J. Pediatr.*, **2020**, *87*, 281–286.
- [15] Cheng, Z.; Shan, J. 2019 Novel Coronavirus: Where We Are and What We Know. *Infection*, **2020**, *48*, 155–163.
- [16] Zou, L.; Ruan, F.; Huang, M. et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med*, **2020**, *382*, 1177–1179.
- [17] World Health Organization, W. Situation reports www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (accessed July 29, 2020).
- [18] Rothe, C.; Schunk, M.; Sothmann, P. Transmission of 2019-NCoV Infection from an Asymptomatic Contact in German. *N Engl J Med*, **2020**, *382*, 970–971.

- [19] Weiss, S.R.; Navas-Martin, S. Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. *Microbiol. Mol. Biol. Rev.*, **2005**, *69*, 635–664.
- [20] Sahin, A.; Erdogan, A.; Mutlu Agaoglu, P.; Dineri, Y.; Cakirci, A.; Senel, M.; Okyay, R.; Tasdogan, A. 2019 Novel Coronavirus (COVID-19) Outbreak: A Review of the Current Literature. *EURASIAN J. Med. Oncol.*, **2020**, *4*, 1–7.
- [21] Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y. et al. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *Lancet*, **2020**, *395*, 497–50.
- [22] Jin, Y.; Cai, L.; Cheng, Z. et al. A Rapid Advice Guideline for the Diagnosis and Treatment of 2019 Novel Coronavirus [2019-NCoV] Infected Pneumonia [Standard Version]. *Mil Med Res*, **2020**, *7*, 4.
- [23] Chen, Z.-M.; Fu, J.-F.; Shu, Q. et al. Diagnosis and Treatment Recommendations for Pediatric Respiratory Infection Caused by the 2019 Novel Coronavirus. *World J Pediatr*, **2020**, *16*, 240–246.
- [24] World Health Organisation. Clinical management of severe acute respiratory infection when novel coronavirus [nCoV] infection is suspected [https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novelcoronavirus-\(ncov\)-infection-is-suspected](https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novelcoronavirus-(ncov)-infection-is-suspected) (accessed July 29, 2019).
- [25] Kupferschmidt, K.; Cohen, J. WHO Launches Global Megatrial Of The Four Most Promising Coronavirus Treatments. *Science (80-.)*, **2020**.
- [26] Smith, T.; Bushek, J.; Prosser, T. COVID-19 Drug Therapy – Potential Options Highlights : Antimicrobials with Potential Activity against SARS-CoV-2 : **2020**.
- [27] McCreary, E.; Pogue, J. COVID-19 Treatment: A Review of Early and Emerging Options. *Oxford Univ. Press*, **2020**, *7*, ofaa105.

- [28] Wang, M.; Cao, R.; Zhang, L.; Yang, X.; Liu, J.; Xu, M.; Shi, Z.; Hu, Z.; Zhong, W.; Xiao, G. Remdesivir and Chloroquine Effectively Inhibit the Recently Emerged Novel Coronavirus (2019-NCoV) in Vitro. *Cell Res.*, **2020**, *30*, 269–271.
- [29] Mulangu, S.; Dodd, L.; Davey, R.J. et al. A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *N. Engl. J. Med.*, **2019**, *381*, 2293–2303.
- [30] Gordon, C.; Tcheshnokov, E.; Feng, J.; Porter, D.; Gotte, M. The Antiviral Compound Remdesivir Potently Inhibits RNA-Dependent RNA Polymerase from Middle East Respiratory Syndrome Coronavirus. *J. Biol. Chem.*, **2020**, *295*, 4773–4779.
- [31] Sheahan, T.; Sims, A.; Leist, S. et al. Comparative Therapeutic Efficacy of Remdesivir and Combination Lopinavir, Ritonavir, and Interferon Beta against MERS-CoV. *Nat. Commun.*, **2020**, *11*, 222.
- [32] Sheahan, T.; Sims, A.; Graham, R. et al. Broad-Spectrum Antiviral GS-5734 Inhibits Both Epidemic and Zoonotic Coronaviruses. *Sci. Transl. Med.*, **2017**, *9*, 396.
- [33] Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Wang, Q.; Xu, Y.; Li, M.; Li, X.; Zheng, M.; Chen, L.; Li, H. Analysis of Therapeutic Targets for SARS-CoV-2 and Discovery of Potential Drugs by Computational Methods. *Acta Pharm. Sin. B*, **2020**, *10*, 766–788.
- [34] Liu, H.; Dastidar, S.; Lei, H.; Zhang, W.; Lee, M.; Duan, Y. Conformational Changes in Protein Function. In *Molecular Modeling of Proteins.. Methods Molecular BiologyTM*; Kukol, A., Ed.; Humana Press, **2008**.
- [35] Pawson, T. Specificity in Signal Transduction: From Phosphotyrosine-SH2 Domain Interactions to Complex Cellular Systems. *Cell*, **2004**, *116*, 191–203.
- [36] Karplus, M.; McCammon, J.A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.*, **2002**, *9*, 646–652.

- [37] Salmaso, V.; Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front. Pharmacol.*, **2018**, *9*, 1–16.
- [38] Kaserer, T.; Beck, K.R.; Akram, M.; Odermatt, A.; Schuster, D.; Willett, P. Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules*, **2015**, *20*, 22799–22832.
- [39] Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. Novel Approach for Efficient Pharmacophore-Based Virtual Screening: Method and Applications. *J. Chem. Inf. Model.*, **2009**, *49*, 2333–2343.
- [40] Richmond, N.J.; Abrams, C.A.; Wolohan, P.R.N.; Abrahamian, E.; Willett, P.; Clark, R.D. GALAHAD: 1. Pharmacophore Identification by Hypermolecular Alignment of Ligands in 3D. *J. Comput. Aided. Mol. Des.*, **2006**, *20*, 567–587.
- [41] Jones, G.; Willett, P.; Glen, R.C. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation. *J. Comput. Aided. Mol. Des.*, **1995**, *9*, 532–549.
- [42] Leach, A.R.; Gillet, V.J.; Lewis, R.A.; Taylor, R. Three-Dimensional Pharmacophore Methods in Drug Discovery. *Journal of Medicinal Chemistry*, **2010**, *53*, 539–558.
- [43] Kumalo, H.M.; Soliman, M.E. Per-Residue Energy Footprints-Based Pharmacophore Modeling as an Enhanced In Silico Approach in Drug Discovery: A Case Study on the Identification of Novel β -Secretase1 (BACE1) Inhibitors as Anti-Alzheimer Agents. *Cell. Mol. Bioeng.*, **2016**, *9*, 175–189.
- [44] Cele, F.N.; Ramesh, M.; Soliman, M.E.S. Per-Residue Energy Decomposition Pharmacophore Model to Enhance Virtual Screening in Drug Discovery: A Study for Identification of Reverse Transcriptase Inhibitors as Potential Anti-HIV Agents. *Drug Des. Devel. Ther.*, **2016**, *10*, 1365–1377.

- [45] Qing, X.; Lee, X.Y.; De Raeymaeker, J.; Tame, J.R.; Zhang, K.Y.; De Maeyer, M.; Voet, A.R. Pharmacophore Modeling: Advances, Limitations, And Current Utility in Drug Discovery. *J. Receptor. Ligand Channel Res.*, **2014**, *7*, 81–92.
- [46] Kaalia, R.; Kumar, A.; Srinivasan, A.; Ghosh, I. An Ab Initio Method for Designing Multi-Target Specific Pharmacophores Using Complementary Interaction Field of Aspartic Proteases. *Mol. Inform.*, **2015**, *34*, 380–393.
- [47] Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics*, **2006**, *22*, 195–201.
- [48] Schwede, T. SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acids Res.*, **2003**, *31*, 3381–3385.
- [49] Kirchdoerfer, R.N.; Ward, A.B. Structure of the SARS-CoV Nsp12 Polymerase Bound to Nsp7 and Nsp8 Co-Factors. *Nat. Commun.*, **2019**, *10*, 2342.
- [50] Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure Validation by C α Geometry: ϕ , ψ and C β Deviation. *Proteins Struct. Funct. Bioinforma.*, **2003**, *50*, 437–450.
- [51] Wiederstein, M.; Sippl, M.J. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res.*, **2007**, *35*, W407–W410.
- [52] Eisenberg, D.; Lüthy, R.; Bowie, J.U. VERIFY3D: Assessment of Protein Models with Three-Dimensional Profiles. *Methods Enzymol.*, **1997**, *277*, 396–404.
- [53] Yang, Z.; Lasker, K.; Schneidman-Duhovny, D.; Webb, B.; Huang, C.C.; Pettersen, E.F.; Goddard, T.D.; Meng, E.C.; Sali, A.; Ferrin, T.E. UCSF Chimera, MODELLER, and IMP: An Integrated Modeling System. *J. Struct. Biol.*, **2012**, *179*, 269–278.

- [54] Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.*, **2004**, *25*, 1605–1612.
- [55] Lung, J.; Lin, Y.; Yang, Y.; Chou, Y.; Shu, L.; Cheng, Y.; Liu, H. Te; Wu, C. The Potential Chemical Structure of Anti-SARS-CoV-2 RNA-dependent RNA Polymerase. *J. Med. Virol.*, **2020**, *2*, 1–5.
- [56] 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.*, **2010**, *31*, 455–461.
- [57] Olotu, F.A.; Agoni, C.; Adeniji, E.; Abdullahi, M.; Soliman, M.E. Probing Gallate-Mediated Selectivity and High-Affinity Binding of Epigallocatechin Gallate: A Way-Forward in the Design of Selective Inhibitors for Anti-Apoptotic Bcl-2 Proteins. *Appl. Biochem. Biotechnol.*, **2018**, *187*, 1061–1080.
- [58] Abdullahi, M.; Olotu, F.A.; Soliman, M.E. Solving the Riddle: Unraveling the Mechanisms of Blocking the Binding of Leukotoxin by Therapeutic Antagonists in Periodontal Diseases. *J. Cell. Biochem.*, **2018**, *119*, 9364–9379.
- [59] Grand, S. Le; Götz, A.W.; Walker, R.C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.*, **2013**, *184*, 374–380.
- [60] Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T.; Cruzeiro, III, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan,

I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. AMBER 2018. *University of California, San Francisco*, **2018**.

[61] Oguntade, S.; Ramharack, P.; Soliman, M.E. Characterizing the Ligand-Binding Landscape of Zika NS3 Helicase-Promising Lead Compounds as Potential Inhibitors. *Future Virol.*, **2017**, *12*, 261–273.

[62] Agoni, C.; Ramharack, P.; Soliman, M.E.S. Synergistic Interplay of The Co-Administration of Rifampin And Newly Developed Anti-TB Drug : Could It Be a Promising New Line of TB Therapy ? *Comb Chem High Throughput Screen*, **2018**, *21*, 453–460.

[63] Agoni, C.; Ramharack, P.; Soliman, M. Co-Inhibition as a Strategic Therapeutic Approach to Overcome Rifampin Resistance in Tuberculosis Therapy: Atomistic Insights. *Future Med. Chem.*, **2018**, *10*, 1665–1675.

[64] Olotu, F.A.; Soliman, M.E.S. Dynamic Perspectives into the Mechanisms of Mutation-Induced P53-DNA Binding Loss and Inactivation Using Active Perturbation Theory: Structural and Molecular Insights toward the Design of Potent Reactivators in Cancer Therapy. *J. Cell. Biochem.*, **2018**, *120*, 951–966.

[65] Salomon-Ferrer, R.; G??tz, A.W.; Poole, D.; Le Grand, S.; Walker, R.C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.*, **2013**, *9*, 3878–3888.

[66] Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.*, **2015**, *11*, 3696–3713.

- [67] Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry*, **2005**, *26*, 1668–1688.
- [68] Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, a; Haak, J.R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, **1984**, *81*, 3684–3690.
- [69] Kräutler, V.; Van Gunsteren, W.F.; Hünenberger, P.H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J. Comput. Chem.*, **2001**, *22*, 501–508.
- [70] Roe, D.R.; Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Com*, **2013**, *9*, 3084–3095.
- [71] Seifert, E. OriginPro 9.1: Scientific Data Analysis and Graphing Software-Software Review. *J. Chem. Inf. Model.*, **2014**, *54*, 1552.
- [72] Massova, I.; Kollman, P.A. Combined Molecular Mechanical and Continuum Solvent Approach (MM-PBSA / GBSA) to Predict Ligand Binding. **2000**, 113–135.
- [73] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Donini, O.; Cieplak, P.; Case, D.A.; Cheatham, T.E. Calculating Structures and Free Energies of Complex Molecules : Combining Molecular Mechanics and Continuum Models. **2000**, *33*, 889–897.
- [74] Miller, B.R.; Mcgee, T.D.; Swails, J.M.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA .Py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.*, **2012**, *8*, 3314–3321.
- [75] Onufriev, A.; Bashford, D.; Case, D.A. Modification of the Generalized Born Model Suitable for Macromolecules. **2000**, 3712–3720.

- [76] Koes, D.R.; Camacho, C.J. ZINCPharmer: Pharmacophore Search of the ZINC Database. *Nucleic Acids Res.*, **2012**, *40*, W409-14.
- [77] Yin, W.; Mao, C.; Luan, X.; Shen, D.-D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; Chang, S.; Xie, Y.-C.; Tian, G.; Jiang, H.-W.; Tao, S.-C.; Shen, J.; Jiang, Y.; Jiang, H.; Xu, Y.; Zhang, S.; Zhang, Y.; Xu, H.E. Structural Basis for Inhibition of the RNA-Dependent RNA Polymerase from SARS-CoV-2 by Remdesivir. *Science (80-.)*, **2020**, *1504*, eabc1560.
- [78] Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.*, **2016**, *17*, 144.
- [79] Lobanov, M.Y.; Bogatyreva, N.S.; Galzitskaya, O. V. Radius of Gyration as an Indicator of Protein Structure Compactness. *Mol. Biol.*, **2008**, *42*, 623–628.
- [80] Olotu, F.A.; Soliman, M.E.S. From Mutational Inactivation to Aberrant Gain-of-Function: Unraveling the Structural Basis of Mutant P53 Oncogenic Transition. *J. Cell. Biochem.*, **2018**, *119*, 2646–2652.
- [81] Karshikoff, A.; Nilsson, L.; Ladenstein, R. Rigidity versus Flexibility: The Dilemma of Understanding Protein Thermal Stability. *FEBS J.*, **2015**, *282*, 3899–3917.
- [82] Issahaku, A.R.; Agoni, C.; Kumi, R.O.; Olotu, F.A.; Soliman, M.E.S. Lipid-Embedded Molecular Dynamics Simulation Model for Exploring the Reverse Prostaglandin D2 Agonism of CT-133 towards CRTH2 in the Treatment of Type-2 Inflammation Dependent Diseases. *Chem. Biodivers.*, **2020**, *17*, e1900548.
- [83] Agoni, C.; Salifu, E.Y.; Munsamy, G.; Olotu, F.A.; Soliman, M. CF3-Pyridinyl Substitution on Antimalarial Therapeutics: Probing Differential Ligand Binding and Dynamical Inhibitory Effects of a Novel Triazolopyrimidine-Based Inhibitor on Plasmodium Falciparum Dihydroorotate Dehydrogenase. *Chem. Biodivers.*, **2019**, *16*, e1900365.

- [84] Malleshappa Gowder, S.; Chatterjee, J.; Chaudhuri, T.; Paul, K. Prediction and Analysis of Surface Hydrophobic Residues in Tertiary Structure of Proteins. *ScientificWorldJournal.*, **2014**, *2014*, 971258.
- [85] Ali, S.; Hassan, M.; Islam, A.; Ahmad, F. A Review of Methods Available to Estimate Solvent-Accessible Surface Areas of Soluble Proteins in the Folded and Unfolded States. *Curr. Protein Pept. Sci.*, **2014**, *15*, 456–476.
- [86] Lu, S.; Wagaman, A.S. On Methods for Determining Solvent Accessible Surface Area for Proteins in Their Unfolded State. *BMC Res. Notes*, **2014**, *7*, 602.
- [87] Badichi Akher, F.; Farrokhzadeh, A.; Olotu, F.A.; Agoni, C.; Soliman, M.E.S. The Irony of Chirality-Unveiling the Distinct Mechanistic Binding and Activities of 1-(3-(4-Amino-5-(7-Methoxy-5-Methylbenzo[*b*] Thiophen-2-Yl)-7 H -Pyrrolo[2,3- *d*] Pyrimidin-7-Yl)Pyrrolidin-1-Yl)Prop-2-En-1-One Enantiomers as Irreversible Covalent FGFR4. *Org. Biomol. Chem.*, **2019**, *17*, 1176–1190.
- [88] Skelton, A.; Maharaj, Y.; Soliman, M. Targetbound Generated Pharmacophore Model to Improve the Pharmacophore-Based Virtual Screening: Identification of G-Protein Coupled Human CCR2 Receptors Inhibitors as Anti-Inflammatory Drugs. *Cell. Mol. Bioeng.*, **2014**, *7*, 45–57.
- [89] Ncube, N.B.; Ramharack, P.; Soliman, M.E.S. An “All-In-One” Pharmacophoric Architecture for the Discovery of Potential Broad-Spectrum Anti-Flavivirus Drugs. *Appl. Biochem. Biotechnol.*, **2018**, *185*, 799–814.
- [90] Agoni, C.; Ramharack, P.; Soliman, M.E.S. Allosteric Inhibition Induces an Open WPD-Loop: A New Avenue towards Glioblastoma Therapy. *RSC Adv.*, **2018**, *8*, 40187–40197.

CHAPTER 6

Published Article

Human Rhinovirus Inhibition through Capsid “Canyon” Perturbation: Structural Insights into the Role of a Novel Benzothiophene Derivative

Clement Agoni^a, Pritika Ramharack^a, Geraldene Munsamy^a, Mahmoud E. S. Soliman^{a*}

^aMolecular 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

<http://soliman.ukzn.ac.za>

ABSTRACT

The challenge in targeting human rhinoviruses (HRV) over the years has been attributed to the wide variety in HRV serotypes. Nonetheless, the search for therapeutic agents against HRV continues unabated. These efforts have been augmented by the recent discovery of the novel benzothiophene derivative shown to inhibit HRV viral replication. Bound to subtype HRV-B14, the compound showed similar inhibitory activity as Pleconaril, a known capsid inhibitor. However, the molecular and structural basis of this inhibition remains unclear. In this *in silico* report, residue interaction network analysis revealed that the binding of the benzothiophene derivative into the “canyon” region of the active site of HRV-B14 distorts its initially extensively networked and compact residue profile. This was characterized by fewer inter-residue hydrogen bonds, reduced van der Waals interactions, and increased residue flexibility. Interestingly, however, the binding of this benzothiophene derivative decreased the flexibility of the north-south wall around the canyon region possibly impeding the “breathing motion” of HRV-B14, hence its inhibition. Atomistic insights also revealed the cruciality of Tyr152 towards inhibitor binding at HRV-B14. This was justified by the amino acid’s high intermolecular interaction with both inhibitors. Findings provide important structural insights in the inhibitory activity the novel benzothiophene derivative and reaffirm its promising potential as an alternative capsid inhibitor towards common cold therapy upon further experimental validation.

Keywords: common cold, human rhinovirus, molecular dynamics simulation, residue interaction network

1.0 INTRODUCTION

Coming from the *picoviridae* family, rhinoviruses (HRVs) are the most prominent causative agent of upper-respiratory infections [1]. First discovered in 1950, these viruses have evolved into at least 160 strains, which have been clustered into 3 subgroups (A, B and C) [2]. Although these viruses generally only cause acute infections, cases have been reported where HRV has been implicated in exacerbated chronic respiratory disease, immunodeficiencies, as well as chronic obstructive pulmonary diseases [3].

Human HRV infections begin with the binding and translocation of the 30nm diameter, non-enveloped virion into the cell. The genetic material is then released into the cytoplasm where the host ribosome translates the positive-sense, single-stranded RNA into a polyprotein that is cleaved to form 11 new viral proteins, eventually maturing into a new viral particle [4].

The HRV polyprotein consists of 4 capsid proteins, VP 1-4, which encloses the viral RNA, as well as 7 non-structural proteins, including proteins 2A-C and 3A-D [1]. The capsid proteins are arranged in a unique icosahedral symmetrical structure, containing 60 copies of each protein [5]. The 3 larger proteins, VP1-3, are located at the external surface of the virus, whereas, VP4 lines the inner surface interfacing VP1-3 and the RNA. Crystallographic studies have indicated that VP1-3 comprises of 8-stranded, antiparallel beta-barrel folds and VP4 is an extended polypeptide [6]. This pentameric protomer structure (figure 6.1) contains structural channels, which are known as “canyons” around its apex.

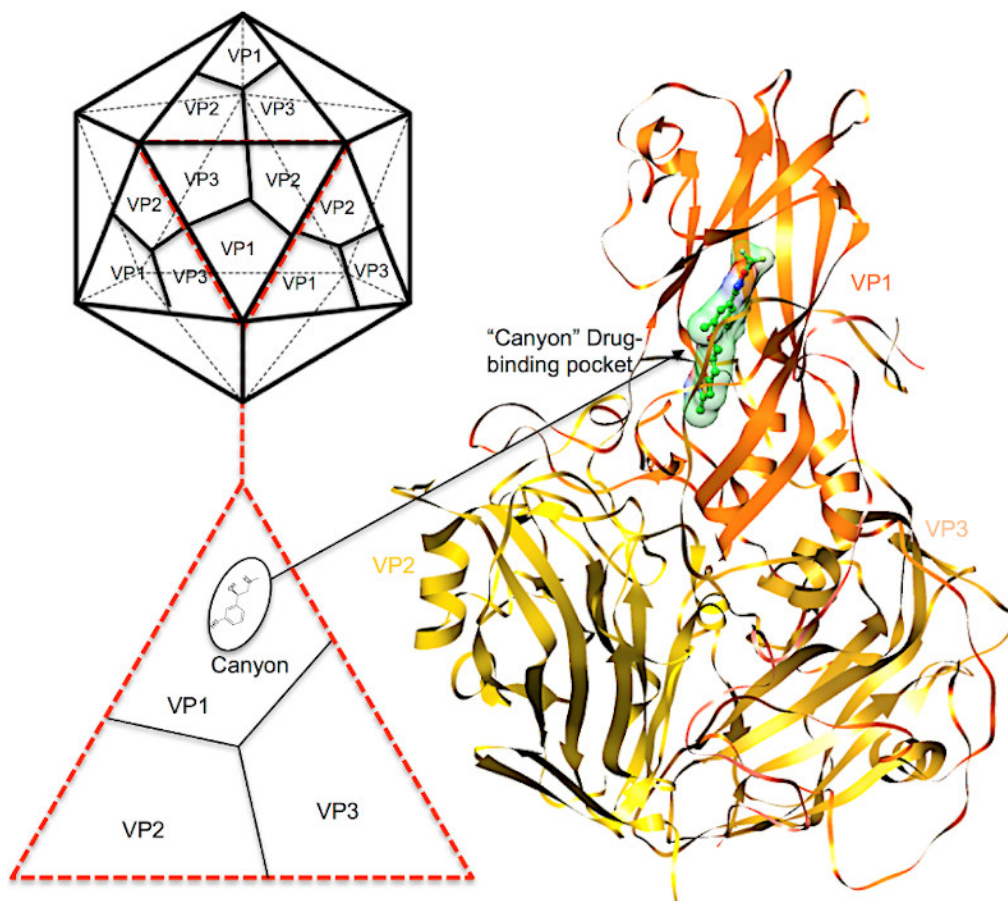


Figure 6.1: Structure of icosahedral viral capsid demonstrating the unique VP1-VP3 domain orientations and the characteristic “canyon” drug-binding pocket.

These “canyons” are used for cell-receptor binding and uncoating during entry into the cell [7,8]. Various sub-groups of HRV utilize small molecules, known as “pocket factors”, that bind in the “canyon” until a receptor becomes available. This hydrophobic binding pocket (Leu25, Ala21, Leu106, Phe124, Tyr128, Iso130, Tyr152, Val176, Tyr197, Met221, Asp219, and Pro172) found within the “canyon” has become a crucial target in several antiviral strategies against HRV [9].

Over recent years, several lead molecules have been established against HRV, with capsid inhibitors such as Pleconaril and vapedavir showing potential to inhibit viral uncoating by stabilizing the cleft region [7]. These drugs, however, are yet to be approved as inhibitors of the

virus [9]. A study by Kim *et al* (2017) put forward a series of benzothiophene derivatives that were effective against both HRV subgroups A and B [7]. Of these compounds, Isopropyl 3-Methyl-6-{{2-(methylcarbamoyl)pyridin-4-yl}oxy}-benzo[b]thiophene-2-carboxylate (compound 6g) demonstrated similar characteristics to Pleconaril, thus augmenting its use as a capsid inhibitor. The report by Kim *et al* (2017) also established that compound 6g binds to the Pleconaril binding site of HRV-B14 *via* molecular docking while also highlighting key interacting residues. Although their work identified compound 6g as a potential HRV capsid inhibitor, no available study currently exists to explain the structural mechanism by which it inhibits the capsid.

In this investigation, we aim to develop on the above research by using a range of *in silico* tools to discover the structural mechanism of inhibition of HRV upon administration of the compound 6g. The binding landscape at the capsid's "canyon" region upon binding of compound 6g will also be assessed, contrasted with Pleconaril. This computational approach will facilitate a further understanding of the machinery of the capsid protein in HRV therapy.

2.0 COMPUTATIONAL METHODS

2.1 System Preparation

The X-ray crystal structures of HRV-B14 was retrieved from the Protein Data Bank with PDB codes 1NCQ. Using Molecular Modeling Viewer (MMV)[10] and UCSF Chimera [11], the structures were prepared for a 130ns Molecular Dynamics (MD) simulation. The 2D structure of compound 6g was modeled using the MarvinSketch software [12]. Using the UFF [13] forcefield, energy optimization was performed on the 2D structure using Avogadro 1.2.0 software [13]. A 3D format of the structure was then saved for molecular docking and MD simulation. In preparing for

MD simulation hydrogen atoms and AM1-BCC charges were added to the 3D structure of compound 6g. Likewise, all co-crystallized non-standard residues on the 3D structures of HRV-B14 were all removed to minimize computational cost. In all three systems were prepared for MD simulation; unbound HRV-B14, Pleconaril-HRV-B14 complex and compound 6g-HRV-B14 complex.

2.1.1 Molecular Docking

The molecular docking protocol implemented in this study utilized AutoDock Vina tool [14] incorporated in UCSF Chimera with default docking parameters. Compound 6g was docked into the Pleconaril binding pocket of HRV-B14. In preparing for molecular docking, compound 6g was supplemented with Gasteiger charges followed by non-polar hydrogen bonds which were merged to the carbon atoms. Compound 6g was then docked into the Pleconaril binding pocket by defining a grid box with a spacing of 1 Å and size of 13.90 × 13.58 × 6.65 pointing in x=36.34, y=4.79 and z=123.93 directions. The best-docked pose was then subjected to 130ns MD simulations.

2.1.2 Molecular Dynamic (MD) Simulations

Molecular dynamic (MD) simulations offer a robust tool to explore the physical movements of biological systems, providing an atomistic view of interactions that occur between atoms and molecules on a molecular level[15]. The insight extracted from MD simulations enable a deeper perspective into the dynamical evolution of biological systems and for these reasons, this computational technique was employed in this study [15–18]. The MD simulations were performed using the GPU version of the PMEMD engine supported by the AMBER 18 package

[19]. This was done in accordance with in-house standard MD simulation protocols employed in several of our previous reports [16,20,21]. The atomic partial charge for compound 6g and Pleconaril was determined using ANTECHAMBER using the General Amber Force Field (GAFF) protocol [22]. Each system was solvated using the Leap module of AMBER 18 implicitly within an orthorhombic box of TIP3P water molecules such that all atoms were within 10Å of any box edge. The Leap module was also employed to neutralize all systems by the addition of Na⁺ or Cl⁻ as counter ions. An initial minimization was carried out for 2000 steps with an applied restraint potential of 500 kcal/mol. This was closely followed by a full minimization of 1000 steps carried out by a conjugate gradient algorithm in the absence of all restraints. All systems were gradually heated from 0K to 300K for 50ps to ensure that all systems maintained a fixed number of atoms and a fixed volume. The solute molecules within each system were imposed with a potential harmonic restraint of 10kcal/mol and collision frequency of 1ps. Following the heating protocol that each system was subjected to, an equilibration step estimating 500ps for all systems was conducted with a constant operating temperature of 300K. An isobaric-isothermal (NPT) ensemble was mirrored by the addition of a number of atoms, and the pressure was also kept constant 1 bar using the Berendsen barostat.

MD simulations on all systems were run for a total time of 130ns. In each simulation, the SHAKE algorithm was interpolated to constrict the bonds of hydrogen atoms. The step size of each simulation was 2fs, employing the SPFP precision model. Each simulation coincided with an isobaric-isothermal ensemble (NPT), with randomized seeding, constant pressure of 1 bar, a pressure-coupling constant of 2ps, a temperature of 300K, and Langevin thermostat with a collision frequency of 1ps.

2.2 Post-Dynamic Analysis

The coordinates of the free HRV-B14 and bound complexes were then saved after every 1ps. The CPPTRAJ module employed in the AMBER 18 suit was employed for post dynamic analysis. The Root Mean Square Deviation (RMSD) and thermodynamic energy estimation of each system were then investigated.

2.2.1 Residue Interaction Network Exploration (RIN)

RIN analysis of the studied systems was performed using a snapshot generated from each system at the end of the 130ns MD simulation. The network of interactions with the selected snapshot was then obtained by uploading the .pdb file format of the snapshot onto the online platform RING server [23]. The generated network is downloaded in .xml format for visualization and on Cytoscape [24] and the RINalyzer plugin[25] was employed to visualize generated RIN. Residues of the HRV-B14 are represented by nodes whilst non-covalent interaction elicited between residues are denoted by edges. Interactions highlighted includes; hydrogen bond interactions, van der Waals interactions, ionic interaction amongst others. Variations in elicited interactions are represented by a unique color for each interaction. Incorporated in Cytoscape is the NetworkAnalyzer [26] plugin, which was used to explore the node degree and node connectivity. Closeness centrality and shortest path betweenness of the simulated models were also analyzed using the RINalyzer plugin.

2.2.2 Binding Free Energy Calculations

The Molecular Mechanics/PB Surface Area method (MM/PBSA) [27] was employed to estimate the binding free energy of each of the systems. Binding free energy averaged over 130000 snapshots which were extracted from the 130ns trajectory.

The free binding energy (ΔG) computed by this method for each molecular species (complex, ligand and receptor) can be represented as:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = E_{\text{gas}} + G_{\text{sol}} - TS \quad (2)$$

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

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{SA}} \quad (4)$$

$$G_{\text{SA}} = \gamma \text{SASA} \quad (5)$$

The term E_{gas} denotes the gas-phase energy, which consists of the internal energy E_{int} ; Coulomb energy E_{ele} and the van der Waals energies E_{vdw} . The E_{gas} was directly estimated from the FF14SB force field terms. Solvation free energy, G_{sol} , was estimated from the energy contribution from the polar states, G_{GB} , and non-polar states, G . 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 Å, whereas the polar solvation, G_{GB} , contribution was estimated by solving the GB equation. S and T denote the total entropy of the solute and temperature respectively.

3.0 RESULTS AND DISCUSSION

3.1 Exploring the Impact of a Drug-filled “Canyon” Region on the Structural Dynamics of HRV-B14

The capsid protein of HRV has been deemed a fascinating viral target due to its unique structural features. The capsid protein, comprising of VP1-4, contains a structural channel, known as a “canyon” region. This “canyon” region (Leu25, Ala21, Leu106, Phe124, Tyr128, Ile130, Tyr152, Val176, Tyr197, Met221, Asp219, and Pro172) mediates cell receptor binding and fusion with the host cell. In physiological conditions, small molecules, known as pocket factors, bind to the “canyon” region to stabilize the capsid until the virus locates a cell receptor [28,29]. In this study, the stability of the HRV-B14 was investigated upon binding of the novel benzothiophene derivative, Isopropyl 3-Methyl-6-{{2-(methylcarbamoyl)pyridin-4-yl}oxy}-benzo[b]thiophene-2-carboxylate (compound 6g). This was contrasted against a known capsid inhibitor, Pleconaril. The stability of the MD simulation was determined by measuring the variations in the root mean square deviation (RMSD) of the C- α atoms across the 130ns trajectory (figure S1). RMSD calculations also allowed for an assessment of the reliability of the simulated models. It was evident from the analysis that the simulated models all achieved stability hence any further analyses performed on the MD trajectories were reliable. The lower RMSD of the bound HRV-B14 complexes relative to the unbound conformation suggests the binding of both compound 6g and Pleconaril induced a level of stability on HRV-B14 with respect to their C- α atoms. To ascertain the flexibility and conformation dynamics of the HRV-B14 upon the binding of compound 6g and Pleconaril, root mean square fluctuation (RMSF) of individual residues were calculated (figure 6.2).

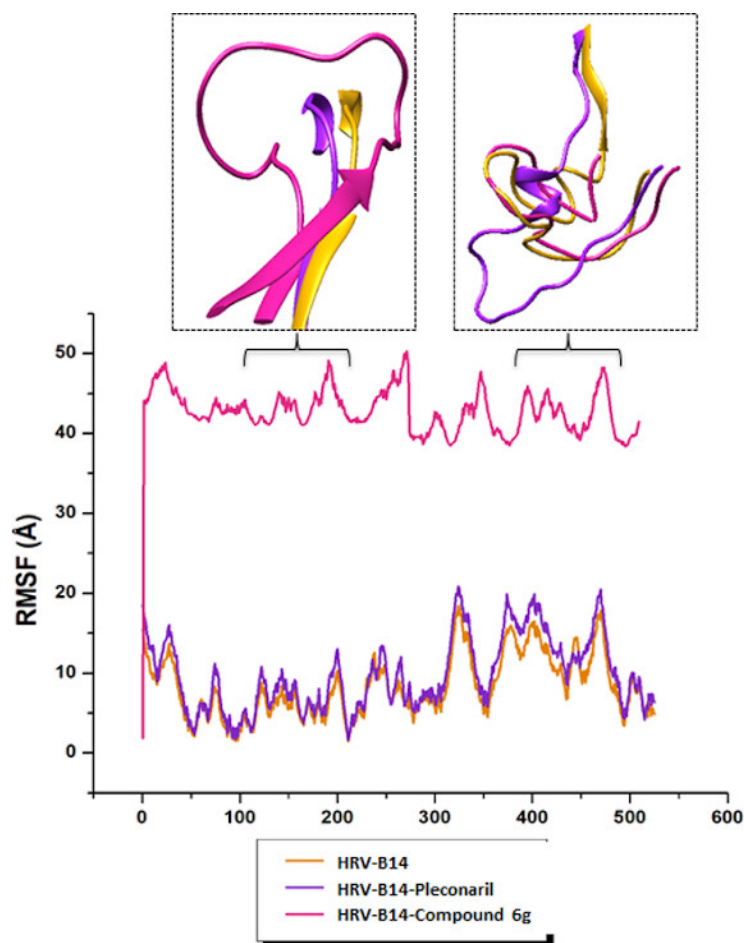


Figure 6.2: Root mean square fluctuations of HRV Capsid protein. Increased fluctuations are seen in the inhibitor bound conformations

Comparative analysis revealed that the compound 6g bound HRV-B14 exhibited a higher fluctuation of residues relative to the Pleconaril bound model. This implies the inhibitory activity of compound 6g resulted in an increase in residue flexibility, which could possibly interfere with the physiological activity of HRV-B14. Also, relative to the unbound HRV-B14, both inhibitor-bound complexes exhibited higher flexibility of residues, suggesting their HRV-B14 inhibitory activity is generally characterized by a distortion an originally stiff residue conformation. A calculation of the radius of gyration (Rg) was also performed to give insights into the distribution of atoms from the center of mass of HRV-B14 over the course of the MD simulations (figure S2).

Rg calculations also allowed for an assessment of the compactness of the HRV-B14 upon inhibitor binding. The highest Rg was observed in the compound 6g bound HRV-B14, suggesting a less compact protein structure relative to the other systems. This less compact protein structure observed in the compound 6g bound complex was consistent with the relatively higher residue flexibility observed in the RMSF calculations. In all, it could be presumed that the inhibitor binding in the canyon region of HRV-B14 induced residue flexibility and as well as a less compact protein conformation which could have impeded its functions.

3.2 Novel benzothiophene derivative binding in “Canyon” region induces residue network perturbations

The dynamics of the web of interactions elicited between residues in the active site of HRV-B14, particularly “canyon” region, prior to inhibitor binding and after inhibitor binding revealed essential atomistic insights that could form the basis HRV-B14 inhibition [30–32]. RIN as a post-MD analysis technique has been explored extensively in previous reports for analysis of the residue-residue interactions for several simulation models, including mutant and wild protein structures [16,33,34]. Notable parameters assessed with RIN node degree, short path length, betweenness, and closeness centrality. Node degree described the number of connections between nodes is whereas the compactness of the protein network is described by the diameter (D) [35,36]. The betweenness and closeness centrality also revealed crucial residues within the residue interaction network of HRV-B14 [35,36]. The short path length parameter of analyses from the RIN provides insights on the flexibility of residue and takes which is described by interconnective edges between respective nodes [35,36].

Based on the residue network generated from average snapshots of each of the 130ns MD simulated systems, the residue-residue interactions of residues that constituted the HRV-B14 binding site were investigated. As shown in figure 6.3, prior to inhibitor binding, residues in the active site region were shown to possess an extensive broad network of interactions amongst each other, engaging in numerous van der Waals (~45 interactions) and hydrogen bond interactions (~32 interactions). The conspicuous extensive network shown in the figure portrays a tight and compact residue network, with residues strongly interacting with each other to facilitate viral uncoating. This was characterized by a high node degree, notable betweenness, and closeness centrality of binding site residues within the unbound HRV-B14. The shortness in length of the connective edges also affirmed a compact residue network within the active site, consistent with the compact overall structure as observed in the R_g and RMSF calculations.

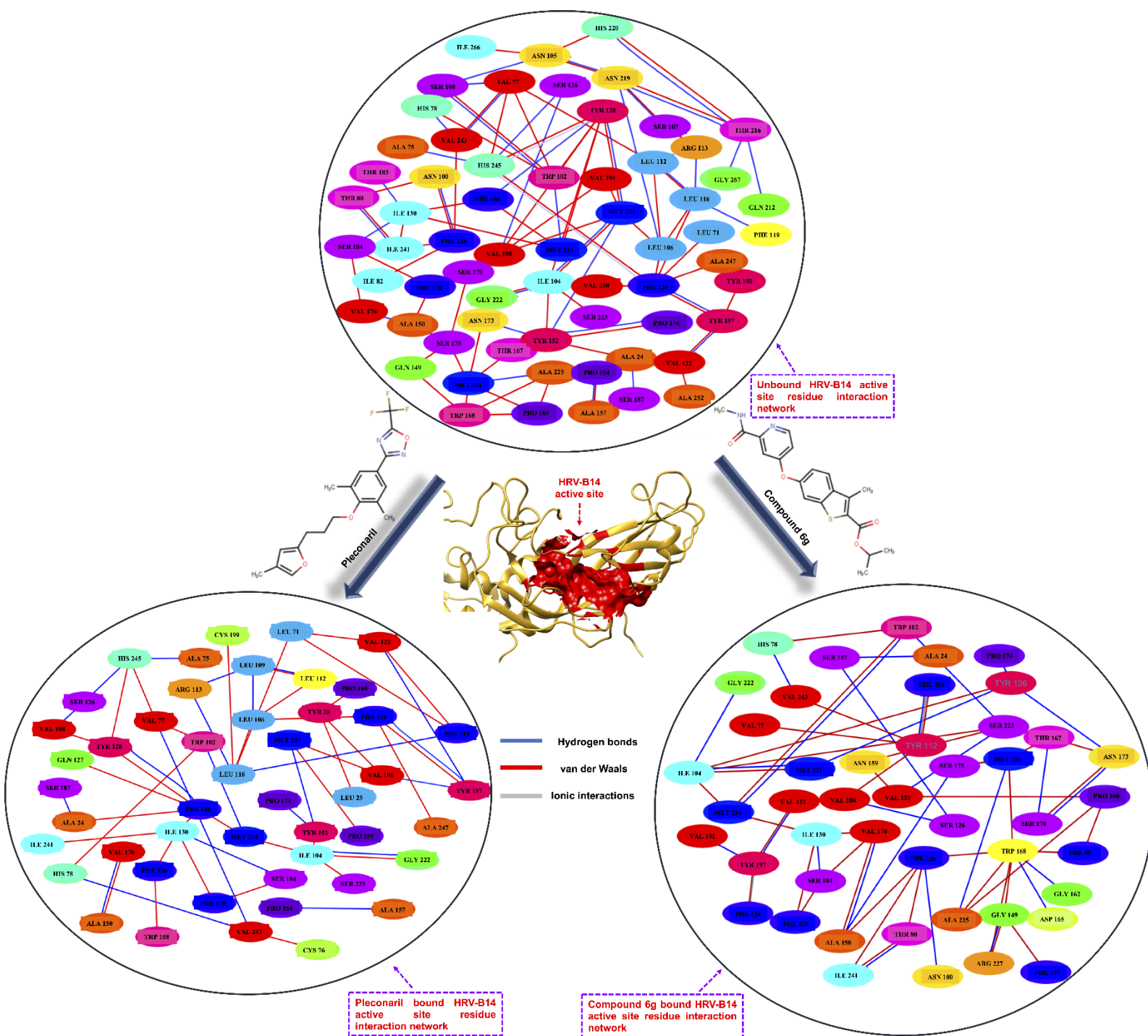


Figure 6.3: Residue Interaction Network of HRV-14 active site residues upon binding of compound 6g, Pleconaril and when unbound. The network shows an extensively tight and compact active site residue network in the unbound conformation which becomes distorted residue network with reduced interactions upon binding of both compound 6g and Pleconaril.

However, upon inhibitor binding, there was an obvious perturbation of the compact residue network of HRV-B14. This perturbation was characterized by a decrease in the number of residue-residue interactions. Both the Pleconaril and the compound 6g bound system elicited approximately 30 van der Waals interactions and roughly 22 hydrogen bond interactions amongst neighboring residues within the active site. A clear increase in the length of connective edges, a decreased node degree, as well as less prominence in closeness and betweenness within the pocket relative to the bound conformation, was also noted. This coupled with the fewer interactions confers with a less compact and a disoriented residue interaction network relative to the unbound system, consistent with less compact overall protein structure observed in Rg and RMSF calculations. There also an increase in short path lengths amongst active site residues of the bound HRV-B14 relative to the unbound conformation. It could, therefore, be presumed that the binding of an inhibitor into the “canyon” region within the active site of HRV-B14 induces a structural perturbation of the residue interaction network of the region, which could, in turn, interfere with the functions of the protein. It was observed that the binding of both compound 6g and Pleconaril imposed a similar impact on the residue network profile suggesting a possible similarity in their inhibitory activity.

3.3 Inhibitor binding impedes North-South wall motion

The “breathing motion” of the HRV-B14 remains crucial for its biological activity and requires conformational flexibility of the canyon region [4]. Therefore, impeding the flexible motion of the canyon region will block essential interactions that occurs at the base of the canyon. To access the flexibility of the North-South Wall, distance analysis was calculated to assess the motion between the north and south wall relative to the effectiveness of each inhibitor as depicted in figure 6.4.

The average distance between the north and south wall calculated from the simulations were 9.90Å, 31.74Å and 32.21Å for the HRV-B14-compound 6g complex, HRV-B14-Pleconaril complex and the unbound HRV-B14 and systems respectively as shown in Table 6.1.

Table 6.1: Measurement of the distance by which the north and south wall moves, measured in Angstroms. (Δ^a Change between maximum and minimum distance)

	HRV-B14	COMPOUND 6G	PLECONARIL
MAXIMUM	38.98	13.33	36.45
MINIMUM	28.97	7.86	28.24
AVERAGE	32.21	9.90	31.74
Δ^a	10.01	5.47	8.21

Overall, the difference between the maximum and minimum distance throughout the 130ns MD simulation in the inhibitor-bound HRV-B14 were shortened relative to the unbound system. As shown, the compound 6g bound system displayed a substantially lower change in distance as opposed to the apo and HRV-B14-Pleconaril systems This suggested that bound conformation exhibited reduced flexibility between the north and south wall located at the at base of the canyon. This suggests that the inhibitory activity of compound g probably allows its effective wedging into the ‘canyon’ region, holding the protein together in order to elicit its therapeutic activity. The shorter distance between the walls in the compound 6g bound complex as the simulation proceeded is indicative of a decline in flexibility of the canyon region, since these walls enclose the canyon region. This decreased flexibility as result of the binding of compound g could possibly impede the “breathing motion” of HRV-B14, hence its inhibition.

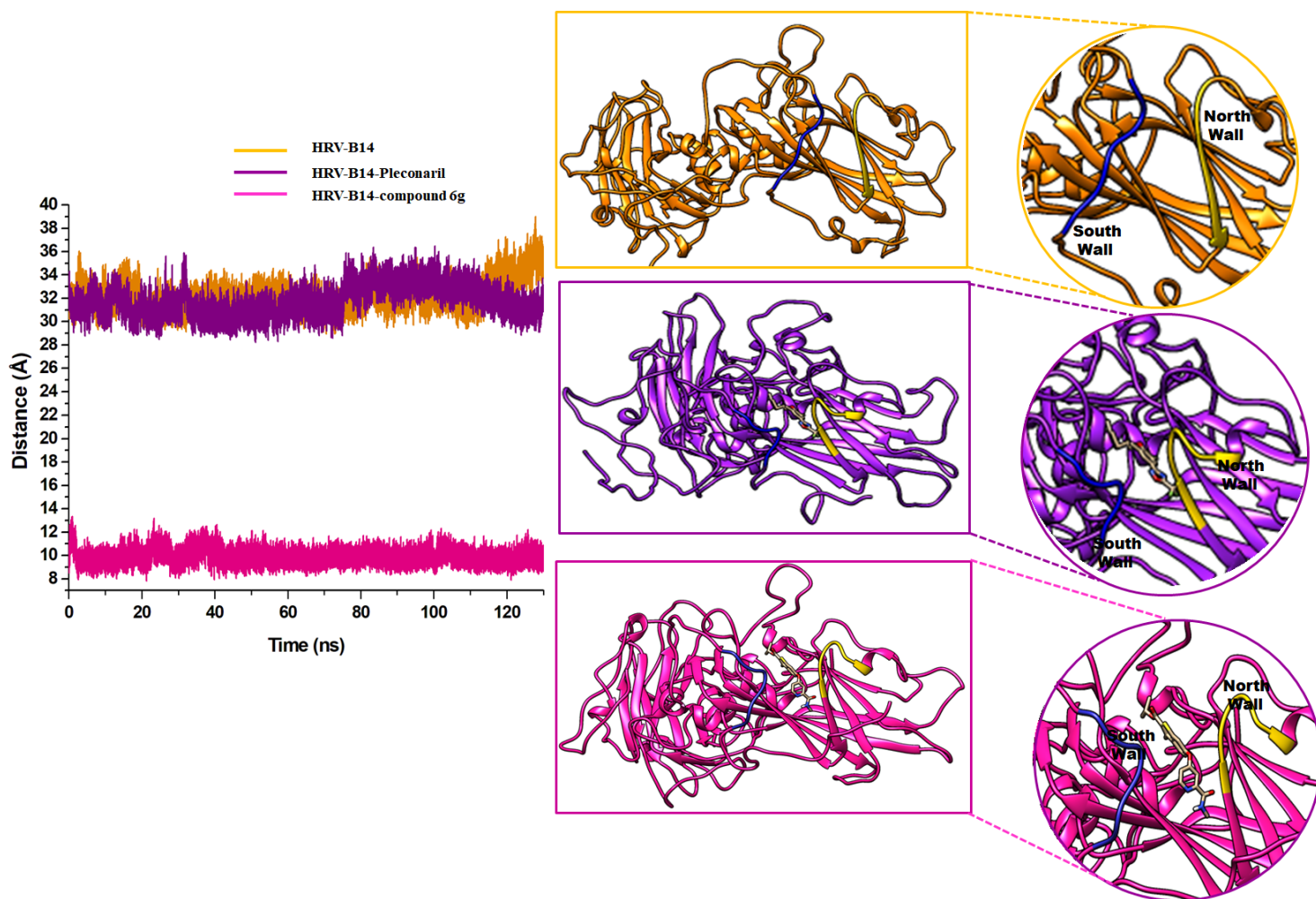


Figure 6.4: Plot of the estimated distances between the north and south wall as well as a visual representation of the flexibility of walls at the base of canyon region of HRV-B14 (orange), HRV-B14-Pleconaril (purple) and HRV-B14-compound 6g (pink)

3.4 Comparative binding free energy profiles of novel benzothiophene derivative and Pleconaril

The therapeutic binding of an inhibitor to any pharmacological target usually underpins the basis of activity of that inhibitor, as such an estimation of the binding affinity of the inhibitor is of critical importance [37]. To ascertain the binding affinity of compound 6g towards establishing the basis of its experimentally determined inhibitory prowess against HRV-B14, we calculated the binding

free energy of compound 6g using the MM/PBSA based approach. Binding free energy of the novel compound 6g was directly compared with the binding free energy of Pleconaril, the known HRV-B14 inhibitor. The estimated binding free energies, even though not absolute values as experimental values, are still reliable because the relative binding affinity between congeneric inhibitors can be computed. This is because binding free energies can be calculated relatively and efficiently using thermodynamic cycles that facilitate a small perturbation of phase space, in contrast to many other quantities of pharmaceutical interest that require much more computational resources [37]. As shown in table 6.2, compound 6g exhibited a total binding free energy of -28.87kcal/mol, whilst Pleconaril exhibited binding free energy of -45kcal/mol. The relatively higher binding free energy of Pleconaril could be attributed to the increased van der Waals (-54.50kcal/mol) and electrostatic interactions (-6.10kcal/mol) it elicited with HRV-B14 as shown in the MM/PBSA calculations. The relatively favorable estimated binding free energy of compound 6g again highlights its promising ability as a potential capsid inhibitor such as Pleconaril.

Table 6.2: MM/PBSA-based binding free energy profiles of compound 6g and Pleconaril

Systems	Energy components (kcal/mol)				
	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
Compound 6g	-37.43± 0.42	-0.70 ± 0.04	-38.13±0.43	9.26 ± 0.11	-28.87±0.33
Pleconaril	-54.50 ± 0.05	-6.10 ± 0.04	-60.60± 0.07	15.60± 0.04	-45.00 ± 0.05

ΔE_{ele} = electrostatic energy; ΔE_{vdW} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG = gas phase free energy.

3.5 Novel benzothiophene derivative shows strong inhibitory per-residue energy-based interactions with HRV-B14

To establish the atomistic basis of the estimated binding free energies with particular focus on the energy contributions of residues that make up the “canyon” region, we calculated the per-residue energy composition of each active site residue of HRV-B14 towards the binding of compound 6g relative Pleconaril. Analyzing per-residue energy contributions towards the binding of compound 6g was performed in as performed in comparison with the per-residue energy contributions of residues that interacted with Pleconaril , since Pleconaril has been shown to bind to the “canyon” and stabilize the protein, preventing the capsid from opening up as well as inhibiting the entry of viruses into the host cell [38]. Exploring the energy contributions of residues that constitute the binding region will be very critical to the inhibitory activity of compound 6g.

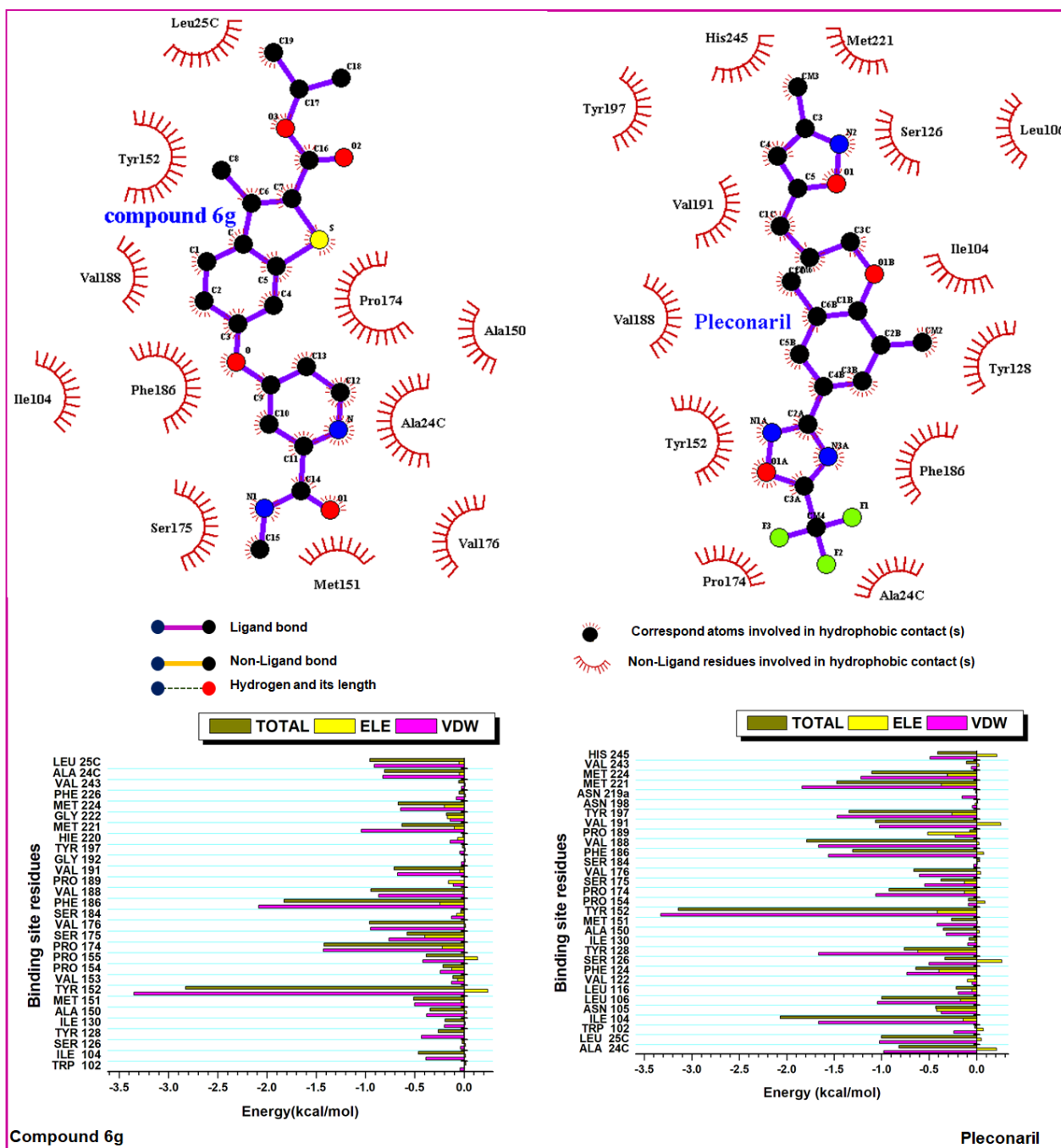


Figure 6.5: Per-residue decomposition analyses revealing individual energy contributions of HRV-B14 active site residues to the binding and stability of compound 6g and Pleconaril. Active site residue that contributed the most to binding of both inhibitors was Tyr152.

As shown in figure 6.5, residues that contributed the most to the total binding of Pleconaril included; Ile104(-2.07kcal/mol), Tyr152(-3.14kcal/mol), Phe186(-1.30kcal/mol), Val188(-1.79kcal/mol), Val191(-1.06kcal/mol), Tyr197 (-1.34kcal/mol), Met221(-1.47kcal/mol), and Met224(-1.10kcal/mol). Likewise, as shown in figure 5, residues that contributed the most to total binding free energy of compound 6g include Pro174(-1.42kcal/mol), Phe186 (-1.83kcal/mol), Leu25C(-0.95kcal/mol), Ala24C(-0.81kcal/mol), Tyr152(-2.83kcal/mol), and Val191(0.71kcal/mol). Interestingly, it could be observed that there were residues that consistently interacted in both pockets, viz; Tyr152, Val191, and Phe186 were notably the most recurrent residues that contributed prominently to the total binding energy of both compound 6g and Pleconaril. It could, therefore, be inferred that the inhibition of HRV-B14 possibly depends on the interaction of the given inhibitor with these three residues. Amongst these three residues, Tyr152, which coincidentally was the highest energy contributing residue in both the compound 6g and Pleconaril bound systems, further signifies its essentiality in the overall, therapeutic inhibition of HRV-B14. Comparatively, residues in the Pleconaril bound complex generally elicited stronger energies relative to the compound 6g bound complex, which possibly influenced the overall higher binding free energy of Pleconaril as observed. Ultimately, the ability of compound 6g to interact with similar residues as Pleconaril with high energies, further reaffirms its ability to inhibit HRV-B14 in a similar fashion.

4.0 CONCLUSION

By employing advanced computational techniques and MD simulations, this study aimed to discover the structural mechanism of inhibition of HRV-B14 upon administration of the novel benzothiophene derivative, Isopropyl 3-Methyl-6-{{2-(methylcarbamoyl)pyridin-4-yl}oxy}-

benzo[b]thiophene-2-carboxylate (compound 6g). The binding landscape at the capsid's "canyon" region was assessed upon binding of compound 6g using Pleconaril, a known capsid inhibitor. An initial conformational assessment revealed that the binding of compound 6g conferred stability to the conformational structure of HRV-B14 relative to the unbound conformation through a calculation of the root mean square deviation of the C- α atoms of each simulated system. An estimation of the individual residue fluctuations of the entire HRV-B14 structure upon binding of compound 6g relative to Pleconaril further revealed that the binding of compound 6g induced increased residue fluctuations, which probably contributed to the inhibition of HRV-B14. A further exploration of individual residue behavior over the course of the 130ns MD simulation by employing the RIN analysis revealed that binding of an inhibitor into the "canyon" region within the active site of HRV-B14 distorts the initially extensively networked residues of the binding site, which could, in turn, possibly favored the inhibitory activities of Pleconaril and compound 6g. The ability of compound 6g to also perturb the residue interaction network of the HRV-B14 active site in a similar fashion as the known capsid inhibitor, Pleconaril, affirms its promising role as a potential novel capsid inhibitor. The flexibility between the north and south wall which enclose the active site of HRV-B14 was also analysed by calculating the distance between walls over the course of the simulation. Distances between the walls in the inhibitor bound systems were shorter in the 130ns MD simulation relative to the unbound conformation, with the shortest distance average distance 9.94Å occurring in the compound 6g bound system. This decreased flexibility as result of the binding of compound g could have possibly interfering with the "breathing motion" of HRV-B14. Comparative binding free energy analysis using the MM/PBSA-based approached also revealed a total binding free energy of -28.87kcal/mol and -45kcal/mol for compound 6g and Pleconaril respectively, with the higher binding of Pleconaril attributing to its corresponding high

electrostatic and van der Waals interactions. The favorable estimated binding free energy of compound 6g enforces its role as a promising capsid inhibitor. Atomistic insights into the energy contributions of HRV-B14 active site residues using per-residue energy decomposition approach also revealed that the inhibitory activity of both Pleconaril and compound 6g involved engaging in interactions with some particular residues within the active site, notably Tyr152. Tyr152 contributed the most towards the total binding of both Pleconaril and compound 6g, suggesting its key role in HRV-B14 inhibition. The structural and atomistic perspectives provided on the inhibitory prowess of the benzothiophene derivative, compound 6g towards HRV-B14 further reaffirms its promising potential as capsid inhibitor, as could therefore be further explored experimentally.

5.0 CONFLICT OF INTEREST

Authors declare no conflict of interest.

6.0 ACKNOWLEDGEMENT

The authors appreciate the College of Health Sciences, University of KwaZulu-Natal for financial and infrastructural support while we also thank the Center for High-Performance Computing (CHPC, www.chpc.ac.za) Cape-Town, South Africa for providing computational resources.

7.0 REFERENCES

- [1] Jacobs SE, Lamson DM, St. George K, et al. Human Rhinoviruses. *Clin. Microbiol. Rev.* 2013;26:135–162.
- [2] Casanova V, Sousa FH, Stevens C, et al. Antiviral therapeutic approaches for human rhinovirus infections. *Future Virol.* 2018;13:505–518.
- [3] Lewis-Rogers N, Bendall ML, Crandall KA. Phylogenetic relationships and molecular

adaptation dynamics of human rhinoviruses. *Mol. Biol. Evol.* 2009;26:969–981.

[4] Reisdorph N, Thomas JJ, Katpally U, et al. Human rhinovirus capsid dynamics is controlled by canyon flexibility. *Virology.* 2003;314:34–44.

[5] Panjwani A, Strauss M, Chou JJ, et al. Capsid Protein VP4 of Human Rhinovirus Induces Membrane Permeability by the Formation of a Size-Selective Multimeric Pore. *PLoS Pathog.* 2014;10:1–12.

[6] Wang L, Smith DL. Capsid structure and dynamics of a human rhinovirus probed by hydrogen exchange mass spectrometry. *Protein Sci.* 2005;14:1661–1672.

[7] Kim J, Jung YK, Kim C, et al. A Novel Series of Highly Potent Small Molecule Inhibitors of Rhinovirus Replication. *J. Med. Chem.* 2017;60:5472–5492.

[8] Roy A, Post CB. Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity. *Proc. Natl. Acad. Sci.* 2012;109:5271–5276.

[9] Rotbart HA. Antiviral therapy for enteroviruses and rhinoviruses. *Antivir. Chem. Chemother.* 2000;11:261–271.

[10] Kusumaningrum S, Budianto E, Kosela S, et al. The molecular docking of 1,4-naphthoquinone derivatives as inhibitors of Polo-like kinase 1 using Molegro Virtual Docker. *J. Appl. Pharm. Sci.* 2014;4:47–53.

[11] Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004;25:1605–1612.

[12] ChemAxon. Marvin Sketch [Internet]. 2013. Available from: <https://www.chemaxon.com/products/marvin/>.

[13] Hanwell MD, Curtis DE, Lonie DC, et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* 2012;4.

- [14] Trott O, Olson AJ. AutoDock Vina. *J. Comput. Chem.* 2010;31:445–461.
- [15] Adcock SA, McCammon JA. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* 2006. p. 1589–1615.
- [16] Agoni C, Ramharack P, Soliman MES. Synergistic Interplay of The Co-administration of Rifampin And Newly Developed Anti-TB Drug : Could It Be a Promising New Line of TB Therapy ? 2018; 21:453-460.
- [17] Agoni C, Ramharack P, Soliman M. Co-inhibition as a strategic therapeutic approach to overcome rifampin resistance in tuberculosis therapy: atomistic insights. *Future Med. Chem.* 2018;10:1665–1675.
- [18] Agoni C, Ramharack P, Soliman MES. Allosteric inhibition induces an open WPD-loop: a new avenue towards glioblastoma therapy. *RSC Adv.* 2018;8:40187–40197.
- [19] Nair PC, Miners JO. Molecular dynamics simulations: from structure function relationships to drug discovery. *silico Pharmacol.* 2014;2:1–4.
- [20] Abdullahi M, Olotu FA, Soliman ME. Solving the riddle: Unraveling the mechanisms of blocking the binding of leukotoxin by therapeutic antagonists in periodontal diseases. *J. Cell. Biochem.* 2018;
- [21] Akher FB, Farrokhzadeh A, Olotu FA, et al. The Irony of Chirality- Unveiling the distinct mechanistic binding and activities of 1-(3-(4-Amino-5-(7-methoxy-5-methylbenzo[b]thiophen-2-yl)-7H-pyrrolo[2,3-d]pyrimidin-7-yl)pyrrolidin-1-yl)prop-2-en-1-one enantiomers as irreversible covalent FGFR4. *Org. Biomol. Chem.* 2019;1176–1190.
- [22] Wang J, Wang W, Kollman PA, et al. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations Correspondence to. *J. Chem. Inf. Comput. Sci.*
- [23] Piovesan D, Minervini G, Tosatto S. The RING 2.0 web server for high quality residue

interaction networks. *Nucleic Acids Res.* 2016;44:W367-74.

[24] Shannon P, Markiel A, Owen Ozier 2, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;2498–2504.

[25] Doncheva NT, Klein K, Domingues FS, et al. Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* 2011;36:179–182.

[26] Assenov Y, Ramírez F, Schelhorn SESE, et al. Computing topological parameters of biological networks. *Bioinformatics.* 2008;24:282–284.

[27] Ylilauri M, Pentikäinen OT. MMGBSA as a tool to understand the binding affinities of filamin-peptide interactions. *J. Chem. Inf. Model.* 2013;53:2626–2633.

[28] Phelps DK, Speelman B, Post CB. Theoretical studies of viral capsid proteins. *Curr. Opin. Struct. Biol.* 2000;10:170–173.

[29] Phelps DK, Post CB. Molecular dynamics investigation of the effect of an antiviral compound on human rhinovirus. *Protein Sci.* 2009;8:2281–2289.

[30] Amitai G, Shemesh A, Sitbon E, et al. Network analysis of protein structures identifies functional residues. *J. Mol. Biol.* 2004;344:1135–1146.

[31] Vishveshwara S, Ghosh A, Hansia P. Intra and inter-molecular communications through protein structure network. *Curr. Protein Pept. Sci.* 2009;10:146–160.

[32] del Sol A, Fujihashi H, Amoros D, et al. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* 2006;2.

[33] Moonsamy S, Bhakat S, Walker RC, et al. Single Active Site Mutation Causes Serious Resistance of HIV Reverse Transcriptase to Lamivudine: Insight from Multiple Molecular Dynamics Simulations. *Cell Biochem. Biophys.* 2016;74:35–48.

[34] Ndagi U, Mhlongo NN, Soliman ME. The impact of Thr91 mutation on c-Src resistance to

UM-164: molecular dynamics study revealed a new opportunity for drug design. *Mol. BioSyst.* 2017;13:1157–1171.

[35] Yan W, Zhou J, Sun M, et al. The construction of an amino acid network for understanding protein structure and function. *Amino Acids.* 2014;46:1419-1439.

[36] Xue W, Jin X, Ning L, et al. Exploring the molecular mechanism of cross-resistance to HIV-1 integrase strand transfer inhibitors by molecular dynamics simulation and residue interaction network analysis. *J Chem Inf Model.* 2012;53:210-222.

[37] Cournia Z, Allen B, Sherman W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. 2017; 5:2911-2937.

[38] Zhang Y, Simpson AA, Ledford RM, et al. Structural and Virological Studies of the Stages of Virus Replication That Are Affected by Antirhinovirus Compounds. 2004;78:11061–11069.

CHAPTER 7

Submitted Article

Dual-Targeting of Human *N*-Myristoltransferase Subtypes 1/2 Halts Common Cold Pathogenesis: Atomistic Perspectives from The Case Of IMP-1088

Clement Agoni^a, Elliasu Y. Salifu^a, Mahmoud E Soliman^a

^aMolecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University
of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

<http://soliman.ukzn.ac.za>

*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 8048, Fax: +27 (0) 31 260 7872

ABSTRACT

Background: The pharmacological inhibition of human *N*-myristoyltransferase (HsNMT) has emerged as an efficient non-cytotoxic strategy to completely thwart the replication process of rhinovirus toward common cold treatment. This approach has potentiated the discovery of IMP-1088, a novel compound that inhibits both HsNMT1 and HsNMT2.

Methods/Results: An augmentative computational and nanosecond-based analyses reveal that the steady and consistent interactions of IMP-1088 with specific residues; Tyr296, Phe190, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182, and Asn246, shared within the binding pockets of both HNMT subtypes, in addition to peculiar structural changes account for its dual inhibitory potency.

Conclusion: Findings unveil atomistic and structural perspectives that could form the basis for the design of novel dual-acting inhibitors of *N*-myristoyltransferase towards common cold therapy.

Keywords: human *N*-myristoyltransferase, common cold, compound IMP-1088, Molecular Dynamics simulation

1.0 INTRODUCTION

Respiratory tract infections are among the leading cause of severe illness and death globally, associated with high incidence rates and consequential economic implications [1,2]. Common in both adults and children, respiratory tract infections categorized into upper respiratory infections (URT) and lower respiratory infections [1]. Available reports suggest that 6 to 8 episodes of respiratory tract infections occur per child per year, with similar rates occurring in developed and developing countries [3,4]. Most respiratory tract infections are moderately mild and usually restricted to the URT [4]. Nevertheless, URT infections may spread downwards and cause severe in infants and young children, and in some cases result in life-threatening infections [5].

Accounting for more than 50% of URT infections [6,7], Human rhinoviruses (HRVs) are exclusively responsible for the common cold in humans, together with other life-threatening respiratory infections such as cystic fibrosis, chronic obstructive pulmonary disease (COPD) and asthma [8,9]. Common cold constitutes an acute, self-limiting viral infection of the URT involving the nose, sinuses, pharynx, and larynx [10].

Rhinoviruses are members of the picornaviridae family, consisting of viral pathogens such as poliovirus, foot-and-mouth disease virus, coxsackievirus, hepatitis A virus, and enterovirus 71 [9]. Rhinovirus is spread either directly or indirectly by contact with secretions from an infected person or aerosol of the secretions [7].

Despite their pathogenic involvements, there have been few therapeutic interventions towards HRV. This is widely attributed to the serotype diversity of the virus, which requires a broad-spectrum of vaccines for thorough efficacy [11,12]. Also, RV has been observed to elicit rapid resistance to existing inhibitors due to its high replication and mutation rates [12]. All these have compounded the existing challenges of treating or suppressing RV infections.

In an infected host cell, the RNA genome of RV usually develops into an infectious virion during the viral life cycle by a series of cellular processes [13]. These processes initiate by a translation of the RV RNA genome into a single polyprotein by the ribosomes of the host cell [13]. This results in the formation of capsid precursors by viral proteases, which are further processed into three capsid proteins, viz.; VP0, VP1, and VP3 [14]. The capsid proteins subsequently trigger a cascade of protein self-assembly leading to the development of infectious virions. The cleavage of VPO into VP4 and VP2 in the intact viral capsid culminates in the final maturation step [14].

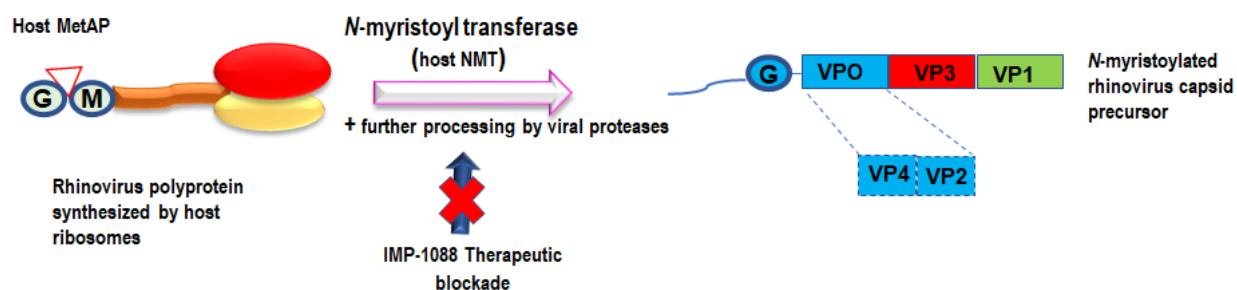


Figure 7.1: Schematic showing of the crucial role of *N*-myristoyl transferase in the viral life cycle. This also highlights how infectious Rhinovirus polyproteins are generated in the cells of an infected host [9]

In many picornaviruses, VP0 is encoded at the N terminus of the viral polyprotein and is subsequently *N*-myristoylated by the host cell's *N*-myristoyltransferase (NMT) [9]. However, in humans and other higher eukaryotic organisms, NMT is expressed as NMT1 and NMT2 proteins in many tissues [9,15].

Mutagenesis studies on poliovirus revealed that the *N*-myristoylation of NMT plays a crucial role in viral assembly and infectivity[15–19]. Also, RV replicates inside a host cell by invading host NMT and subsequently develops a capsid (shell) that protects its genetic material [20]. Therefore, targeting host cell NMT presents an opportunity for inhibiting the infectivity of the virus.

Recent studies reveal that a newly discovered compound, IMP-1088, exerted potent dual inhibitory activities against Human *N*-myristoyltransferase 1 and 2 (HsNMT 1/2) by halting viral replication,

which ultimately results in the prevention of viral infectivity and capsid assembly without causing toxicity to the host cell [9]. Using a resolved X-ray crystal structure of HsNMT1/2 in complex with IMP-1088, it was revealed compound IMP-1088 reportedly formed an extensive network of interactions at the substrate pocket of its target proteins (HsNMT1/2), which potentiates its unprecedented efficacy [20]. However, since these structural insights provided in their report are deduced from single snapshots of interactions from the X-ray crystal complexes which may be inadequate to provide a thorough and dynamic perspective into the mechanistic binding mechanistic activities of compound IMP-1088 towards HsNMT1/2, we employ more rigorous computational methods including molecular dynamics (MD) simulations, sequence analysis, and cheminformatics to provide additional molecular insights into the structural dynamics of HsNMT1/2 upon compound IMP-1088 binding which could account its dual binding activity. MD simulation allows for the mimicking of the dynamical events of molecular systems as a function of time, with a molecular description of all biological components (inhibitor and protein) in a solvent space [21,22]. MD simulations would enable a more detailed elucidation of molecular events and inhibitory mechanisms that characterize the dual inhibitory prowess of compound IMP-1088. We also employ *in silico* approaches to assess the physicochemical and pharmacokinetic profiles of compound IMP-1088. Findings would augment previous experimental efforts and also reveal further important structural perspectives that could advance drug development with respect to the treatment of respiratory tract infections, most especially those mediated by RVs.

2.0 COMPUTATIONAL METHODOLOGY

2.1 System preparation

The X-ray crystal structures of HsNMT1 and HsNMT2 were obtained from RSCB Protein Data Bank with code 5MU6 and 4C2X, respectively. 5MU6 is a dimer with two exact chains (A and B) complexed with compound IMP-1088 and myristoyl CoA; as such, we deleted chain B and prepared our model systems with only chain A to reduce computational cost. The HsNMT2-IMP-1088 complex, on the other hand, was prepared by docking the 2D structure of compound IMP-1088 into the substrate-binding pocket of HsNMT2. Prior to molecular dynamics (MD) simulations, additional co-crystallized molecules such as crystal water, which were not applicable to this study, were removed using the graphical user interface of UCSF Chimera [23].

2.2 Molecular dynamics (MD) simulations

This was performed using the GPU version of AMBER 18 with an integrated PMEMD module [24], according to standard MD simulation protocols, which have been employed extensively in our previous reports[24–28]. IMP-1088 was parameterized using the ANTECHAMBER module of AMBER18 in which its atomic partial charges were generated. Parameterization with the ANTECHAMBER module was performed using the restrained electrostatic potential (RESP) and the General AMBER force field (GAFF) protocol [29]. HsNMT1 and HsNMT2 were subsequently parameterized using the FF14SB AMBER force field [30]. Protonation of histidine residues was subsequently performed using the pdb4amber files at a constant pH to prepare the enzyme structures for the LEAP process that followed. The Leap module integrated with AMBER18 was then employed for the purpose of neutralization of the entire system and for the addition of hydrogen atoms. Neutralization was carried out by adding an equivalent number of

any of the counter ions, Na⁺, Cl⁻. Afterward, the topology and parameter files of IMP-1088, HsNMT1, HsNMT2, and complexes of the enzymes with the IMP-1088 were then saved. The Leap module also allowed for the addition of water molecules with a TIP3P orthorhombic box size of 7Å to explicitly solvate the entire system [31,32]. Prepared systems were then taken through a two-phase minimization process in which the initial minimization of 2000 steps was performed at 500kcal/mol restraint potential. The second minimization referred to as a full minimization involved 1000 steps at a steepest descent with no restraint. This was followed by the gradual thermalization of the systems with a temperature range of 0-300K for 50ps after which each system was equilibrated for 500ps while the temperature and pressure were kept constant at 300K and 1bar, respectively using the Berendsen barostat [33]. This was followed by MD production runs of 100ns for each system during which the SHAKE algorithm [34] was used to constrict all atomic hydrogen bonds. The process of MD simulation involved a 1fs time step initiation process while at 1ps, coordinates of simulated files were saved. All trajectories generated over the course of the simulation were subsequently analyzed using the PTRJ and CPPTRAJ modules of AMBER18 [35]. Plots to interpret findings from the generated MD trajectories were also created using Microcal Origin [36], while UCSF Chimera was used to visualize and create relevant images.

2.3 Thermodynamics calculations

The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) [37–39] method was used to calculate binding free energy. The binding free energy (ΔG_{bind}) was calculated from the following equations:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} + G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = \Delta G_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S, \quad (2)$$

Where ΔG_{bind} is considered to be the summation of the gas phase and solvation energy terms less the entropy ($T\Delta S$) term

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} \quad (3)$$

ΔE_{gas} is the sum of the AMBER force field internal energy terms ΔE_{int} (bond, angle, and torsion), the covalent van der Waals (ΔE_{vdw}) and the non-bonded electrostatic energy component (ΔE_{elec}).

The solvation energy is calculated from the following equation:

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

$$G_{\text{non-polar}} = \gamma \text{SASA} + b \quad (5)$$

The polar solvation contribution is represented as G_{GB} and while $G_{\text{non-polar}}$ is the non-polar solvation contribution. $G_{\text{non-polar}}$ is calculated from the solvent assessable surface area (SASA), obtained by means of a 1.4 Å water probe radius. The surface tension constant (c) was set to 0.0072 kcal/mol and b to 0 kcal/mol [40]. Per-residue decomposition analyses were also carried out to estimate individual energy contribution of residues of the substrate pocket towards the affinity and stabilization of compound IMP-1088. This was aimed at providing atomistic insights in the dual-targeting ability of compound IMP-1088 towards HsNMT1/2, considering the fact that the prominent residual energy contributions could highlight essential residues. In addition to the per-residue energy contributions of the binding site residues, we also examined the sequence similarity of these binding site residues in HsNMT1/2. This was done using the sequence alignment component of UCSF Chimera software.

3.0 RESULTS AND DISCUSSION

3.1 Sequence profiling of HsNMT1/2 hydrophobic pockets with IMP-1088

The ability of a small molecule compound to pharmacologically inhibit VPO *N*-myristoylation by targeting HsNMT of host cells can be considered a breakthrough towards common cold therapy. Inhibition of *N*-myristoylation interferes with the assembly of virions and, as a resultant, causes suppression of the infectivity and prevents RV replication. Reports of the dual-targeting ability of HsNMT1/2 by compound IMP-1088 leaves room to explore the molecular and structural bases behind its dual inhibitory prowess. As such, we analysed the sequences of amino acid residues that constitute the compound IMP-1088 binding pockets in both HsNMT1 and HsNMT2, since these residues directly interacted with compound IMP-1088 and could, therefore, influence its function and overall binding affinity. A variation or similarity in amino acid sequences of the binding pockets of two different enzymes could account for the ability of a particular small molecule inhibitor to either bind to both or be selective towards one. With regards to the binding of compound IMP-1088, the binding pocket sequence analysis permitted a thorough insight into the molecular basis that could account for the ability of compound IMP-1088 to target both HsNMT1 and HsNMT2. As shown in figure 7.2, binding site residues were almost but the same across both enzymes IMP-1088. Some of the common binding site residues across both enzyme binding pockets include; Tyr296, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182, Tyr180, Asp183, Asp184, Phe188, Arg189, Phe190, Asp191, Tyr192, Asn246, Tyr420, Ser421, Asn452, Ala453 and Leu454 amongst many others. This suggested a possible similarity in interaction dynamics with bound inhibitors. The almost identical binding pocket of both enzymes could also confirm an

analogous binding mode of any bound compounds such as IMP-1088 since interaction dynamics would be similar.

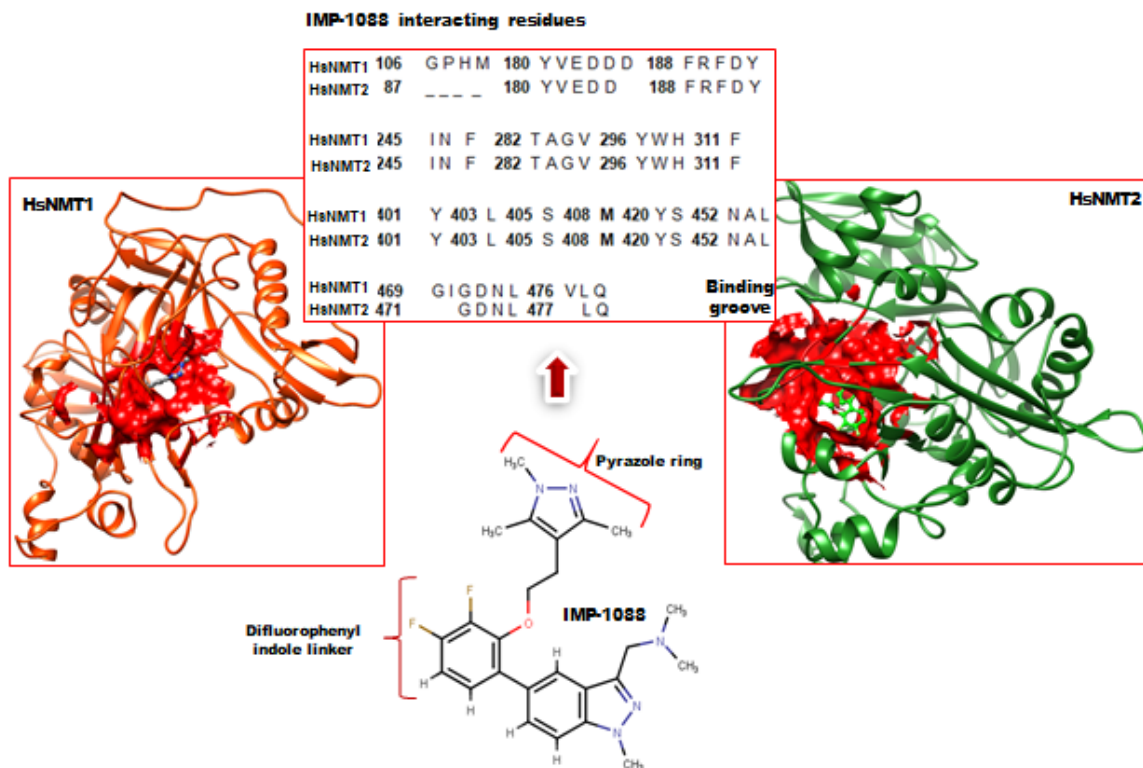


Figure 7.2: Showing the sequence alignment of binding sites residues of HsNMT1 and HsNMT2 and a 2D structure of compound IMP-1088, highlighting essential moieties. Sequence alignment reveals a similarity in binding site residue of both enzyme subtypes.

3.2 IMP-1088 binding pocket stability favors dual binding activity

To understand the conformational dynamics of compound IMP-1088 upon binding to HsNMT1 and HsNMT2, we monitored the relative stability of compound IMP-1088 within the binding pockets throughout the simulation. This was done by a calculation of the root mean square deviation of compound IMP-1088 in both pockets. As shown in figure 7.3A, compound IMP-1088 was relatively more stable in HsNMT1 (1.42Å) than in HsNMT2 (1.94Å). Relative stability within HsNMT1 is suggestive of a steady conformation, which could have allowed for favorable and

sustained interactions with many residues within the binding pocket. This stable conformation was further observed from a superimposition of three representative snapshots (initial, optimized, and final) of compound IMP-1088 in both systems over the course of the simulation. As shown in figure 7.3B and 7.3C, compound IMP-1088 maintained an almost similar conformation throughout the simulation in HsNMT1 relative to varying conformations within HsNMT2. The stable motions of compound IMP-1088 within the binding pocket of HsNMT2 could have favored interactions with residues within the binding pocket.

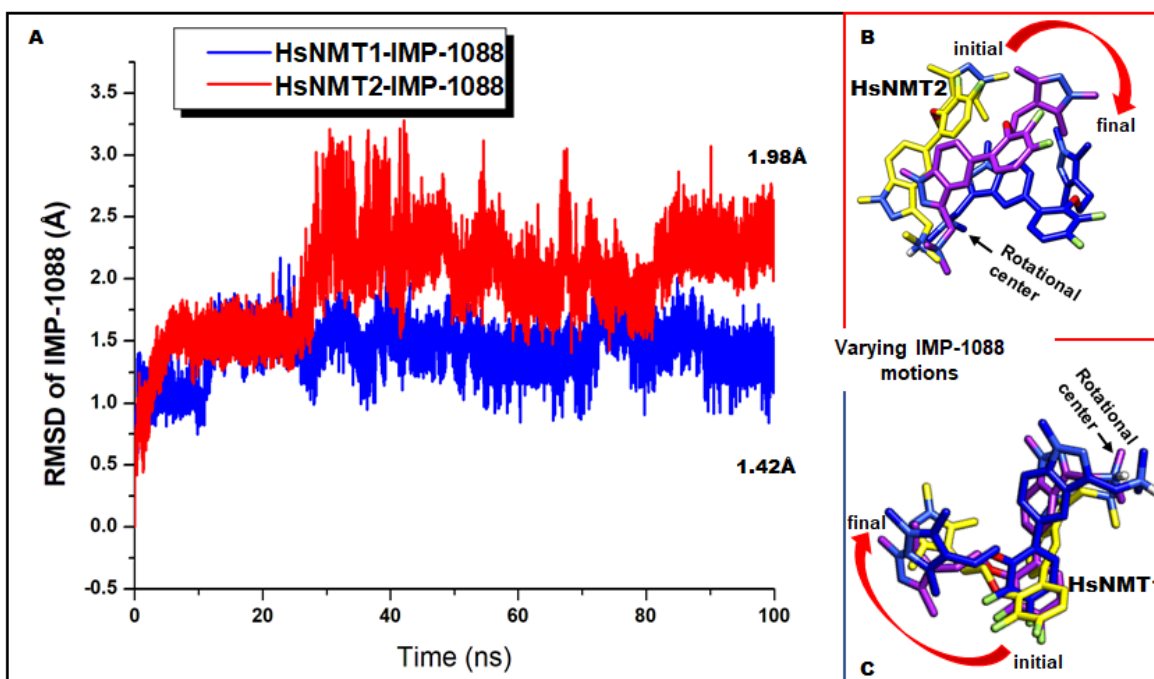
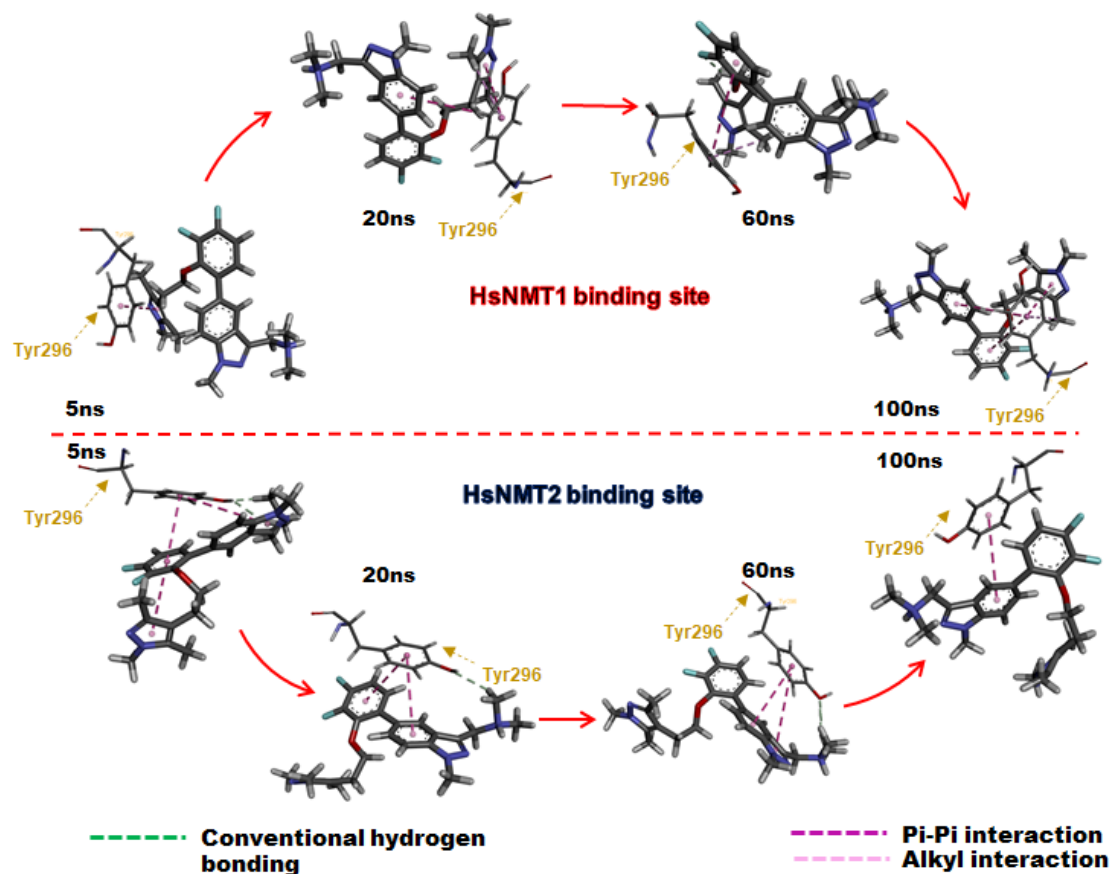


Figure 7.3: A) $C\alpha$ RMSD plot showing comparative stability and motions of IMP-1088 at the binding site HsNMT1 (blue) and HsNMT2 (red) over the 100ns simulation B) The differential positioning of IMP-1088 in the binding site of HsNMT2 and a structural comparison of the initial (yellow), optimized (purple) and final snapshot at 100 ns (blue) C) The differential positioning of IMP-1088 in the binding site of HsNMT1 and a structural comparison of the initial (yellow), optimized (purple) and final snapshot at 100 ns

3.3 Peculiar interactions of Tyr296 with IMP-1088 in both HsNMT1/2 reaffirms role IMP-1088

As reported by Mousnier *et al.*, [9] interaction of compound IMP-1088 with Tyr296 in HsNMT1 led to its displacement and subsequent burial of the compound IMP-1088 resulting in increased interactions. This was reaffirmed in this report by the observed sustained interactions of compound IMP-1088 with Tyr296 throughout the MD simulation period. As shown in figure 7.4, it was observed that from the beginning of the MD simulation, at 5ns and 20ns, Tyr296 elicits a π - π stacked interaction with the pyrazole ring of compound IMP-1088 in the HsNMT1 complex. As the simulation proceeds, this π - π stacked interaction is shown to change into an alkyl interaction at about 60ns, as shown in figure 7.4. The alkyl interaction is subsequently maintained until the end of the MD simulations at 100ns. In comparison with its interaction dynamics in the HsNMT2, Tyr296 is shown to engage in π - π stacked interactions with the difluorophenyl indazole linker of IMP-1088 throughout the simulation. The replication of these crucial interactions with Tyr296 in both subtypes further enforces the role of Tyr296 in the dual inhibitory activity of compound IMP-1088. A time-based monitoring of the interactions that IMP-1088 engaged in over the course of the simulation revealed a gradual expansion of the interaction network as the simulation continued in both enzymes even though there was a larger pool of interactions within HNMT1 relative to HNMT2 as shown in figure 7.4. Also, it was also observed that the binding of compound IMP-1088 in both enzymes involved consistent interactions with particular residues other residues; Phe190, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182, and Asn246 across the simulation period as shown in figure S1 and S2. These residues consistently reoccurred throughout the simulation period in both systems, engaging in various strong interactions with compound IMP-1088. It could, therefore, be inferred that the inhibitory activity of compound IMP-1088 probably

relied on interaction with these particular residues, hence could also possibly underpin the dual targeting ability of compound IMP-1088.



5ns		20ns		60ns		100ns	
HsNMT1	HsNMT2	HsNMT1	HsNMT2	HsNMT1	HsNMT2	HsNMT1	HsNMT2
Asp183	Gly470	Glu182	Asn451	Asp183	His298	His298	His298
Glu182	Gly472	Asp183	Gly470	Asp471	Gly470	Gly470	Tyr420
Phe188	Asp471	His298	Phe190	Gly470	Phe311	Asp183	Tyr296
His298	Glu182	Phe311	Gly472	Phe188	Gly472	Phe188	Leu474
Phe190	Phe188	Asn451	Asp471	His298	Asp471	Asp471	Leu403
Tyr296	Asn473	Ala452	Glu182	Phe190	Phe190	Gly106	Tyr180
Trp297	Asp183	Phe190	Phe311	Tyr420	Glu182	Asn451	Tyr192
Ala452	Phe190	Ser421	Asp183	Asn451	Phe188	Ala452	Gly284
Asn451	Arg189	Tyr420	Arg189	Ala452	Arg189	Leu453	Val181
Ser421	Thr282	Val494	Val181	Leu453	Asp183	Leu495	Phe190
Tyr420	Gly284	Leu453	Tyr401	Val494	Val181	Gln496	Asp183
Val494	Leu474	Tyr401	Leu495	Gln496	Ser405	Thr282	Arg189
Leu453	Gln496	Gln496	Phe196	Asn246	Leu474	Asn246	Ser405
Tyr401	Leu495	Leu495	Tyr192	Leu495	Gln496	Phe247	Phe188
Leu474	Leu495	Tyr180	Leu403	Thr282	Leu495	Tyr192	Asp471
Val181	Leu453	Phe247	Gln496	Phe247	Tyr401	Tyr180	Phe311
Leu495	Tyr401	Asn246	Tyr296	Ala283	Leu403	Ala283	Gly470
Gln496	Leu403	Tyr192	Tyr420	Tyr192	Asn451	Leu474	
Asn246	Tyr192	Thr282		Gly284	Tyr420	Gly284	
Thr282	Tyr420	Ala283		Leu474	Tyr296	Val181	
Tyr192	Tyr296	Gly284		Val181	Tyr192	Trp297	
Tyr180	Val181	Tyr296		Tyr296		Tyr296	
Gly284	Asn451					Phe190	
Phe247	Ala283					Tyr420	

Figure 7.4: Comparative time-based analysis of the interactions of IMP-1088 with the binding pocket of HsNMT1 and HsNMT2 with particular emphasis on interaction with Tyr296 chain. Shown in the table are

individual residues engaged in various intermolecular interactions with IMP-1088. Residues highlighted in red are those that consistently interacted with both HsNMT1 and HsNMT2 throughout the simulation

3.4 Quantifying energy contribution profiles of binding site residues

After the observed similarity in interacting residues in both HsNMT1 and HsNMT2 upon binding of IMP-1088, we proceeded to quantify the specific energy contributions of the individual residues towards the overall binding within the pockets. This was performed by calculation of per-residue based energy decomposition using the MMGBSA incorporated approach. As shown in figure 7.5, prominent energy contributions of active site residues were observed, with most having total energy contributions <-1 , confirming their cruciality to overall total binding of compound IMP-1088[41]. Comparatively, residues within the binding pocket of HsNMT1 contributed more towards the total binding of compound IMP-1088 relative to the binding site residues of HsNMT2, which could largely be attributed to the stronger interaction that's existed between those residues and compound IMP-1088. It was observed that the residues that consistently interacted with compound IMP-1088 in both HsNMT1 and HsNMT2 complex (Tyr296, Phe190, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182, and As246) were among the highest energy contributing residues with total energies being <-1 . Their prominence in total energy contributions establishes how essential their roles are in the activity of compound IMP-1088 in dual targeting of HsNMT1 and HsNMT2.

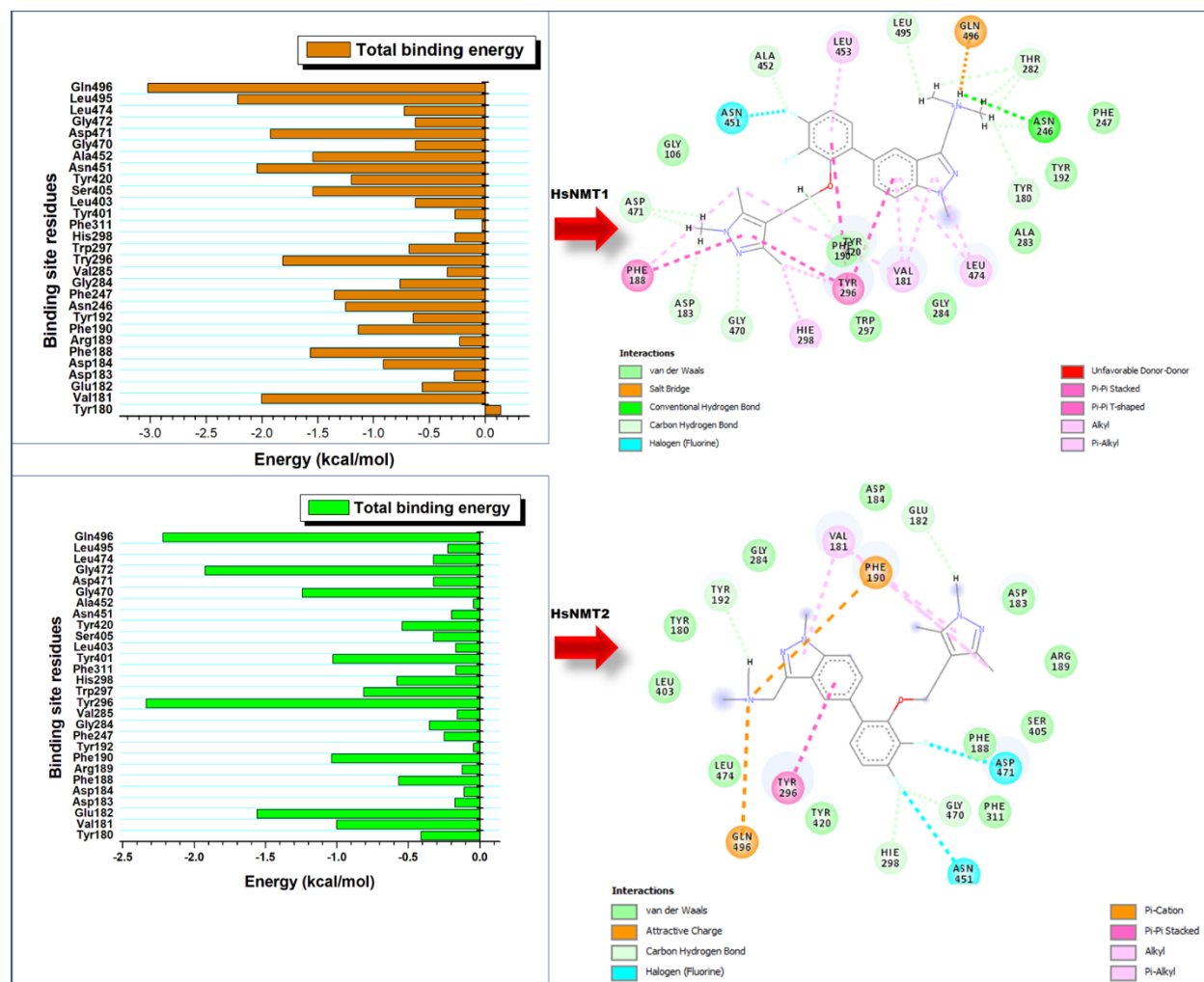


Figure 7.5: Per-residue energy plot of residues that consistently interacted with both HsNMT1 and HsNMT2 over the course of the 100ns MD simulation and their corresponding ligand-residue interaction plots highlighting specific interactions involved

3.5 Comparative binding free energy analysis of compound IMP-1088 upon binding to both HsNMT1/2

Having established that compound IMP-1088 exhibits dual-targeting ability on HsNMT1 and HsNMT2 due to similarity of interacting residues within the respective binding pockets, we proceeded to estimate its total binding free energy towards both enzymes. This was to further assess whether the similarity of interacting residues could confer a similarity in binding affinity. Using the MM/GBSA approach, binding free energy of compound IMP-1088 in complex with

HsNMT1 and HsMNT2 was calculated as presented in Table 7.1. A total binding free energy of -35.20kcal/mol and -30.42kcal/mol was estimated for compound IMP-1088 in the HsNMT1-IMP-1088 and HsNMT2 complex, respectively. Although very close, compound IMP-1088 exhibited higher total binding free energy towards HsNMT1 relative to HsNMT2. This higher binding free energy could be attributed to the expanded network of stronger interactions that compound IMP-1088 engaged in with HsMNT1 over the simulation period relative to the fewer interactions observed in the HsNMT2-IMP-1088 complex. This observed similarity in total binding free energy could also be attributed to similarity in prominent energy contributing residues (Tyr296, Phe190, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182, and As246) as observed in the per-residue energy decomposition.

Table 7.1: MM/GBSA-based binding free energy profile of compound IMP-1088

Complexes	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
HsNMT1-IMP-1088	-44.97 ±0.33	-47.24±0.39	-82.21±0.27	54.03±0.36	-35.20±0.15
HsNMT2-IMP-1088	-48.69± 0.18	-208.43 ±0.81	-247.13±0.13	216.50±0.40	-30.42±0.1

ΔE_{ele} = electrostatic energy; ΔE_{vdW} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG = gas phase free energy.

3.6 Comparative structural impacts of IMP-1088 binding on HsNMT1 and HsNMT2

The binding of any small molecular compound, in this case, compound IMP-1088, always tends imposing structural changes to the respective binding targets. This imposed structural implications could lead to a significant change in the structural architecture of the given target, which can

consequentially influence the overall function of the given target. Although the potency and drug worthiness of compound IMP-1088 have been thoroughly established by earlier experimental reports and *in silico* insights from this report, it was important to further delve into the impact IMP-1088 binding on the structural architecture of both HsNMT1 and HsNMT2. This was to reveal significant structural perturbations and architectural disorientations that could have accounted for inhibitory competence of compound IMP-1088 on both HsNMT1 and HsNMT2 by contrasting with their unbound forms.

To reveal these structural occurrences, we measured parameters such as the C- α root mean square deviation (RMSD), root mean square fluctuation (RMSF) and C- α radius of gyration (RoG) and solvent accessible surface area (SASA) from the combined trajectories that were generated from the 100ns MD simulation. Generated plots from the analysis are presented in figure S3. Insight from the C- α RMSD plots provided a perspective on the stability of the simulated models, a crucial parameter in the overall functionality of an enzyme, which can largely be influenced by the binding of a ligand or a small molecule inhibitor.

Relatively high C- α atom deviations for a given simulated system is usually consistent with structural instability, whereas lower atomic deviation depicts a relatively more stable system. As shown in figure S3, the simulated HsNMT1/2 systems both achieved convergence at ~20ns and ~25ns, respectively. Comparatively, the compound IMP-1088 bound conformations of HsNMT1 and HsMNT2 both exhibited lower C- α atoms deviations relative to their respective unbound conformations, signifying higher structural stability upon compound IMP-1088 binding with average C-a RMSD of 3.26Å and 6.09Å, respectively. The unbound HsMNT1 and HsMNT2 exhibited average C- α RMSD values of 4.02Å and 6.81Å, respectively. Overall it could be deduced that the binding of compound IMP-1088 induced a level of stability on both HsNMT1 and

HsNMT2, thus impeding the deviation of respective C- α atoms. The observed lower C- α RMSD in both HsNMT1 and HsNMT2 also suggests a similarity in the mode by which compound IMP-1080 influences the stability of HsNMT1 and HsNMT2. Nonetheless, there was a relatively higher deviation in the HsNMT2 system compared to the HsNMT1 system, suggesting that compound IMP-1088 induced higher structural stability on HsNMT2. The flexibility of individual residues of a protein could influence the overall function of the protein since residue motions could inform how close they are to bound small molecule inhibitors in order to engage in essential interactions required to influence the activity of the given target. As such, we monitored the root mean square fluctuation of HsNMT1 and HsNMT2 in their bound and unbound conformations. This was performed to also deduce if the binding of compound IMP-1088 somewhat had an impact on the mobility of individual amino acid residues that make up HsNMT1 and HsNMT2, and as such, could be attributed to its inhibitory prowess as experimentally predicted.

Insights from the comparative RMSF plots shown in figure 7.6 revealed that overall, the binding of compound IMP-1088 decreased the flexibility of both HsNMT1 and HsNMT2, signifying a similarity in its influence on residue motions. This similarity of influence on residue motions could possibly form the structural insight that accompanies the ability of compound IMP-1088 to dual-target both HsNMT1 and HsNMT2. On average, unbound HsNMT1 and HsNMT2 exhibited RMSF values of 9.64Å and 10.36Å, respectively whereas the bound versions had lower average RMSF of 7.51Å and 10.27Å, respectively.

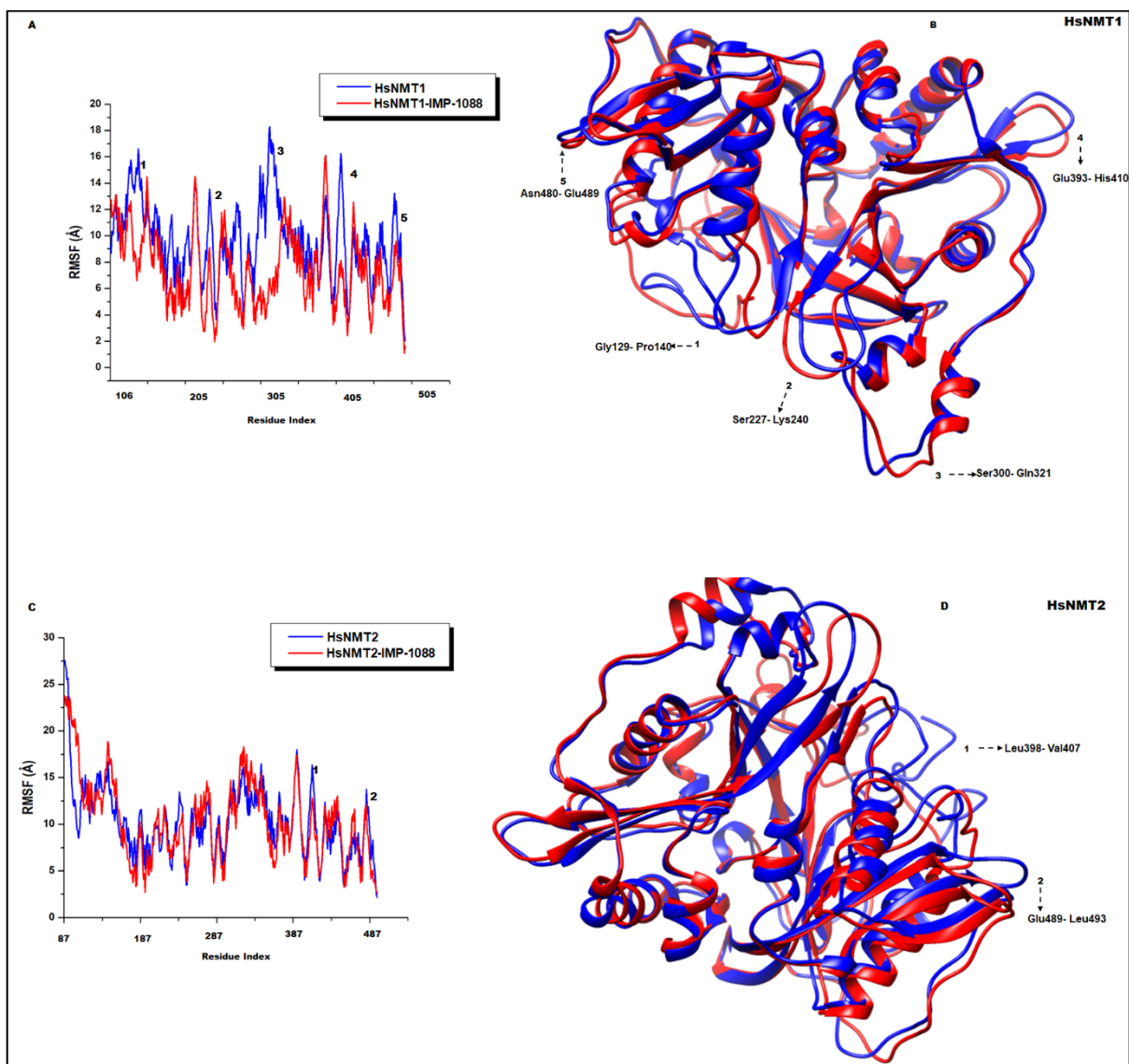


Figure 7.6: Characteristic secondary structure fluctuation among the unbound (blue) and IMP-1088-bound (red) HsNMT1/2. A) and C) RMSF plot showing regions with notable fluctuation which are indicated numerically. B) and D) Superimposed structures of the bound and apo conformation HsNMT1 and HsNMT2 showing regions with high fluctuations as indicated in 6A.

To further explore the structural impact of compound IMP-1088 binding on HsNMT1 and HsNMT2, we also assessed the compactness of the enzymes in the presence and absence of compound IMP-1088 throughout the 100ns simulation. This was assessed by a calculation of the radius of gyration of the C- α atoms of each enzyme subtype from the generated MD trajectories. The comparative RoG plots are presented in figure S1 with the unbound enzymes exhibiting the

highest RoG of 21.85Å for HsNMT1 and 22.43Å for HsNMT2. The compound IMP-1088 bound system exhibited relatively lower RoG of 21.77Å and 22.41Å for HsNNT1 and HsMNT2, respectively. The lower RoG in the inhibitor bound systems were consistent with very compact enzyme conformation, congruous with the decreased flexibility, as shown in the RMSF calculations. This rigid or compact conformation of both enzymes suggested a restricted atomistic mobility within the enzymes, which could have inhibited crucial interactions required for the function of these enzymes in catalyzing RV myristoylation. The impact of the binding of compound IMP-1088 on the folding and unfolding of the HsNMT1 and HsNMT2 was further investigated by a calculation of the solvent accessible-surface area (SASA) of the bound complexes. This provided insights on how individual amino acid residues of the enzymes are exposed to hydrophilic surface since this could influence the interaction dynamics of each amino acid residues to bound ligands. The unbound HsNMT1 and HsNMT2 systems exhibited an average SASA of 18333Å² and 19857Å², respectively, whereas the compound IMP-1088 bound models also exhibited an average SASA of 17501Å² and 19146Å², respectively as shown in figureS3. The lower average SASA values of the compound IMP-1088 bound *N*-myristoyltransferases infers that IMP-1088 binding induced a burial or folding of the individual amino acid residues of the enzymes and thus the residues were less surface-exposed relative to the unbound enzyme conformations. The folding of the amino acid residues upon IMP-1088 binding as showcased in the SASA calculation is consistent with the compact structure and decreased residue flexibility of the bound conformations as observed in the RMSF and RoG calculations. Altogether, it could be inferred that compound IMP-1088 inhibited the HsNMT1 and HsNMT2 by impeding atomistic residue fluctuation, thus inducing a compact structural conformation which culminated into a burial of hydrophobic residues required for the function of the HsNMT1 and HsNMT2.

3.7 *In silico* Physicochemical and ADMET profiling of compound IMP-1088

Available *in silico* approaches allow for the prediction of the physicochemical and pharmacokinetic properties of chemical compounds prior to rigorous experimental methods. Thus, these approaches provide an avenue to assess the drug-likeness of the compound and how its physicochemical properties could influence its absorption, distribution, metabolism, and excretion of the chemical compound in human systems prior to the application of experimental approaches. Following the Lipinski's rule of five (LRO5) [42] and using the online platform, SwissADME [43] we predict the physicochemical and pharmacokinetic descriptors of compound IMP-1088 in order to validate its drug-likeness as presented in Table 7.2. The descriptors, according to the LRO5 include molecular weight (MW) [≤ 500 Da], octanol-water partition coefficient [$\log P \leq 5$], H-bond donors (HBD) [≤ 5] and H-bond acceptors (HBA) [≤ 10]. As estimated, compound IMP-1088 had a molecular weight of 453.53g/mol with 33 heavy atoms, 7 rotatable bonds, 6 hydrogen bonds acceptors, and a molar refractivity of 126.59. Accordingly, a small MW of 453.53g/mol [≤ 500 Da] suggests that compound IMP-1088 potentiates low toxicity risk [44,45]. MW also determines the bioavailability of molecules because a large-sized compound would be impeded during transportation, absorption, and uptake by cells, which could reduce therapeutic potency [46,47]. Hence, the small MW of compound IMP-1088 would favor cellular uptake with little or no impediment to its transport and delivery at target sites. Effective delivery of the compound to the target site could consequently ensure that an adequate concentration of compound IMP-1088 reaches the surface of intestinal epithelial cells for efficient absorption. The favorable MW of compound IMP-1088 also correlated with the moderated number of rotatable bonds (7 rotatable bonds) it possessed since a higher molecular weight of a compound is shown to directly correlate with a higher number of rotatable bonds. The lower number of rotatable bonds confers some level

of molecular rigidity on the structure of compound IMP-1088, which could enhance efficient delivery to target sites for further absorption. A highly flexible molecular structure may engage in unwanted interactions as it is being delivered to the target site. In all, compound IMP-1080 shows no violations of the rule of five, confirming its potential drug likeliness in humans with less toxicity.

Table 7.2: The Summarized ADMET Profile of compound IMP-1088

IMP-1088							
Molecular formula	Molecular weight	Lipophilicity (LogP _{o/w})	Water Solubility (LogS)	GIT absorption	BBB Permeability	Bioavailability score	Lipinki's rule
C ₂₅ H ₂₉ F ₂ N ₅ O	454.3 g/mol	4.48	-5.35 (moderately soluble)	High	Yes	0.55	Yes

The lipophilicity of compound IMP-1080 as a physiochemical feature has significant effects on the solubility, absorption, distribution, metabolism, and excretion properties as well as pharmacological activity. Using SwissADME, an octanol-water partition coefficient (log P_{o/w}) of 4.48 was predicted for compound IMP-1088. A favorable log P_{o/w} usually ranges between -0.4 to 5.6, with extremely high values suggesting a decrease in aqueous solubility with a consequential decrease in absorption rates. In contrast, values below -0.4 depict compounds with lower membrane permeability and poor absorption [47,48]. With a favorable predictive lipophilic value of 4.48 for compound IMP-1088, it implies that compound IMP-1088 exhibits considerable solubility attributes with a tendency of being adequately absorbed in human cells. This could, in

turn, improve the compound IMP-1088 bioavailability and permeability across membranes. In all, based on the pharmacokinetic features explored, compound IMP-1088 possesses a promising potential in its journey towards becoming a one-top therapeutic agent that inhibits both HsNMT1 and HsNMT2.

4.0 CONCLUSION

The enzyme *N*-myristoyltransferase is involved in the transfer of myristate from myristoyl Co-enzyme A to the N terminus of a range of proteins during protein translation. Two variants of this protein are expressed in humans, namely HsNMT1 and HsNMT2. The *N*-myristoylation process of VPO in poliovirus from a mutagenesis report proved its essentiality in the assembly of capsids and the infectivity of the virus. As such, therapeutic targeting of this process in humans may provide a viable therapeutic option in the treatment of viral infections. The ability of compound IMP-1088 to double target HsNMT1 and HsNMT2 in host cells has proven to be a very significant discovery towards the search of a lasting treatment of common cold. Using molecular dynamics simulation and an array of computer-aided tools, the present study, for the first time, sought to explore and provide structural and atomistic insight into the dual-targeting ability of compound IMP-1088 towards HsNMT1 and HsNMT2. Exploration of the pattern of the amino acid sequences that interacted directly with compound IMP-1088 in both HsNMT1/2 revealed striking similarity of interacting amino acid sequences even though HsNMT1 and HsNMT2 only possessed an overall sequence similarity of 83.2%. Also, consistent interaction with Tyr296 resulted in the displacement of the Tyr296 chain which enhanced the burial of compound IMP-1088 within the hydrophobic pocket. The burial within the hydrophobic pocket allowed for the expansion of the interaction network in both complexes as simulation persisted. Quantification of the energy

contributions of each residue to the binding of compound IMP-1088 showed that these recurring residues contributed the highest energies < -1 kcal/mol, further establishing their essentiality in the dual acting ability of compound IMP-1088. Estimated total binding free energy of compound IMP-1088 in HsNMT1 and HsNMT2 revealed near-similar ΔG s of -35.20 kcal/mol and -30.42 kcal/mol, respectively whilst the higher binding free energy exhibited in HsNMT1 could be attributed to stronger active site interactions. An assessment of the conformational dynamics of HsNMT1 and HsNMT2 revealed that compound IMP-1088 exerted a similar pattern of structural inactivation, which entails reduced atomistic fluctuation and structural compactness, which culminated into the burial of surface-exposed residues. Finally, an assessment of pharmacological and pharmacokinetic profiles of compound IMP-1088 revealed a favorable predictive lipophilic value of 4.48 and a general agreement with the Lipinski's rule of five, implying compound IMP-1088 exhibits a considerable solubility attribute, which can consequentially improve its bioavailability and permeability across membranes. Findings from this report unveils essential atomistic insights from a structural perspective that support the experimentally reported inhibitory prowess of compound IMP-1088 against HsNMT1 and HsNMT2 while providing the basis for the design of improved dual-acting inhibitors.

5.0 CONFLICTS OF INTEREST

Authors declare no financial and intellectual conflict of interest

6.0 ACKNOWLEDGEMENTS

The authors acknowledge the School of Health Science, University of KwaZulu-Natal, Westville campus for financial assistance, and The Centre of High-Performance Computing (CHPC, www.chpc.ac.za), Cape Town, RSA, for computational resources.

Executive Summary

Introduction

Human rhinoviruses (HRVs) account for more than 50% of upper respiratory tract infections. These pathogenic organisms are exclusively responsible for common cold in humans together with other life-threatening respiratory infections such as asthma, cystic fibrosis and chronic obstructive pulmonary disease (COPD).

N-myristoylation, catalyzed by human *N*-myristoyltransferase (HNMT) plays a crucial role in viral assembly and infectivity, and as such therapeutic targeting the HNMT in a host cell presents an opportunity towards inhibiting the infectivity of the rhinovirus.

IMP-1088 was recently reported to possess potent dual inhibitory activities against HNMT1/2 thereby halting viral replication and ultimately, elicit the prevention of viral infectivity and capsid assembly without causing toxicity to the host cell.

However, the structural and molecular basis of this dual inhibitory potency remain inconclusive. This study therefore attempts to provide essential structural and atomistic insights that explain the dual-targeting of IMP-1008 while unraveling its pharmacological propensities

Computational Methods

This study employed in-house pre-molecular dynamics simulation preparation protocols as reported in our previous reports. The X-ray crystal structures of Human *N*-myristoyltransferase 1 and 2 (HNMT 1/2) were obtained from RSCB Protein Data Bank with code 5MU6 and 4C2X respectively. HNMT1 was retrieved co-crystallized with the novel inhibitor IMP-1088 while the HNMT2-IMP-1088 complex, on the other hand, was prepared by docking the 2D structure of IMP-1088 into its substrate binding pocket.

A 100ns MD simulation was conducted using the AMBER18 Suite with its CPPTRJ and PTRAJ modules employed analyzing generated MD trajectories.

Results/Discussion

Sequence and nanosecond-based analyses identified Tyr296, Phe190, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182, and Asn246 as residues common within the binding pockets of both HsNMT1 and HsNMT2 subtypes whose consistent interactions with IMP-1088 could underpin the basis for its dual inhibitory potency

Consistent interaction of IMP-1088 with Tyr296 in both enzymes resulting in the displacement of the Tyr296 chain and enhancing the burial of compound IMP-1088 within the hydrophobic pockets is also confirmed

Structural changes upon IMP-1088 binding reveal a characteristic impeding of residue fluctuations, structural compactness, and a consequential burial of crucial hydrophobic residues, features required for HsNMT1/2 functionality.

A quantification of the energy contributions of prominent interacting residues to the binding of IMP-1088 in both HNMT1/2 showed that these residues contributed the highest energies <-1

kcal/mol, further establishing the essentially of these residues in the dual-targeting prowess of IMP-1088

Estimated total binding of IMP-1088 in HNMT1/2 revealed near-similar ΔG s of -35.20kcal/mol and -30.42kcal/mol respectively while the higher binding exhibited in HNMT1 could be attributed to stronger active site interactions.

Conclusions

Dual-targeting activity of IMP-1088 is mediated by its interaction with specific residues (Tyr296, Phe190, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182 and As246) accompanied by peculiar structural changes.

Findings unveil interesting atomistic and structural perspectives into dual inhibitory activity of IMP-1088 against HNMT1/2 while providing the basis for the design of novel dual-acting inhibitors of *N*-myristoltransferase towards the treatment of common cold.

7. REFERENCES

- [1] NHS, Centre for Clinical Practice at NICE. Respiratory tract infections – antibiotic prescribing. NICE Clin. Guidel. 2008.
- [2] Hayward G, Thompson MJ, Perera R, et al. Corticosteroids for the common cold. Cochrane Database Syst. Rev. 2015.
- [3] Lessler J, Reich NG, Brookmeyer R, et al. Incubation periods of acute respiratory viral infections: a systematic review. Lancet Infect. Dis. 2009.
- [4] Whitman JH. Upper respiratory tract infections. Clin. Fam. Pract. 2004.
- [5] Hersh AL, Jackson MA, Hicks LA. Principles of Judicious Antibiotic Prescribing for Upper Respiratory Tract Infections in Pediatrics. Pediatrics. 2013;

- [6] Blaas D, Fuchs R. Mechanism of human rhinovirus infections. *Mol Cell Pediatr.* 2016;3:21.
- [7] Heikkinen T, Järvinen A. The common cold. *Lancet.* 2003.
- [8] Mirabelli C, Scheers E, Neyts J. Novel therapeutic approaches to simultaneously target rhinovirus infection and asthma/COPD pathogenesis. *F1000Research.* 2017;6:1860.
- [9] Mousnier A, Bell AS, Swieboda DP, et al. Fragment-derived inhibitors of human N-myristoyltransferase block capsid assembly and replication of the common cold virus. *Nat Chem.* 2018;10:599–606.
- [10] Allan GM. *Prob23.* 2014;186:190–199.
- [11] van der Schaar HM, Leyssen P, Thibaut HJ, et al. A novel, broad-spectrum inhibitor of enterovirus replication that targets host cell factor PI4KIII β . *Antimicrob Agents Chemother.* 2013;
- [12] Thibaut HJ, De Palma AM, Neyts J. Combating enterovirus replication: State-of-the-art on antiviral research. *Biochem. Pharmacol.* 2012.
- [13] Jiang P, Liu Y, Ma H-C, et al. Picornavirus Morphogenesis. *Microbiol Mol Biol Rev.* 2014;
- [14] Liu Y, Hill MG, Klose T, et al. Atomic structure of a rhinovirus C, a virus species linked to severe childhood asthma. *Proc Natl Acad Sci.* 2016;
- [15] Marc D, Drugeon G, Haenni A, et al. Role of myristoylation of poliovirus capsid protein VP4 as determined by site-directed mutagenesis of its N-terminal sequence. *EMBO J.* 1989;8:2661–2668.
- [16] Bell AS, Mills JE, Williams GP, et al. Selective inhibitors of protozoan protein N-myristoyltransferases as starting points for tropical disease medicinal chemistry programs. *PLoS Negl Trop Dis.* 2012;6.
- [17] Marc D, Girard M, Van Der Werf S. A Gly1 to Ala substitution in poliovirus capsid protein VP0 blocks its myristoylation and prevents viral assembly. *J Gen Virol.* 1991;

- [18] Marc D, Masson G, Girard M, et al. Lack of myristoylation of poliovirus capsid polypeptide VP0 prevents the formation of virions or results in the assembly of noninfectious virus particles. *J Virol.* 1990;
- [19] Moscufo N, Simons J, Chow M. Myristoylation is important at multiple stages in poliovirus assembly. *J Virol.* 1991;
- [20] Corbic Ramljak I, Stanger J, Real-Hohn A, et al. Cellular N-myristoyltransferases play a crucial picornavirus genus-specific role in viral assembly, virion maturation, and infectivity. *PLoS Pathog.* 2018.
- [21] Salmaso V. Exploring Protein Flexibility During Docking to Investigate ligand-Target Recognition. University of Padova, Padova.; 2018.
- [22] De Vivo M, Masetti M, Bottegoni G, et al. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* 2016. p. 4035–4061.
- [23] Goncalves V, Brannigan JA, Thinon E, et al. A fluorescence-based assay for N-myristoyltransferase activity. *Anal Biochem.* 2012;
- [24] Salomon-Ferrer R, Gotz AW, Poole D, et al. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J Chem Theory Comput.* 2013;9:3878–3888.
- [25] Olotu FA, Agoni C, Adeniji E, et al. Probing Gallate-Mediated Selectivity and High-Affinity Binding of Epigallocatechin Gallate: a Way-Forward in the Design of Selective Inhibitors for Anti-apoptotic Bcl-2 Proteins. *Appl Biochem Biotechnol [Internet].* 2018;187:1061–1080. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30155742><http://link.springer.com/10.1007/s12010-018-2863-7>.

- [26] Oguntade S, Ramharack P, Soliman ME. Characterizing the ligand-binding landscape of Zika NS3 helicase-promising lead compounds as potential inhibitors. *Future Virol* [Internet]. 2017;12:261–273. Available from: <http://www.futuremedicine.com/doi/10.2217/fvl-2017-0014>.
- [27] Agoni C, Ramharack P, Soliman MES. Synergistic Interplay of The Co-administration of Rifampin And Newly Developed Anti-TB Drug : Could It Be a Promising New Line of TB Therapy ? *Comb Chem High Throughput Screen*. 2018;21:453–460.
- [28] Agoni C, Ramharack P, Soliman M. Co-inhibition as a strategic therapeutic approach to overcome rifampin resistance in tuberculosis therapy: atomistic insights. *Future Med Chem* [Internet]. 2018;10:1665–1675. Available from: <https://doi.org/10.4155/fmc-2017-0197>.
- [29] Wang J, Wang W, Kollman PA, et al. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations Correspondence to. *J. Chem. Inf. Comput. Sci*.
- [30] Maier JA, Martinez C, Kasavajhala K, et al. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 2015;11:3696–3713.
- [31] Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79:926–935.
- [32] Case D a., T.E. Cheatham I, Darden T, et al. The Amber biomeolecular simulation programs. *J Comput Chem* [Internet]. 2005;26:1668–1688. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/jcc.20290/pdf>.
- [33] Berendsen HJC, Postma JPM, van Gunsteren WF, et al. Molecular dynamics with coupling to an external bath. *J Chem Phys* [Internet]. 1984;81:3684–3690. Available from: <http://link.aip.org/link/JCPSA6/v81/i8/p3684/s1&Agg=doi%5Cnpapers2://publication/doi/10.1063/1.448118>.
- [34] Kräutler V, Van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve

distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem.* 2001;22:501–508.

[35] Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Com.* 2013;9:3084–3095.

[36] Seifert E. OriginPro 9.1: Scientific data analysis and graphing software – Software review. *J Chem Inf Model.* 2014;54:1552.

[37] Massova I, Kollman PA. Combined molecular mechanical and continuum solvent approach (MM-PBSA / GBSA) to predict ligand binding. 2000;113–135.

[38] Kollman PA, Massova I, Reyes C, et al. Calculating Structures and Free Energies of Complex Molecules : Combining Molecular Mechanics and Continuum Models. 2000;33:889–897.

[39] Miller BR, McGee TD, Swails JM, et al. MMPBSA.py: An efficient program for end-state free energy calculations. *J Chem Theory Comput.* 2012;8:3314–3321.

[40] Sitkoff D, Sharp KA, Honig B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J Phys Chem [Internet].* 1994;98:1978–1988. Available from: <http://pubs.acs.org/doi/abs/10.1021/j100058a043>.

[41] Akher FB, Farrokhzadeh A, Olotu FA, et al. *Biomolecular Chemistry.* 2019;1176–1190.

[42] Lipinski CA, Lombardo F, Dominy BW, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 2012;64:4–17.

[43] Daina A, Michielin O, Zoete V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep.* 2017;7.

- [44] Struck S, Schmidt U, Gruening B, et al. Toxicity versus potency: elucidation of toxicity properties discriminating between toxins, drugs, and natural compounds. *Genome Inf.* 2008;20:231–42.
- [45] Veber D, Johnson S, Cheng H, et al. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem.* 2002;45:2615–23.
- [46] Gademann K, Sieber S. Chemical interference of biological systems with natural products. *Chim.* 2011;65:835–8.
- [47] Remko M, Boháč A, Kováčiková L. Molecular structure, pKa, lipophilicity, solubility, absorption, polar surface area, and blood brain barrier penetration of some antiangiogenic agents. *Struct Chem.* 2011;22:635–48.
- [48] Remko M. Molecular structure, lipophilicity, solubility, absorption, and polar surface area of novel anticoagulant agents. *J Mol Struct Theochem.* 2009;916:76–85.

CHAPTER 8

Published Article

The Dual-Targeting Activity of The Metabolite Substrate of Para-Amino Salicylic Acid In The Mycobacterial Folate Pathway: Atomistic And Structural Perspectives

Clement Agoni^a, Pritika Ramharack^a, Eliasu Y. Salifu^a, Mahmoud E. S. Soliman^{a*}

^aMolecular 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

<https://soliman.ukz.ac.za>

ABSTRACT

Therapeutic targeting of folate biosynthetic pathway has recently been explored as a viable strategy in the treatment of tuberculosis. The metabolite substrate of Para-amino salicylic acid (PAS-M) reportedly dual-targets dihydrofolate reductase (DHFR) and flavin-dependent thymidylate synthase (FDTS), two essential enzymes in folate biosynthetic pathway. However, the molecular mechanisms and structural dynamics of this dual inhibitory activity of PAS-M remain elusive.

Molecular dynamics simulations revealed that binding of PAS-M towards DHFR is characterized by a recurrence of strong conventional hydrogen bond interactions between a peculiar DHFR binding site residue (Asp27) and the 2-amino-decahydropteridin-4-ol group of PAS-M. Similarly, the binding of PAS-M towards FDTS also involved consistent strong conventional hydrogen bond interactions between some specific residues (Tyr101, Arg172, Thr4, Gln103, Arg87 and Gln106) and the 2-amino-decahydropteridin-4-ol group, thus establishing the cruciality of the group. Structural dynamics of the bound complexes of both enzymes revealed that, upon binding, PAS-M is anchored at the entrance of hydrophobic pockets by a strong hydrogen bond interaction while the rest of the structure gains access to deeper hydrophobic residues to engage in favorable interactions. Further analysis of atomistic changes of both enzymes showed increased C- α atom deviations as well as an increase in C- α atoms radius of gyration consistent with structural disorientations. These conformational changes possibly interfered with the biological functions of the enzymes and hence their inhibition as experimentally reported. Structural insights provided could open up a novel paradigm of structure-based design of multi-targeting inhibitors of biological targets in the folate biosynthetic pathway towards tuberculosis therapy.

Keywords: *Mycobacterium tuberculosis*, tuberculosis, dihydrofolate reductase, flavin-dependent thymidine synthase, folate biosynthetic pathway, Molecular dynamics simulation

1.0 INTRODUCTION

As the leading cause of death from a single infectious disease, Tuberculosis (TB) is caused by *Mycobacterium Tuberculosis* (Mtb) [1]. In 2017, TB caused an estimated range of 1.2–1.4 million deaths among HIV-negative people and an additional 266000–335000 deaths from HIV-positive co-infected TB patients [1]. In a high TB burden country such as South Africa, the high mortality rates associated with TB is compounded by the co-infection with human immunodeficiency virus (HIV) [1–3]. In 2017 alone, South Africa contributed about 3% of the total global TB cases [1,4].

In other to curb this global menace, several chemotherapeutic options against the mycobacterium have been widely explored over the years with reasonably high success rates towards treatment and cure of tuberculosis [5–7]. Nonetheless, these efforts have been challenged by treatment failure, poor adherence and the development of drug-resistant strains largely due to the long duration of the treatment regimen of a minimum of six months [8–10].

Drug resistance in TB therapy, in particular, has contributed massively to the associated high mortality rates of TB such that in 2017 a range of 483000–639000 people developed rifampin resistant TB (RR-TB) alone [1,11–13]. With a high incidence of resistance in recent years, there have been reports of a subsequent relapse of the disease as well [14,15]. This has therefore led to a rise in the search for new therapeutic alternatives possibly targeting new enzymatic biomolecules implicated in various essential pathways. One of such high-value drug target therapeutic alternatives is the bacterial folate cycle, which is utilized in a wide range of bacterial metabolic processes, including DNA replication and repair [16].

During the folate biosynthetic pathway, dihydrofolate reductase (DHFR) is utilized as a catalyst to convert dihydrofolate to tetrahydrofolate. The tetrahydrofolate is acted on by serine hydroxymethyl transferase (SHMT) to form N5, N10 Methylene tetrahydrofolate. The tetrahydrofolate is further utilized as a precursor for thymidine synthesis. During this process, thymidylate synthase (TSase) enzyme converts N5,N10-methylene-tetrahydrofolate (MTHF) and 2'-deoxyuridine-5'-monophosphate (dUMP) to 2'-deoxythymidine-5'-monophosphate (dTMP) and dihydrofolate (DHF). The dTMP is then utilized in bacterial DNA synthesis. Subsequent to dTMP production, the DHF can then be recycled through DHFR in the folate pathway mentioned above [17] (Figure 8.1). Interestingly, it was also found that flavin-dependent thymidylate synthase (FDTS) also allowed for dTMP production by converting DHF to MTHF. This provided a dual role by producing dTMP, as well as reducing DHF to MTHF [17].

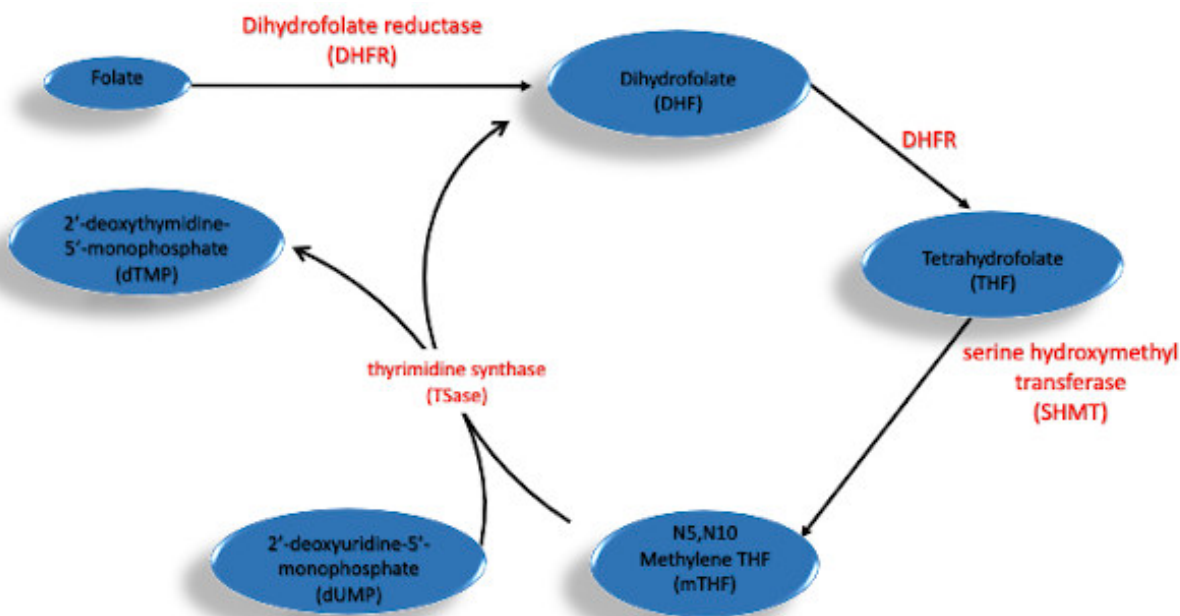


Figure 8.1: Schematic of the folate biosynthetic pathway

Of the plethora of studies directed toward identifying inhibitors against Mtb replication, Para-amino salicylic acid (PAS) is the only effective antifolate that is currently in use. The use of PAS in combination therapy has been evidenced to potentiate the anti-bacterial properties of isoniazid and streptomycin, especially against Mtb [18]. A Recent study by Haijan *et al* (2019) identified PAS as one of the very few drugs that are currently used in drug-resistant tuberculosis and further established its use as an alternative substrate in folate biosynthesis. Subsequent to PAS metabolism, the substrate is converted to PAS-M, which has shown inhibitory characteristics against both DHFR, although it exhibited poor binding affinity [19]. Haijan *et al* further explored the possibility of PAS-M binding to FdTS due to the structural similarities between PAS-M and mTHF, since mTHF is able to bind to FdTS. Their report through differential scanning calorimetry (DSC) revealed that PAS-M increases in melting temperature (T_m) stability implying a stabilizing effect of PAS-M and therefore and hence confirming its binding to Mtb FdTS. This confirmed dual-enzymatic binding portrayed by PAS-M may be used to overcome Mtb drug resistance by inhibition of more than one rate-limiting enzyme. It may also allow decreased adverse drug interactions due to combination therapy against tuberculosis, as well as HIV/tuberculosis co-infections.

Although experimental studies have defined PAS-M as an inhibitory lead molecule against DNA replication and repair of Mtb, there is missing evidence on the structural mechanism by which PAS-M inhibits the above-mentioned enzymes. To further progress through the rational drug design process, as well as to allow for the optimization of this lead molecule, the structural inhibitory mechanism of the drug is required. In this study, we therefore employ molecular dynamics simulations (MD) coupled with binding free energy calculations, to provide additional molecular and structural insights associated with the mechanistic binding of PAS-M towards both

DHFR and FDTS. We also elucidate the recognition mechanisms that characterize the selective binding of PAS-M towards mycobacterial DHFR and FDTS. Molecular dynamics simulations is a reliable computational technique extensively explored to mimic the dynamical events of molecular systems as a function of time, with a molecular description of all biological components (ligand and protein) in a solvent space [20]. Findings from this study will also provide important atomistic and structural perspectives that could aid in the design of novel multi-targeting TB therapy.

2.0 COMPUTATIONAL METHODOLOGY

2.1 System preparation and MD simulations

The X-ray crystal structures of DHFR and FDTS were retrieved from the RSCB Protein Data Bank with codes 6DDW and 2DQ7 respectively [21,22]. These structures were then prepared for molecular dynamics (MD) simulations using UCSF Chimera [23] and the Molecular Molegro Viewer (MMV)[24]. For the purpose of this study all nonstandard molecules co-crystallized with the enzyme structures were removed in preparation for the MD simulation to reduce computational cost and missing residues were added using Modeller, a structural refinement tool incorporated in UCSF Chimera [25]. The 3D structure of PAS-M was retrieved from the PDB structure 6DDW in which it was co-crystallized with DHFR and prepared for molecular docking into the binding of 2DQ7 and subsequently MD simulation. Molecular docking of PAS-M into the FDTS binding pocket was performed using AutoDock Vina [26] incorporated in UCSF Chimera with default docking parameters. In preparing for the molecular docking, PAS-M was supplemented with Gasteiger charges followed by non-polar hydrogen bonds which were merged to the carbon atoms prior to docking. The binding site was defined using coordinates obtained from co-crystalized

inhibitor with FDTS. A grid box with a spacing of 1 Å and size of 11.27 × 11.08 × 8.88 pointing in x=-2.03, y=7.21 and z=-11.06 directions). The best-docked pose for each ligand was then subjected to 100ns molecular dynamic simulations. Altogether, 4 systems comprising of unbound and PAS-M bound forms of DHFR and FDTS were subjected to MD simulations using the Amber18 software package [27]. Protein optimization and explicit solvation were carried out using the integrated LEAP module while the AMBER FF14SB forcefield was employed to define protein parameters[28]. The systems were partially minimized for 2500 steps with a restraint potential of 500kcal/mol Å² followed by full minimization of 10000 steps. The systems were gradually heated from 0-300K using a Langevin thermostat in a canonical ensemble (NVT) [29]. Equilibration was also carried out without restraints at a temperature of 300k in an NPT ensemble for 1000ps while atmospheric pressure was maintained at 1bar using the Berendsen barostat [30]. This was followed by MD production runs of 200ns for each system during which the SHAKE algorithm was used to constrict all atomic hydrogen bonds [31]. The integrated CPPTRAJ and PTRAJ modules [32] of AMBER18 were used to analyze resulting coordinates and trajectories while obtained data were plotted using Microcal Origin software [33]. UCSF Chimera was also used to visualize and analyze structural events. These are in accordance with our in-house MD simulation protocol that has been previously reported [34,35].

2.2 Binding free energy calculations

In order to estimate the binding interactions of PAS-M to DHFR and FDTS, binding free energy calculations were carried out using the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method [36,37]. This approach has been widely employed in several reports [38–

41]and has proven to be reliable in measuring binding free energies involved in protein-ligand complex formation. Moreover, MM/PBSA is mathematically represented as follows:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}} \quad (2)$$

$$G_{\text{sol}} = G_{\text{GB/PB}} + G_{\text{SA}} \quad (3)$$

$$G_{\text{SA}} = \gamma \text{SASA} \quad (4)$$

Where van der Waals and electrostatic interactions are represented as E_{vdw} and E_{ele} while E_{gas} denote gas-phase energy and E_{int} as internal energy. The solvation free energy, denoted by G_{sol} represents. G_{sol} is obtained by a summation of the polar and nonpolar contribution states. To obtain the polar solvation contribution, $G_{\text{GB/PB}}$, the GB/PB equation solved. Using a 1.4Å water probe radius, the nonpolar solvation contribution, G_{SA} , was calculated from the solvent accessible surface area (SASA). Per-residue decomposition analyses were also carried out to estimate individual energy contribution of binding site residues to the stabilization and affinity of masitinib. This could provide more insights into the basis of the pan-inhibition exhibited by masitinib since high residual energy contributions could depict crucial residues.

3.0 RESULTS AND DISCUSSION

3.1 Assessing structural stability among the simulated models

It was important to ensure that the unbound and bound systems converged and were structurally stable during the 100ns MD production run to avoid disrupted motions and simulation artifacts. Therefore, C- α Root Mean Square Deviation (RMSD) was employed to measure structural stability across the studied systems. This parameter is used to estimate deviations within the

backbone atoms of the respective protein forms. A high RMSD value depicts an increased deviation in backbone atoms while on the other hand, a low RMSD value implies reduced motions in backbone atoms and could, in turn, explain structural stability[42–44]. The RMSD plots for DHFR and FDTS systems showed convergence at the beginning of the MD simulation run until ~50ns where distinct separation in atomistic motions occurred among PAS-M-bound and unbound enzyme forms (Figure S1). As shown in figure S1, the binding of PAS-M to DHFR was characterized by an increase in deviation among the backbone atoms (1.65Å) relative to its unbound conformation which exhibited relative structural stability with a lower average RMSD of 1.29Å. Presumably, disruption of the backbone atoms in DHFR upon inhibitor binding could highlight the mechanistic inhibitory activity of PAS-M, since an induced loss of structural integrity in a protein correlate with loss of functionality [45–47]. In contrast to its influence of increase C- α atom deviation on DHFR, the binding of PAS-M on FDTS was characterized by a decrease in C- α atom deviation with an average RMSD of 1.65Å relative to higher average RMSD of 3.30Å. It could therefore be inferred that although the binding of PAS-M generally induced stabilized systems with average RMSD values less than 2Å, the was contrasting effect on the stability of both DHFR and FDTS which could be attributed to difference in amino acids sequences of the enzymes.

3.2 Exploring the mechanism of PAS-M binding towards DHFR and FDTS

The ability of a drug molecule to elicit its therapeutic (inhibitory) function is centered on its activity at the binding site of its target protein. Moreover, inter-molecular interactions with catalytic/allosteric site residues contribute significantly to ligand selectivity, binding affinity and stabilization. Therefore, it was essential to investigate the binding of PAS-M to DHFR and FDTS in order to understand the molecular basis of its dual inhibition. Firstly, we estimated binding free

energies involved in the interaction of PAS-M with the two enzymes using the MM/PBSA-based approach. Our results revealed that PAS-M bound favorably with DHFR and FDTS with ΔG values of -41.46kcal/mol and -20.72kcal/mol respectively (Table 8.1). The observed prominent difference in the estimated ΔG values could indicate dissimilar modes of binding that possibly underlie the observed difference in the impact of PAS-M binding on the respective stability of both enzymes. This may likewise be attributed to the varying prominent difference in the amino acid sequences of both enzymes. In addition, relatively high electrostatic and van der Waals interactions were involved in the binding of PAS-M towards DHFR could have accounted for its higher binding affinity relative to FDTS (Table 8.1).

Table 8.1: MM/PBSA-based binding free energy profile of PAS-M

Systems	Energy components (kcal/mol)				
	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
PAS-M-DHFR	-46.69 ±0.11	-77.73±0.49	-124.43±0.43	82.97±0.27	-41.46±0.20
PAS-M-FDTS	-35.50± 0.98	-37.56 ±0.61	-73.06±0.52	52.34±0.36	-20.72±0.19

ΔE_{ele} = electrostatic energy; ΔE_{vdW} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG = gas phase free energy.

3.3 Probing the interaction dynamics of PAS-M at DHFR and FDTS binding site

Furthermore, molecular visualization was used to obtain insights into the activity of PAS-M at the respective binding pockets of DHFR and FDTS with respect to its affinity, stabilization and dual

inhibitory activity. Also, this was fundamental to the identification of key binding site residues and their corresponding interactions with PAS-M. As shown in Figure 8.2 and 8.3, certain residues of the hydrophobic deep pocket elicited strong interactions with PAS-M, which accounted for its stability and high affinity towards both enzymes. These interactions vary from strong hydrogen bond interactions to general hydrophobic interactions.

To understand the interaction dynamics of PAS-M at the binding pockets of DHFR and FDTS, we noted certain residues that interacted consistently, steadily and intermittently across the 100ns MD simulation time. These modes of interactions could relatively provide necessary details and insights into the molecular basis and mechanisms of PAS-M selectivity and high-affinity binding towards the two proteins. As shown in figure 8.2, among all interactions caused by PAS-M-binding towards DHFR, there was a recurrence of certain residues that consistently formed strong interactions with PAS-M during the simulation. Notable amongst these residues is Asp27 which consistently maintained a strong hydrogen bond interaction with the amino group of the 2-amino-decahydropteridin-4-ol group of PAS-M. Other hydrogen bonded residues observed included; Arg60 at 10ns and Arg23 at 100ns. These hydrogen bond interactions in addition to the numerous hydrophobic interactions with other binding pockets residues could have collectively contributed to the favorable binding free energies calculated. Moreover, PAS-M was anchored at the entrance of the hydrophobic pocket by a strong hydrogen bond with Arg60 at 10ns and Arg23 at 100ns. This anchorage over the simulation period could allowed for its steady interactions of PAS-M with other hydrophobic residues as observed while also permitting its deep access of other hydrophobic residues. Access to the deep hydrophobic pocket allowed for favorable interactions with other hydrophobic residues which could have contributed to the total binding free energy of PAS-M towards DHFR. Altogether, these strong hydrogen bond interaction at the entrance of the

hydrophobic pocket and with the deeply embedded residues could suggestively have accounted for favorable affinity binding of PAS-M towards DHFR.

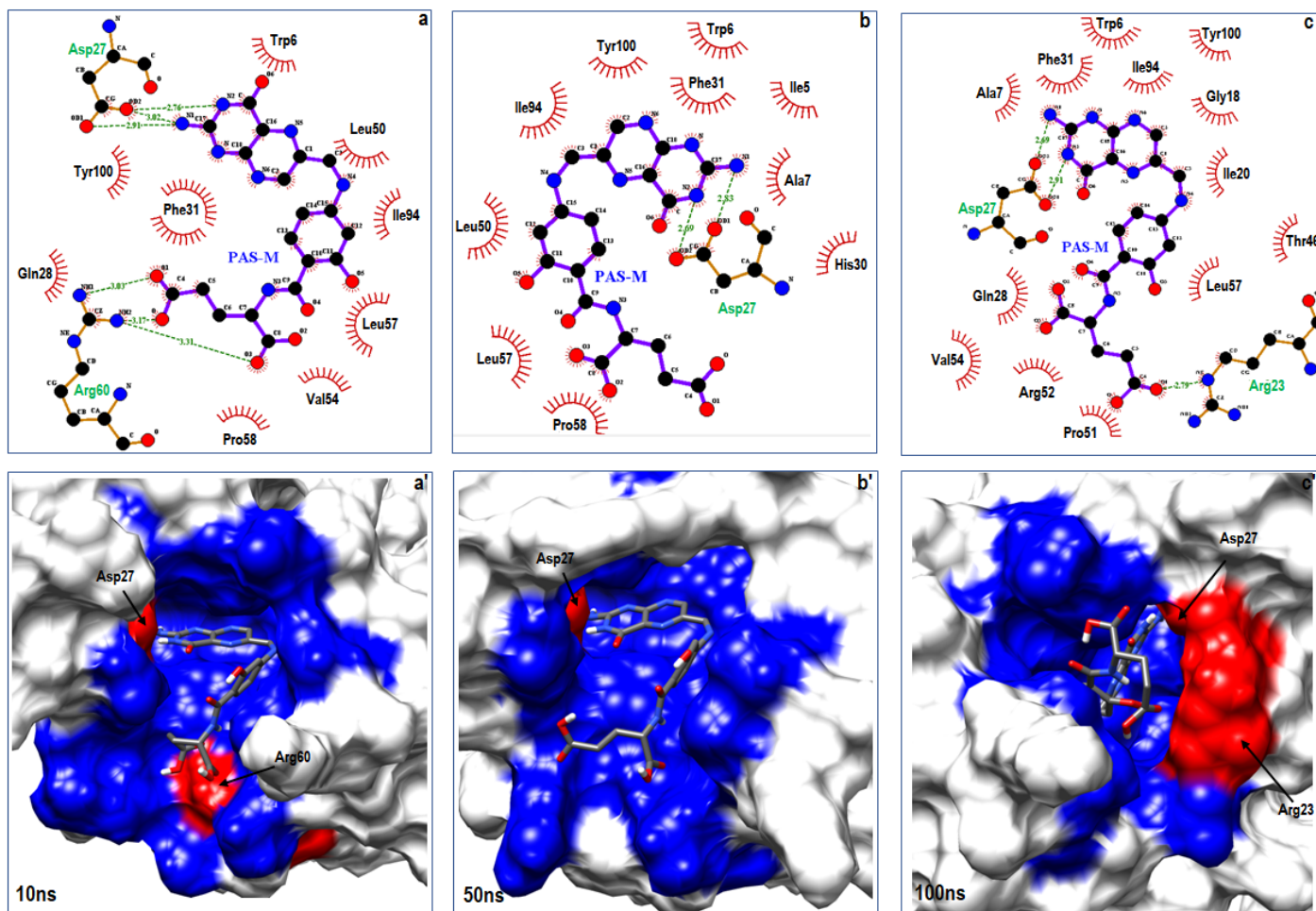


Figure 8.2: Structural positioning and orientation of PAS-M at the hydrophobic grooves of DHFR depicting residues within the binding pockets at 10 ns [a], 50 ns [b] and 100 ns [c]. Hydrophobic surface representations of PAS-M-bound within the DHFR binding pockets (blue), highlighting hydrogen bond interacting residues at 10 ns [a¹], 50 ns [b¹] and 100 ns [c¹].

Similarly, the binding of PAS-M towards FDTS was also characterized consistent strong hydrogen bond interactions with some specific residues over the course of the simulation. Notable amongst these residues are Tyr101, Arg172, Thr4, Gln103, Arg87 and Gln106. Tyr101 in particular recurred throughout the simulation period. Similarly, PAS-M is anchored at the entrance of hydrophobic pocket by strong hydrogen bond interactions while the rest of the compound gains

access to deeper hydrophobic residues for corresponding favorable interactions. These consistent and steady interactions with these residues could have equally accounted for the favorable binding free energies towards FDTS as estimated.

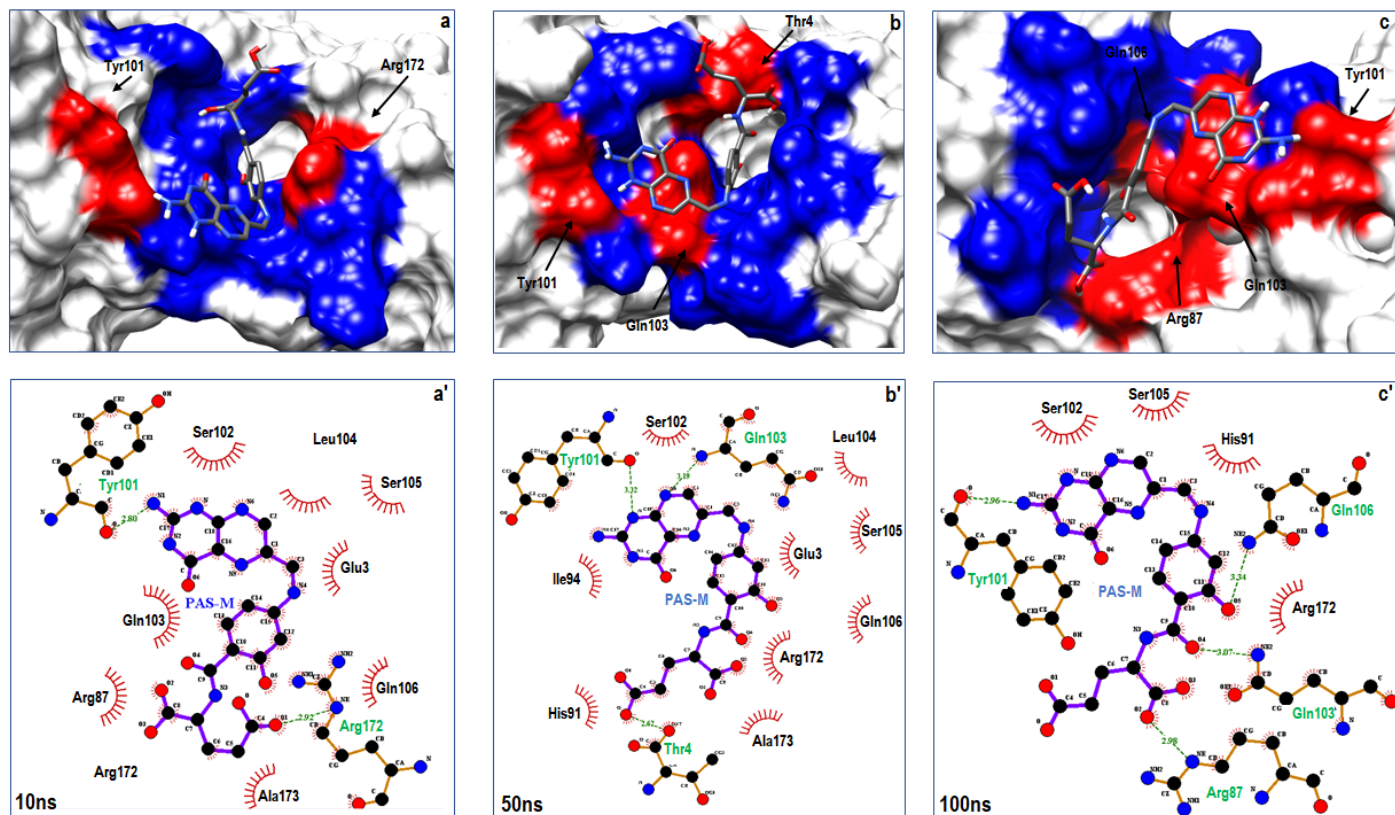


Figure 8.3: Hydrophobic surface representations of PAS-M-bound within the FDTS binding pockets (blue), highlighting hydrogen bond interacting residues at 10 ns [a], 50 ns [b] and 100 ns [c]. Structural positioning and orientation of PAS-M at the hydrophobic grooves of FDTS depicting residues within the binding pockets at 10 ns [a¹], 50 ns [b¹] and 100 ns [c¹].

Taken together, residues that consistently interacted with PAS-M towards the binding of both DHFR (Arg27) and FDTS (Tyr101) over the simulation period were shown to consistently interact with the amino moiety on the 2-amino-decahydropteridin-4-ol group of PAS-M. This could imply that interactions of this amino group could be crucial for high-affinity binding, and stability of PAS-M in the hydrophobic binding pockets of both enzymes and could as well account for the dual binding activity of PAS-M towards DHFR and FDTS.

The critical roles of the interactions engaged in by the amino moiety on the 2-amino-decahydropteridin-4-ol group were further revealed *via* per-residue decomposition analyses, which provided a quantitative insight into the energy contributions of individual residues that interacted with PAS-M at the DHFR and FDTS binding pockets over the 100ns MD simulation duration, particularly Arg27 and Tyr101 respectively. As presented in figure 8.4, residues which formed the strongest bonds contributed the highest energies towards the binding PAS-M to either of the enzymes as expected. In the PAS-M-DHFR complex as shown figure 3a, residues Asp27, Arg60 and Arg23 contributed the highest energies towards binding with total energies of -58.07kcal/mol, -54.29kcal/mol and -166.07kcal/mol respectively. This reaffirms the prominence of these residues in the inhibition of DHFR. Likewise, as shown in figure 3a', residues, Tyr101, Arg172, Gln103, Arg87 and Glu3 contributed the most as well towards the binding of FDTS with total energies of -170.37kcal/mol, -171.66kcal/mol, -43.20kcal/mol, -164.41kcal/mol and -103.88kcal/mol. Also, it was revealed that binding site residues in both DHFR and FDTS also contributed significant electrostatic energies towards the total binding of PAS-M which could accounted for its overall favorable binding in both pockets as estimated.

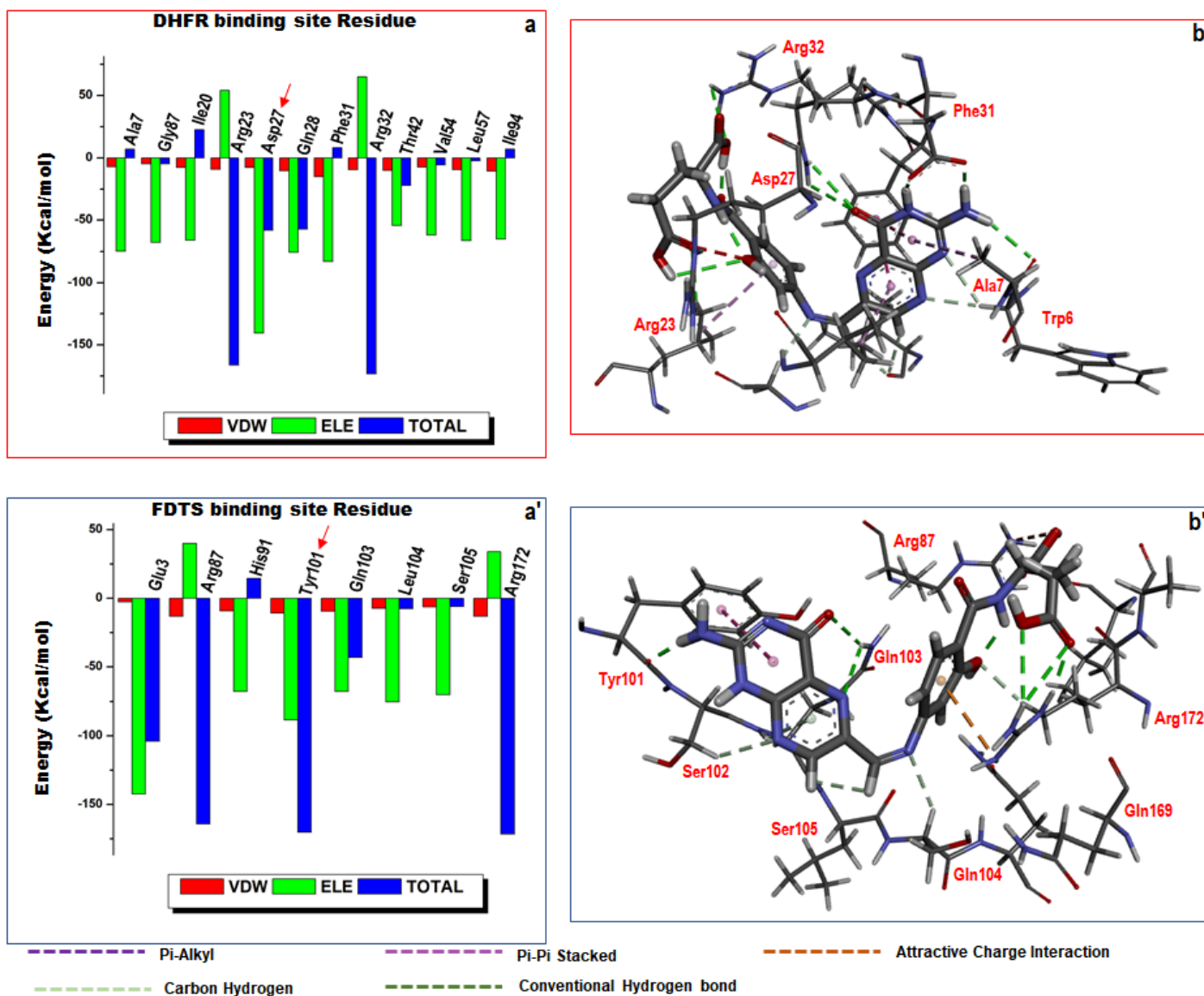


Figure 8.4 : Per-residue decomposition plots showing individual energy contributions to the binding and stabilization of PAS-M at the hydrophobic grooves of DHFR [a] and FDTS [a']. This reveals that energy contributions were highest in residues that formed hydrogen bond interactions with PAS-M. Corresponding inter-molecular interactions are shown in DHFR [b] and FDTS [b'].

3.4 Structural dynamics of hydrophobic deep pockets upon PAS-M binding

In furtherance to the peculiarities of the dual acting activity of PAS-M towards DHFR and FDTS as revealed in the corresponding residue interaction dynamics over the simulation period, we proceeded to investigate the dynamics of the hydrophobic deep pockets with respect to the dual and favorable binding of PAS-M. In other words, while we first investigated the dynamics of PAS-M at the respective binding pockets of both enzymes, we further explored the dynamics of their hydrophobic deep pocket residues upon the binding of PAS-M across the 100ns MD simulation time. This investigative approach necessitated the use of metrics such as C- α root mean square deviation (RMSD) and C- α radius of gyration (RoG) to describe the stability and compactness among hydrophobic pocket residues. According to previous studies, the stability and compactness of these residues could influence their corresponding interactions with bound inhibitors and hence can be employed to describe per-residue fluctuations across a protein's secondary structure while C- α RoG can be used to predict structural compactness, perturbations and atomistic mobility [35,46,48]. As shown on figure 8.5, the binding of PAS-M towards DHFR increased the deviation of C- α atoms of binding site residues within DHFR with an average RMSD of 2.57Å relative to the unbound conformation which exhibited a lower average RMSD of 1.29Å. Similarly, the binding of PAS-M towards FDTS also induced an increased deviation of C- α atoms deviation of active site residues with an average RMSD of 8.09Å relative to the unbound conformation which also exhibited an average RMSD of 3.43Å. It can therefore be implied that the binding of PAS-M to both enzymes is characterized by increased instability of the binding site residues which could have contributed to its favorable inhibition of both enzymes as experimentally reported [19]. Likewise, the instability of the binding site residues due to the inhibitory presence of PAS-M have distorted the original conformation of the enzymes required for their biological functions.

Presumably, systemic structural C- α atoms deviation of active site could highlight a possible mechanism by which PAS-M exhibit its dual-inhibitory activities against these enzymes, that in turn result in the disruption of their biological activities.

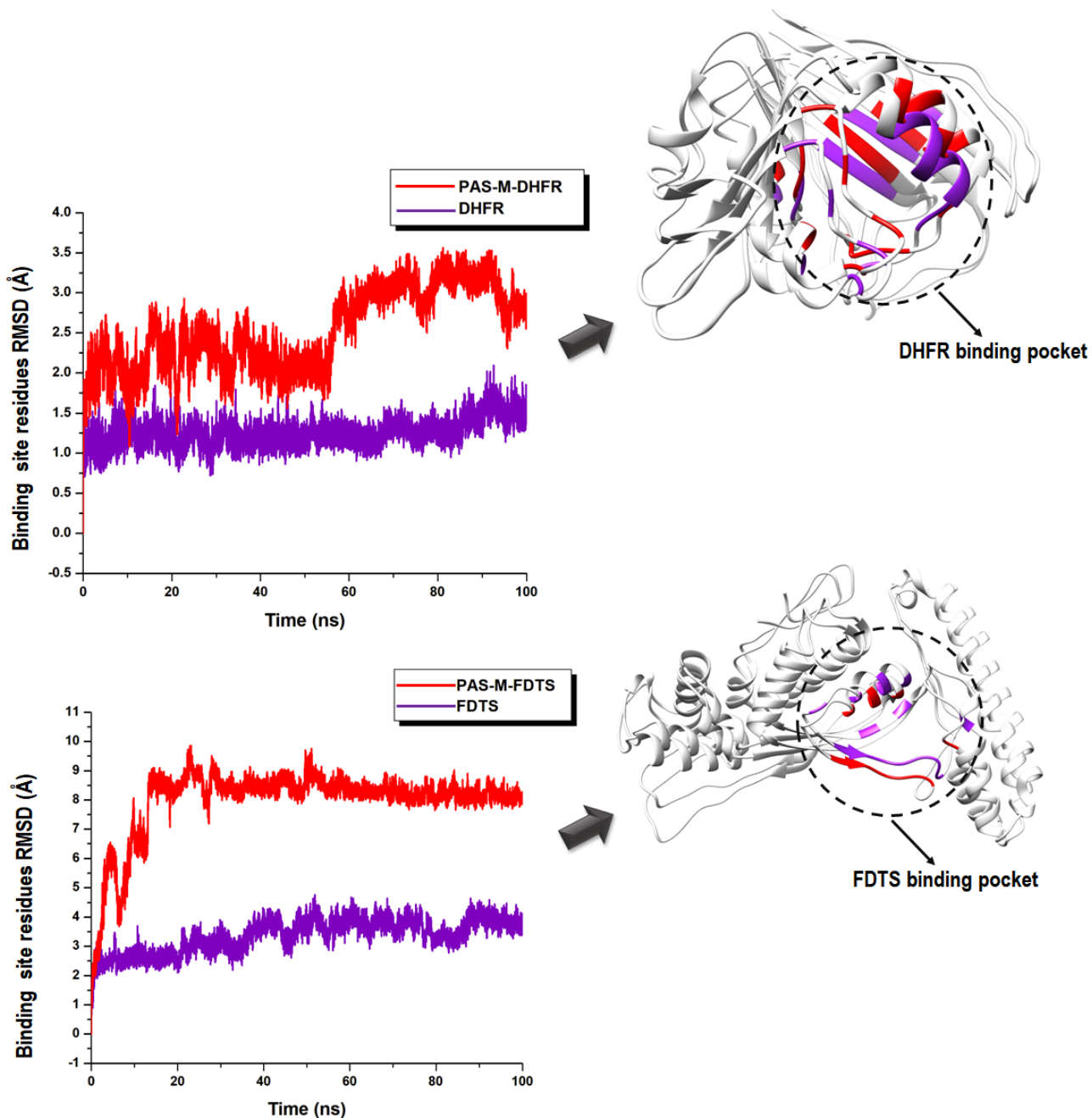


Figure 8.5: Comparative C- α RMSD plots showing the degree of instability and disruption induced by PAS-M at the binding pocket of DHFR and FDTS. Also shown is the binding site superposition of unbound (violet) and PAS-M-bound (red) forms of the two enzymes.

This presumption was further verified by calculating the C- α radius of gyration, which is able to depict the structural compactness of the binding pockets residues in both DHFR and FDTS across the MD simulation time of 100ns. As shown in figure 8.6, binding site residues of both enzymes exhibited increased radius of gyration of C- α atoms with an average RoG of 9.82Å and 9.65Å for DHFR and FDTS respectively. Active site residues of the unbound DHFR and FDTS exhibited lower average RoG of 8.24Å and 8.26Å. It could therefore imply that the binding of PAS-M also decreased the compactness of these residues, consistent with the increased deviation of C- α atoms as estimated relative to the unbound conformations.

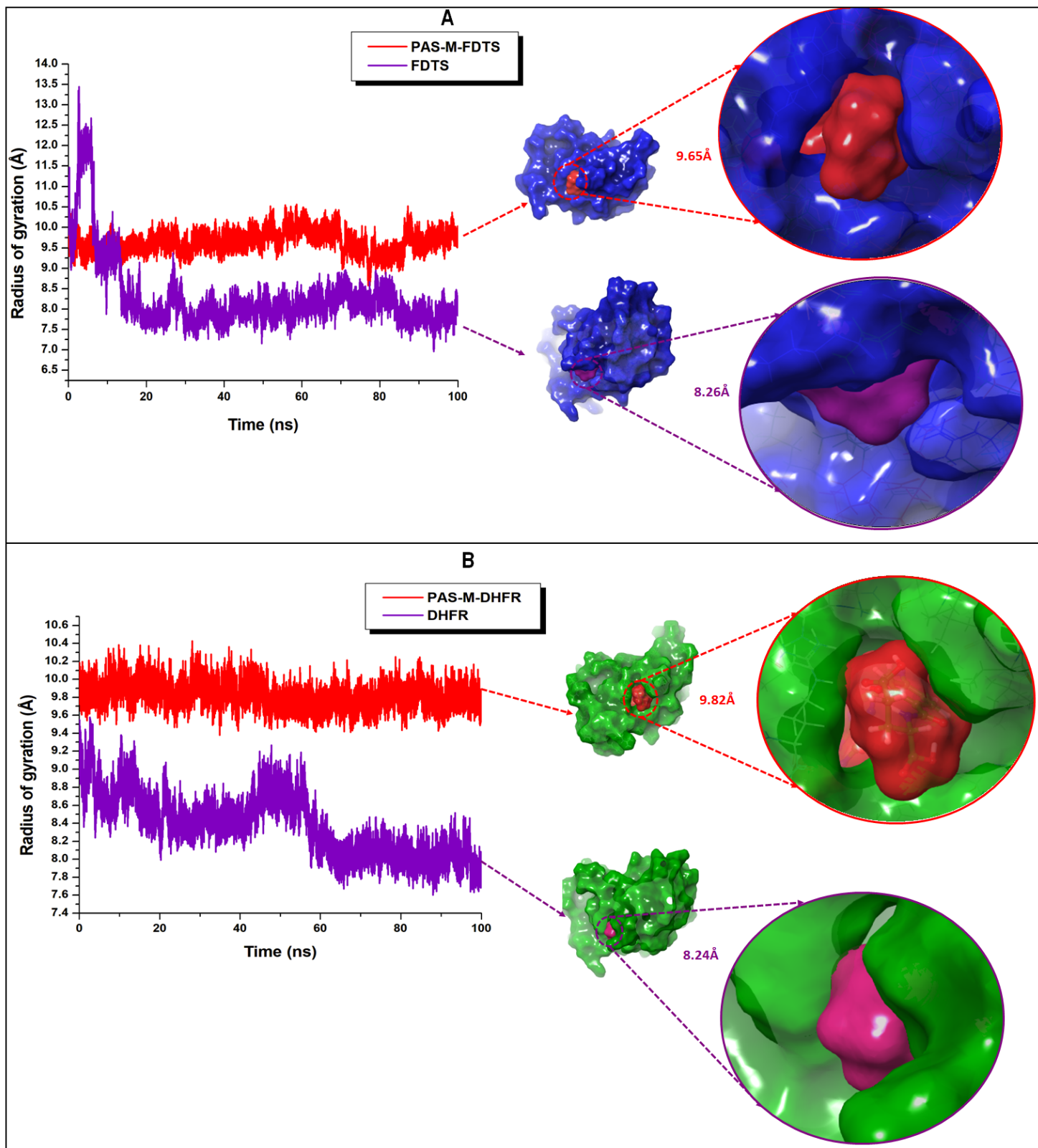


Figure 8.6: A) Comparative compactness of PAS-M binding site residues of FDTS in the presence and absence of PAS-M. Binding site residues in the unbound FDTS were more compact relative to the PAS-M bound conformation with averagely lower RoG of 8.26 Å. B) Comparative compactness of PAS-M binding

site residues of DHFR in the presence and absence of PAS-M. Binding site residues in the unbound DHFR were more compact relative to the PAS-M bound conformation with averagely lower RoG of 8.24 Å

Taken together, we could suggest that PAS-M selectively targets binding pockets residues of both enzymes and disrupts their original conformation by increasing their C- α atom deviations as well as their C- α atoms radius of gyration. These conformational changes could in turn have interfered with the biological functions of these enzymes and hence their inhibition as experimentally reported [19].

4.0 CONCLUSION

Employing advanced computational techniques and MD simulations, this study aims to provide molecular perspectives into the dual inhibitory mechanisms of PAS-M towards DHFR and FDTS which are crucial in the folate biosynthetic pathway. Differential binding of PAS-M to these enzyme targets was measured using the MM/PBSA method, which revealed favorable interactions with ΔG values of -41.46kcal/mol and -20.72kcal/mol towards DHFR and FDTS respectively. The favorable binding free energies as estimated confirms the possible inhibitory activity of PAS-M towards DHFR and FDTS. Investigations into the interaction dynamics of PAS-M at the binding pockets of DHFR and FDTS revealed crucial structural insights that could have favored its dual binding activity. The interaction of PAS-M with DHFR was characterized by a recurrence of certain residues that consistently formed strong interactions with PAS-M during the simulation, notably Asp27 which consistently maintained a strong hydrogen bond interaction with the amino group of the 2-amino-decahydropteridin-4-ol group of PAS-M. Similarly, the binding of PAS-M towards FDTS involved consistent strong hydrogen bond interactions with some specific residues over the course of the simulation, notably, Tyr101, Arg172, Thr4, Gln103, Arg87 and Gln106.

Interestingly, upon binding to either enzyme, PAS-M is anchored at the entrance of hydrophobic pocket by a strong hydrogen bond interaction while the rest of the compound gains access to deeper hydrophobic residues for corresponding favorable interactions. Quantification of the energy contributions through per-residue energy decomposition analysis further revealed that, these same residues contributed the most towards total binding of PAS-M, further establishing their cruciality to enzyme inhibition. Analysis of the conformational changes of both enzymes upon PAS-M binding revealed increased C- α atom deviations as well as an increase in their C- α atoms radius of gyration. These conformational changes could in turn have interfered with the biological functions of these enzymes and hence their inhibition as experimentally reported. Identification of the recurring residues and their crucial roles could open up a novel paradigm of structure-based design of highly selective multiple targeting inhibitors of the foliate biosynthetic pathway in the treatment of tuberculosis, amidst the rise in TB drug resistance. A further exploration of the 2-amino-decahydropteridin-4-ol group of PAS-M could serve as a baseline for the design of new therapeutic agents in TB therapy.

5.0 REFERENCES:

- [1] World Health Organization. *Global Tuberculosis Report 2018*; **2018**.
- [2] Weld, E.; Dooley, K. State-of-the-Art Review of HIV-TB Coinfection in Special Populations. *Clin Pharmacol Ther*, **2018**, *104*, 1098–1109.
- [3] Bell, L.; Noursadeghi, M. Pathogenesis of HIV-1 and Mycobacterium Tuberculosis Co-infection. *Nat. Rev. Microbiol.*, **2018**, *16*, pages 80–90.
- [4] Pietersen, E.; Ignatius, E.; Streicher, E.M.; Mastrapa, B.; Padanilam, X.; Pooran, A.; Badri,

M.; Lesosky, M.; Van Helden, P.; Sirgel, F.A.; Warren, R.; Dheda, K. Long-Term Outcomes of Patients with Extensively Drug-Resistant Tuberculosis in South Africa: A Cohort Study. *Lancet*, **2014**, *383*, 1230–1239.

[5] Tiberi, S.; Munoz-Torrico, M.; Duarte, R.; Dalcolmo, M.; D'Ambrosio, L.; Migliori, G, B. New Drugs and Perspectives for New Anti-Tuberculosis Regimens. *Pumonology*, **2018**, *24*, 86–98.

[6] Davis, A.; Meintjes, G.; Wilkinson, R.J. Treatment of Tuberculous Meningitis and Its Complications in Adults. **2018**.

[7] Sotgiu, G.; Centis, R.; Migliori, G.B. Tuberculosis Treatment and Drug Regimens. **2015**, 1–12.

[8] Councils, R.E.A.-B.M.R. *Controlled Clinical Trial of Four Short-Course (6-Month) Regimens of Chemotherapy for Treatment of Pulmonary Tuberculosis*; **1974**; Vol. ii.

[9] Yee, D.; Valiquette, C.; Pelletier, M.; Parisien, I.; Rocher, I.; Menzies, D. Incidence of Serious Side Effects from First-Line Antituberculosis Drugs among Patients Treated for Active Tuberculosis. *Am J Respir Crit Care Med*, **2003**, *167*, 1472–1477.

[10] Al-humadi, H.W.; Al-saigh, R.J.; Al-humadi, A.W. Addressing the Challenges of Tuberculosis : A Brief Historical Account. **2017**, *8*, 1–10.

[11] Weis, S.; Slocum, P.; Blais, F.; King, B.; Nunn, M.; Matney, G.; Gomez, E.; Foresman, B. The Effect of Directly Observed Therapy on the Rates of Drug Resistance and Relapse in Tuberculosis. *N Engl J Med*, **1994**, *330*, 1179–1184.

[12] Ormerod, L. Multidrug-Resistant Tuberculosis (MDR-TB): Epidemiology, Prevention and Treatment. *Br Med Bull*, **2005**, *73–74*, 17–24.

[13] Nguyen, L. Antibiotic Resistance Mechanisms in M. Tuberculosis: An Update. *Arch*

Toxicol, **2016**, *90*, 1585–1604.

[14] Zhao, Y.; Xu, S.; Wang, L.; Chin, D.; Wang, S.; Jiang, G.; Xia, H.; Zhou, Y.; Li, Q.; Ou, X.; Pang, Y.; Song, Y.; Zhao, B.; Zhang, H.; He, G.; Guo, J.; Wang, Y. National Survey of Drug-Resistant Tuberculosis in China. *N Engl J Med*, **2012**, *366*, 2161–2170.

[15] Kurz, S.G.; Furin, J.J.; Medicine, S.; Bark, C.M. HHS Public Access. **2017**, *30*, 509–522.

[16] Minato, Y.; Thiede, J.M.; Kordus, L.; Mcklveen, E.J.; Turman, B.J.; Baughn, D. Mycobacterium Tuberculosis Folate Metabolism and the Mechanistic Basis for Para - Aminosalicylic Acid Susceptibility and Resistance. **2015**, *59*, 5097–5106.

[17] Choi, M.; Karunaratne, K.; Kohen, A. Flavin-Dependent Thymidylate Synthase as a New Antibiotic Target. *Molecules*, **2016**, *21*.

[18] Rengarajan, J.; Sasseti, C.M.; Naroditskaya, V.; Sloutsky, A.; Bloom, B.R.; Rubin, E.J. The Folate Pathway Is a Target for Resistance to the Drug Para-Aminosalicylic Acid (PAS) in Mycobacteria. *Mol. Microbiol.*, **2004**, *53*, 275–282.

[19] Hajian, B.; Scocchera, E.; Shoen, C.; Krucinska, J.; Viswanathan, K.; G-Dayanandan, N.; Erlandsen, H.; Estrada, A.; Mikušová, K.; Korduláková, J.; Cynamon, M.; Wright, D. Drugging the Folate Pathway in Mycobacterium Tuberculosis: The Role of Multi-Targeting Agents. *Cell Chem. Biol.*, **2019**, *26*, 781–791.e6.

[20] Salmaso, V.; Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front. Pharmacol.*, **2018**, *9*, 1–16.

[21] Williams, N.K.; Lucet, I.S.; Peter Klinken, S.; Ingley, E.; Rossjohn, J. Crystal Structures of the Lyn Protein Tyrosine Kinase Domain in Its Apo- And Inhibitor-Bound State. *J. Biol. Chem.*, **2009**, *284*, 284–291.

[22] Kinoshita, T.; Matsubara, M.; Ishiguro, H.; Okita, K.; Tada, T. Structure of Human Fyn

Kinase Domain Complexed with Staurosporine. *Biochem. Biophys. Res. Commun.*, **2006**, *346*, 840–844.

[23] Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.*, **2004**, *25*, 1605–1612.

[24] Kusumaningrum, S.; Budianto, E.; Kosela, S.; Sumaryono, W.; Juniarti, F. The Molecular Docking of 1,4-Naphthoquinone Derivatives as Inhibitors of Polo-like Kinase 1 Using Molegro Virtual Docker. *J. Appl. Pharm. Sci.*, **2014**, *4*, 47–53.

[25] Eswar, N.; Webb, B.; Marti-Renom, M.A.; Madhusudhan, M.S.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Protein Sci.*, **2007**, *Chapter 2*, Unit 2.9.

[26] Trott, O.; Olson, A.J. AutoDock Vina. *J. Comput. Chem.*, **2010**, *31*, 445–461.

[27] Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, III, T.E.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. AMBER 2018, **2018**.

[28] Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.*, **2015**, *11*, 3696–3713.

[29] Grest, G.S.; Kremer, K. Molecular Dynamics Simulation for Polymers in the Presence of

a Heat Bath. *Phys. Rev. A*, **1986**, *33*, 3628–3631.

[30] Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, a; Haak, J.R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, **1984**, *81*, 3684–3690.

[31] 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.*, **1977**, *23*, 327–341.

[32] Roe, D.R.; Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. **2013**.

[33] Seifert, E. OriginPro 9.1: Scientific Data Analysis and Graphing Software – Software Review. *J. Chem. Inf. Model*, **2014**, *54*, 1552.

[34] Abdullahi, M.; Olotu, F.A.; Soliman, M.E. Dynamics of Allosteric Modulation of Lymphocyte Function Associated Antigen-1 Closure-Open Switch: Unveiling the Structural Mechanisms Associated with Outside-in Signaling Activation. *Biotechnol. Lett.*, **2017**, *39*, 1843–1851.

[35] Olotu, F.A.; Soliman, M.E.S. From Mutational Inactivation to Aberrant Gain-of-Function: Unraveling the Structural Basis of Mutant P53 Oncogenic Transition. *J. Cell. Biochem.*, **2018**, *119*, 2646–2652.

[36] Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.*, **2011**, *51*, 69–82.

[37] Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol. Inform.*, **2012**, *31*, 114–122.

[38] Agoni, C.; Ramharack, P.; Soliman, M. Co-Inhibition as a Strategic Therapeutic Approach

to Overcome Rifampin Resistance in Tuberculosis Therapy: Atomistic Insights. *Future Med. Chem.*, **2018**, *10*, 1665–1675.

[39] Abdullahi, M.; Olotu, F.A.; Soliman, M.E. Allosteric Inhibition Abrogates Dysregulated LFA-1 Activation: Structural Insight into Mechanisms of Diminished Immunologic Disease. *Comput. Biol. Chem.*, **2018**, *73*, 49–56.

[40] Lawal, M.; Olotu, F.A.; Soliman, M.E.S. Across the Blood-Brain Barrier: Neurotherapeutic Screening and Characterization of Naringenin as a Novel CRMP-2 Inhibitor in the Treatment of Alzheimer's Disease Using Bioinformatics and Computational Tools. *Comput. Biol. Med.*, **2018**, *98*, 168–177.

[41] 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*, **2017**, 1–16.

[42] Maiorov, V.N.; Crippen, G.M. Significance of Root-Mean-Square Deviation in Comparing Three-Dimensional Structures of Globular Proteins. *J. Mol. Biol.*, **1994**, *235*, 625–634.

[43] Pitera, J.W. Expected Distributions of Root-Mean-Square Positional Deviations in Proteins. *J. Phys. Chem. B*, **2014**, *118*, 6526–6530.

[44] Galindo-Murillo, R.; Roe, D.R.; Cheatham, T.E. Convergence and Reproducibility in Molecular Dynamics Simulations of the DNA Duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta - Gen. Subj.*, **2015**, *1850*, 1041–1058.

[45] Wallnoefer, H.G.; Lingott, T.; Gutiérrez, J.M.; Merfort, I.; Liedl, K.R. Backbone Flexibility Controls the Activity and Specificity of a Protein-Protein Interface: Specificity in Snake Venom Metalloproteases. *J. Am. Chem. Soc.*, **2010**, *132*, 10330–10337.

[46] Karshikoff, A.; Nilsson, L.; Ladenstein, R. Rigidity versus Flexibility: The Dilemma of

Understanding Protein Thermal Stability. *FEBS J.*, **2015**, 282, 3899–3917.

[47] Rehn, A.; Moroni, E.; Zierer, B.K.; Toppel, F.; Morra, G.; John, C.; Richter, K.; Colombo, G.; Buchner, J. Allosteric Regulation Points Control the Conformational Dynamics of the Molecular Chaperone Hsp90. *J. Mol. Biol.*, **2016**, 428, 4559–4571.

[48] Lobanov, M.Y.; Bogatyreva, N.S.; Galzitskaya, O. V. Radius of Gyration as an Indicator of Protein Structure Compactness. *Mol. Biol.*, **2008**, 42, 623–628.

CHAPTER 9

Published Article

Halting ionic shuttle to disrupt the synthetic machinery – Structural and molecular insights into the inhibitory roles of Bedaquiline towards *Mycobacterium tuberculosis* ATP synthase in the treatment of tuberculosis

Elliasu Y. Salifu^a, Clement Agoni^a, Fisayo A. Olotu^a, Yussif M. Dokurugu^b, Mahmoud E. S.

Soliman^{a*}

^aMolecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

^bCollege of Pharmacy & Pharmaceutical Sciences, Florida Agricultural & Mechanical University, FAMU, Tallahassee, FL 32307, USA.

*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 8048, Fax: +27 (0) 31 260 7872

Abstract

Therapeutic targeting of the ATP machinery of *Mycobacterium tuberculosis* (*Mtb*) has recently presented a potent and alternative measure to halt the pathogenesis of tuberculosis. This has been potentiated by the development of Bedaquiline (BDQ), a novel small molecule inhibitor that selectively inhibits mycobacterial F₁F_o-ATP synthase by targeting its rotor c-ring, resulting in the disruption of ATP synthesis and a consequential cell death. Although, the structural resolution of the mycobacterial C₉ ring in complex with BDQ provided a first-hand detail of BDQ interaction at the c-ring region of the ATP synthase, there still remains a need to obtain essential and dynamic insights into the mechanistic activity of this drug molecule towards a crucial survival machinery of *Mtb*. As such, for the first time, we report an atomistic model to describe the structural dynamics that explicate the experimentally reported antagonistic features of BDQ in halting ion shuttling by the mycobacterial c-ring, using molecular dynamic (MD) simulation and MM/PBSA methods. Results showed that BDQ exhibited a considerably high ΔG while it specifically maintained high-affinity interactions with Glu65_B and Asp32_B, blocking their crucial roles in proton binding and shuttling which is required for ATP synthesis. Moreover, the bulky nature of BDQ induced a rigid and compact conformation of the rotor c-ring which impedes the essential rotatory motion that drives ion exchange and shuttling. In addition, the binding affinity of a BDQ molecule was considerably increased by the complementary binding of another BDQ molecule, which indicates that an increase in BDQ molecule enhances inhibitory potency against *Mtb* ATP synthase. Taken together, findings provide atomistic perspectives into the inhibitory mechanisms of BDQ coupled with insights that could enhance the structure-based design of novel ATP synthase inhibitors towards the treatment of tuberculosis.

Keywords: Tuberculosis, *Mycobacterium tuberculosis*, Bedaquiline, Molecular Dynamic Simulation, Ionophores.

1.0 Introduction

Tuberculosis (TB), as an infectious disease, has existed for a millennium but still remains a major global health concern. TB is caused by the bacillus, *Mycobacterium tuberculosis* (*Mtb*) and currently ranks above HIV/AIDS as the leading cause of death globally from a single infectious agent [1, 2]. The global incidence of TB in 2017 alone was estimated to about 10 million recorded cases in which the WHO South Asia region recorded the highest incidence of 44% followed by the WHO Africa region with about 25% of total global incidence. South Africa remains prominent in its contribution to total global TB burden, contributing about 3% in 2017 alone to the total global incidence. Nonetheless, there has been a steady decline in TB in many southern African countries, including South Africa between years 2010-2017. The co-infection of TB with HIV has also contributed immensely to its current global mortality rate. For instance, in 2017, over 50% of TB case were co-infected with HIV in Southern African countries. However, there has also been a steady decline death from TB-HIV co-infection as well in the last few years due to an increase coverage of antiretroviral therapy in high burden regions. The major impediment to the treatment of TB over the years remains the development of resistance of the mycobacterium to current therapeutic option. A suitable treatment regimen for TB majorly involves a combinatorial therapeutic approach that constitute at least four drugs namely isoniazid, rifampin, pyrazinamide and ethambutol [3]. Strains resistant to only one first-line anti-TB agents are categorized and mono-resistant TB, of which Rifampin resistance is the most common. On the other hand, resistance to both rifampicin and isoniazid is referred as multi-drug resistant TB (MDR-TB) and has been categorized as poly-resistant TB. MDR-TB cases are treated with second-line TB drugs which are usually more toxic and less effective as compared to the first-line agents [4, 5]. MDR-TB in which the mycobacterium also develops resistance to any fluoroquinolone and at least one of the three second line injectables (capreomycin, kanamycin and amikacin) is further referred to

as extensively drug resistant tuberculosis (XDR-TB). In recent years, the most virulent form of resistant TB has been classified as totally drug resistant TB (TDR-TB) in strains are resistant to all the first second line TB drugs [4, 6, 7]. Hence, there has been a continual surge in the identification of potent and highly effective therapeutic molecules that target crucial elements (such as protein) involved in the pathogenesis and virulence of *Mtb* [8, 9]. In this regards, targeting the rotor ring of *Mtb*'s F_1F_0 -ATP synthase (henceforth F_1F_0) has been explored over the years as a viable therapeutic option considering the crucial the role it plays in ATP synthesis of the mycobacterium [3]. The F_1F_0 -ATP synthase is a macromolecular, membrane-embedded protein complex in *MTB* that uses the transmembrane electrochemical ion (H^+ or Na^+) gradient to convert adenosine diphosphate (ADP) and inorganic phosphate (P_i) into adenosine triphosphate (ATP) by a rotary mechanism [10–13] during respiration. The membrane-embedded F_0 domain of the complex harbors the rotor ring of the F-type ATP synthase which usually consists of identical copies of c-subunits that form an hourglass-shaped cylinder with a central pore (c-ring) in the bacteria [14]. F_1F_0 -ATP synthase shuttles ions across the membrane and thereby powers the synthesis of ATP within the three catalytically active sites of the F_1 headpiece [10]. A report by Andries *et al.*, in 2005, identified bedaquiline (BDQ) , via phenotypic screening, as an antitubercular agent which was subsequently approved by the FDA in 2012 as a member of the chemical class of diarylquinolines to inhibit F_1F_0 in mycobacteria [3, 15]. The inhibitory activity of BDQ demonstrates that metabolism and energy generation is a promising new target space in inhibiting mycobacteria. By acting as an ionophore, BDQ blocks shuttling H^+ and K^+ ions across membranes [16]. The recently published structure of the c-ring from *Mycobacterium phlei* with bound BDQ suggested the selectivity of BDQ for only mycobacterial species as against other prokaryotic and eukaryotic organisms [17].

According to Preiss *et al.*, BDQ has a high specificity and affinity towards the mycobacterial c-ring of *M. phlei* where it binds in a lock and key fashion, with a virtually complete sequence conservation of this region, which suggests an identical surface profile and binding site geometry possibly in all mycobacteria, and most importantly, *M. tuberculosis* [17]. The binding of BDQ to the c-ring rotor of the F_o of mycobacterial ATP synthase results in an eventual inhibition of ATP synthesis leading to mycobacterial death [3, 17–20]. Although the crystal structure of F₁F₀ in complex with BDQ has been resolved experimentally, there has been little or no insights into the mechanisms, structural dynamics and conformational changes that accompanies the binding of BDQ. Moreover, these insights could be crucial to the development of novel small molecule inhibitors of mycobacterial F₁F₀ ATP synthase, most especially in the treatment of TB. Figure 9.1 shows a 3D representation of F₁F₀-ATP synthase subunit c in complex with BDQ.

The application of computational (*in silico*) methods in providing structural insights into the binding of small molecule inhibitors has gained prominence in the drug discovery process in recent years. Insights from computational reports can aid in the optimization of existing small molecule compounds to yield enhanced and improved novel small molecule inhibitors and eventually improve the overall efficiency of the drug discovery process [21, 22]. Computational approaches have been employed in the study of several conditions and for varying purposes including mutational studies and the identification of new inhibitors. In mutations studies computer aided techniques have aided in unravelling the structural dynamics of mutant proteins as well as revealing the impact of mutations on the activity of bound inhibitors [23–27] while techniques have focused on the identification of novel compounds from large pool of existing databases[28]. However, in the present report, molecular dynamics simulation coupled with other computational

approaches was employed to provide, for the first time, atomistic insights into the reported ionophoric features of BDQ relative to the inhibition of ATP synthesis in *Mtb*.

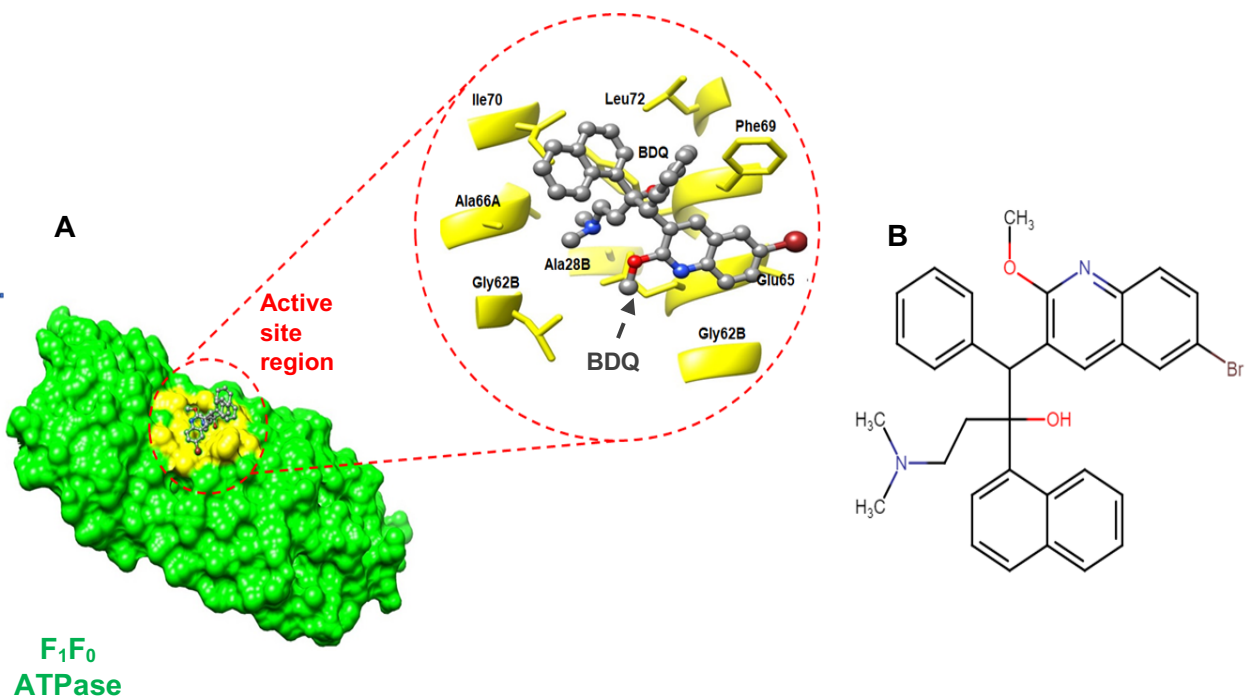


Figure 9.1: [A] Surface depiction of the mycobacterial c-ring F₁F₀-ATP synthase (**green**) bound by BDQ (**grey**). Also, active site and BDQ-interacting residues are depicted in yellow coloration and clearly shown in red inset. [B] 2D structure of BDQ showing constituent functional groups.

2.0 Computational methodology

2.1 System preparation

The X-ray crystal structure of the c-ring rotor (F₀ portion) of mycobacterial ATP synthase was obtained from RSCB Protein Data Bank with code 4V1F [17], which was already complexed with Bedaquiline (ID:5388906). This complex was then prepared for simulation using UCSF Chimera [29] and the Molecular Molegro Viewer (MMV) [30], a technical procedure that basically involves the removal of co-crystallized molecules not relevant to the study such as crystal waters. The

structure comprised of three chains A, B and C with BDQ bound to all three, however only the central region of the three BDQ molecules within the crystal structure was visible whereas parts of the other two molecules were partially visible. In this report our focus was restricted to the interaction of a single molecule of BDQ with the F_1F_0 structure. As such the two partially visible BDQ molecules were taken off leaving just the full visible molecule in the middle to reduce computational cost. Eventually, two systems which comprised of a BDQ-bound F_1F_0 and an unbound (apo) form of the ATP synthase were set up for the molecular dynamic (MD) simulation. An additional system was setup comprising two molecules of BDQ bound to F_1F_0 using UCSF Chimera-integrated AutoDock Vina module and the Molecular Molegro Viewer (MMV) to depict a dual BDQ-bound F_1F_0 ATP synthase. This was necessary to investigate the inhibitory prowess of dual BDQ molecules on ion-shuttling as compared to single-bound systems, a rationale that was aimed at providing important atomistic perspectives based on the report by Preiss *et al* in which they stated that an increased concentration of BDQ could result in the binding of the membrane exposed ion binding sites of the c-ring of ATP synthase by one or more molecules of BDQ with enhanced inhibitory potency [17].

2.2 Molecular dynamics (MD) simulations

This was performed using the GPU version of AMBER 14 with an integrated PMEMD module [31], according to standard simulation protocols, which has been previously adopted in previous studies and enumerated as follows [32–36]. Parameterization of the inhibitor was carried out using the ANTECHAMBER module wherein atomic partial charges (AM1BCC) gaff, using the bcc charge scheme were generated [37]. The FF14SB AMBER force field [38] was then used to parameterize the protein. Using the LEAP module, hydrogen atoms were also added, while the

entire system was neutralized by the addition of counter ions (Na^+ , Cl^-) followed by a subsequent generation of ligand, protein and complex topologies as well as parameter files. The systems were explicitly solvated with water using the TIP3P orthorhombic box size of 7\AA , which enclosed all atoms of the protein [39]. Before running the LEAP module, the `pdb4amber` script was executed to protonate the histidine residues at a constant pH (cpH). This was appointed to automatically modify the protein system for use with LEAP. Both complexes were minimized initially for 2000 minimization steps applying a restraint potential of 500kcal/mol and then fully minimized for another 1000 steps of steepest descent without restraint. These minimization steps were run on 2 computational nodes with 48 cores on a 48 hours wall-time each. Likewise, at these stages, a fixed protein algorithm was used to place restraints on the entire protein and hold it within the TIP3P water box. This was followed by the gradual thermalization of the systems with a temperature range of $0\text{-}300\text{K}$ for 50ps . Similar to initial minimization steps, thermalization was performed using 2 nodes, 48 cores and walltime of 48 hours. At this stage, restraints were placed on the ligands and proteins altogether to prevent their release from the TIP3P solvent box. Afterwards, the equilibration process was initialized using similar MPI run as indicated above and performed for 500ps while the temperature and pressure were kept constant at 300K and 1bar respectively using the Berendsen barostat [40]. This was followed by MD production runs of 100ns , which correlated with an `nstlim` of 50000 steps for each system during which the SHAKE algorithm [41] was used to constrict all atomic hydrogen bonds. Also, the `iwrap` algorithm was equated to 1 and used to hold the simulated systems in place due to thermodynamically induced atomic vibrations that characterize the production runs. Also, a cut off distance of 12\AA was set for non-bonded van der Waals interactions while the default PME method for summation carried out long-range adjustments for occurring electrostatic interactions. The MD simulation was initiated using a time

step of 1fs and coordinates saved at 1ps interval followed by subsequent analysis of trajectories using the integrated PTRJ and CPPTRAJ module [42]. Visualization of the complexes and data plots were carried out using the graphical user interface of UCSF chimera [43], and Microcal Origin analytical software [44] respectively. Time-scale dynamics and conformational events were visually depicted by retrieving selected trajectories and corresponding snapshots which were analyzed graphically on the graphical interface of LIGPLOT to generate ligand interaction plots. These were further complimented using graphical tools such as BIOVIA discovery studio to properly depict structural and molecular occurrences [45][46].

2.3 Binding Energy Calculations

The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) [47–50] method was used to calculate binding free energy. The binding free energy (ΔG_{bind}) was calculated from the following equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} + G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = \Delta G_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S, \quad (2)$$

where ΔG_{bind} is considered to be the summation of the gas phase and solvation energy terms less the entropy ($T\Delta S$) term

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} \quad (3)$$

ΔE_{gas} is the sum of the AMBER force field internal energy terms ΔE_{int} (bond, angle and torsion), the covalent van der Waals (ΔE_{vdw}) and the non-bonded electrostatic energy component (ΔE_{elec}).

The solvation energy is calculated from the following equation:

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

$$G_{\text{non_polar}} = \gamma \text{SASA} + b \quad (5)$$

The polar solvation contribution is represented as G_{GB} and $G_{non-polar}$ is the non-polar solvation contribution and is calculated from the solvent assessable surface area (SASA), obtained by the use of a 1.4 Å water probe radius. The surface tension constant (c) was set to 0.0072 kcal/mol and b to 0 kcal/mol [51]. Per-residue decomposition analyses were also carried out to estimate individual energy contribution of binding site residues towards the affinity and stabilization of BDQ. This was aimed at providing atomistic insights in the ionophoric properties of BDQ towards the inhibition of ATP synthase considering the fact that prominent residual energy contributions could highlight essential residues.

3.0 Results and Discussion

3.1 Conformational dynamics of *Mycobacterium tuberculosis*'s F_1F_0 synthase upon BDQ binding

Up till date, no current experimental data reveals the crystal structure for the rotor ring of *Mtb*'s F_1F_0 synthase thereby accounting for insufficient insights on the inhibitory activity of BDQ towards the inhibition of ATP synthesis in *Mtb*. As such, it was expedient that we employed the non-pathogenic *Mycobacterium phlei* model so as to give a clear molecular insight on the mechanistic activity and efficacy of BDQ. This is feasible because the c-ring of *M. phlei* shares a very high sequence homology and identity (83.7%) with its *M. tuberculosis* variant, particularly in the transmembrane region, where the drug was proposed to bind [3, 52]. Furthermore, the minimum inhibitory concentration (MIC) reported for *M. phlei* was 0.05 mg/ml which is basically identical to that reported for *M. tuberculosis* (0.06 mg/ml) [52, 53], suggesting the identical mode of interaction between the drug and the rotor ring in these two species [17]. MD simulation, as a reliable computer aided drug design approach employed in this study, allowed for a time-bound analysis of the inhibitory impact of BDQ upon binding to *Mtb*'s F_1F_0 synthase in an effort to

impede ATP synthesis. To ensure the structural stability of the studied systems and trustworthiness of the findings from the resulting trajectories, an initial assessment of system convergence was performed, which could depict a reliable simulation run. As observed, the systems converged early in the simulation period (~20ns) after which they were a notable separation in motions of the C α atoms among the unbound (Apo) and BDQ-bound systems, further signifying that the assumptions drawn from the simulated systems were reliable and depict a near accurate representation of the structural events that actually characterize the inhibitory activity of BDQ. From the RMSD plots, we also evaluated the impact of the binding of BDQ on the overall stability of F₁F₀ synthase, considering how essential, the stability of a protein is to its function. The effect of BDQ on the stability of F₁F₀ synthase as shown in the plot (Figure 9.2A), was very evident as there was an obvious decrease in c- α RMSD after the first 20ns relative to the unbound conformation. The markedly lower average RMSD of the bound conformation was suggestive of a reduced atomistic deviation and a structurally stable F₁F₀ synthase, as induced by the binding of BDQ. The relatively stable conformation could have ensued due to favorable steady interactions of key active site residues with BDQ. A further analysis of the inhibitory impact of BDQ on the structural conformation of F₁F₀ synthase was performed by an estimation of the root mean fluctuation (RMSF) of the individual residues. Overall, the unbound F₁F₀ synthase exhibited the highest fluctuation of amino acids with an average RMSF value of 10.9 Å relative to the BDQ-bound F₁F₀ synthase. However, there was a notable increase in the flexibility of residues at region “60-75” in the bound system relative to the unbound conformation (Figure 9.2B). Interestingly, this region encompasses the ion binding site on the rotor-stator region of F₁F₀ synthase. This could suggest that although there was a slight decrease in average fluctuation of all residues of F₁F₀ synthase in the presence of BDQ, there was rather a consequential increase in the flexibility of residues that

directly interacted with BDQ with a $<5\text{\AA}$ radius relative to lower flexibility observed in the unbound conformation. We further monitored the compactness or tightness of the c-ring which has been experimentally reported to play a crucial role in ion exchange in the bacterial ATP synthase [17, 54–56] by calculating the radius of gyration (Rg) of its C α atoms [35, 57]. As shown in Figure 9.2D, it was observed that the Rg of binding site residues in the BDQ-bound F₁F₀ synthase were lowered entirely during the 100ns simulation indicative of a highly compact conformation relative to the unbound conformation.

Consequentially, the lower Rg indicates a relatively tighter conformation [35, 57] which would make it impossible for BDQ to bypass the barrier created by the pocket thereby resulting in an eventual blockage of the rotation of the c-ring followed by a consequential blockage of the ion exchange. Moreover, a disruption in the ion exchange mechanisms could consequently halt ATP synthesis posing a significant impact on the general bio-energetic metabolism in *Mtb*, which is ultimately fatal to the survival of mycobacteria [20, 58]. Taken together, the binding of BDQ lowered the structural activity in the F₁F₀ synthase model, which could impede or interfere with its critical role in the ATP synthesis machinery.

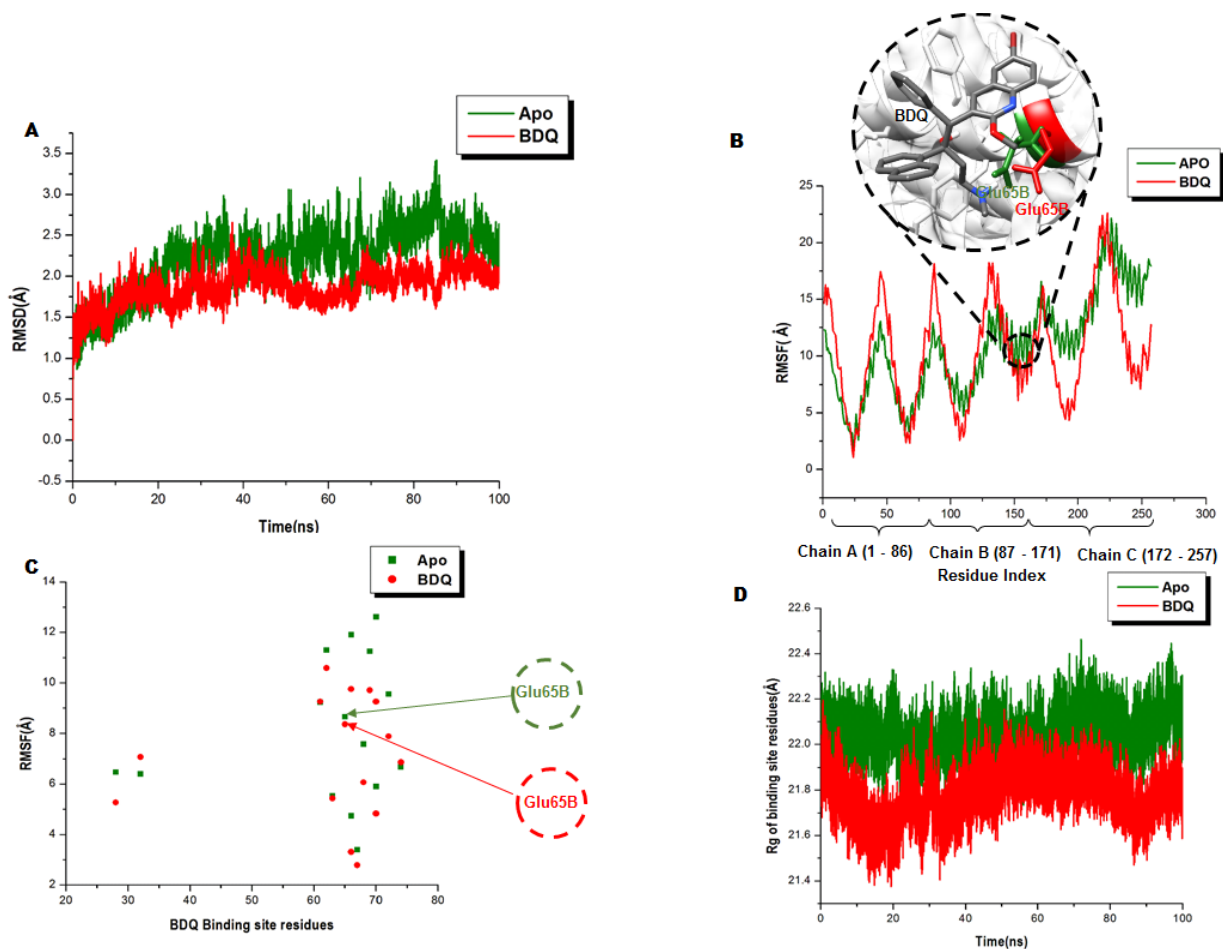


Figure 9.2: Conformational dynamics of F_1F_0 ATP synthase upon binding of BDQ. **A)** Plot of the RMSD showing a gradual increase in RMSD and subsequent steadying of the systems after about 20ns. **B)** RMSF showing the comparative fluctuations of Glu65_B in the bound (BDQ) and the unbound (Apo) systems. **Note:** The scale used is from teap numbering of residues, F_1F_0 of *M. phlei* has three chains of repeated residues as such the RMSF plot shows these chains in brackets and the position of the Glu65_B in the two systems. **C)** Scatter plots showing fluctuations of the binding site residues among the bound and unbound systems. Indicated in **green** (unbound) and **red** (BDQ-bound) circles are the main residues (Glu65_B) responsible for ion exchange in both systems **D)** Radius of gyration (Rg) of the binding site residues showing a low Rg due to the binding of BDQ resulting in a highly compact F_1F_0 structure.

3.2 Dual-molecule BDQ binding increases binding affinity

The concentration-dependent increase in the inhibitory activities of BDQ towards F_1F_0 synthase, as reported by Preiss *et al.*, necessitated the use of the MM/PBSA approach to study the mechanistic single-mode and dual-mode binding and inhibitory potencies of BDQ to the target

protein [17]. This was to provide atomistic perspectives on the findings by Preiss *et al.*, in which it was revealed that one or more BDQ molecules could approach and bind to the c-ring of ATP synthase depending on the concentration [17] and that the more the BDQ molecules bound the stronger the inhibitory potency. To this effect a third system was examined, which comprised of two molecules of BDQ bound to the c-ring of the ATP synthase to depict an increase in concentration of BDQ. Moreover, free binding energies of the BDQ singly- and dual-bound F₁F₀ ATP synthase was evaluated using the MM/PBSA method to obtain a quantitative perspective into the mechanistic high-affinity binding and stability of the compounds at the target site of F₁F₀ synthase. Differential ΔG_{bind} free binding energies for BDQ when bound to F₁F₀ ATP synthase as a single molecule as well as double molecules were estimated and presented in Table 9.1. As estimated, ΔG was higher in the dual bound system as compared to the singly bound F₁F₀ system. The singly bound system had a ΔG value of -29.41kcal/mol while it was considerably higher in the dual system with a value of -38.13kcal/mol. Therefore, we can presume that binding affinity would possibly be higher when there is an increase in the number of BDQ molecules bound to F₁F₀ ATP synthase in agreement with earlier findings that reported improved BDQ potency when the concentration is increased [17]. Furthermore, high-affinity binding in both BDQ-bound systems was evidenced by the estimated high ΔE electrostatic energies as presented in Table 9.1, indicative of the formation of strong interactions that mediated the binding and stability of BDQ in both single and dual states.

Table 9.1: MM/GBSA-based energy profiles of BDQ-singly and double-bound F₁F₀ ATP synthase.

Systems	Energy components (kcal/mol)				
	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
BDQ	-30.86±0.07	-47.49±0.35	-58.35±0.36	48.93±0.32	-29.41±0.06
2BDQ	-43.03± 0.24	-61.77 ±0.39	-40.8±0.55	65.68±0.41	-39.13±0.21

ΔE_{ele} = electrostatic energy; ΔE_{vdw} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG = gas phase free energy.

3.3 Time-bound analysis of the interaction of BDQ at the c-ring F_1F_0 synthase

To further corroborate the mechanistic binding of BDQ molecules, we performed a thorough elucidation of complementary interactions that occurred at the F_1F_0 active site when singly- or dually bound. This entailed a thorough molecular visualization of these interactions over the simulation period followed by the estimations of the residual energy contributions of residues at the c-ring of F_1F_0 synthase towards the high-affinity binding and stabilization of BDQ molecules. This was also important so as to be able to identify key residues that play crucial roles in the BDQ binding. As observed in the singly bound system, certain residues of the hydrophobic groove, which constitutes chains A and B, showed strong interactions with BDQ, and could underlie the basis for its stability and high affinity. Firstly, we investigated key residues and their roles involvement in the binding and stabilization of BDQ over the simulation period. This provided a closer look into the interaction dynamics between certain c-ring residues of F_1F_0 synthase and BDQ where it was revealed that some residues interacted in a consistent, steady and a recurrent manner across the 100ns of simulation period. These residues include; Glu65_B, Tyr68_B and Ile70_A. Notably, these residues formed strong high affinity interactions with constituent BDQ moieties from the beginning of the simulation as shown at 5ns until the end of the simulation at 100ns (Figure 9.3). Interestingly, among these recurrent residues was Glu65_B, which has been shown to play a critical role in the binding of BDQ as evidenced by the systematic occurrence of strong hydrogen, ionic and salt bridge interactions at different periods during the simulation. According to previous reports, Glu65_B is a conserved carboxylate harbored at the ion binding site and plays critical roles that majorly involve the binding and shuttling of protons [14, 17, 54, 59]. Therefore,

the steady engagement of BDQ with Glu65_B could depict a situation whereby the primary role of Glu65_B in ion transport is persistently blocked by BDQ, indicative of its mechanistic activity in disrupting the microbial (*Mtb*) process of ATP synthesis. Likewise, the persistence of Tyr68_B and Ile70_A in forming strong hydrophobic interaction with BDQ over the simulation period could suggest an augmentative role in enhancing the high-affinity binding and inhibitory activity of BDQ. These modes of interactions could relatively provide essential details and atomistic perspectives into the ionophoric inhibitory property of BDQ.

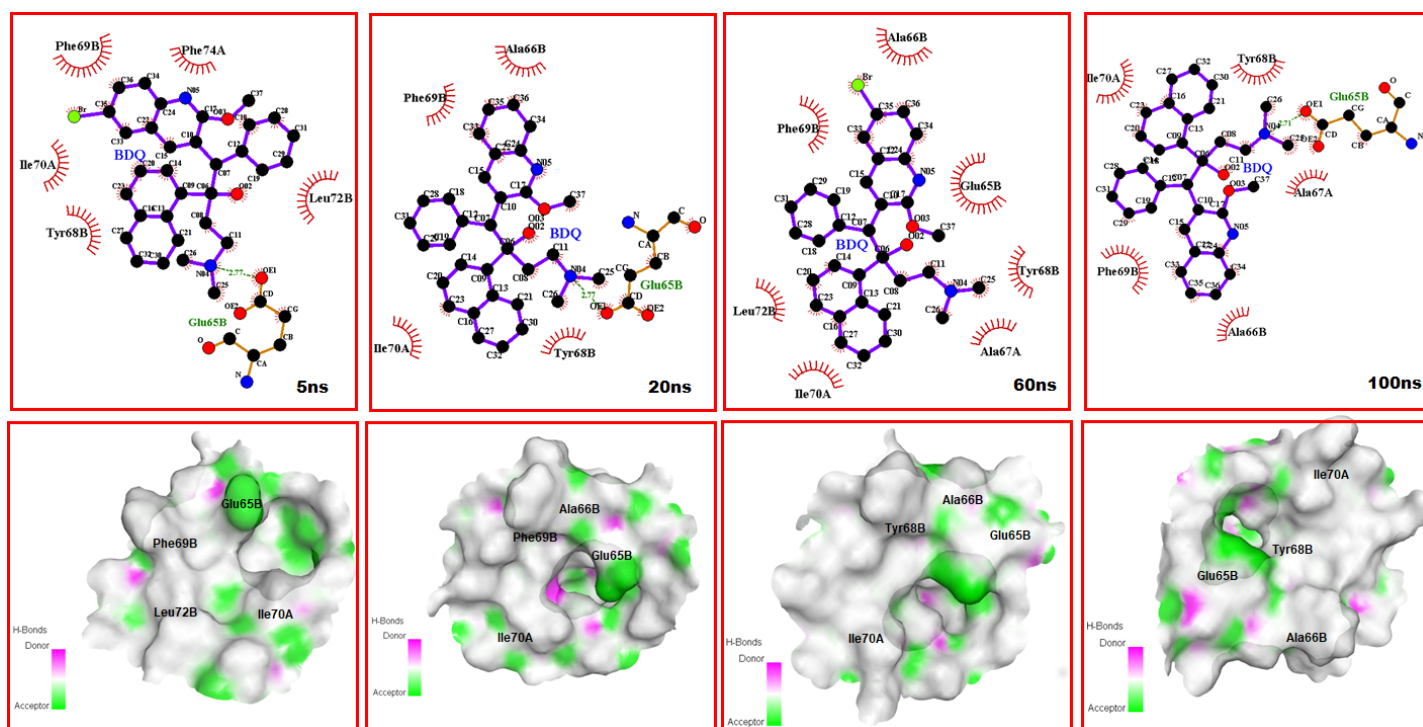


Figure 9.3: Structural positioning and orientation of BDQ at the hydrophilic c rotor-stator region of F₁F₀ synthase BLK depicting steady, consistent and recurrent interaction of BDQ with the c-ring residues. Surface representations of BDQ-bound c-ring in the F₁F₀ synthase are shown in []. The degree of H-bonds ranging from H-bond donor (→ magenta) to H-bond acceptor (→ light green).

To further delve into the nature of the binding site residue interactions, we quantified the energetics of the individual residues to the total binding affinity of BDQ with particular emphasis on the electrostatic and van der Waal energy contributions. In doing so, we estimated the individual energy contributions of each active site residue using the MM/GBSA-integrated per-residue decomposition method. Generally, residues with interaction energies lower than -0.8 kcal/mol are regarded as crucial in the molecular recognition, high-affinity binding and stability of small molecule compounds [60]. As estimated, Phe69_B (-2.65 kcal/mol), Tyr68_B (-2.34 kcal/mol), Ala66_A (-1.11 kcal/mol), Glu65_B (-1.04 kcal/mol) and Leu72_B (-0.98 kcal/mol) exhibited the high energy contributions which, could suggest their high importance to the inhibitory activity of BDQ. However, among these residues, Phe69_B showed the highest total binding energy contribution of -2.65kcal/mol. This could have been due to the reported conformational change it exhibits in order to avoid steric clashes with the bulky BDQ, which could contribute to the optimal and favorable positioning of BDQ, and in turn account for the relatively high total energy involved. This crucial involvement of Phe69_B could be substantiated by its intermittent interactions with BDQ over the simulation time, which could be vital for active site stability and optimal positioning. Interestingly, two of the prominent residues that contributed to the binding of BDQ were the conserved carboxylates; Glu65_B and Asp32_B, which exhibited the highest electrostatic contributions of -47.37kcal/mol and -19.42kcal/mol respectively as shown in Figure 9.4. Earlier reports have indicated that the electrostatic interactions of these two residues with the DMA group of BDQ could account for the measured salt effects in ATP synthesis in *M. smegmatis* IMVs in the presence of BDQ [19, 61]. As such, the prominent involvements of these two residues via electrostatic interactions, over the course of the simulation, as earlier estimated corroborates earlier experimental reports, which further emphasizes the antagonistic mechanisms the essential roles of

these conserved carboxylates in the binding of BDQ. Prominent van der Waals energies were also contributed towards the binding of BDQ by certain binding site residues which could as well favor the overall inhibitory activity of BDQ. Notable residues with van der Waals contributions included Gly62_B, Glu65_B, Ala66_A, Ala67_A, Tyr68_B, Phe69_B, Phe74_A and Leu72_B, which are consistent with previous experimental reports [17].

Using the averaged structure of the BDQ-F₁F₀ ATP synthase complex, a closer look at the specific nature of interaction of the binding site residues revealed that Glu65_B exhibited important hydrogen, ionic, salt-bridge and amide-Pi stacked interactions with BDQ, which could account for its high-affinity binding and stabilization (Figure 9.4). In addition, interactions elicited by BDQ with other active site residues included Pi-Pi T-shaped interaction with Phe69_B, two Pi-alkyl interactions with Ile70_A and a Pi-alkyl interaction with Ala66_A. Overall, this blockade mechanism elicited by BDQ towards key residues of F₁F₀ ATP synthase c-ring such as Glu65_B could interrupt the processes of ion exchange and shuttling which are crucial to the synthesis of ATP as mediated by the target protein.

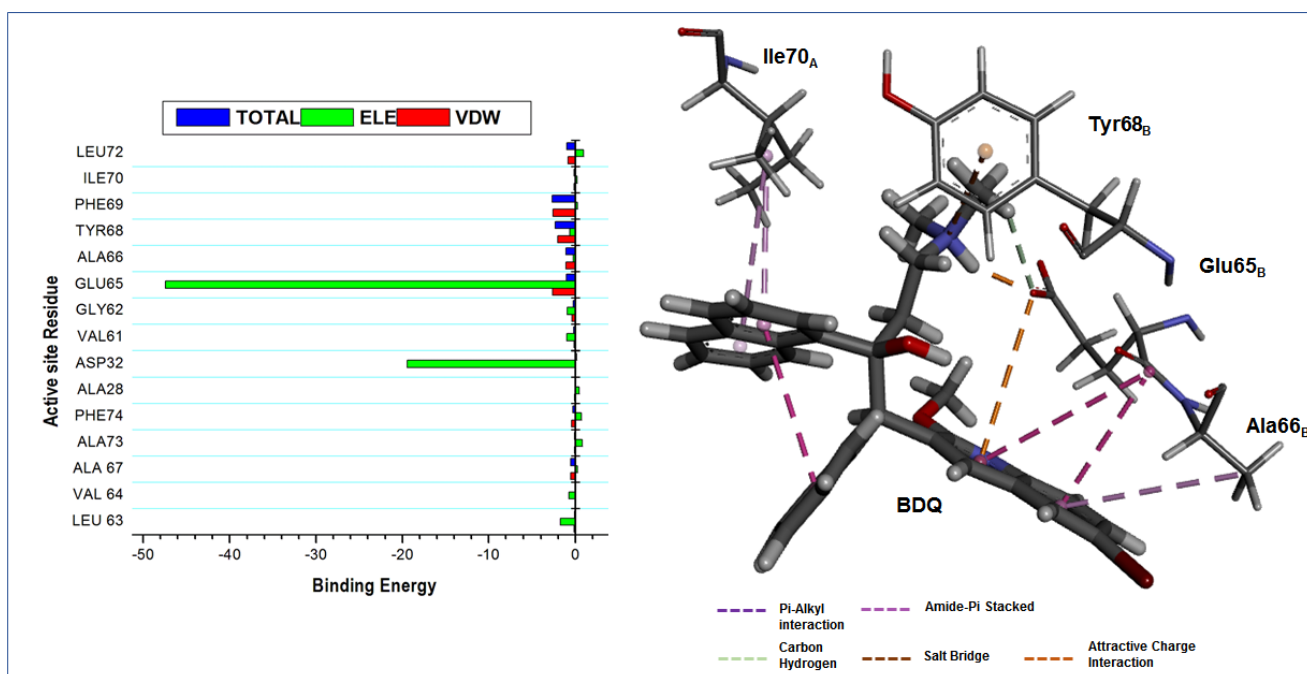


Figure 9.4: Per-residue decomposition plots showing individual energy contributions to the binding and stabilization of BDQ at the c-ring of F₁F₀ synthase. Energy contributions were highest in conserved carboxylates (Glu65_B, Asp32_B) and Phe69_B. Corresponding inter-molecular interactions between BDQ and the hydrophilic c rotor-stator region residues in F₁F₀ synthase is shown as well.

4.0 Conclusion

The uninterrupted synthesis of ATP is crucial for the survival of organisms, which is mediated by the enzyme ATP synthase. This is as well a crucial machinery that has been recently identified as an attractive drug target against *Mtb* largely due to a significant variation in the regulatory mechanisms amongst bacterial and eukaryotic organisms [3, 18, 62]. Among the notable therapeutic agents that target F-type ATP synthase is BDQ, a diarylquinoline that has been experimentally reported to inhibit mycobacterial ATP synthesis by binding to its hydrophilic c rotor-stator region. The main aim of this study was to provide atomistic insights into the inhibitory mechanisms of the ionophore; BDQ, when bound to F₁F₀ ATP synthase using MD simulations and MM/PBSA methods.

Differential assessment of the single and dual binding affinities of BDQ toward F₁F₀ ATP synthase using the MMGBSA free energy calculations revealed that a higher and a more favorable binding of -38.13kcal/mol occurred in the dual molecule bound system relative to the single molecule bound system with a binding free energy of -29.41kcal/mol. This could agree with earlier reports which revealed that increased concentrations of BDQ enhances inhibitory potency against the ATP synthetic machinery of *Mtb*. A time-scale and nanosecond based visualization of the interaction dynamics between BDQ and the c-ring residues of F₁F₀ ATP synthase revealed a consistent and steady high-affinity involvement of Glu65_B, a conserved carboxylate residues involved in ion transfer, an interaction that could explain the mechanistic inhibitory activity of BDQ since this act could block ion shuttling at the c-ring, which could in turn stall the operation of ATP synthase.

Further investigations of the involvement of individual residues at the c-ring region further implicated Asp32_B and Phe69_B, indicative of their essentiality to the binding and stability of BDQ. Decomposition of the ΔG revealed that Glu65_B and Asp32_B had high energy contributions which further conformed their importance as stipulated in a previous study which revealed that the electrostatic interactions mediated by these residues account for measured salt effect in the synthesis of ATP in *M. smegmatis*.

Conformational dynamics of the entire F₁F₀ ATP synthase and the c-ring BDQ binding site residues were also investigated over the course of the simulation by calculating the RMSF and Rg of their C α atoms. It was observed that although there was a slightly decreased average fluctuation of all residues of F₁F₀ synthase in the presence of BDQ, there was rather an increased flexibility of the residues that directly interacted with BDQ with a <5Å radius relative to lower flexibility in the unbound conformation which could have been induced by BDQ. A lower Rg of the c-ring suggested a tighter conformation which could have led to a blockage of c-ring rotation with a subsequent halt in the ion exchange required for ATP synthesis. Findings from this report provide atomistic perspectives into the inhibitory properties of BDQ towards ATP synthesis when bound to the c-ring of F₁F₀ synthase. These useful insights could serve as the basis for the structure-based design of novel and next generation of small molecule inhibitors of ATP synthase in the treatment of TB. Findings from this study also presents a pathway for further investigations into possible resistance to BDQ by mutation of the F₁F₀ ATP synthase due to the numerous resistant strains in *Mycobacterium tuberculosis*.

5.0 Conflict of interest

Authors declare no conflicts of interest.

6.0 Acknowledgment

We would like to acknowledge the Center for High Computing Performance (CHPC) (www.chpc.ac.za), Cape Town for the resources and technical support.

7.0 References

1. World Health Organization. (2018). *Global tuberculosis report 2018*.
2. Sakula, A. (1982). Robert Koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax*, 37(4), 246–51.
3. Andries, K., Verhasselt, P., Guillemont, J., Göhlmann, H. W. H., Neefs, J. M., Winkler, H., ... Jarlier, V. (2005). A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science*, 307(5707), 223–227. doi:10.1126/science.1106753
4. Shah, N. S., Wright, A., Bai, G. H., Barrera, L., Boulahbal, F., Martín-Casabona, N., ... Cegielski, J. P. (2007). Worldwide emergence of extensively drug-resistant tuberculosis. *Emerging Infectious Diseases*, 13(3), 380–387. doi:10.3201/eid1303.061400
5. Ndjeka, N., Conradie, F., Schnippel, K., Hughes, J., Bantubani, N., Ferreira, H., ... Pillay, Y. (2015). Treatment of drug-resistant tuberculosis with bedaquiline in a high HIV prevalence setting: An interim cohort analysis. *International Journal of Tuberculosis and Lung Disease*. doi:10.5588/ijtld.14.0944
6. Klopper, M., Warren, R. M., Hayes, C., Gey van Pittius, N. C., Streicher, E. M., Muller, B., ... Trollip, A. P. (2013). Emergence and spread of extensively and totally drug-resistant tuberculosis, South Africa. *Emerg Infect Dis*, 19(3), 449–455.

doi:10.3201/EID1903.120246\rr10.3201//EID1903.120246

7. Council, I. (n.d.). *TB for Nurses in the Care 3rd Edition*.
8. Caño-muñiz, S., Anthony, R., Niemann, S., & Alffenaar, J. C. (2018). New Approaches and Therapeutic Options for Mycobacterium. *Clinical microbiology reviews*, *31*(1), 1–13.
9. Zhang, Y., Post-martens, K., & Denkin, S. (2006). New drug candidates and therapeutic targets for tuberculosis therapy *REVIEWS*, *11*(1).
10. Walker, J. E. (2013). The ATP synthase: the understood, the uncertain and the unknown. *Biochemical Society Transactions*, *41*(1), 1–16. doi:10.1042/BST20110773
11. Boyer, P. D. (1997). the Atp Synthase—a Splendid Molecular Machine. *Annual Review of Biochemistry*, *66*(1), 717–749. doi:10.1146/annurev.biochem.66.1.717
12. Junge, W., Sielaff, H., & Engelbrecht, S. (2009). Torque generation and elastic power transmission in the rotary F O F 1-ATPase. *Nature*, *459*(7245), 364–370. doi:10.1038/nature08145
13. Noji, H., Yasuda, R., Yoshida, M., & Kinosita, K. (1997). Direct observation of the rotation of F1-ATPase. *Nature*. doi:10.1038/386299a0
14. Meier, T., Polzer, P., Diederichs, K., Welte, W., & Dimroth, P. (2005). Structure of the Rotor ring of F-Type Na⁺ sup +\$-ATPase from *Mycobacter tararicus*\fP. *Science*, *308*(April), 659–662.
15. D., J. (2013). Tuberculosis Success. *Nat Rev Drug Discov*, *12*, 175–176.
16. Hards, K., McMillan, D., Schurig-Briccio, L., Gennis, R., Lill, H., Bald, D., & Cook, G. (2018). Ionophoric effects of the antitubercular drug bedaquiline. *Proceedings of the National Academy of Sciences*, *28*(115), 7326–7331.
17. Preiss, L., Langer, J. D., Yildiz, Ö., Eckhardt-Strelau, L., Guillemont, J. E. G., Koul, A., & Meier, T. (2015). Structure of the mycobacterial ATP synthase Forotor ring in complex with the

anti-TB drug bedaquiline. *Science Advances*, 1(4), 1–8. doi:10.1126/sciadv.1500106

18. Koul, A., Dendouga, N., Vergauwen, K., Molenberghs, B., Vranckx, L., Willebrords, R., ... Andries, K. (n.d.). Diarylquinolines target subunit c of mycobacterial ATP synthase, 323–324. doi:10.1002/prot.21376
19. Haagsma, A. C., Abdillahi-ibrahim, R., Wagner, M. J., Krab, K., Vergauwen, K., Guillemont, J., ... Bald, D. (2009). Selectivity of TMC207 towards Mycobacterial ATP Synthase Compared with That towards the Eukaryotic Homologue \square , 53(3), 1290–1292. doi:10.1128/AAC.01393-08
20. Tran, S. L., & Cook, G. M. (2005). The F1Fo-ATP synthase of *Mycobacterium smegmatis* is essential for growth. *Journal of Bacteriology*. doi:10.1128/JB.187.14.5023-5028.2005
21. Khurshid Ahmad, M. H. (2015). Drug Discovery and In Silico Techniques: A Mini-Review. *Enzyme Engineering*, 04(01), 1–4. doi:10.4172/2329-6674.1000123
22. Tanwar, G. (2019). Gain of native conformation of Aurora A S155R mutant by small molecules, (April 2018), 1–11. doi:10.1002/jcb.28387
23. Daghestani, M., Purohit, R., Daghestani, M., Daghistani, M., & Warsy, A. (2019). Molecular dynamic (MD) studies on leptin receptor gene and associated variations in the anthropometric and metabolic profiles of Saudi women, (Md), 1–15.
24. Purohit, R. (2014). Role of ELA region in auto-activation of mutant KIT receptor: A molecular dynamics simulation insight. *Journal of Biomolecular Structure and Dynamics*, 32, 1033–1046.
25. Da Silva, P. E. A., & Palomino, J. C. (2011). Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: Classical and new drugs. *J Antimicrob Chemother*, 66(7), 1417–30.

26. Rajendran, V., Gopalakrishnan, C., & Purohit, R. (2016). Impact of point mutation P29S in RAC1 on tumorigenesis. *Tumor Biology*. doi:10.1007/s13277-016-5329-y
27. Rajendran, V., Gopalakrishnan, C., Sethumadhavan, R., & Nadu, T. (2017). Pathological role of a point mutation (T315I) in BCR-ABL1 protein - A, (July). doi:10.1002/jcb.26257
28. Purohit, R. (2018). Screening of Potential Inhibitor against Coat Protein of Apple Chlorotic Leaf Spot Virus. *Cell Biochemistry and Biophysics*, 273–278. doi:10.1007/s12013-017-0836-z
29. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. doi:10.1002/jcc.20084
30. Kusumaningrum, S., Budianto, E., Kosela, S., Sumaryono, W., & Juniarti, F. (2014). The molecular docking of 1,4-naphthoquinone derivatives as inhibitors of Polo-like kinase 1 using Molegro Virtual Docker. *Journal of Applied Pharmaceutical Science*, 4(11), 47–53. doi:10.7324/JAPS.2014.4119
31. Grand, S. Le, Götz, A. W., & Walker, R. C. (2013). SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Computer Physics Communications*, 184(2), 374–380. doi:http://dx.doi.org/10.1016/j.cpc.2012.09.022
32. Olotu, F. A., Agoni, C., Adeniji, E., Abdullahi, M., & Soliman, M. E. (2018). Probing Gallate-Mediated Selectivity and High-Affinity Binding of Epigallocatechin Gallate: a Way-Forward in the Design of Selective Inhibitors for Anti-apoptotic Bcl-2 Proteins. *Applied Biochemistry and Biotechnology*. doi:10.1007/s12010-018-2863-7
33. Oguntade, S., Ramharack, P., & Soliman, M. E. (2017). Characterizing the ligand-binding landscape of Zika NS3 helicase-promising lead compounds as potential inhibitors. *Future Virology*, 12(6), 261–273. doi:10.2217/fvl-2017-0014

34. Agoni, C., Ramharack, P., & Soliman, M. E. S. (2018). Synergistic Interplay of The Co-administration of Rifampin And Newly Developed Anti-TB Drug : Could It Be a Promising New Line of TB Therapy ?, 1–8. doi:10.2174/1386207321666180716093617
35. Agoni, C., Ramharack, P., & Soliman, M. (2018). Co-inhibition as a strategic therapeutic approach to overcome rifampin resistance in tuberculosis therapy: atomistic insights. *Future Medicinal Chemistry*, 10(14), 1665–1675. doi:10.4155/fmc-2017-0197
36. Olotu, F. A., & Soliman, M. E. S. (2018). Dynamic perspectives into the mechanisms of mutation-induced p53-DNA binding loss and inactivation using active perturbation theory: Structural and molecular insights toward the design of potent reactivators in cancer therapy. *Journal of Cellular Biochemistry*. doi:10.1002/jcb.27458
37. Salomon-Ferrer, R., G??tz, A. W., Poole, D., Le Grand, S., & Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *Journal of Chemical Theory and Computation*, 9(9), 3878–3888. doi:10.1021/ct400314y
38. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696–3713. doi:10.1021/acs.jctc.5b00255
39. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., ... Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. doi:10.1002/jcc.20290
40. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, a, & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*,

81(1984), 3684–3690. doi:10.1063/1.448118

41. Kräutler, V., Van Gunsteren, W. F., & Hünenberger, P. H. (2001). A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry*, 22(5), 501–508. doi:10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V
42. Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.
43. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J Comput Chem*, 25, 1605–1612. doi:10.1002/jcc.20084
44. Seifert, E. (2014). OriginPro 9.1: scientific data analysis and graphing software-software review. *Journal of chemical information and modeling*, 54(5), 1552. doi:10.1021/ci500161d
45. BIOVIA, D. S. (2016). Discovery Studio 2016 Client. *San Diego: Dassault Systèmes*.
46. Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*. doi:10.1021/ci200227u
47. Massova, I., & Kollman, P. A. (2000). Combined molecular mechanical and continuum solvent approach (MM-PBSA / GBSA) to predict ligand binding, (i), 113–135.
48. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., ... Cheatham, T. E. (2000). Calculating Structures and Free Energies of Complex Molecules : Combining Molecular Mechanics and Continuum Models, 33(12), 889–897.
49. Miller, B. R., Mcgee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). MMPBSA . py : An E ffi cient Program for End-State Free Energy Calculations.

doi:10.1021/ct300418h

50. Onufriev, A., Bashford, D., & Case, D. A. (2000). Modification of the Generalized Born Model Suitable for Macromolecules, 3712–3720.
51. Sitkoff, D., Sharp, K. A., & Honig, B. (1994). Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models, 1978–1988.
52. Koul, A., Andries, K., Jonge, M. R. De, & Koymans, L. H. M. (2007). A Computational Model of the Inhibition of Mycobacterium tuberculosis ATPase by a New Drug Candidate R207910, 980, 971–980. doi:10.1002/prot
53. Huitric, E., Verhasselt, P., Andries, K., & Hoffner, S. E. (2007). In Vitro Antimycobacterial Spectrum of a Diarylquinoline ATP Synthase Inhibitor □, 51(11), 4202–4204. doi:10.1128/AAC.00181-07
54. Pogoryelov, D., Krah, A., Langer, J. D., Yildiz, Ö., Faraldo-Gómez, J. D., & Meier, T. (2010). Microscopic rotary mechanism of ion translocation in the F(o) complex of ATP synthases. *Nature Chemical Biology*, 6(12), 891–899. doi:10.1038/nchembio.457
55. Hakulinen, J. K., Klyszejko, A. L., Hoffmann, J., Eckhardt-Strelau, L., Brutschy, B., Vonck, J., & Meier, T. (2012). Structural study on the architecture of the bacterial ATP synthase F_o motor. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1203971109
56. Allegretti, M., Klusch, N., Mills, D. J., Vonck, J., Kühlbrandt, W., & Davies, K. M. (2015). Horizontal membrane-intrinsic α -helices in the stator a-subunit of an F-type ATP synthase. *Nature*. doi:10.1038/nature14185
57. Lobanov, M. Y., Bogatyreva, N. S., & Galzitskaya, O. V. (2008). Radius of gyration as an indicator of protein structure compactness. *Molecular Biology*, 42(4), 623–628. doi:10.1134/S0026893308040195

58. Koul, A., Vranckx, L., Dhar, N., Göhlmann, H. W. H., Özdemir, E., Neefs, J. M., ... Bald, D. (2014). Delayed bactericidal response of *Mycobacterium tuberculosis* to bedaquiline involves remodelling of bacterial metabolism. *Nature communications*. doi:10.1038/ncomms4369
59. Matthies, D., Zhou, W., Klyszejko, A. L., Anselmi, C., Yildiz, Ö., Brandt, K., ... Meier, T. (2014). High-resolution structure and mechanism of an F/V-hybrid rotor ring in a Na⁺-coupled ATP synthase. *Nature Communications*, 5(May 2015). doi:10.1038/ncomms6286
60. Fang, J., Wu, P., Yang, R., Gao, L., Li, C., Wang, D., ... Du, G. H. (2014). Inhibition of acetylcholinesterase by two genistein derivatives: Kinetic analysis, molecular docking and molecular dynamics simulation. *Acta Pharmaceutica Sinica B*, 4(6), 430–437. doi:10.1016/j.apsb.2014.10.002
61. Haagsma, A. C., Podasca, I., Koul, A., Andries, K., Guillemont, J., Lill, H., & Bald, D. (2011). Probing the interaction of the diarylquinoline TMC207 with its target mycobacterial ATP synthase. *PLoS ONE*, 6(8). doi:10.1371/journal.pone.0023575
62. Krah, A. (2015). Linking structural features from mitochondrial and bacterial F-type ATP synthases to their distinct mechanisms of ATPase inhibition. *Prog Biophys Mol Biol*, 119, 94–102.

CHAPTER 10

Published Article

Triple *Mycobacterial* ATP-synthase mutations impedes Bedaquiline binding: Atomistic and structural perspectives

Elliasu Y. Salifu^a, **Clement Agoni^a**, Fisayo A. Olotu^a, Mahmoud E. S. Soliman^{a*}

^aMolecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

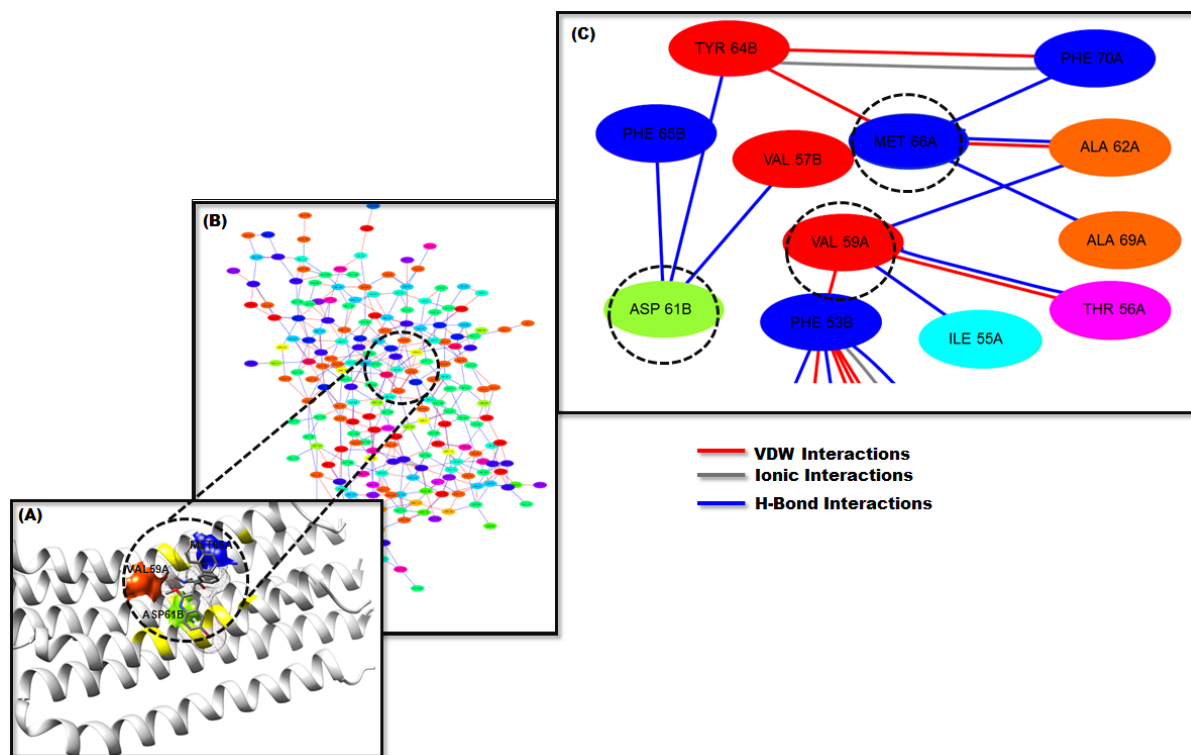
*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

<http://soliman.ukzn.ac.za>

Telephone: +27 (0) 31 260 8048, Fax: +27 (0) 31 260 7872

Graphical Abstract



Abstract

Bedaquiline (BDQ) has demonstrated formidable bactericidal activity towards *Mycobacterium tuberculosis* (*Mtb*) in the treatment of multi-drug resistant (MDR) and extensively drug resistant (XDR) tuberculosis (TB). BDQ elicits its therapeutic function by halting the ionic shuttle of *Mtb* via mycobacterial F₁F₀ ATP-synthase blockade. However, triple mutations (L59V, E61D and I66M) at the ligand-binding cavity characterize emerging BDQ-resistant strains thereby restraining the potentials embedded in this anti-microbial compound, particularly in MDR/XDR-TB therapy. In this report, the effects of these triple mutations on BDQ-*Mtb* F₁F₀ ATP-synthase binding were investigated using molecular dynamics, free energy binding and residue interaction network (RIN) analyses.

The highlight of our findings is the drastic reduction in BDQ binding affinity (ΔG) in the triple mutant protein, which was caused by a systemic loss in high-affinity interactions primarily mediated by L59, E61 and I66. While wildtype L59 and I66 formed pi-alkyl interactions with BDQ at the F₁F₀ ATP-synthase binding site, E61 elicited conventional (O--HO) bond. Upon transition, V59 and I66 were devoid of interactions with BDQ while D61 existed in a weaker non-conventional (C--HO) bond. Likewise, these mutations distorted the binding site and overall structural architecture of F₁F₀ ATP-synthase in the presence of BDQ as revealed by the RIN and conformational analyses.

Insights from this study could serve as a starting point for the structure-based design of novel inhibitors that could overcome mutational setbacks posed by BDQ-resistant strains in MDR/XDR-TB treatment.

Keywords: *Mycobacterium tuberculosis*, Multi-Drug resistant TB, Bedaquiline, F₁F₀-ATP synthase, Molecular dynamic simulations

1.0 Introduction

Tuberculosis (TB) is currently the leading cause of death from an infectious disease worldwide, killing more people than Human immunodeficiency virus/ Acquired immunodeficiency syndrome (HIV/AIDS) [1]. Eradication of TB still remains a challenge despite the availability of vaccines as well as drug treatments. In 2017, 10 million people fell ill with TB, among which 1.6 million died from the disease (including 0.3 million among people with HIV) [1]. The continual emergence of resistant *Mycobacterium tuberculosis* (*Mtb*) strains further complicates existing treatment options and consequently poses a serious global threat toward curtailing TB epidemic.

Resistance over the years exists in different forms, ranging from Multidrug-Resistant TB (MDR-TB) to Extensive Drug Resistance TB (XDR-TB). While the former is mediated by an *Mtb* strain resistant to isoniazid and rifampin; the two most potent first-line anti-TB drugs [2], the latter describes resistance to the fluoroquinolones or injectable second-line anti-TB drugs such as capreomycin, viomycin and ciprofloxacin (CIP) [3]. Moreover, the most virulent form of resistance is described as the Totally Drug Resistant TB (TDR-TB) which categorizes strains that are resistant to both first-line and second-line anti-TB therapy [4].

The 2018 WHO Global TB report on MDR-TB revealed an estimated 588,000 cases globally with South Africa alone accounting for about 14,000 cases. MDR-TB strains are historically difficult to treat compared to drug susceptible TB, which poses a global threat towards disease eradication. As such there is the need to explore new therapeutic options in this field of research. Emerging reports have revealed that Bedaquiline (BDQ), a recently approved anti-TB drug, is potent in the treatment of MDR-TB and XDR-TB.

Bedaquiline, an antibiotic drug belonging to the chemical class of diarylquinolines, was recently granted approval by the Food and Drugs Administration (FDA)[2]. As an anti-TB agent, BDQ has

shown that targeting energy metabolism is a new and efficient approach for inhibiting *Mtb* activity [5, 6]. Moreover, structural and inhibitory mechanisms that underlie BDQ mode of action at atomistic level have been recently elucidated in report from our research group, supportive of its potency. BDQ functions by targeting F₁F₀ ATP synthase, an essential enzyme in ATP synthesis [7]. Further to this, BDQ has been known to cause mycobacterial cell death through a phenomenon known as uncoupling in which it functions as a H⁺/K⁺ ionophore, causing transmembrane pH and potassium gradients to be equilibrated [8].

The F₁F₀ ATP synthase is a macromolecular membrane embedded protein complex which employs electrochemical ion gradient for the conversion of adenosine diphosphate (ADP) and inorganic phosphate into adenosine triphosphate (ATP) by a rotary mechanism [9–12]. Moreover, the F_o domain of the F₁F₀ harbors the rotor ring where BDQ is known to bind. In bacteria, it consists of identical copies of c-subunits forming an hourglass-shaped cylinder with a central pore known as the c-ring [13].

In a recent study, the crystal structure of the c-ring from *Mycobacterium phlei* (*M. phlei*) in complex with BDQ was resolved, indicating that BDQ cannot bind to the ATP synthase of non-mycobacterial species [14]. Also BDQ reportedly forms a large number of hydrophobic, hydrophilic and electrostatic interactions, which explains the high affinity binding demonstrated by BDQ towards the *M. phlei* c-ring [14].

Despite the widely reported potency of BDQ in the treatment of MDR-TB via F₁F₀ ATP synthase inhibition, the emergence of mutations has been related to BDQ-resistant *Mtb* [6, 15, 16]. The occurrence of a triple mutation; L59V, E61D and I66M, reportedly interferes with the direct binding of BDQ limiting its therapeutic efficiency [6].

This necessitates the development of novel therapeutic options that could help surmount BDQ resistance in the treatment of XDR/MDR-TB. Moreover, detailed understanding into the mechanisms by which triple mutation impedes BDQ activity is highly crucial, hence the design of this study. We employed molecular dynamics and binding analyses to obtain structural and molecular insights into the triple L59V, E61D and I66M mutational effects of BDQ on F_1F_0 ATP synthase. Findings would form an important basis for the design and development of novel non-resistant *Mtb* F_1F_0 ATP synthase inhibitors.

2.0 Computational Methodology

2.1 Homology modelling of F_1F_0 ATP synthase of Mycobacterial tuberculosis

The unavailability of an X-ray crystal structure for *Mtb* F_1F_0 ATP synthase necessitated the use of homology modeling (HM) technique to obtain its 3-D structure from a template *M. phlei* structure with PDB entry 4V1G. The FASTA sequence of the target protein used for HM was retrieved from Uniprot database with entry C1AMU9, having ~90% sequence similarity with *M. phlei* F_1F_0 ATP synthase. The HM of the *Mtb* F_1F_0 ATP synthase was based on established in-house protocols as employed in our previous reports [17].

Structural validation of the obtained model was done using MolProbity webserver [18] with a Ramachandran plot that provided visual representations of energetically allowed or stable regions. Further validation of the modelled structure was done via PROSA[19] and PROCHECK[20] analysis. The F_1F_0 structure modelled comprised of three identical chains A, B, and C, with each chain representing a single subunit C.

2.2 Molecular Docking

Molecular docking was performed on the modelled *Mtb* F₁F₀ ATP synthase subunit C structure to predict the binding conformation and affinity of the BDQ within the active site region. The active site was determined by using the match maker functionality on UCSF Chimera [21] to superimpose the built homology model with the X-ray crystal structure of the template *M. phlei*. Molecular docking was then carried out using the AutoDock Vina software [23]. The grid box that defines the binding site region of the F₁F₀ was generated using the AutoDock Vina functionality on UCSF Chimera. BDQ was then docked to the active site region which constituted chains A and B. The grid box for the docking of BDQ had size and center coordinates of $x = (14, -168.70)$, $y = (11, 58.11)$ and $z = (10, 29.70)$ respectively.

2.3 System Preparation

The modelled structure of *Mtb* was then prepared for molecular dynamic simulation (MD) using UCSF chimera [24] and Molecular Molegro Viewer (MMV) [25], to eliminate molecules that are irrelevant to the current study. Three-point mutations 59(Leu → Val), 61(Glu → Asp) and 66(Ile → Met) were carried on the protein structure (Figure 10.1) using the integrated Dunbrack rotamer method of UCSF Chimera [26]. Altogether, BDQ-bound non-mutated (wildtype, *wt*) and mutant (*mt*) systems were obtained and set up for 150ns simulation time in order to clearly distinguish mutational effects on BDQ activities.

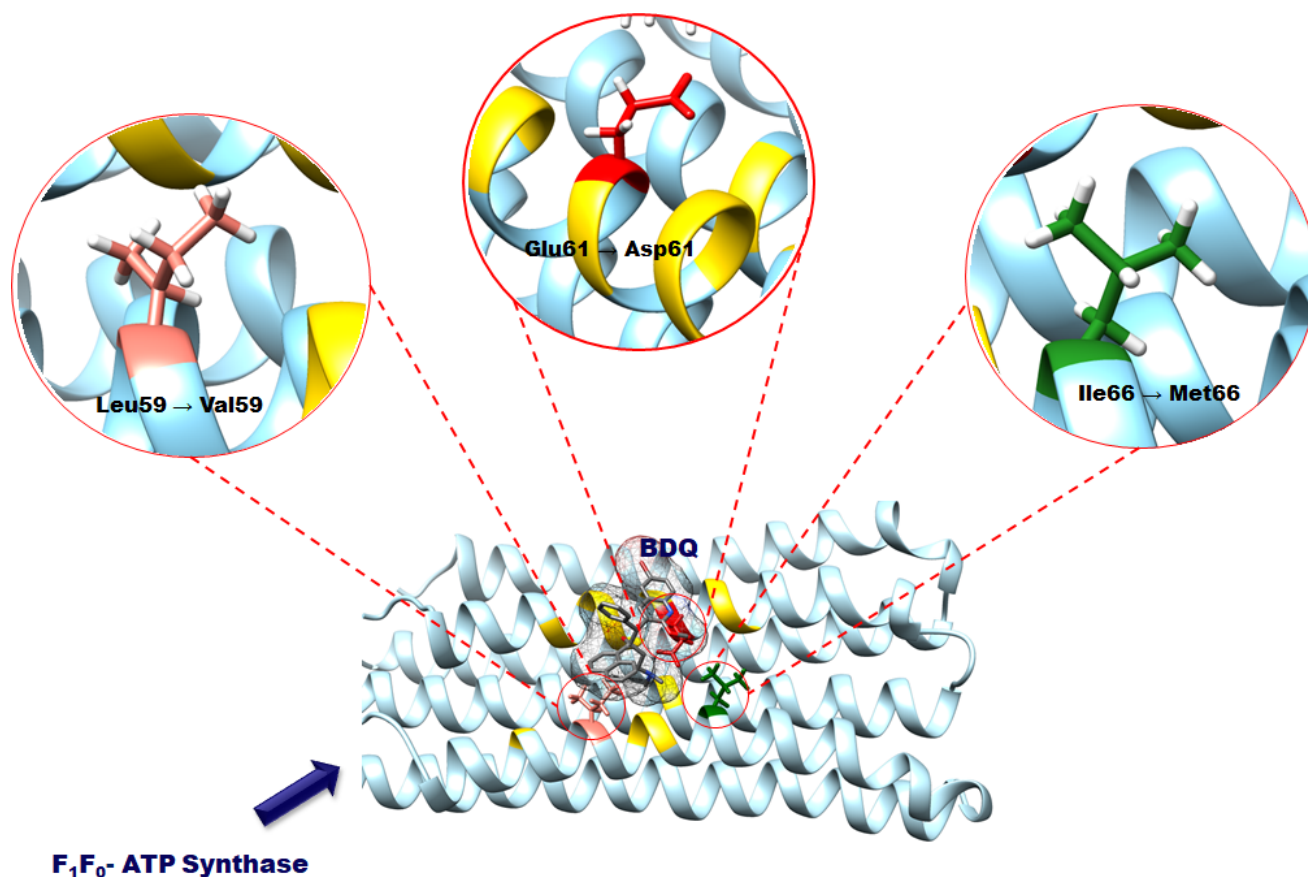


Figure 10.1: 3-D structure of *Mtb* F₁F₀ ATP synthase (light blue) bound by BDQ (grey) at the active site (yellow) and showing the three mutations 59(Leu → Val), 61(Glu → Asp) and 66(Ile → Met).

Molecular dynamics (MD) simulations

MD simulation was performed using the GPU version of AMBER18 with an integrated PMEMD module [27], according to in-house standard simulation protocols, as employed in several of our previous studies [28–32]. The MD simulation was initiated using a time step of 1 fs and coordinates saved at 1 ps interval followed by subsequent analysis of trajectories using the integrated PTRJ and CPPTRAJ module [33]. Visualization of the complexes and data plots were carried out using the graphical user interface of UCSF chimera [21] and Microcal Origin analytical software [34].

2.5 Thermodynamic calculations

The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) [35–38] method was used to calculate binding free energy of the BDQ in the wild and mutant *Mtb* F₁F₀ ATP- synthase.

The binding free energy (ΔG_{bind}) was calculated from the following equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = \Delta G_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S, \quad (2)$$

Where ΔG_{bind} is considered to be the summation of the gas phase and solvation energy terms less the entropy ($T\Delta S$) term

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} \quad (3)$$

ΔE_{gas} is the sum of the AMBER force field internal energy terms ΔE_{int} (bond, angle and torsion), the covalent van der Waals (ΔE_{vdw}) and the non-bonded electrostatic energy component (ΔE_{elec}).

The solvation energy is calculated from the following equation:

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

$$G_{\text{non-polar}} = \gamma \text{SASA} + b \quad (5)$$

The polar solvation contribution is represented as G_{GB} and $G_{\text{non-polar}}$ is the non-polar solvation contribution and is calculated from the solvent assessable surface area (SASA), obtained by the use of a 1.4 Å water probe radius. The surface tension constant (γ) was set to 0.0072 kcal/mol and b to 0 kcal/mol [39]. Per-residue decomposition analyses were also carried out to estimate differential energy contributions of binding site residues towards BDQ affinity and stabilization in

both systems. This was important to further monitor triple mutational effects on BDQ binding complementarity.

3.0 Results and Discussion

3.1 Structural validation of the 3D Homology model of Mtb F₁F₀ ATP-synthase

The Molprobit web server was employed in validating the homology model built in this study and the results presented in a Ramachandran plot as shown in figure S1. From the plot 95.2% of all residues were in favored region while 98.3% were in the allowed region leaving 4 residues as outliers. Interestingly these four outliers did not include residues of the active site region when the homology model was assessed for its active site residues by superimposing with the template *M. phlei* structure (4V1G) on Chimera Multi-align viewer [21]. The results show a level of precision of the dihedral angles and backbone in three dimensional space. In addition to this, the PROSA web server was used to measure the total model quality, as it affirms if built model was within the acceptable X-ray and NMR studies. The Z-score for the modelled ATP synthase was -1.88, as shown in Figure S2, which implies a good quality model for use. Further validation was also carried out using PROCHECK which identified 92.6% of the residues to be in the favored region, 5.7% of the residues in the additionally allowed regions while 1.1% in the generously allowed region and 0.6% of the residues in the disallowed region (figure S3). Taken together, these validation protocols established the reliability of the 3D structure model used in this study.

3.2 Triple mutation induces loss of crucial interactions and consequential energy contributions

The binding affinity of inhibitors to their respective targets is largely influenced by complementary interactions occurring within the binding sites, particularly with certain site residues. As such point-mutational changes or disorders among these residues could alter the binding mechanisms of high-affinity inhibitors and in turn affect their therapeutic functionalities [40, 41].

At first, we measured the involvement of those residues (*wt*Leu59_A, *wt*Ile66_A and *wt*Glu61_B) in the binding of BDQ prior to mutations. This was done by evaluating differential energy contributions of BDQ-interactive site residues across *wt* and *mt* F₁F₀ ATP synthase. Overall energies were decomposed into van der Waals and electrostatic energy including their overall contributions to total energy. This was performed using MM/GBSA-based per-residue energy decomposition method. As estimated, Leu59_A, Ile66_A and Glu61_B contributed total energies of -2.60kcal/mol, -1.21kcal/mol and -1.38kcal/mol respectively to BDQ binding in the *wt* form (Figure 5.2).

Interestingly, we observed that the mutational change in residues (*wt*Leu59_A → *mt*Val59_A; *wt*Ile66_A → *mt*Met66_A; and *wt*Glu61_B → *mt*Asp61_B) accounted for a reduction in total energies where *mt*Val59_A, *mt*Met66_A and *mt*Asp61_B had energies of -1.69kcal/mol, -0.82kcal/mol and -0.69kcal/mol respectively. These energy differences could account for weaker BDQ interactions given that *wt*Leu59_A, *wt*Ile66_A and *wt*Glu61_B played critical roles in the binding and stabilization of BDQ consistent with its high-affinity inhibitory activities. In a similar fashion, vdW and electrostatic energies were lower in the mutated residues compared to their *wt* forms as presented in Table 10.1. Relatively, an increase in positive value for the vdW energy contributions in *mt*Met66 could be more unfavorable towards BDQ binding.

Additional insights were provided by elucidating the nature and types of bonds that occur at the *Mtb* F₁F₀ ATP synthase binding site relative to the BDQ activities in both *wt* and *mt* variants.

Table 10.1: Calculated final decomposition energies for all residues of interest.

Energy components (kcal/mol)	Residues					
	Leu59_A → Val59_A		Ile66_A → Met66_A		Glu61_B → Asp61_B	
vdW	-2.77	-1.99	-1.40	-0.62	-2.12	-0.74
ELE	-1.97	-1.68	0.86	0.96	-7.52	-5.20
TOTAL	-2.60	-1.69	-1.21	-0.82	-1.38	-0.69

As shown in Figure 10.2, *wt*Leu59 and *wt*Ile66 mediated a pi-alkyl interaction with BDQ; interactions that could be crucial for its stabilization at the binding site. However, these interactions were lost when mutated to Val59_A and Met66_A, events that could possibly affect BDQ stability and binding affinity. Similarly, there was the occurrence of a high-affinity hydrogen bond between *wt*Glu61_B and BDQ.

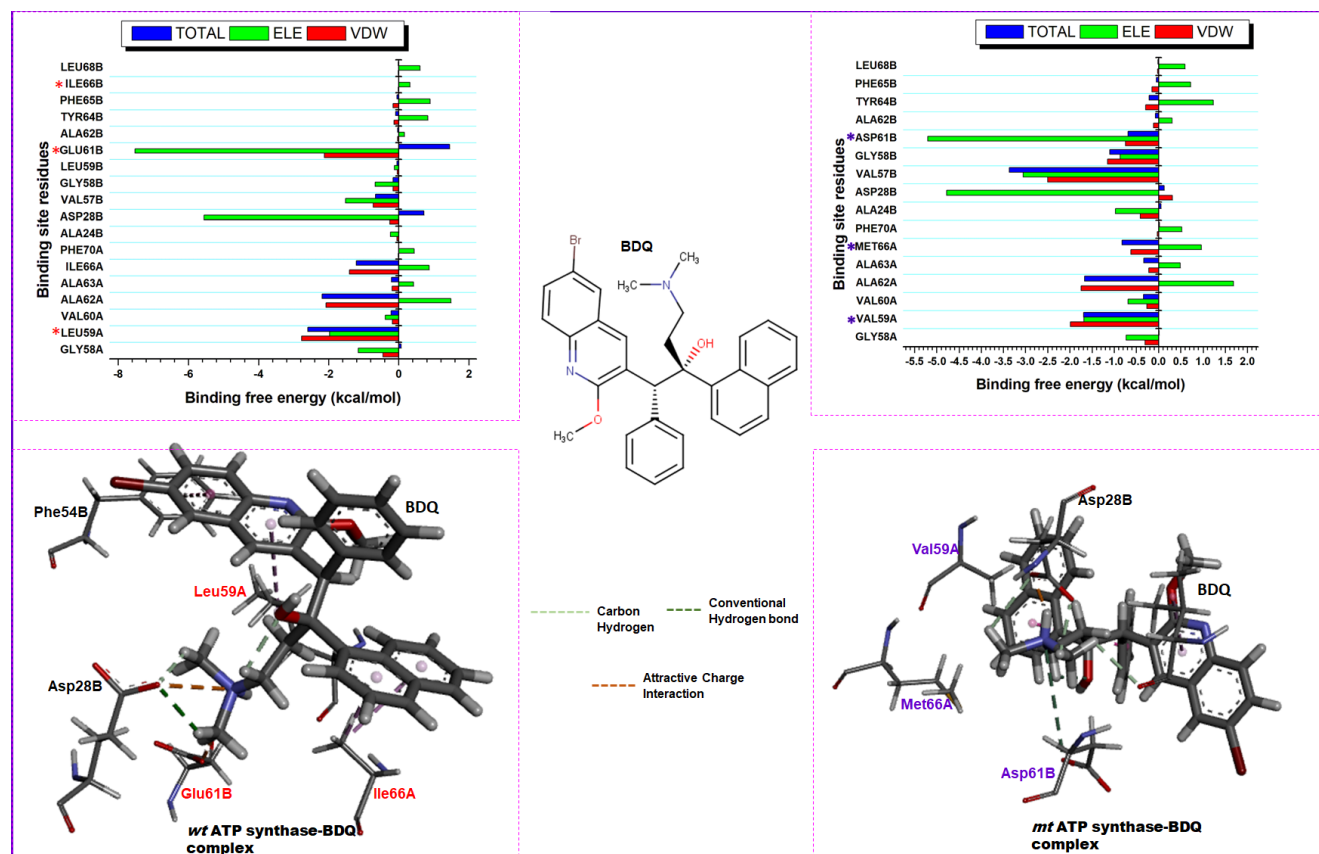


Figure 10.2: Per-residue decomposition plots showing individual energy contributions to the binding of BDQ with much emphasis on the residues in the wild ATP synthase-BDQ system and the corresponding mutants in the mutant ATP synthase-BDQ system. Corresponding inter-molecular interactions between the wild and mutant systems is shown as well.

Moreover, while this was an OH--O H-bond type (conventional), *mt*Asp61_B mediated an unconventional CH--O bond with BDQ which is less strong compared to the conventional types. This correlates with a reduction in ELE energy among both residues as shown in Table 10.1. Common to both is the attractive charge interaction mediated by Asp28B in both *wt* and *mt* variants as shown in the per-residue plot in Figure 10.2.

Mutational-induced alterations in the primary or wildtype interactions mediated by BDQ could also correlate with the reduction in total energies observed in the mutant system. Hence, we can

infer that the triple mutation of *wt*Leu59_A, *wt*Ile66_A and *wt*Glu61_B destabilizes their critical involvement in BDQ binding, which could in turn account for a loss in therapeutic activity.

3.3 Triple mutation impacts BDQ binding by disorienting residue interaction network profile of *Mtb* F₁F₀ ATP synthase

The structural integrity of any enzyme largely depends on the inter-residue interactions existing between the individual residues that form the basic unit of that particular enzyme. As such any perturbation or disorientation of the residue interaction profile of the enzyme, *Mtb* F₁F₀ ATP synthase could possibly influence its biological activity and hence curtail its involvement in ATP synthesis [42, 43]. An exploration of residue-residue interaction network within *Mtb* F₁F₀ ATP synthase as performed in this report could also provide essential structural insights into the impact of BDQ binding on its structural integrity as well as reveal crucial atomistic insights into the impact of the induced triple mutation. Individual residues of *Mtb* F₁F₀ ATP synthase are represented by nodes while specific inter-residue interactions are described by edges shown in figure 10.3 and figure 10.4. Predominant interactions that exist between residues included hydrogen bonds, van der Waals interactions and ionic interactions [44]. To explore the structural impact of the induced triple mutation on BDQ binding, specific parameters described as centrality measures were taken into consideration to provide insights on the residue network profile of *Mtb* F₁F₀ ATP in its wild and mutant conformation. The assessed parameters included diameter density, betweenness, short path length, node degree and closeness centrality. The number of connections between nodes is described as node degree whereas the compactness of the enzyme network is described by the diameter (D) [43, 44]. Likewise, the sparsity of the network profile measured between 0 and 1 is described by the density parameter whereas the betweenness and closeness centrality revealed essential residues within the residue interaction network of *Mtb* F₁F₀ ATP synthase [43, 44]. The

short path length parameter provides insights on the residue flexibility and takes into consideration the interconnective edges between respective nodes[43, 44]. Results as shown in figure 10.3 for the wild conformation suggests the essentiality of (Leu59, Ile66 and Glu61) in the overall integrity of *Mtb* F₁F₀ ATP synthase considering the extensive interactions each of these residues engaged in with neighboring residues. These residues exhibit centrality and coordination by forming various interactions, notably, hydrogen bonds, van der Waals and ionic interactions with neighboring residues as shown in figure 10.3. As observed, there was also a high node degree as well as prominent betweenness and closeness centrality in the wild conformation. The observed shortness of length of the connective edges confers with a compact and relatively stable residue conformation of the BDQ bound wild conformation. The roles of these residues in imposing stability and compactness further highlights their cruciality in *Mtb* F₁F₀ ATP for which their mutation could impact inhibitor binding.

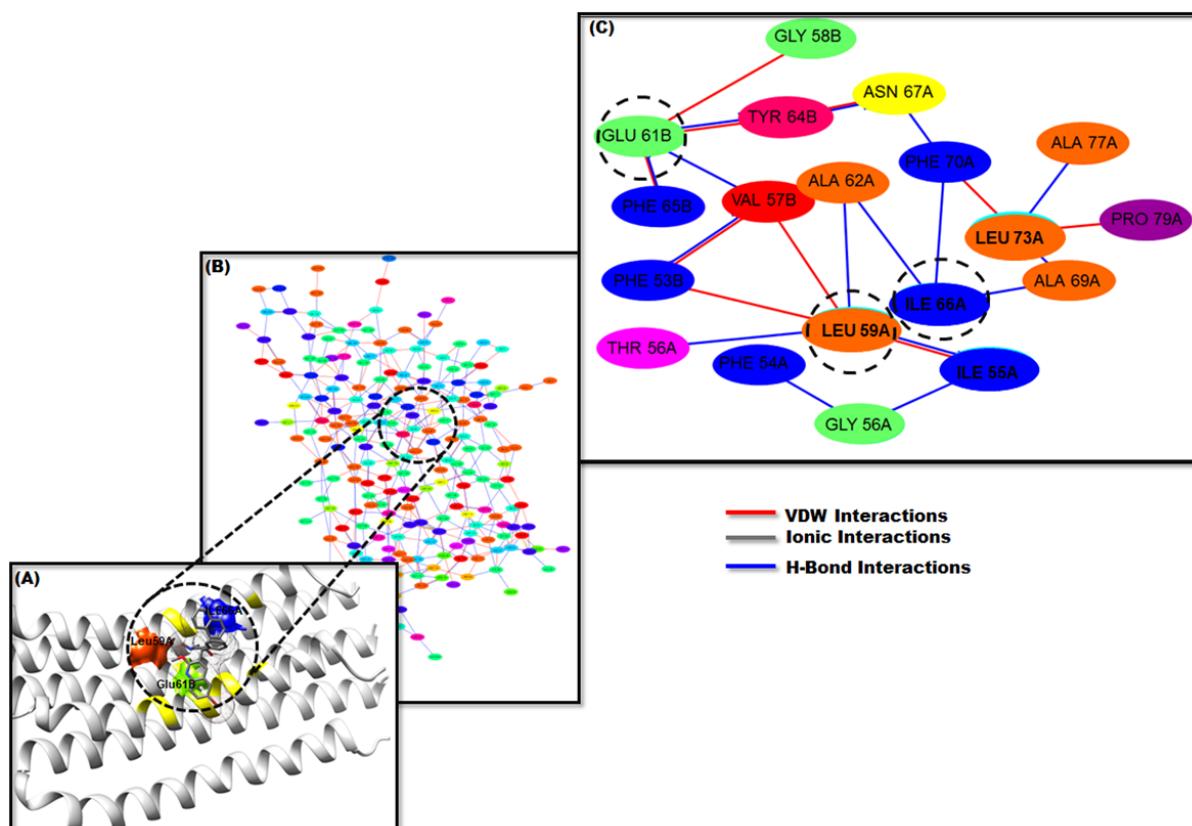


Figure 10.3: Residue Interaction network analysis of wild *Mtb* F₁F₀ ATP synthase in complex with BDQ. (A) Circled insert shows hotspot residues (Leu59A, Glu61B and Ile66A) which are subsequently mutated. (B) Network and topology representation highlighting both hotspot residues with the corresponding interacting neighboring residues. (C) Closer network view showing the residual interaction network and centrality of the hotspot residues.

In contrast to wild conformation, the mutant bound conformation exhibited varying residue interaction network profile as shown figure 10.4. Which could largely be attributed to the induced triple mutation. Although the mutated residues-maintained interactions with neighboring residues there was an apparent reduction in node degree, much longer path lengths between nodes and a corresponding reduction connectivity of nodes relative to the wild conformation. These corresponds with a relatively loose and less compact residue interaction network profile of the mutant conformation. The occurrence of these observed residue network distortions even in the

binding of BDQ suggests that, regardless of the BDQ binding, the structural integrity of ATP synthase is still disoriented. This could eventually influence BDQ binding affinity and its consequential inhibitory potency.

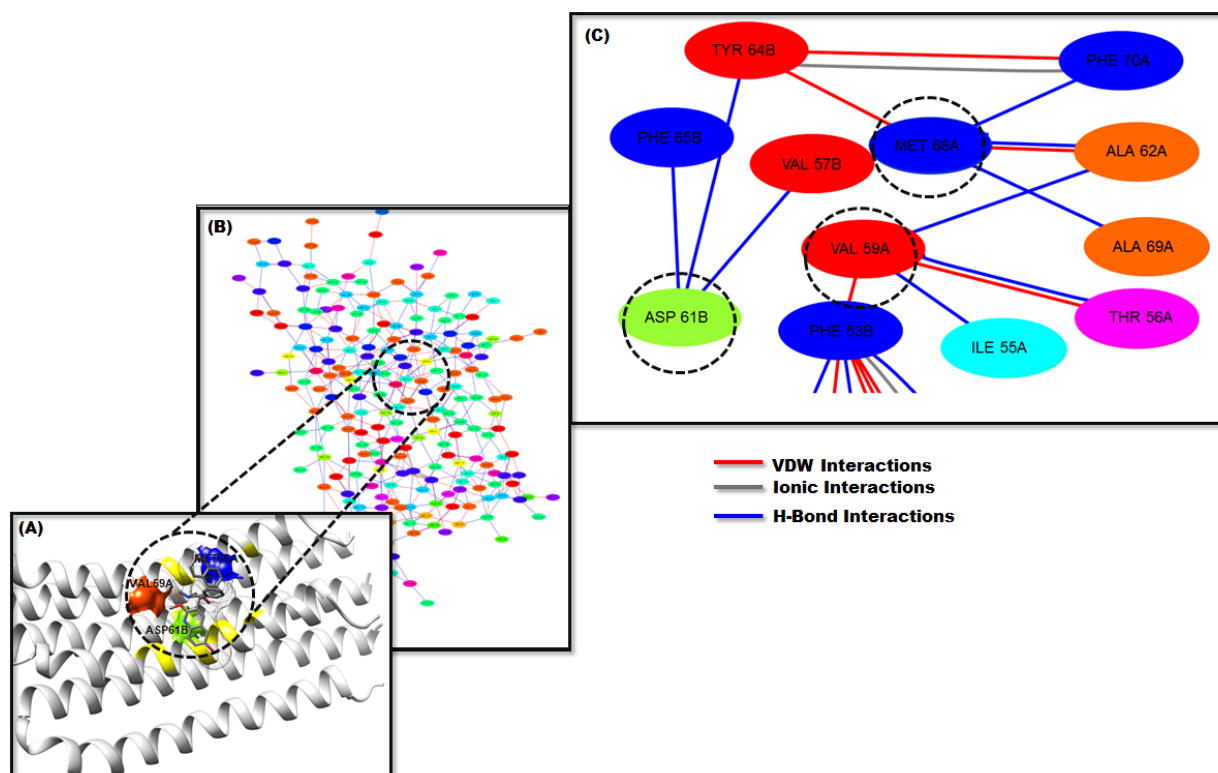


Figure 10.4: Residue Interaction network analysis of mutant *Mtb* F₁F₀ ATP synthase in complex with BDQ. (A) Circled insert shows hotspot mutant residues (Val59A, Asp61B and Met66A). (B) Network and topology representation highlighting both hotspot residues with the corresponding interacting neighboring residues. (C) Closer network view showing the residual interaction network and centrality of the hotspot residues.

To further explore the impact of the triple mutation residues interaction profile of *Mtb* F₁F₀ ATP, the hydrogen bond patterns of active site residues of the wild and mutant bound conformation of *Mtb* F₁F₀ ATP was estimated from the generated 150ns MD trajectory. The formation of hydrogen bonds between residues in a protein is very essential in the conformational dynamics of the protein structure [45]. As shown in the comparative hydrogen bonds plot in figure 10.5A, the wild bound

conformation of *Mtb* F₁F₀ ATP exhibited the highest average hydrogen bonds of 94 H-bond interactions while the mutant showcased an estimated average hydrogen bonds of 92 interactions. The decrease in hydrogen bond interaction in the mutant bound conformation was consistent with the distorted residue interaction network of the mutant complex as shown in figure 10.4. The distorted residue network profile could have impeded the formation of hydrogen bonds, hence lower average hydrogen bonds as estimated. Upon mutation, Glutamate 61(Glu), the main residue involved in ion exchange in mycobacteria [46], was replaced with Aspartate (Asp). Even though both Glu and Asp are polar and negatively charge hydrophilic amino acids that are essential in the formation of hydrogen bonds the shorter side chain of Asp could have informed its varying structural impact on *Mtb* F₁F₀ ATP. Similarly in the Leu59Val mutation, although both Leucine and Valine are aliphatic hydrophobic amino acids, buried in the core of the enzyme structure, the non-reactive side chain on Valine [46] which rarely involves in protein function could have accounted for the lower hydrogen bond formation. More so, unlike most amino acids which possess only one non-hydrogen substituent attached to the α -carbon, Valine contains two non-hydrogen substituents. This property of Valine could have also influenced its formation of hydrogen interactions with neighboring residues and a consequential decrease in hydrogen bond formation within the active site of the mutant *Mtb* ATP synthase. Mutation of Ile66 with methionine (Met) could have also influenced the decreased hydrogen bond formation in the mutant bound complex. Even though both are favorable to be substituted by each other due to their hydrophobic nature, the sulphur atom on methionine possibly impeded its connection with hydrogen atoms but rather favored its interaction with methyl groups and consequently limiting its reactivity within *Mtb* F₁F₀ ATP synthase. Taken together, the triple mutation structurally distorted

the conformation integrity of *Mtb* F₁F₀ ATP synthase and consequential reduced its overall hydrogen bonding and probably the binding affinity of BDQ.

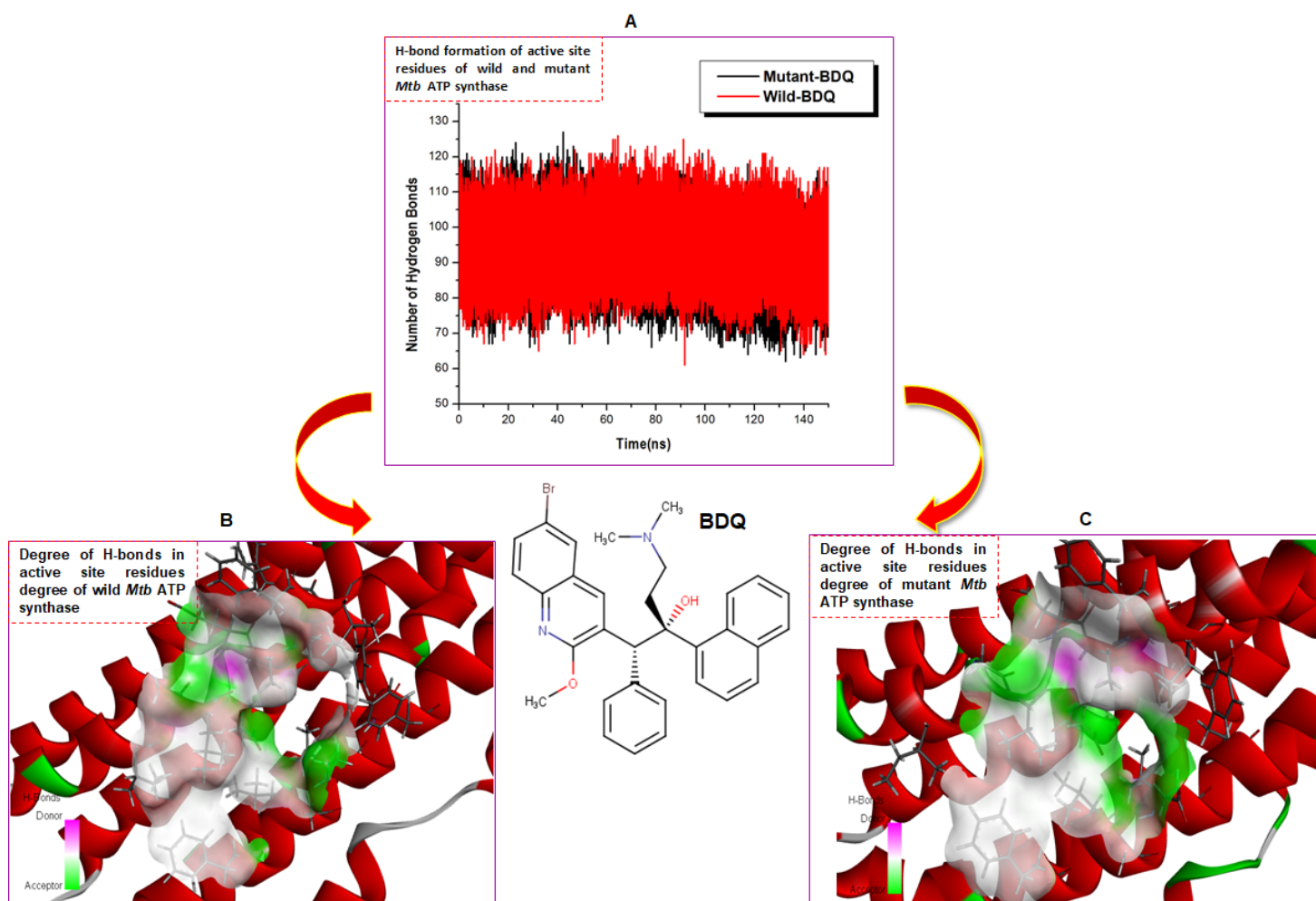


Figure 10.5: (A) Comparative hydrogen bond plots of within active site region of the wild and mutant bound *Mtb* ATP synthase. (B) 3D representation of the degree of hydrogen bond donors (magenta) and acceptors (light green) within the active of site of the wild bound *Mtb* ATP synthase. (C) 3D representation of the degree of hydrogen bond donors (magenta) and acceptors (light green) within the active of site of the mutant bound *Mtb* ATP synthase.

3.4 Triple mutation interferes with BDQ binding affinity on *Mtb* F₁F₀ ATP synthase

Binding free energy, calculated using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) [35–38] method was estimated to assess any differences in the binding of BDQ to

Mtb F₁F₀ ATP synthase, in the presence and absence of the induced triple mutation. This provided insights on whether the mutation could probably influence BDQ binding and its overall inhibitory potency. Results obtained from the calculation reveals that the total binding energy of BDQ was higher in the wild *Mtb* F₁F₀ ATP conformation, with a total binding free energy of -30.10kcal/mol relative to a lower energy of -18.47kcal/mol in the mutant conformation as shown in Table 10.2. The lower binding of BDQ in the mutant structure could be attributed to the observed distortion in the residue interaction network of the mutant *Mtb* F₁F₀ ATP, characterized by increased residue flexibility, which could have impeded the favorable interaction crucial residues with BDQ.

Table 10.2: Estimated Binding free energy profiles of BDQ to mutant and wild *Mtb* F₁F₀ ATP synthase

Systems	Energy components (kcal/mol)				
	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
Mutant-BDQ	-36.01 ±0.06	-167.73±0.19	-203.7±0.19	173.64±0.17	-18.47±0.05
Wild-BDQ	-27.67± 0.05	-144.79 ±0.23	-172.5±0.23	153.98±0.23	-30.10±0.07

ΔE_{ele} = electrostatic energy; ΔE_{vdW} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG = gas phase free energy.

3.5 Exploration of the structural and conformational dynamics of wild and mutant *Mtb* F₁F₀ ATP synthase upon binding of BDQ

As reported experimentally, mutations on *Mtb* F₁F₀ ATP synthase causes direct structural interference and subsequently affects the binding of BDQ [14]. In light of this we assessed the stability and structural dynamics of the wild and mutant *Mtb* F₁F₀ ATP synthase to establish the effects of the triple mutations. To ascertain the impact of the triple mutation on structural stability of *Mtb* F₁F₀ ATP synthase, the root mean square deviation (RMSD) of its C- α atoms were calculated over the simulation period. As shown in figure 10.6A all simulated models were well stabilized and model was observed to have achieved convergence after about 20ns. The average deviations computed for C- α atoms of the mutant BDQ bound *Mtb* F₁F₀ ATP synthase was 2.83Å while the wild bound conformation exhibited an average RMSD value of 1.80Å. This suggested a more stable conformation in the wild conformation compared with the mutant which was relatively unstable. It could therefore be inferred that the residue network disorientation and residue flexibility in the mutant complex as observed in the RIN analysis possibly conferred the observed conformational instability on *Mtb* F₁F₀ ATP synthase relative to the wild bound complex. This could have in turn influenced the overall binding of BDQ considering the fact that the increased deviation of the C- α atoms in the mutant conformation could have equally impeded the interaction of crucial residues required for the binding of BDQ. To further assess the conformational dynamics of bound *Mtb* F₁F₀ ATP synthase in the wild and mutant forms, fluctuations of individual residues were assessed by calculating the root mean square fluctuation of corresponding C- α atoms as shown in figure 10.6B. The RMSF calculations also gave insights on the flexibility of each residues as was observed in the RIN analysis. Comparatively, the bound mutant complex exhibited a slightly higher average RMSF of 12.5Å while the bound wild

conformation exhibited a lower RMSF of 11.0Å. This higher flexibility of residues is also consistent with the increase in short path length of residues as observed in the RIN analysis further highlighting the impact of the mutation on residue flexibility as well as a consequential effecting on BDQ binding. In all, although the difference in average flexibility was minimal between the two models, the slightly higher flexibility in bound wild conformation, it confers with earlier assertion that mutation perturbed the conformation dynamics of *Mtb* F₁F₀ ATP synthase and hence influence BDQ binding. With the observed increased flexibility in the mutant complex, we further assessed the compactness of *Mtb* F₁F₀ ATP synthase by calculating the radius of gyration (RoG) of its C- α atoms over the simulation period. A compact or rigid enzyme structure is consistent with reduced residue flexibility of the enzyme [47]. As shown in figure 10.6C the mutant complex exhibited a slightly higher average RoG 20.4Å relative to the wild complex which exhibited an average RoG of 20.3Å. Though minimal, the slight less compact conformation of the mutant *Mtb* F₁F₀ ATP synthase complex due to its relatively higher RoG conferred with its slightly higher overall enzyme flexibility as observed in the RMSF and RIN analysis. In summary the effect of the induced triple mutations on F₁F₀ ATP synthase was definitive in causing structural alterations characterized by the high C- α deviations, increased residue flexibility and less compact enzyme structure. These structural alterations could in turn have interfered with BDQ binding, hence its lower binding free energy as estimated in the mutant complex.

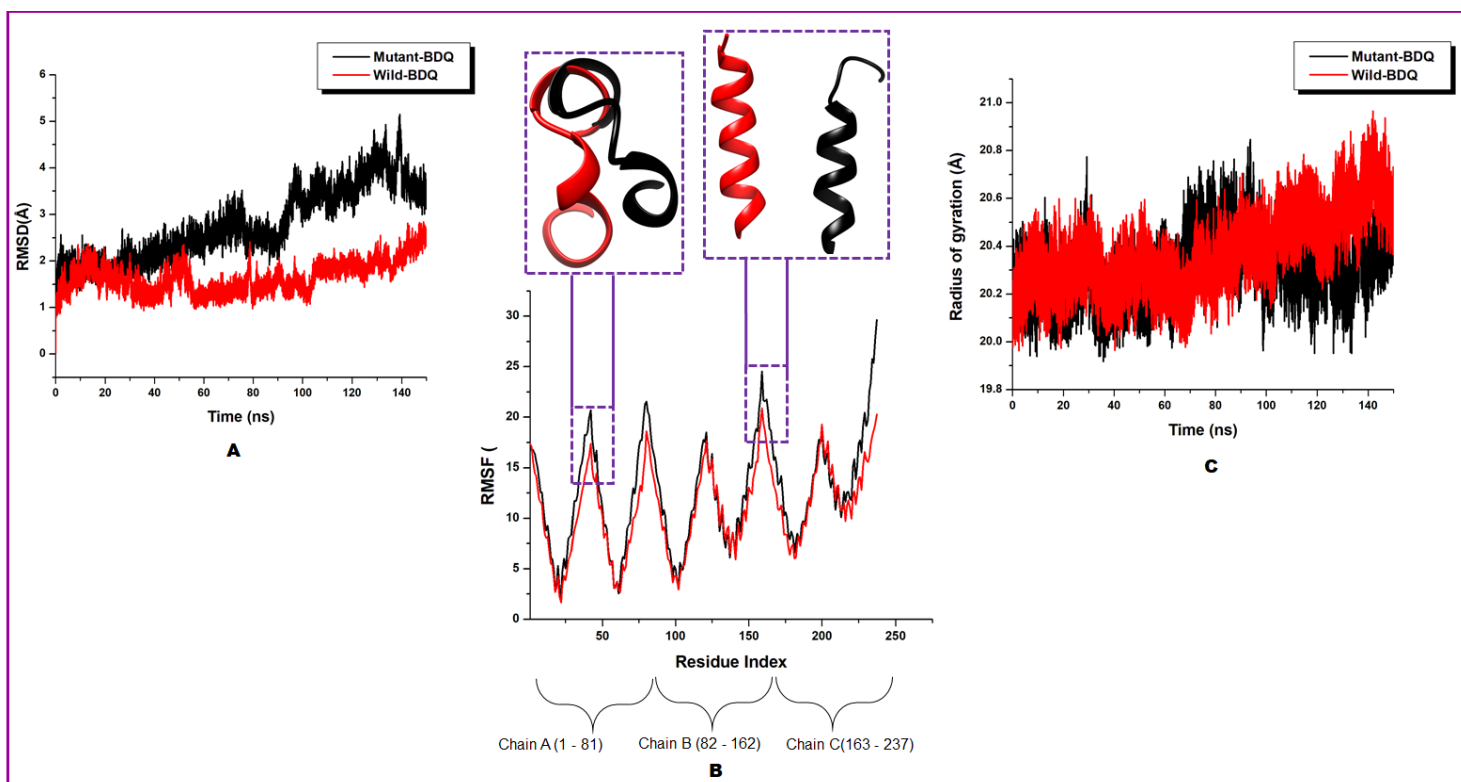


Figure 10.6: **A)** Comparative C- α RMSD plots wild and mutant BDQ bound *Mtb* ATP synthase. **B)** RMSF plots showing the degree of residue fluctuations that occurred on both wild and mutant *Mtb* ATP synthase over the simulation period. Inset shows 3D representation of some prominent residue fluctuation regions (region 37-67 and 160-175). **C)** C- α Rg plot showing distinct structural compactness and rigidity of wild and mutant bound *Mtb* ATP synthase.

4.0 Conclusion

Multi-Drug resistance TB remain a public health concern and the continuous search for various therapeutic approaches that will counter this menace is on the rise. The recent approval of BDQ for the treatment of MDR-TB has shown great advancement toward attenuating the disease. By inhibiting ATP production in *Mtb*, BDQ deprives the microorganism of energy to carry out metabolic activities leading to cell death. However emerging reports suggests resistance to BDQ in *Mtb* to be attributed to some notable mutations {59(Leu \rightarrow Val), 61(Glu \rightarrow Asp) and 66(Ile \rightarrow Met)} in the amino acid sequence of the mycobacterium. In this study we employed computational

methods to provide insights into the effect of mutations on the inhibitory prowess of BDQ by ascertaining any mutation induced structural perturbations as well as the residue interaction network of *Mtb* F₁F₀ ATP synthase. Per-residue energy decomposition analysis revealed that, upon mutation, there was reduction in the total energies contributed by the residues of interest towards the binding of BDQ with a corresponding loss of crucial interactions such as hydrogen and pi-alkyl interactions that were formed with BDQ in the wild complex. A further exploration of the impact of the mutation revealed, the triple mutations induced, altered BDQ binding by distorting the structural integrity of *Mtb* F₁F₀ ATP synthase. This was characterized by an increase in short path lengths of edges between nodes as observed in the RIN analysis and a corresponding increase in individual residue flexibility as shown with increased C- α atom fluctuation and deviation of the mutant complex relative to the wild complex. The disorganization of the structurally compact wild *Mtb* F₁F₀ ATP synthase complex into a more flexible conformation upon mutation was further exhibited by a decrease in hydrogen bond formation in the mutant complex relative to the wild complex. These findings provide structural and atomistic insights into the experimental reports that indicate that these triple mutations could interfere with BDQ binding and consequently prevent its ionophoric role of halting ion transfer as an anti-TB agent. Conclusions from this report could also aid in the design of new *Mtb* F₁F₀ inhibitors that could overcome resistance in MDR-TB.

5.0 Acknowledgements

The authors acknowledge the School of Health Science, University of KwaZulu-Natal, Westville campus for financial assistance, and The Centre of High-Performance Computing (CHPC, www.chpc.ac.za), Cape Town, Republic of South Africa, for computational resources.

6.0 Conflict of Interest

Authors declare no conflict of interests.

7.0 References

1. World Health Organization. (2018). *Global tuberculosis report 2018*.
2. Deoghare, S. (2013). Bedaquiline: A new drug approved for treatment of multidrug-resistant tuberculosis. *Indian Journal of Pharmacology*, 45(5), 536. doi:10.4103/0253-7613.117765
3. Gandhi, N. R., Nunn, P., Dheda, K., Schaaf, H. S., Zignol, M., van Soolingen, D., Bayona, J. (2010). Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *The Lancet*. 375(9728): 1830-43 doi:10.1016/S0140-6736(10)60410-2
4. WHO. "Totally Drug-Resistant TB ".*WHO consultation on the diagnostic definition and treatment options*. (2012).
5. Ndjeka, N., Conradie, F., Schnippel, K., Hughes, J., Bantubani, N., Ferreira, H., ... Pillay, Y. (2015). Treatment of drug-resistant tuberculosis with bedaquiline in a high HIV prevalence setting: An interim cohort analysis. *International Journal of Tuberculosis and Lung Disease*.19(8):979-8 doi:10.5588/ijtld.14.0944
6. Andries, K., Verhasselt, P., Guillemont, J., Göhlmann, H. W. H., Neefs, J. M., Winkler, H., ... Jarlier, V. (2005). A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science*, 307(5707), 223–227. doi:10.1126/science.1106753
7. Hards, K., Robson, J. R., Berney, M., Shaw, L., Bald, D., Koul, A., Cook, G. M. (2015). Bactericidal mode of action of bedaquiline. *Journal of Antimicrobial Chemotherapy*.70(7):202-37 doi:10.1093/jac/dkv054
8. Hards, K., McMillan, D. G. G., Schurig-Briccio, L. A., Gennis, R. B., Lill, H., Bald, D., &

- Cook, G. M. (2018). Ionophoric effects of the antitubercular drug bedaquiline. *Proceedings of the National Academy of Sciences*, *115*(28), 7326–7331. doi:10.1073/pnas.1803723115
9. Walker, J. E. (2013). The ATP synthase: the understood, the uncertain and the unknown. *Biochemical Society Transactions*. *41*(1):1-16 doi:10.1042/BST20110773
 10. Boyer, P. D. (1997). the Atp Synthase a Splendid Molecular Machine. *Annual Review of Biochemistry*, *66*(1), 717–749. doi:10.1146/annurev.biochem.66.1.717
 11. Junge, W., Sialaff, H., & Engelbrecht, S. (2009). Torque generation and elastic power transmission in the rotary F₀F₁-ATPase. *Nature*, *459*(7245), 364–370. doi:10.1038/nature08145
 12. Noji, H., Yasuda, R., Yoshida, M., & Kinosita, K. (1997). Direct observation of the rotation of F₁-ATPase. *Nature*. *386*(6622):299-302 doi:10.1038/386299a0
 13. Meier, T., Polzer, P., Diederichs, K., Welte, W., & Dimroth, P. (2005). Structure of the Rotor ring of F-Type Na⁺-ATPase from *Ilyobacter tararicus*. *Science*, *308*(5722), 659- 662.
 14. Preiss, L., Langer, J. D., Yildiz, Ö., Eckhardt-Strelau, L., Guillemont, J. E. G., Koul, A., & Meier, T. (2015). Structure of the mycobacterial ATP synthase Forotor ring in complex with the anti-TB drug bedaquiline. *Science Advances*, *1*(4), 1–8. doi:10.1126/sciadv.1500106
 15. Segala, E., Sougakoff, W., Jarlier, V., & Petrella, S. (2012). Synthase : New Insights into the Binding of Synthase C-Ring Structure New Mutations in the Mycobacterial ATP Synthase : New Insights into the Binding of the Diarylquinoline TMC207 to the ATP Synthase. *56*(5):2326-34 doi:10.1128/AAC.06154-11
 16. Petrella, S., Cambau, E., Chauffour, A., Andries, K., Jarlier, V., Sougakoff, W., & Curie, M. (2006). Genetic Basis for Natural and Acquired Resistance to the Diarylquinoline R207910 in Mycobacteria, *50*(8), 2853–2856. doi:10.1128/AAC.00244-06
 17. Ramharack, P., & Soliman, M. E. S. (2018). Zika virus NS5 protein potential inhibitors:

an enhanced in silico approach in drug discovery. *Journal of Biomolecular Structure and Dynamics*. 36(5):1118-1133 doi:10.1080/07391102.2017.1313175

18. Gooch, J. W. (2011). Ramachandran Plot. In *Encyclopedic Dictionary of Polymers*. doi:10.1007/978-1-4419-6247-8_14641

19. Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(Web Server issue), W407–W410. doi:10.1093/nar/gkm290

20. Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M., & IUCr. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291. doi:10.1107/S0021889892009944

21. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J Comput Chem*, 25, 1605–1612. doi:10.1002/jcc.20084

22. Huang, B. (2009). MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. *OMICS: A Journal of Integrative Biology*. 13(4):325-30 doi:10.1089/omi.2009.0045

23. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). AutoDock-related material Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *Comput. Chem. J. Comput. Chem*, 19(28), 1639–1662. doi:10.1002/jcc.20634

24. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. doi:10.1002/jcc.20084

25. Kusumaningrum, S., Budianto, E., Kosela, S., Sumaryono, W., & Juniarti, F. (2014). The

- molecular docking of 1,4-naphthoquinone derivatives as inhibitors of Polo-like kinase 1 using Molegro Virtual Docker. *Journal of Applied Pharmaceutical Science*, 4(11), 47–53. doi:10.7324/JAPS.2014.4119
26. Dunbrack, R. L. (2002). Rotamer libraries in the 21st century. *Current Opinion in Structural Biology*, 12(4):430-40 doi:10.1016/S0959-440X(02)00344-5
27. Grand, S. Le, Götz, A. W., & Walker, R. C. (2013). SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Computer Physics Communications*, 184(2), 374–380. doi:http://dx.doi.org/10.1016/j.cpc.2012.09.022
28. Olotu, F. A., Agoni, C., Adeniji, E., Abdullahi, M., & Soliman, M. E. (2018). Probing Gallate-Mediated Selectivity and High-Affinity Binding of Epigallocatechin Gallate: a Way-Forward in the Design of Selective Inhibitors for Anti-apoptotic Bcl-2 Proteins. *Applied Biochemistry and Biotechnology*. 187(3):1061-1080 doi:10.1007/s12010-018-2863-7
29. Oguntade, S., Ramharack, P., & Soliman, M. E. (2017). Characterizing the ligand-binding landscape of Zika NS3 helicase-promising lead compounds as potential inhibitors. *Future Virology*, 12(6), 261–273. doi:10.2217/fvl-2017-0014
30. Agoni, C., Ramharack, P., & Soliman, M. E. S. (2018). Synergistic Interplay of The Co-administration of Rifampin And Newly Developed Anti-TB Drug : Could It Be a Promising New Line of TB Therapy ?, 21(6) 453–60. doi:10.2174/1386207321666180716093617
31. Agoni, C., Ramharack, P., & Soliman, M. (2018). Co-inhibition as a strategic therapeutic approach to overcome rifampin resistance in tuberculosis therapy: atomistic insights. *Future Medicinal Chemistry*, 10(4). doi:10.4155/fmc-2017-0197
32. Olotu, F. A., & Soliman, M. E. S. (2018). Dynamic perspectives into the mechanisms of mutation-induced p53-DNA binding loss and inactivation using active perturbation theory:

Structural and molecular insights toward the design of potent reactivators in cancer therapy. *Journal of Cellular Biochemistry*. 120(1):951-966 doi:10.1002/jcb.27458

33. Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.

34. Seifert, E. (2014). OriginPro 9.1: scientific data analysis and graphing software-software review. *Journal of chemical information and modeling*, 54(5), 1552. doi:10.1021/ci500161d

35. Massova, I., & Kollman, P. A. (2000). Combined molecular mechanical and continuum solvent approach (MM-PBSA / GBSA) to predict ligand binding, 18(1), 113–135.

36. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., ... Cheatham, T. E. (2000). Calculating Structures and Free Energies of Complex Molecules : Combining Molecular Mechanics and Continuum Models, 33(12), 889–897.

37. Miller, B. R., Mcgee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). MMPBSA . py : An E ffi cient Program for End-State Free Energy Calculations. 8(9):3314-21 doi:10.1021/ct300418h

38. Onufriev, A., Bashford, D., & Case, D. A. (2000). Modification of the Generalized Born Model Suitable for Macromolecules, 104(15): 3712–3720.

39. Sitkoff, D., Sharp, K. A., & Honig, B. (1994). Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models, 98(7): 1978–1988.

40. Ricatti, J., Acquasaliente, L., Ribaud, G., De Filippis, V., Bellini, M., Llovera, R. E., ... Mucignat-Caretta, C. (2019). Effects of point mutations in the binding pocket of the mouse major urinary protein MUP20 on ligand affinity and specificity. *Scientific Reports*. 9(1):300 doi:10.1038/s41598-018-36391-3

41. Aldeghi, M., Gapsys, V., & De Groot, B. L. (2018). Accurate Estimation of Ligand Binding

Affinity Changes upon Protein Mutation. *ACS Central Science*. 4(12):1708-18
doi:10.1021/acscentsci.8b00717

42. Hu, G., Yan, W., Zhou, J., & Shen, B. (2014). Residue interaction network analysis of Dronpa and a DNA clamp. *J Theor Biol*, 348, 55-64.

43. Yan, W., Zhou, J., Sun, M., Chen, J., Hu, G., & Shen, B. (2014). The construction of an amino acid network for understanding protein structure and function. *Amino Acids*, 46, 1419-1439.

44. Xue, W., Jin, X., Ning, L., Wang, M., Liu, H., & Yao, X. (2012). Exploring the molecular mechanism of cross-resistance to HIV-1 integrase strand transfer inhibitors by molecular dynamics simulation and residue interaction network analysis. *J Chem Inf Model*, 53, 210-222.

45. Sheu, S.-Y., Yang, D.-Y., Selzle, H. L., & Schlag, E. W. (2003). Energetics of hydrogen bonds in peptides. *Proceedings of the National Academy of Sciences*. 100(22):12683-7
doi:10.1073/pnas.2133366100

46. Betts, M. J., & Russell, R. B. (2007). Amino-Acid Properties and Consequences of Substitutions. In *Bioinformatics for Geneticists: A Bioinformatics Primer for the Analysis of Genetic Data: Second Edition*. doi:10.1002/9780470059180.ch13

47. Abdullahi, M., Olotu, F. A., & Soliman, M. E. (2018). Allosteric inhibition abrogates dysregulated LFA-1 activation: Structural insight into mechanisms of diminished immunologic disease. *Computational Biology and Chemistry*, 73, 49–56.
doi:10.1016/j.compbiolchem.2018.02.002

CHAPTER 11

Published Article

Allosteric Inhibition Induces an Open WPD-Loop: A New Avenue Towards Glioblastoma Therapy

Clement Agoni^a, Pritika Ramharack^a, Mahmoud E. S. Soliman^{a*}

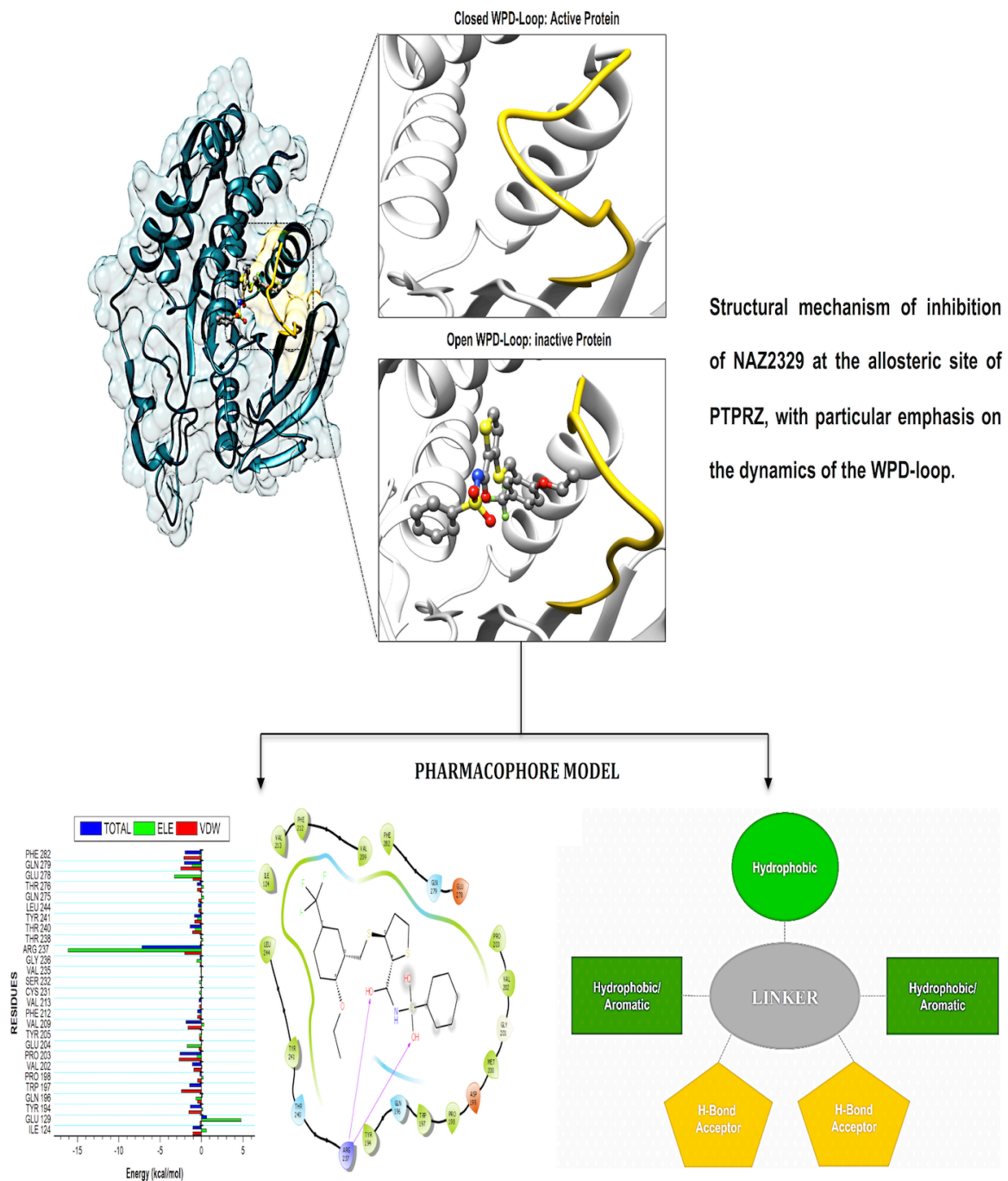
^aMolecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University
of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 8048, Fax: +27 (0) 31 260 7872

GRAPHICAL ABSTRACT



ABSTRACT

The mobility of loops around the catalytic site of a protein remains crucial to its activity. Dynamics of the WPD-loop is an essential determinant of the catalytic activity of tyrosine-protein phosphatase zeta, an implicated protein in glioblastoma cells. The WPD-loop assumes a closed conformation upon substrate binding in order to position its catalytic aspartate to participate in catalysis. Herein, we explore the impact of NAZ2329, a recently identified allosteric inhibitor of tyrosine-protein phosphatase zeta, on the atomic flexibility of the WPD-loop. The druglikeness of NAZ2329 was assessed using the SwissADME online tool. The enzymatic complex was then subjected to conformational simulations using the AMBER molecular dynamics software. Structural analysis revealed that NAZ2329 induced an open conformation of the crucial WPD-loop, consequently impeding enzyme activity even upon substrate binding. Based on the molecular interactions between of NAZ2329 and tyrosine-protein phosphatase zeta, a pharmacophore model was generated to exhibit the important functional moieties of NAZ2329. These findings provide an insightful molecular and structural mechanism in targeting tyrosine-protein phosphatase zeta as a therapeutic intervention for glioblastoma. We believe that this optimized pharmacophoric model will aid in the design of improved anti-tyrosine phosphatase agents, thus allowing for increased patient adherence.

Keywords: Glioblastoma therapy, pharmacophore, molecular simulations, WPD-loop, conformational dynamics

1. INTRODUCTION

Glioblastoma is classified as the most aggressive and frequently diagnosed central nervous system malignancy, with an annual incidence of 100,000 new cases, globally ¹. Adding to the complexity of the disease, glioblastoma is associated with poor prognosis and low survival rates. Glioblastomas are tumors that arise from astrocytes in the cerebral cortex, but may be found anywhere in the brain or spinal cord ². The tumors are highly vascularized to allow oxygen and nutrients to pass through, thus increasing its growth ³. Currently, there are no preventative therapies or cure for the disease, with only surgical resection and adjuvant chemotherapy being available ⁴.

There have been numerous studies elucidating the development of new therapeutic approaches, including an article that investigated the receptor type tyrosine-protein phosphatase zeta (PTPRZ) as a target in glioblastoma cells ⁵. The PTPRZ enzyme plays a crucial role in regulating protein tyrosine phosphorylation, thereby leading to the survival of the glioblastoma cells and promoting the growth of tumors ⁶. Due to this mechanism of action, the PTPRZ enzyme is a crucial target in the design of efficient inhibitors of against glioblastoma ³. Structurally, active site architecture of PTPRZ is made up the Phosphate binding P-loop, the catalytic acid/base aspartate containing WPD-loop, the conserved glutamine containing Q-loop, the pTyr-recognition pTyr-loop and the multiple conserved residues containing E-loop ⁷. As a vital component of the catalytic process of PTPRZ, the WPD-loop serves as a flexible gate to catalytic site which is shown to assume a closed conformation in active protein form and an open conformation in inactive protein form (Figure 11.1) ⁸.

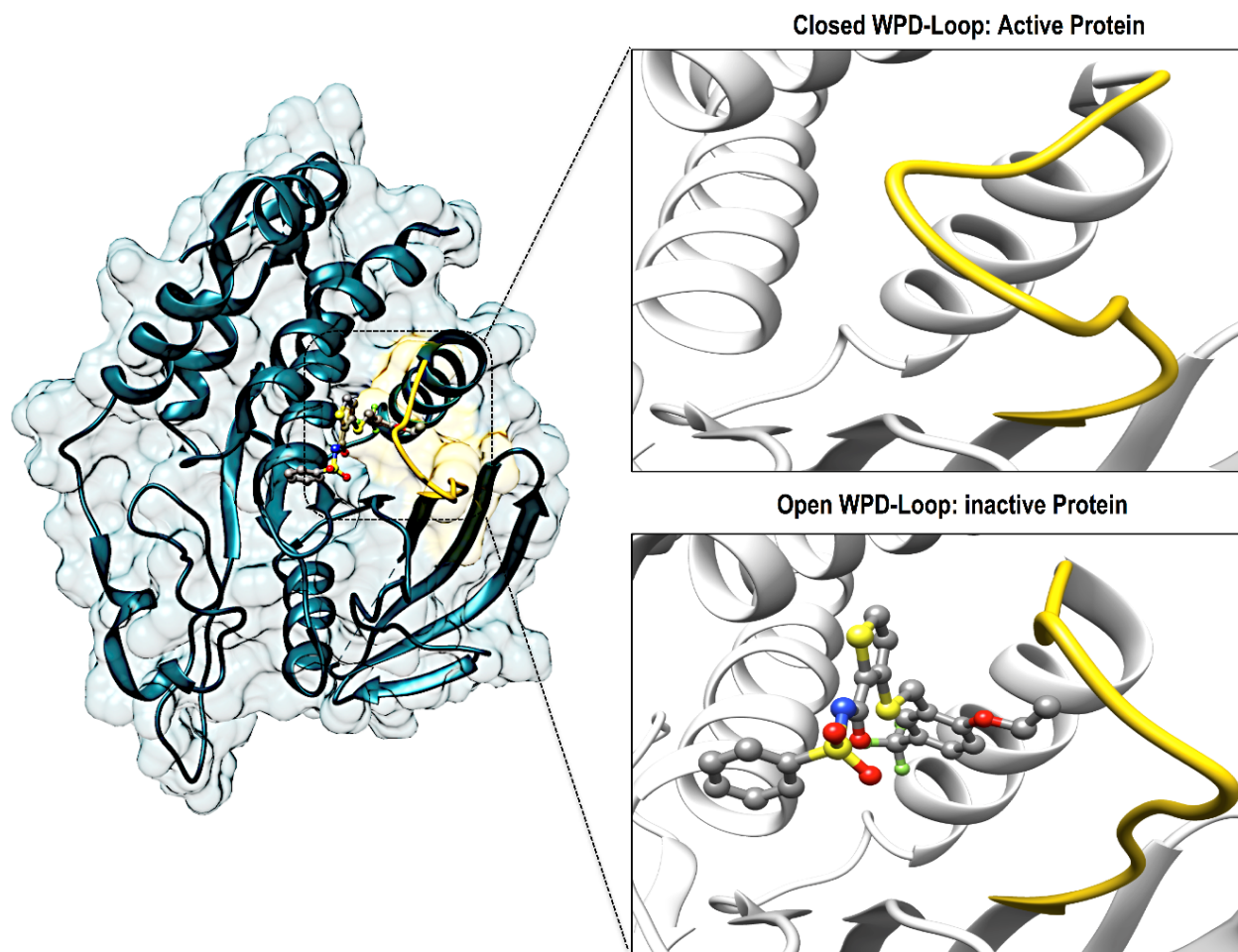


Figure 11.1: Graphical representation of the WPD-loop (yellow) of PTPRZ. The active form of the protein is associated with a closed WPD-loop upon binding of the natural substrate, whereas, an open conformation of results in an inactive protein.

Upon binding of a substrate to the active site the, the WPD-loop assumes a close conformation to position its catalytic aspartate to participate in catalysis ⁷. Therefore a compromised WPD-loop mobility can substantially decrease the catalytic activity of the overall PTPRZ protein as was reported by studies with mutated tryptophan hinge residue ^{9,10}.

There have been multiple phosphotyrosine competitive inhibitors that have been designed to inhibit the PTPRZ enzyme. These compounds, however, have failed to successfully attenuate

cancer due to the drugs inability to permeate the cell wall ¹¹. Another challenge when designing effective inhibitors of the enzyme is its highly conserved, positively charged active site. Due to these reasons, the enzyme was long known as an “undruggable target” ⁵.

Over recent years, targeting the allosteric site of the PTP family of enzymes has proven to be successful ¹². In 2017, a cell-permeable small molecule, NAZ2329, was identified to allosterically inhibit both the PTPRZ and PTPRG, thus mitigating the tumorigenicity in glioblastoma cells ¹³. This allosteric inhibition strategy has since then shown to be extremely promising in glioblastoma drug design.

Here we assess, through the use of *in silico* tools, the structural mechanism of inhibition of NAZ2329 at the allosteric site of PTPRZ, with particular emphasis on the dynamics of the WPD-loop. Also, a pharmacophore model was design based on the binding profile and structure activity relationship of NAZ2329 with PTPRZ. This pharmacophore model approach will facilitate the design of small molecule inhibitors that will not only target PTPRZ and PTPRG but will be applicable to other tyrosine phosphatases as well.

2. COMPUTATIONAL METHODS

2.1 Exploring the Drug Likelihood of NAZ2329

SwissADME ¹⁴, an online software, was used to assess the physicochemical descriptors, the pharmacokinetic features and the drug-worthy nature of NAZ2329. In computing the lipophilicity and polarity of NAZ2329, the “Brain Or Intestinal Estimated permeation, (BOILED-Egg)” method was employed ¹⁴.

2.2 Preparation of PTPRZ System

To identify the most suitable crystal structure for molecular dynamic simulations, all PTPRZ structures available on RSCB Protein Data Bank ¹⁵ were assessed. Of the eight structures, only four were of human origin. As our study focused on the NAZ2329 molecule bound to the allosteric site of PTPRZ, the PDB code 5AWX was chosen as the most suitable crystal structure. Based on the literature accompanying the crystal structure, it was noted that the free enzyme demonstrated a closed WPD-loop, whilst an open WPD-loop was depicted in the NAZ2329-bound complex. This justified the use of both the bound crystal structure as well as the free enzyme (PDB codes: 5AWX and 5H08, respectively) ¹⁶.

To validate the binding site of PTPRZ, the NAZ2329 molecule was extracted from the complex using the UCSF Chimera software package ¹⁷. Molecular docking was then carried out using the Autodock Vina plugin on the Chimera software, where default settings were used. The grid box was defined around the following key residues Glu1981, Arg1939, Gly1938, Pro1905, Glu1898, Trp1899, Val1911 and Glu1977, which covered the allosteric region of PTPRZ. The X, Y and Z centre values of the grid box were defined as 30.58, -2.11 and 43.21, respectively, whereas, the X, Y and Z size dimensions were defined as 10.8, 14.3 and 11.4, respectively. Docking results, of binding affinity -9.7 kcal/mol, indicated significantly similar binding poses of the original crystal structure to the docked NAZ2329 (Figure S1). This validated the docking pose, thus allowing further analysis to be carried out. The structures of PTPRZ and NAZ2329 were then prepared and missing residues modeled using the UCSF Chimera software package ¹⁷.

2.3 Molecular Dynamic (MD) Simulations

Molecular dynamic (MD) simulations present a vigorous tool to explore the physical movements of atoms and molecules, and in so doing, unveiling dynamical evolution of biological systems. Using the GPU version of the PMEMD engine provided with the AMBER package, MD simulation was carried out in which the FF14SB variant of the AMBER force field ¹⁸.

Atomic charges for NAZ2329 were created by employing the Restrained Electrostatic Potential (RESP) and the General amber Force Field (GAFF) in ANTECHAMBER. Neutralization of all systems and addition of all hydrogen atoms was performed by using the Leap module incorporated in the AMBER 14 package. Na⁺ and Cl⁻ served as counter ions for the neutralization process. As per Leap's default settings, the amino acids were renumbered in a sequential order, so that residue "G1696" was renamed as "G1". Results were thus displayed as per renamed residues. Using TIP3P water molecules of 8Å box size, all systems were implicitly solvated.

Minimizations of the systems were carried out in two separate stages. The first stage involved a 2000 steps minimization with an incorporated restraint potential of 500kcal/mol Å². The second stage involved a 1000 steps full minimization process incorporating a conjugate gradient with no restraint.

All systems were then steadily heating from 0K to 300K over 50ps, in a manner that ensured that all systems such that the systems conserved a fixed atom number and volume. Solutes in the systems were given a potential harmonic restraint. A potential harmonic restraint of 10kcal/mol with a collision frequency of 1.0ps was imposed on solutes in all systems. Equilibrations of all systems were then performed after heating over a 500ps period at a constant operating temperature of 300K. Constant pressure and atom numbers were also ensured by mirroring an isobaric-

isothermal ensemble (NPT). A 1 bar pressure was maintained for all systems using the Berendsen barostat.

A 100ns MD simulation was carried out on all systems in which the SHAKE algorithm was incorporated to constrict bonds of hydrogen atoms. A 2fs simulation step coupled with a SPFP precision model was used. The simulations coincided with isobaric-isothermal ensemble (NPT), with randomized seeding, constant pressure of 1 bar maintained by the Berendsen barostat, a pressure-coupling constant of 2ps, a temperature of 300K and Langevin thermostat with collision frequency of 1.0ps.

2.4 Post-Dynamic Analysis

The PTRAJ module of AMBER14 suit was used to analyze all trajectories generated from coordinates at every 1ps from all simulated systems. Also, the CPPTRAJ module of AMBER14 package was used to analysis of RMSD, RMSF and Radius of Gyration.

2.4.1 Binding Free Energy Calculations

The binding free energy generated in the simulated systems was estimated using the Molecular Mechanics/GB Surface Area approach as has been employed in some of our previous reports¹⁹⁻²¹. The estimated binding free energies may expound the binding mechanism between NAZ2329 and PTPRZ. From the 100ns simulation trajectory, 10000 snapshots were obtained and subsequently used to estimate the binding free energy. By the MMGBSA approach, the binding free energy is estimated for the complex, ligand and receptor as follows;

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

$$\Delta G_{\text{bind}} = E_{\text{gas}} + G_{\text{sol}} - TS \quad (2)$$

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

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{SA}} \quad (4)$$

$$G_{\text{SA}} = \gamma \text{SASA} \quad (5)$$

E_{gas} represents gas-phase energy and is made up of the sum total of the internal energy E_{int} ; Coulomb energy E_{ele} and the van der Waals energies E_{vdw} . G_{sol} represents the solvation energy which is estimated by the sum total of polar state energy contributions, G_{GB} and non-polar energy contributions, G_{SA} . Using a water probe radius of 1.4 Å, G_{SA} is estimated from the solvent accessible surface area (SASA). G_{GB} is determined from the GB equation. Total entropy of solute and temperature is represented as S and T respectively. The accuracy of the estimated relative binding free energies reported here might be enhanced if the terms in the equation (2) are averaged over multiple conformations or MD snapshots²², but this usually dependent on the research interest under consideration. Conducting MD simulations for NAZ2329, PTPRZ and the complex will yield more accurate results in the calculating the binding free energies; it requires greater computational resources, which were not readily available for this study.

3.4.2 Dynamic Cross-correlation Analysis (DCC)

The correlation coefficient of motions between the atoms in PTPRZ over the simulation period were quantified by calculating the dynamic cross correlation matrix²³. DCC was performed using CPPTRAJ module of the AMBER 14 suite. The formula used to describe dynamic cross correlation is given below:

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle)^{\frac{1}{2}}}$$

Where, i and j represents the i^{th} and j^{th} residue respectively. The displacement vectors that correspond to the i^{th} and j^{th} residue respectively is represented as Δr_i or Δr_j . The Origin software²⁴ was then used to plot the DCC matrix.

2.4.3 Principal Component Analysis (PCA)

Principal component analysis (PCA), is a covariance-matrix-based approach that is used to show the displacement of atoms and the dynamics of loops of a protein²⁵. Using the PTRAJ module of the AMBER14 package, solvent water molecules and neutralizing ions added by the Leap module are stripped prior to MD trajectory generation. There is then an alignment of the stripped trajectories against their respective fully minimized structures. PCA was performed for C- α atoms on 1000 snapshots each. The first two principal components and covariance matrices were calculated using scripts developed with our research group. PCA calculations were conducted over 1000 snapshots each for C- α atoms. PC1 and PC2, which represent the first two principal components, are created from the trajectories averaged from the unbound PTPRZ and the bound system. Using Cartesian coordinates of C- α atoms a 2X2 covariance matrix is created. The first two eigen vectors of covariant matrices correspond with the created PC1 and PC2. The Origin software²⁴ was then employed to create PCA plots.

2.5 Pharmacophore Model Creation

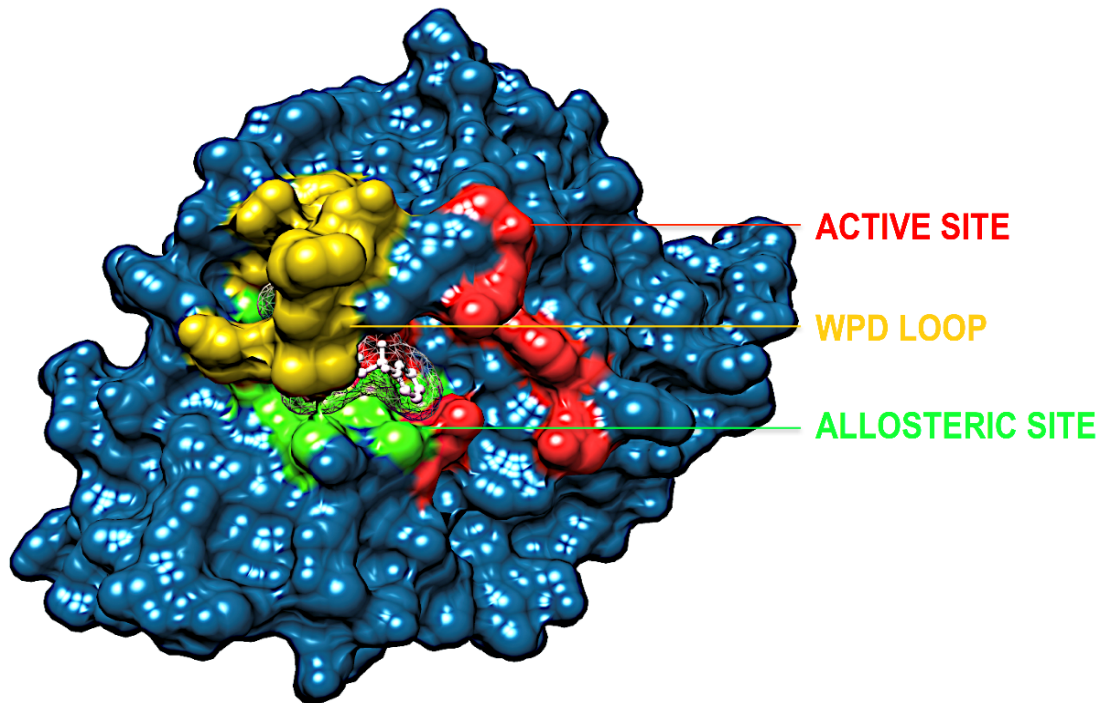
Following the simulation of NAZ2329 at the active site of PTPRZ, per-residue energy decomposition analysis was used to determine the amino acids that contribute the most towards the binding of NAZ2329. In constructing the pharmacophore in this study, the pharmacophoric moieties that exhibited prominent interaction with highest energy contributing amino acids were

used. In validating and generating our model, it was uploaded on the ZincPharmer ²⁶ and LigandScout ²⁷.

3 RESULTS AND DISCUSSION

3.1 Sequence analysis and structural stability of PTPRZ

To understand the structural mechanism of inhibition of PTPRZ, it is important to identify the fundamental structural characteristics of the protein. Based on previous studies, it has been established that the active and allosteric sites may be found adjacent to each other. The WPD-loop, a vital catalytic site regulator, forms an outer cover to part of the allosteric site, indicating that the ligand that binds to this allosteric pocket would govern the dynamics of this loop. Figure 11.2 graphically represents these unique structural features, including the corresponding amino acids for each region.



```

PTPRZ- 1696 GPAIP I KHFPKHVADLHASSGFTEEFEEVQSCTVDLGI TADSSNHPDNKH
PTPRZ- 1746 KNRY INIVAY DHS RVKLAQLAEKDGKLT DY INANYVDGYNRPKAY IAAQG
PTPRZ- 1796 PLKSTAEDFWRMIWEHNVEV IVM I TNLVEKGRRKCDQYWP ADGSEEYGNF
PTPRZ- 1846 LVTQKSVQVLAYYTVRNFTLRNTK I KKGSKGRPSGRVV TQYHYTQWPD
PTPRZ- 1896 GVPEYSLPVLTFVRK AAYAKRHAVGPVVVHCSAGVGRTGTY IVLDSMLQQ
PTPRZ- 1946 IQHEGTVNI FGFLKH I RSQRNYLVQTEEQYV FI HDTLVEA ILSKETEV

```

Figure 11.2: Structural and corresponding sequence representation of the unique regions of the PTPRZ protein; active site-red, allosteric site- green and WPD-loop- yellow (PDB Code: 5H08)

16.

PTPRZ as a protein expressed in the central nervous system is localized in the glial cell where it reportedly mediates cell adhesion signaling events during neurogenesis. The strong expression of PTPRZ glioblastoma ²⁸ renders them as viable anti-cancer therapeutic targets treatment. The inhibitory activity of NAZ2329 as experimentally reported, could potentially influence the conformational stability of PTPRZ ⁵.

As such, the stability of 3-D backbone atoms of the simulated APO and complexed PTPRZ was ascertained by calculating the root mean standard deviation (RMSD) of the generated trajectories over the 100ns simulation period. Calculated RMSD also assessed the convergence of the respective systems as depicted in Figure S1, with each system assuming an energetic plateau after about 22ns. An initial increase in RMSD to 3.4Å in the APO system between 0 to 20ns was observed, illustrating the dynamic conformational changes associated with the expansion of PTPRZ over that period. The expansion allowed infiltration of solvent molecules in hydrophobic pockets. However, after about 20ns, all systems rendered energetically stable. The average RMSD for both APO and complexed PTPRZ was 2.0Å and 1.6Å respectively. With the average RMSD less than 2.0Å, it was deduced that the systems attained conformational stability²⁹⁻³¹. The lower RMSD of the NAZ2329-complex relative to the APO system suggest that the inhibitory activity of NAZ2329 possibly induced conformational dynamics that contributed to the system achieving stability at a lower RMSD.

3.2 Investigation of the dynamic structural features of PTPRZ WPD-Loop

Considering the essential role of the WPD-loop in the catalytic activity of PTPRZ, we explored the atomic flexibility of the residues this loop and the entire protein upon binding NAZ2329. The use of the crystal structures chosen for this particular study was of utmost importance as the binding of NAZ2329 to the allosteric site demonstrated a modification in the crystallized WPD-loop of the protein. To the best of our knowledge, this is the only available crystal structure that

demonstrates binding to the allosteric site of the PTPRZ enzyme, thus validating its use for further analysis.

The conformational dynamics of a protein is usually largely dependent on its basic building blocks, thus amino acids ³², hence understanding the flexibility of amino acids that make up WPD-loop and PTPRZ could reveal important structural insights on the inhibitory activity of NAZ2329. The ligand induced motions that occur upon the binding and interaction of NAZ2329 to binding site residues, thus triggering a significant conformational change that influence the function of the protein ³². As a result of this, we calculated the root mean square fluctuation (RMSF) of the simulated systems illustrated in Figure 11.3.

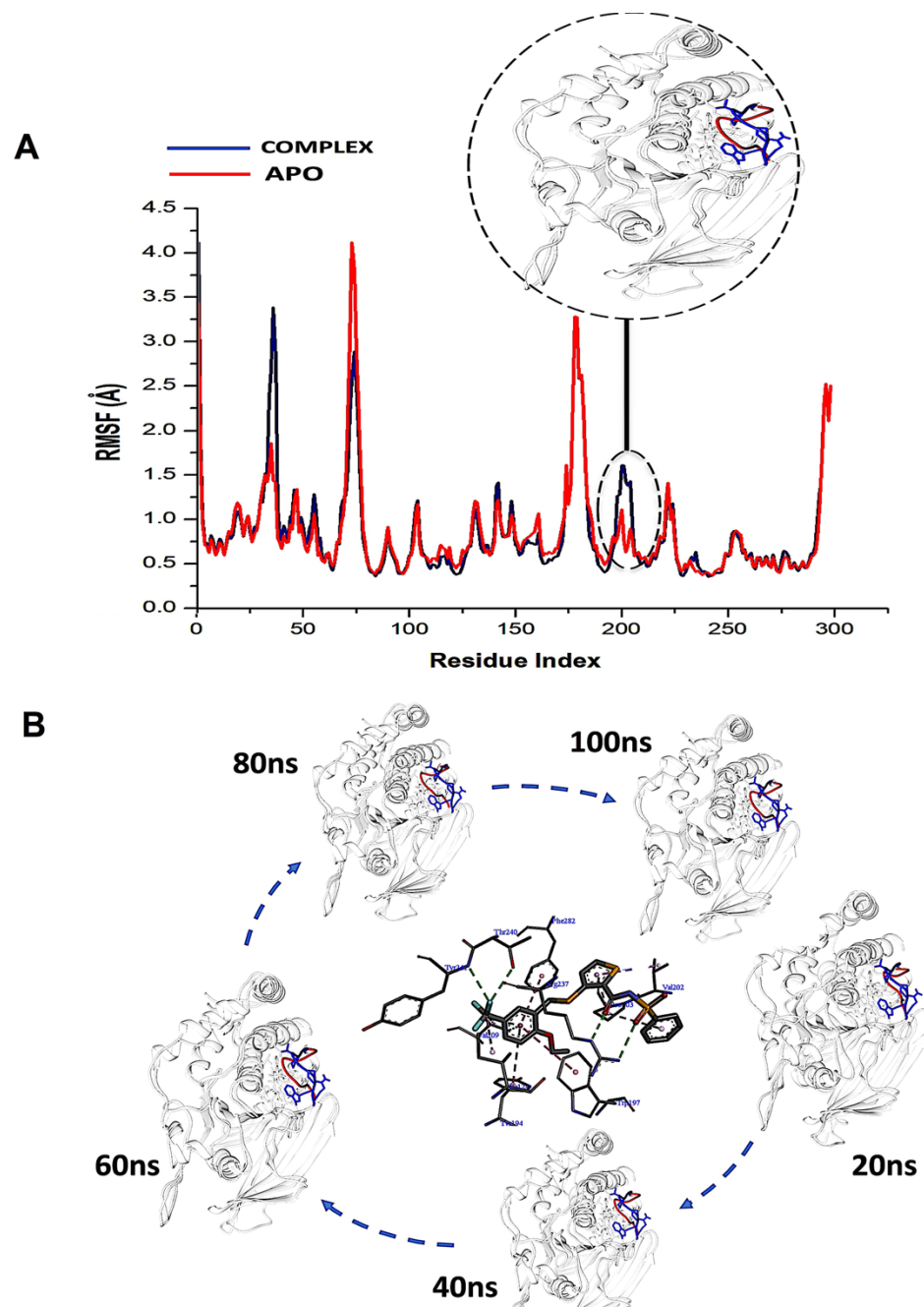


Figure 11.3: (A) Shows the comparative RMSF plot of the unbound (APO-red) and bound PTPRZ (complexed-blue), (B) Systematic NAZ2329 inhibitory activity in opening WPD-loop across the MD simulation period.

Interestingly, residues 198-204 displayed increased fluctuation in the bound system when compared to that of the APO PTPRZ. This particular region of interest encompasses the WPD-loop, a loop that serves as a flexible gate to the catalytic site of PTPRZ. A “closed” conformation of the WPD-loop denotes an active protein, whilst an “open” conformation indicates inactivity of the protein⁸. It could therefore be inferred that the allosteric inhibition of NAZ2329 induced an open conformation of WPD-loop, resulting in an inactive protein conformation of PTPRZ. As shown in Figure 11.3, there was a consistent higher flexibility of the WPD-loop in the presence of NAZ2329 as the simulation proceeded, suggesting that the allosteric inhibitory activity of NAZ2329 could have induced the increased fluctuation of the WPD-loop in order to ensure its continuous “open” conformation and hence a consequent inactivity of PTPRZ. Overall, the unbound APO and complexed protein appeared to be rigid with the exceptions of select regions, being “175-185”, “71-76” and “34-37”.

To further reveal the conformational dynamics that occur on PTPRZ due to the inhibitory action of NAZ2329, DCCM matrix analysis was conducted to determine the occurrence of correlated motions over the simulation period based on the positions of Ca atoms of the PTPRZ³³. High correlated motion, also referred to as positive correlation, ranges from the colour yellow to deep red (+1), while anti-correlated motions, also referred to as negative correlation, ranges from cyan to black (-1). DCCM analysis revealed that binding of NAZ2329 alters the structure conformation of PTPRZ as shown by the changes in the correlated motions and dynamics. There was an overall anti-correlated motion of residues within the WPD-loop (residues 198-204) in the bound system relative to more correlated motion in the APO system. Since anticorrelated motions of a protein may arise from a significant structural external perturbation such as ligand binding^{34,35}, the

observed anticorrelated motion in the bound WPD-loop could be attributed to the allosteric inhibitory prowess of NAZ2329³⁶. This was also consistent with the increased residue fluctuation of the loop in the bound conformation as revealed in RMSF in Figure 11.3 as the simulation proceeded. It is also evident from the correlation matrix that a more widely correlated motion is observed in the APO protein, thus confirming conformational shifts after ligand binding (Figure 11.4). The residues within regions “71-76” and “175-185” of PTPRZ exhibited anti-correlated motions in both the unbound and bound system which conferred with the observed residue fluctuation results.

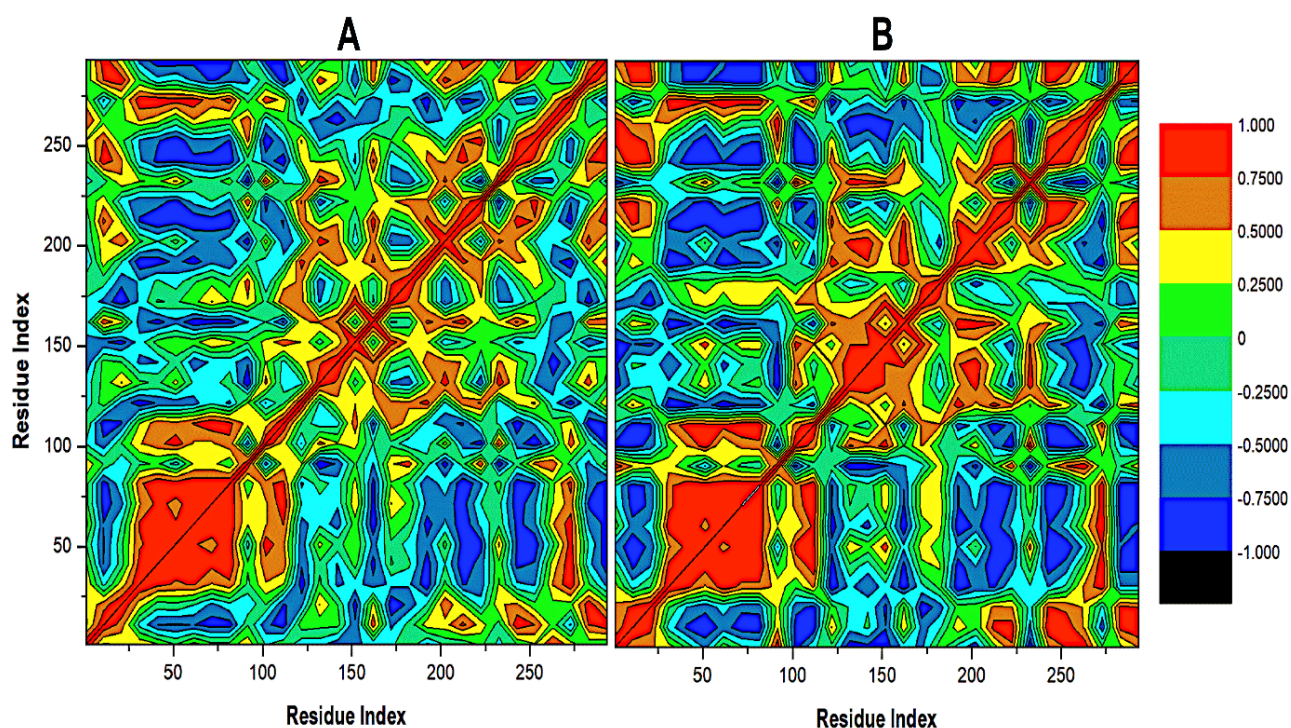


Figure 11.4: Cross-correlation matrices of the fluctuations of C- α atoms in (A) APO and (B) NAZ2329-complex.

Considering the fact that the biological function of a protein is influenced by its conformational dynamics³⁷, we also employed Principal Component Analysis (PCA), an advanced computational

tool to explore the conformational transitions of the APO and complexed PTPRZ over the 100ns simulation as illustrated in Figure 11.5. The clustering method of PCA was employed in this study due to its proven capability of describing varying conformational states generated during an MD simulation. Variations in the conformational states were described by categorizing molecular structures into clusters according to conformational similarities³⁸. Figure 11.5 highlights the overall motional shifts across two principle components in the case of APO and the NAZ2329-complex.

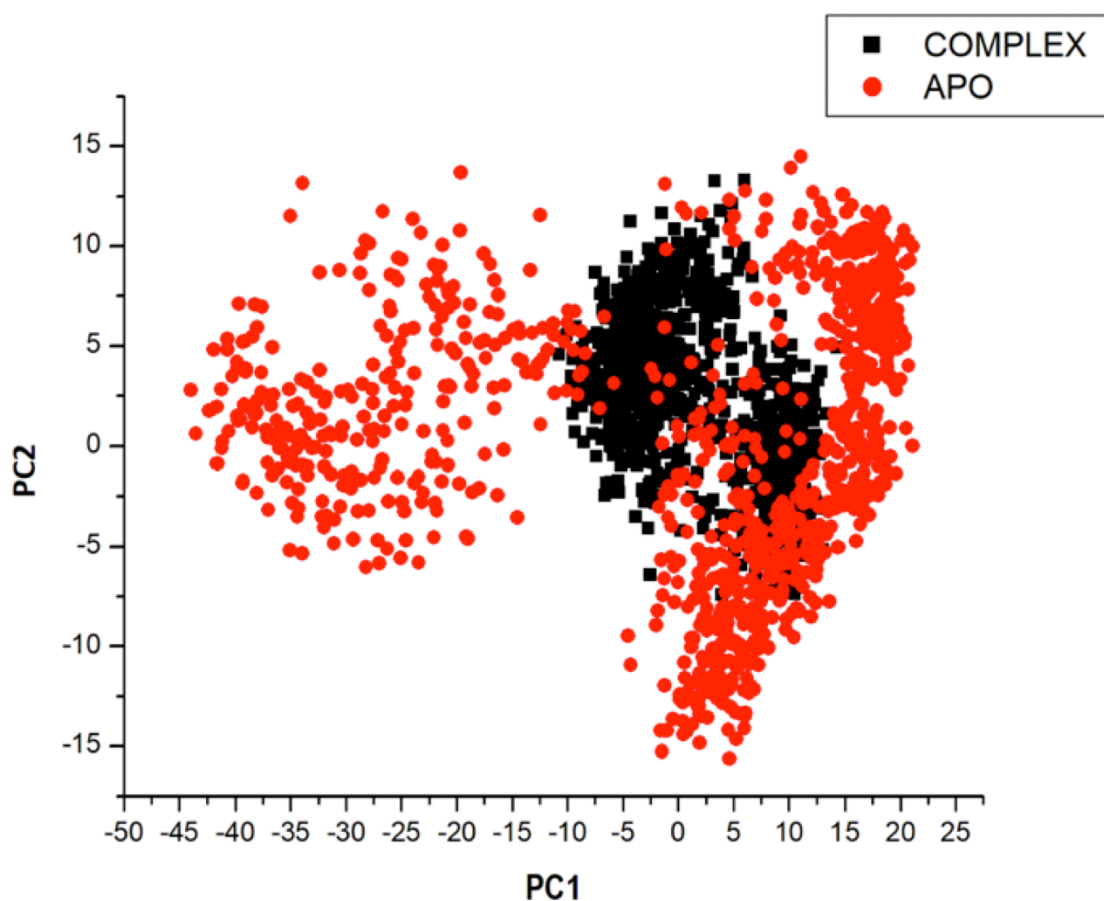


Figure 11.5: PCA projection of the motion of Ca atoms constructed by plotting the first two principal components (PC1 and PC2) in the conformational space with APO (red) and NAZ2329-complex (black) respectively. PC1 and PC2, respectively, represent a covariance matrix after elimination of eigenvectors (rotational movements). Each point between the single-directional

motions represents a unique conformation during the simulation, whereby, similar structural conformations overlap in the graph.

The simulated systems were projected along the directions of first two principal components (PC1 vs. PC2) or eigenvector. The PCA scatterplot generated for the APO and complexed PTPRZ systems shows a significant difference between the two systems along the direction of PC1 and PC2. Overall, the less correlated motion of the APO system confers with the observed higher residue flexibility in the Figure 11.3. This implied that the inhibitory activity of NAZ2329 on PTPRZ triggered conformational dynamics as conferred by the conformational flexibility and dynamic cross-correlation analysis. The variation in structural dynamics prompted us to assess the binding mechanism of NAZ2329 to the allosteric site of PTPRZ.

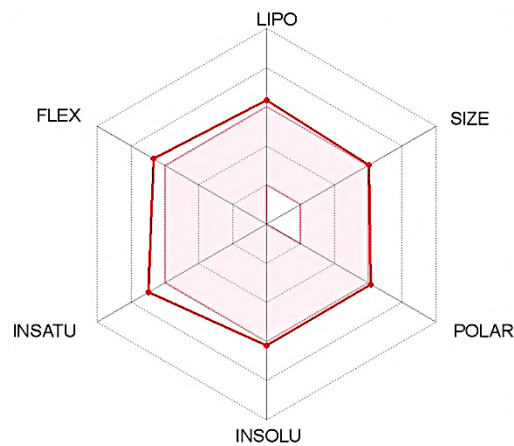
3.3 Exploring the drug likeness of NAZ2329

Having been experimental reported to exhibit inhibitory activity against PTPRZ in mouse models, we used online software SwissADME ¹⁴ to determine the physicochemical descriptors, pharmacokinetic features and drug-worthiness of NAZ2329 (Table 11.1).

Table 11.1: The Swiss ADME Profile of NAZ2329.

NAZ2329								
<i>Molecular Formula</i>	<i>Molecular Weight (g/mol)</i>	<i>Lipophilicity (iLOG P)</i>	<i>Water Soluble</i>	<i>GIT Absorption</i>	<i>BBB Permeability</i>	<i>Bioavailability Score</i>	<i>Synthetic Accessibility</i>	<i>Druglikeness (Lipinski)</i>
C21H18F3NO 4S3	501.56	3.06	Poor	Low	No	0.55	3.77	Yes

“Boiled-Egg” Method Summary



The lipophilicity and polarity of NAZ2329 was computed using SwissADME, based on the “Brain Or Intestinal Estimated permeation, (BOILED-Egg)” method ³⁹. A suitable balance of the pharmacokinetic properties, safety, potency and selectivity is usually paramount in the design of therapeutic chemical agent. Chemical properties such as lipophilicity (LogP), which are assessed by SwissADME, are very essential features to be considered in the interaction between a chemical compound and its biological target. Usually lipophilicity of a chemical agent influences its permeability, hepatic clearance or solubility. A chemical compound with a LogP value that ranges from 2 to 3 exhibits a highly favorable potential of achieving permeability and first pass clearance ⁴⁰. However, NAZ2329 showed a LogP value of 3.06, an indication that NAZ2329 exhibits a less favorable potential of achieving permeability and first pass clearance. Although NAZ2329 exhibited drug-likeness according to Lipinski rules of five ⁴¹, it has poor water solubility, low gastrointestinal absorption, and cannot permeate through the blood-brain-barrier (BBB). A low gastrointestinal absorption may result in a large amount of the drug being excreted resulting in a possible decrease inhibitory activity. The inability of NAZ2329 to pass the BBB as observed suggests the need for its improvement in its design as a therapeutic agent glioblastoma, a malignancy of the central nervous system.

3.4 Estimating the binding mode of NAZ2329 to PTPRZ through Free Energy Calculations

Using the MM/GBSA based approach; the binding free energy of the NAZ2329-complex was estimated over the 100ns MD simulation. This revealed the various energy contributions at the catalytic site. To estimate the energies contributed by individual residues located in the catalytic site, per residue decomposition analysis was performed. A depiction of the individual energy contributions and a map of the interactions between the NAZ2329 and active site residues are

illustrated in Figure 11.6. The estimated relative binding free energy for NAZ2329 is displayed in Table 11.2.

Table 11.2: MM/GBSA-based binding free energy profile of the simulated NAZ2329 complex

Energy Components (kcal/mol)					
	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{solv}	ΔG_{bind}
Complex	-53.88 ± 0.08	-60.72 ± 0.15	-114.60 ± 0.16	61.98 ± 0.14	-52.63 ± 0.09

ΔE_{ele} = electrostatic energy; ΔE_{vdw} = van der Waals energy; ΔG_{bind} = calculated total binding free energy; ΔG_{solv} = solvation free energy.

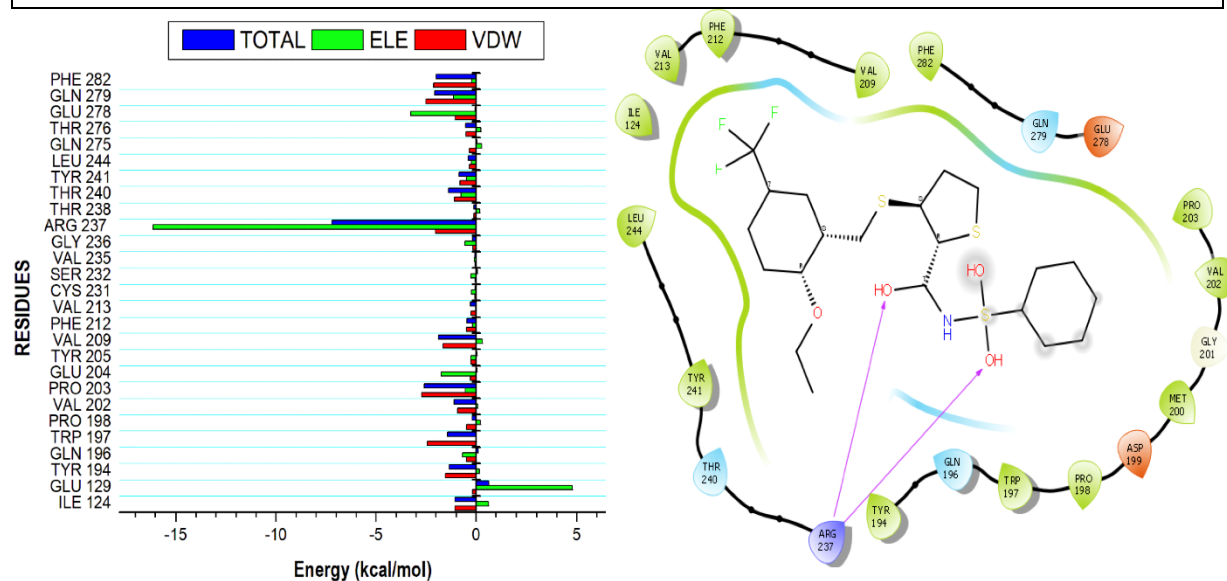


Figure 11.6: Energy contributions of the interacting residues at the NAZ2329 active site.

The residue ligand interaction network illustrates stabilizing hydrophobic interactions pocketing NAZ2329 at the active site. The highest energy contribution was two hydrogen bond interactions shared between ARG 237 with two separate oxygen moieties of NAZ2329.

NAZ2329 showed a relative binding free energy of -52.63 kcal/mol, implying a favorable thermodynamic stability, which conferred with the compact bound protein upon calculation of radius of gyration as shown in Figure S2. Residues that contributed the most to the binding of NAZ2329 include; ARG 237 (-7.201 kcal/mol), PRO 203 (-2.584 kcal/mol) and GLN 279 (-2.064 kcal/mol). From the residue-ligand interaction plot, it could be observed that, ARG 237 formed the most hydrogen bond interactions hence its high-energy contribution to the total binding energy estimated.

3.5 Pharmacophore Model Creation

An informative structural ensemble of steric and electronic features that were necessary to ensure supramolecular interaction of NAZ2329 with PTPRZ was generated (Figure 11.7). This chemical structural scaffold was developed based on the observed NAZ2329-PTPRZ interaction shown in Figure 11.6, while taking into consideration the residues that contributed the most to total binding energy as well. The generated pharmacophore model unveils useful chemical insights that could serve as a starting point for the discovery of improved therapeutic agents that target PTPRZ.

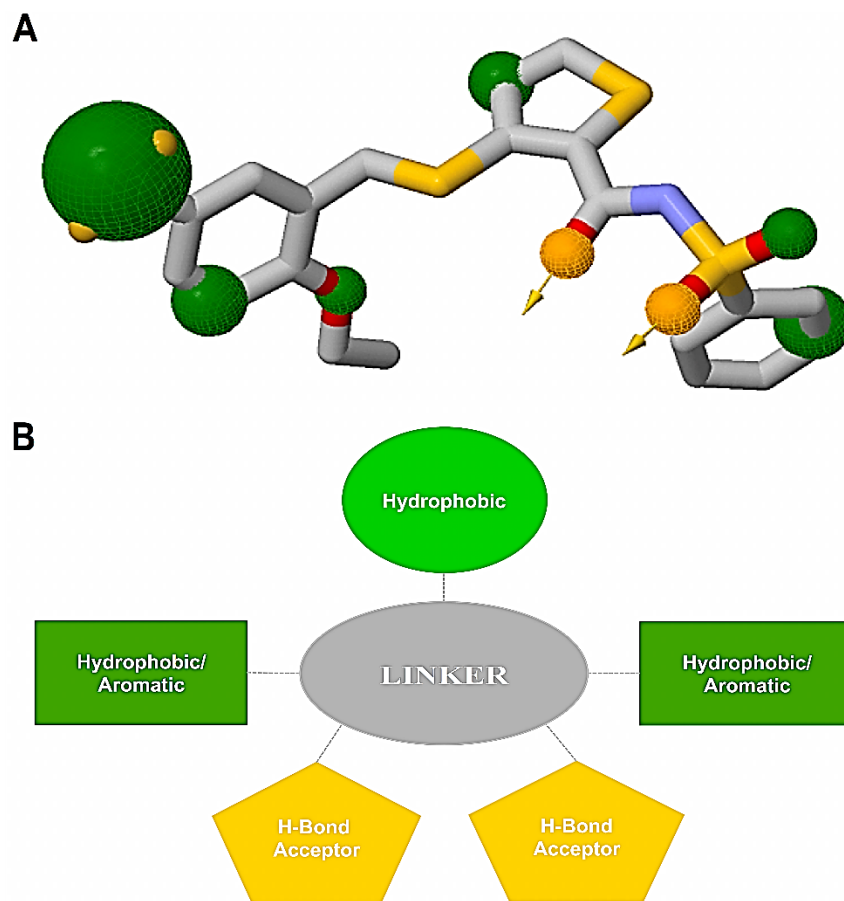


Figure 11.7. (A) Common pharmacophoric features from alignment of the PTPRZ ligand/residue interaction plots, 3-D pharmacophore model generated on ZINCPharmer (Green-Hydrophobic/Aromatic moiety, gold - hydrogen bond donor/acceptor). (B) 2-D representation of the chemical features required for potential PTPRZ inhibitors.

4.0 CONCLUSION

The comprehensive bio-computational analysis employed in this study establishes the structural modifications in PTPRZ subsequent to binding of allosteric inhibitor, NAZ2329. Molecular dynamic simulations divulged profound motional shifts the WPD-loop of the PTPRZ. This flexibility was reported in the RMSF analysis and verified by the graphical investigation of the loop at different time intervals during the simulation. Investigation of the DCC matrix and PCA

led to the deduction that the activity of NAZ2329 on PTPRZ triggered conformational dynamics that may be used to explain the mechanism of inhibition of the protein. Based on previous experimental evidence supporting the inhibitory activity of NAZ2329 and the structural dynamics leading to the design of the pharmacophore in this study, we believe that our conclusions could facilitate the design of small molecules that will not only inhibit PTPRZ and PTPRG, but will be applicable to other tyrosine phosphatases as well. This optimized pharmacophoric approach may also lead to a decline in cross-resistance and may increase patient adherence.

5.0 ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

6.0 HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are the basis of this research.

7.0 CONSENT FOR PUBLICATION

Not applicable.

8.0 CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

9.0 ACKNOWLEDGEMENTS

The authors acknowledge the School of Health Sciences, the University of KwaZulu-Natal, Westville Campus for their financial support. We also acknowledge the Center for High Performance Computing (CHPC, www.chpc.ac.za), Cape Town, for computational resources.

10.0 REFERENCES

- (1) Esmaceli, M.; Stensjøen, A. L.; Berntsen, E. M.; Solheim, O.; Reinertsen, I. The Direction of Tumour Growth in Glioblastoma Patients. *Sci. Rep.* **2018**, *8* (1), 1–6.
- (2) Pearson, J. R. D.; Regad, T. Targeting Cellular Pathways in Glioblastoma Multiforme. *Signal Transduct. Target. Ther.* **2017**, *2* (May), 17040.
- (3) Wang, V.; Davis, D. A.; Veeranna, R. P.; Haque, M.; Yarchoan, R. Characterization of the Activation of Protein Tyrosine Phosphatase, Receptor-Type, Z Polypeptide 1 (PTPRZ1) by Hypoxia Inducible Factor-2 Alpha. *PLoS One* **2010**, *5* (3).
- (4) Lara-Velazquez, M.; Al-Kharboosh, R.; Jeanneret, S.; Vazquez-Ramos, C.; Mahato, D.; Tavanaiepour, D.; Rahmathulla, G.; Quinone-Hinojosa, A. Advances in Brain Tumor Surgery for Glioblastoma in Adults. *Brain Sci.* **2017**, *7* (12), 1–16.
- (5) Fujikawa, A.; Sugawara, H.; Tanaka, T.; Matsumoto, M.; Kuboyama, K.; Suzuki, R.; Tanga, N.; Ogata, A.; Masumura, M.; Noda, M. Targeting PTPRZ Inhibits Stem Cell-like Properties and Tumorigenicity in Glioblastoma Cells. *Sci. Rep.* **2017**, *7* (1), 1–17.
- (6) Navis, A. C.; Van Den Eijnden, M.; Schepens, J. T. G.; Hooft Van Huijsduijnen, R.; Wesseling, P.; Hendriks, W. J. A. J. Protein Tyrosine Phosphatases in Glioma Biology. *Acta Neuropathol.* **2010**, *119* (2), 157–175.
- (7) Tautz, L.; Critton, D. A.; Grotegut, S. Protein Tyrosine Phosphatases: Structure, Function, and Implication in Human Disease. *Methods Mol. Biol.* **2013**, *1053*, 179–221.
- (8) Barr, A. J.; Ugochukwu, E.; Lee, W. H.; King, O. N. F.; Filippakopoulos, P.; Alfano, I.; Savitsky, P.; Burgess-Brown, N. A.; Müller, S.; Knapp, S. Large-Scale Structural Analysis of the Classical Human Protein Tyrosine Phosphatome. *Cell* **2009**, *136* (2), 352–363.
- (9) Hoff, R. H.; Hengge, A. C.; Wu, L.; Keng, Y. F.; Zhang, Z. Y. Effects on General Acid

Catalysis from Mutations of the Invariant Tryptophan and Arginine Residues in the Protein Tyrosine Phosphatase from *Yersinia*. *Biochemistry* **2000**, *39* (1), 46–54.

(10) Keng, Y. F.; Wu, L.; Zhang, Z. Y. Probing the Function of the Conserved Tryptophan in the Flexible Loop of the *Yersinia* Protein-Tyrosine Phosphatase. *Eur. J. Biochem.* **1999**, *259* (3), 809–814.

(11) Combs, A. P.; Yue, E. W.; Bower, M.; Ala, P. J.; Wayland, B.; Douty, B.; Takvorian, A.; Polam, P.; Wasserman, Z.; Zhu, W.; Crawley, M. L.; Pruitt, J.; Sparks, R.; Glass, B.; Modi, D.; McLaughlin, E.; Bostrom, L.; Li, M.; Galya, L.; Blom, K.; Hillman, M.; Gonneville, L.; Reid, B. G.; Wei, M.; Becker-Pasha, M.; Klabe, R.; Huber, R.; Li, Y.; Hollis, G.; Burn, T. C.; Wynn, R.; Liu, P.; Metcalf, B. Structure-Based Design and Discovery of Protein Tyrosine Phosphatase Inhibitors Incorporating Novel Isothiazolidinone Heterocyclic Phosphotyrosine Mimetics. *J. Med. Chem.* **2005**, *48* (21), 6544–6548.

(12) Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randall, M.; McDowell, R. S.; Hansen, S. K. Allosteric Inhibition of Protein Tyrosine Phosphatase 1B. *Nat. Struct. Mol. Biol.* **2004**, *11* (8), 730–737.

(13) Bouyain, S.; Watkins, D. J. The Protein Tyrosine Phosphatases PTPRZ and PTPRG Bind to Distinct Members of the Contactin Family of Neural Recognition Molecules. *Proc. Natl. Acad. Sci.* **2010**, *107* (6), 2443–2448.

(14) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**, *7*.

(15) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Philip, E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Thanki,

N.; Weissig, H.; Westbrook, J. D. The Protein Data Bank. *Biol. Crystallogr.* **2002**, *58*, 899–907.

(16) Fujikawa, A.; Nagahira, A.; Sugawara, H.; Ishii, K.; Imajo, S.; Matsumoto, M.; Kuboyama, K.; Suzuki, R.; Tanga, N.; Noda, M.; Uchiyama, S.; Tomoo, T.; Ogata, A.; Masumura, M.; Noda, M. Small-Molecule Inhibition of PTPRZ Reduces Tumor Growth in a Rat Model of Glioblastoma. *Sci. Rep.* **2016**, *6* (October 2015), 1–16.

(17) Yang, Z.; Lasker, K.; Schneidman-Duhovny, D.; Webb, B.; Huang, C. C.; Pettersen, E. F.; Goddard, T. D.; Meng, E. C.; Sali, A.; Ferrin, T. E. UCSF Chimera, MODELLER, and IMP: An Integrated Modeling System. *J. Struct. Biol.* **2012**, *179* (3), 269–278.

(18) Nair, P. C.; Miners, J. O. Molecular Dynamics Simulations: From Structure Function Relationships to Drug Discovery. *silico Pharmacol.* **2014**, *2* (4), 1–4.

(19) Ylilauri, M.; Pentikäinen, O. T. MMGBSA as a Tool to Understand the Binding Affinities of Filamin-Peptide Interactions. *J. Chem. Inf. Model.* **2013**, *53* (10), 2626–2633.

(20) Agoni, C.; Ramharack, P.; Soliman, M. Co-Inhibition as a Strategic Therapeutic Approach to Overcome Rifampin Resistance in Tuberculosis Therapy: Atomistic Insights. *Future Med. Chem.* **2018**, *10* (14), 1665–1675.

(21) Agoni, C.; Ramharack, P.; Soliman, M. E. S. Synergistic Interplay of The Co-Administration of Rifampin And Newly Developed Anti-TB Drug : Could It Be a Promising New Line of TB Therapy ? **2018**, 1–8.

(22) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461.

(23) Gosu, V.; Choi, S. Structural Dynamic Analysis of Apo and ATP-Bound IRAK4 Kinase. *Sci. Rep.* **2014**, *4* (5748), 1–13.

(24) Seifert, E. OriginPro 9.1: Scientific Data Analysis and Graphing Software – Software

Review. *J. Chem. Inf. Model* **2014**, *54*, 1552.

(25) Martinez, A. M.; Kak, A. C. PCA versus LDA. *Trans. Pattern Anal. Mach. Intell.* **2001**, *23* (2), 228–233.

(26) Koes, D. R. C.; Carlos, J. ZINCPharmer: Pharmacophore Search of the ZINC Database. *Nucleic Acids Res.* **2012**, *40* (1), 409–414.

(27) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, *45* (1), 160–169.

(28) Bouyain, S.; Watkins, D. J. The Protein Tyrosine Phosphatases PTPRZ and PTPRG Bind to Distinct Members of the Contactin Family of Neural Recognition Molecules. *Proc. Natl. Acad. Sci.* **2010**, *107* (6), 2443–2448.

(29) Carugo, O.; Pongor, S. A Normalized Root-Mean-Square Distance for Comparing Protein Three-Dimensional Structures. *Protein Sci.* **2008**, *10* (7), 1470–1473.

(30) Carugo, O. Statistical Validation of the Root-Mean-Square-Distance, a Measure of Protein Structural Proximity. *Protein Eng. Des. Sel.* **2007**, *20* (1), 33–37.

(31) Carugo, O. How Root-Mean-Square Distance (r.m.s.d.) Values Depend on the Resolution of Protein Structures That Are Compared. *J. Appl. Crystallogr.* **2003**, *36* (1), 125–128.

(32) Loeffler, H. H.; Winn, M. D. Ligand Binding and Dynamics of the Monomeric Epidermal Growth Factor Receptor Ectodomain. *Proteins Struct. Funct. Bioinforma.* **2013**, *81* (11), 1931–1943.

(33) Kasahara, K.; Fukuda, I.; Nakamura, H. A Novel Approach of Dynamic Cross Correlation Analysis on Molecular Dynamics Simulations and Its Application to Ets1 Dimer-DNA Complex. *PLoS One* **2014**, *9* (11).

(34) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964–

972.

- (35) Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. Fluctuation and Cross-Correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations. *J. Mol. Biol.* **1995**, *252* (4), 492–503.
- (36) Popovych, N.; Sun, S.; Ebright, R.; Kalodimos, C. Dynamically Driven Protein Allostery. *Nat Struct Mol Biol* **2006**, *13*, 831–838.
- (37) Assadi-Porter, F. M.; Maillet, E. L.; Radek, J. T.; Quijada, J.; Markley, J. L.; Max, M. Key Amino Acid Residues Involved in Multi-Point Binding Interactions between Brazzein, a Sweet Protein, and the T1R2 T1R3 Human Sweet Receptor. *J. Mol. Biol.* **2010**, *398*, 584–599.
- (38) Wolf, A.; Kirschner, K. N. Principal Component and Clustering Analysis on Molecular Dynamics Data of the Ribosomal L11.23S Subdomain. *J. Mol. Model.* **2013**, *19*, 539–549.
- (39) Daina, A.; Zoete, V. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem* **2016**, 1117–1121.
- (40) Ahmed, S. S. S. J.; Ramakrishnan, V. Systems Biological Approach of Molecular Descriptors Connectivity: Optimal Descriptors for Oral Bioavailability Prediction. *PLoS One* **2012**, *7* (7).
- (41) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Setting. *Adv. Drug Deliv. Rev.* **2012**, *64*, 4–17.

CHAPTER 12

Published Article

Tweaking α -galactoceramides: Probing the dynamical mechanisms of improved recognition for invariant natural killer T-cell receptor in cancer immunotherapeutics

Houda Washah^a, Clement Agoni^a, Fisayo A. Olotu^a, Geraldene Munsamy^a, and Mahmoud E. S.

Soliman^{a*}

^aMolecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

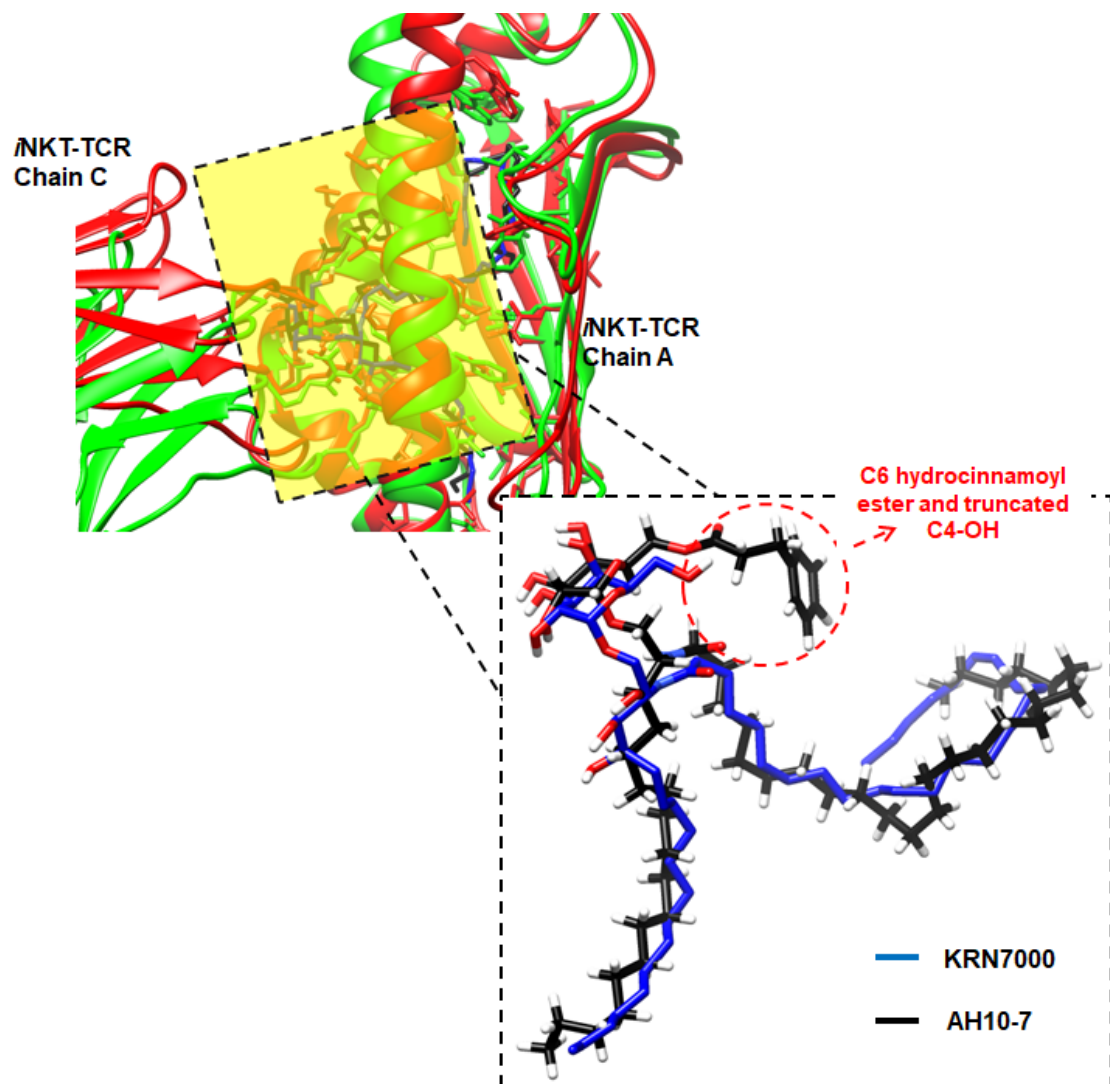
*Corresponding Author: Mahmoud E.S. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 8048, Fax: +27 (0) 31 260 7872

<http://soliman.ukzn.ac.za>

Graphical Abstract



Abstract

Background: The last few decades have witnessed groundbreaking research geared towards immune surveillance mechanisms and have yielded significant improvements in the field of cancer immunotherapy. This approach narrows down on the development of therapeutic agents that either activate or enhance the recognitive function of the immune system to facilitate the destruction of malignant cells. The α -galactosylceramide derivative, KRN7000, is an immunotherapeutic agent that has gained attention due to its pharmacological ability to activate CD1d-restricted invariant natural killer T(*i*NKT) cells with notable potency against cancer cells in mouse models; a therapeutic success was not well replicated in human models. Dual structural modification of KRN7000 entailing the incorporation of hydrocinnamoyl ester on C6'' and C4-OH truncation of the sphingoid base led to the development of AH10-7 which, interestingly, exhibited high potency in human cells.

Objective/Methods: Therefore, to gain molecular insights into the structural dynamics and selective mechanisms of AH10-7 for human variants, we employed integrative molecular dynamics simulations and thermodynamic calculations to investigate the inhibitory activities of KRN7000 and AH10-7 on *h*TCR-CD1d towards activating *i*NKT.

Results: Interestingly, our findings revealed that AH10-7 exhibited higher affinity binding and structural effects on *h*TCR-CD1d, as mediated by the incorporated hydrocinnamoyl ester moiety which accounted for stronger intermolecular interactions with 'non-common' binding site residues.

Conclusions: Findings extracted from this study further reveal important molecular and structural perspectives that could aid in the design of novel α -GalCer derivatives for cancer immunotherapeutics.

Keywords: Cancer immunotherapy, natural killer T-cells, dual modification, α -galactosylceramide.

1.0 Introduction

Immune surveillance mechanisms are physiologically employed in cancer to ameliorate the debilitating effects of the disease [1]. However, there is usually an eventual failure of the immune response due to changes that occur in the microenvironment, which in turn permit tumor growth with a concurrent sabotage of the anti-cancer immune responses [1]. Nonetheless, some immune cells exhibit prominent roles in regulating immune response to cancer, notably, the invariant natural killer T cells *i*NKT, distinguished by their semi-variant T cell receptor (TCR) in humans and mice ($V\alpha 24J\alpha 18$ in humans and $V\alpha 14J\alpha 18$ in mice) [2–5]. As a unique class of lymphocyte involved in both innate and adaptive immune response [6–9], *i*NKT are characterized by their ability to identify lipid-based antigens presented by the MHC class I-like CD1d protein [3,10–13]. Unlike other natural killer cells, *i*NKT are selected by CD1d-expressing thymocytes instead of interacting with class 2 MHC molecules [6,14]. As a sub-type of NKT, *i*NKT are the only CD1d-restricted cells that harbor the invariant TCR and is able to identify the prototype alpha-galactosylceramide inhibitors [5]. Though low in frequency in the peripheral blood, *i*NKTs are dominant in the adipose tissues of both humans and mice. By secreting IL-4, *i*NKT assume an anti-inflammatory role that has been shown to prevent resistance to insulin as well as inflammatory responses in the adipose tissue [15–20]. When activated, *i*NKT cells can also increase the expression of NKTs and consequently increase the stimulation of conventional T cells *via* their cytokine and mature dendritic cells secretion [21,22]. Upon antigenic triggering, *i*NKT cells usually respond in few hours by releasing of Th1, Th2, and Th17 cytokines [23]. The involvement of *i*NKT in the rapid secretion of cytokines as well as in CD40L expression potentiates their prominent roles in immune response [5]. In antitumor immune response, *i*NKTs are directly involved in the killing of tumor cells *via* the induction and release of cytotoxic lymphocytes

(CTLs) coupled with the activation of NK cells and modulation of immunosuppressive tumor microenvironment [24,25]. Hence, considering their crucial roles in antitumor immune response, *i*NKTs have been extensively explored as immunotherapeutic targets [5,26,27]. Current therapeutic approaches that entail *i*NKT cell targeting have focused on stimulating *i*NKT activation with the aid of exogenous agents, an approach referred to as active immunotherapy [24,27]. Alpha-galactosylceramide (α -GalCer), a potent marine sponge sphingolipid antigen was identified in the early 1990s as an exogenous molecule that can activate *i*NKT towards cancer treatment [28–30]. Over the years, α -GalCers have been widely investigated and developed as immunotherapeutic agents and vaccines [26,31]. This feature has resulted in the discovery of KRN7000, the first synthetic α -GalCer by Kobayashi *et al.*, in 1995[32,33] (Figure 12.1). Although KRN7000 showed potent anticancer properties as well as some interesting efficacy against autoimmune and infectious diseases in mouse models [26], its successful translation and advancement into clinical use has been relatively limited [34,35]. A recent phase 1 trial by Nair and Dhodapka in 2017, reported that KRN7000 had limited clinical benefit to humans [35,36]. However, the foremost challenge associated with KRN7000 is its propensity of inducing high levels of helper T cells: T helper 1 (Th1) and T helper 2 (Th2), a phenomenon that accounts for their contradictory function, which subsequently results in an unsuccessful immune response [34,37].

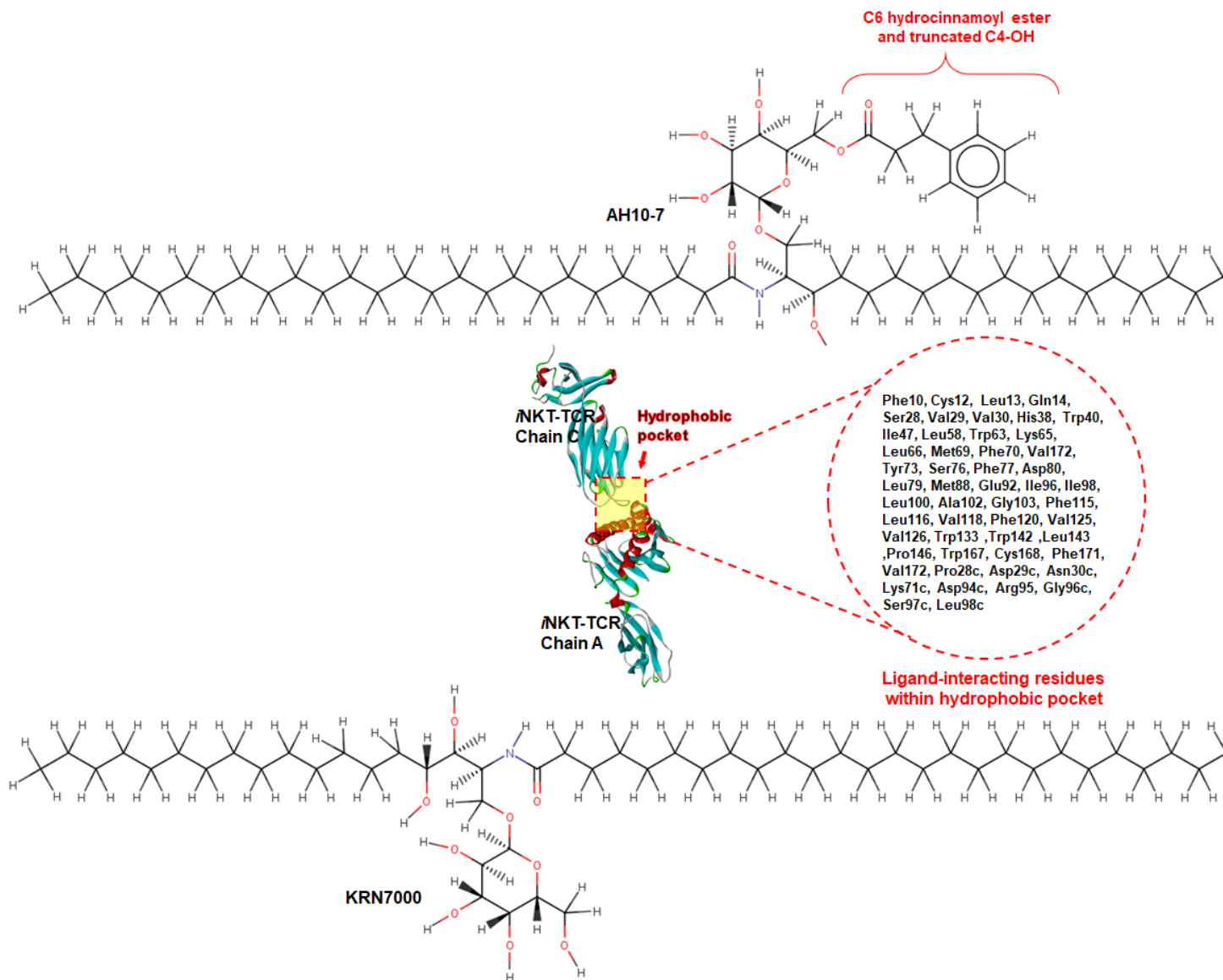


Figure 12.1: 2D structures of AH10-7 (**above**) and KRN7000 (**below**) coupled with the 3D X-ray crystal structure of human *i*NKT-TCR (**middle**). Red inset showcases constituent hydrophobic pocket residues.

Due to the limited success recorded for KRN7000, efforts have been made over the past years to twig the moieties of KRN7000 so as to obtain new α -GalCer derivatives with improved potencies and activities in humans while propagating the release of either Th1 or Th2, but not a concurrent release of both [22,37,38]. The latest of such attempts was reported by Chennamadhavuni *et al.*, where a new α -GalCer derivative, AH10-7 was synthesized by incorporating a C6'' hydro

cinnamoyl ester while removing the C4-OH of the sphingoid base [34] (Figure 1). This unique dual modification resulted in demonstrable AH10-7 potency towards the activation of *i*NKT in both mouse and human models while highly biased towards Th1 secretion. Chennamadhavuni *et al.*, further explored the mechanistic basis of the dual modification of *i*NKT activation *via* X-ray crystallography and molecular modelling. Molecular docking and QM/MM studies showed that the simulated binding pose of AH10-7 reproduced well the 2C12 TCR-*m*CD1d-AH10-7 ternary complex structure obtained by X-ray crystallography with better binding affinity. However, since molecular docking is inadequate in providing a thorough and dynamic perspective into the mechanistic binding and recognition of the dually modified α -GalCer derivative, AH10-7, this current study employs molecular dynamics simulation techniques to provide additional molecular insights into the structural dynamics and mechanistic binding of the novel α -GalCer derivative to human *i*NKT-TCR [39]. Molecular dynamics (MD) simulation mimics the dynamical events of molecular systems as a function of time, with a molecular description of all biological components (ligand and protein) in a solvent space [40]. Hence this approach would enable a detailed elucidation of molecular events and recognition mechanisms that characterize selective binding of AH10-7 [34]. Findings will also reveal important atomistic perspectives that could aid in the design of novel α -GalCer derivatives for cancer immunotherapeutics.

2.0 Computational Methods

2.1 System Preparation

The X-ray crystal structures of 2C12 TCR-CD1d in complex with AH10-7 (PDB:6BNL) and KRN7000 ((PDB code:6BNK) [34] were retrieved respectively from RCSB Protein Data Bank [41] and were used as the starting structures. The retrieved structures exist as a homomultimers

with eight chains, however, only chains (A, C) which contained the ligand binding pockets were used in this study to reduce the computational cost. All non-standard residue molecules and ligands were deleted during the preparation of the receptor structure. Hydrogen atoms were added to the isolated ligands using the UCSF Chimera software package during structure preparation. Software was developed by the Resource for Biocomputing, Visualization, and Informatics (RBVI), University of California, San Francisco, USA [42].

2.2 MD simulations

MD simulations for 2C12 TCR-CD1d complexed distinctly with AH10-7 and KRN7000 were performed. Altogether, three systems comprising of unbound and bound forms of 2C12 TCR-*m*CD1d were subjected to MD simulations using the AMBER14 software package. Software was developed by Case *et al.*, University of California, San Francisco, USA [43]. Protein optimization and explicit solvation were carried out using the integrated LEAP module while the AMBER FF14SB force field was employed to define protein parameters. The systems were partially minimized for 2500 steps with a restraint potential of 500 kcal/mol Å² followed by full minimization of 10000 steps. The systems were gradually heated from 0-300K using a Langevin thermostat in a canonical ensemble (NVT) [44]. Equilibration was also carried out without restraints at a temperature of 300K in an NPT ensemble for 1000 ps while atmospheric pressure was maintained at 1 bar using the Berendsen barostat [45]. This procedure was followed by MD production runs of 100 ns for each system during which the SHAKE algorithm was used to constrict all atomic hydrogen bonds [46]. The integrated CPPTRAJ and PTRAJ modules [47] of AMBER14 were used to analyze resulting coordinates and trajectories while obtained data were plotted using Microcal Origin software, developed by OriginLab Corporation, United States [48].

2.3 Binding free energy calculations

In this study, the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method was used to estimate the differential binding of AH10-7 and KRN7000 to 2C12 TCR-*m*CD1d [49] as employed in many of our previous reports [50–54]. This technique could help predict the differential binding strengths and affinities of both ligands to 2C12 TCR-*m*CD1d. Mathematically, the binding free energy (ΔG) for each complex is represented as:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}} \quad (2)$$

$$G_{\text{sol}} = G_{\text{PB}} + G_{\text{SA}} \quad (3)$$

$$G_{\text{SA}} = \gamma \text{SASA} \quad (4)$$

Van der Waals and electrostatic interactions are represented as E_{vdw} and E_{ele} while E_{gas} denotes gas-phase energy and E_{int} internal energy. The solvation free energy, denoted by G_{sol} represents the solvation free energy and can be decomposed into polar and nonpolar contribution states. The polar solvation contribution, G_{PB} , is determined by solving the PB equation, whereas, G_{SA} , the nonpolar solvation contribution is estimated from the solvent accessible surface area (SASA) determined using a water probe radius of 1.4 Å. Per-residue energy contributions of active sites residues towards the stabilization of AH10-7 and KRN7000 were estimated using data obtained from the MM/PBSA calculations [49]. Energy contributions of these residues give a clue into the essentiality of these residues towards the stabilization of the ligands in the binding pocket and their consequential favorable activity of these ligands [49].

3.0 Results and Discussion

AH10-7 successfully demonstrated its potency in human cells which earned it a significant advantage over earlier derivatives such as KRN7000 [34]. Investigated model systems in this report depict unbound human TCR-CD1d, a KRN7000-bound TCR-CD1d and a AH10-7 bound TCR-CD1d taken through 100 ns MD simulation to explore impact of the dual modification which characterize AH10-7 relative to its enhanced potency over earlier derivative, KRN7000 in human models. An initial sequence analysis as shown in Figure 12.2 showcased a large similarity in sequence of residues that interacted directly with both AH10-7 and KRN7000 due to similarity in structure. Nonetheless there were notable variations in the interaction with distinct residues observed particularly in the AH10-7 complex, which could partly be attributed to the additional interactions mediated by the incorporated hydrocinnamoyl ester moiety. The similarity in sequence of the interacting residues also suggests similar binding modes of the both AH10-7 and KRN7000.

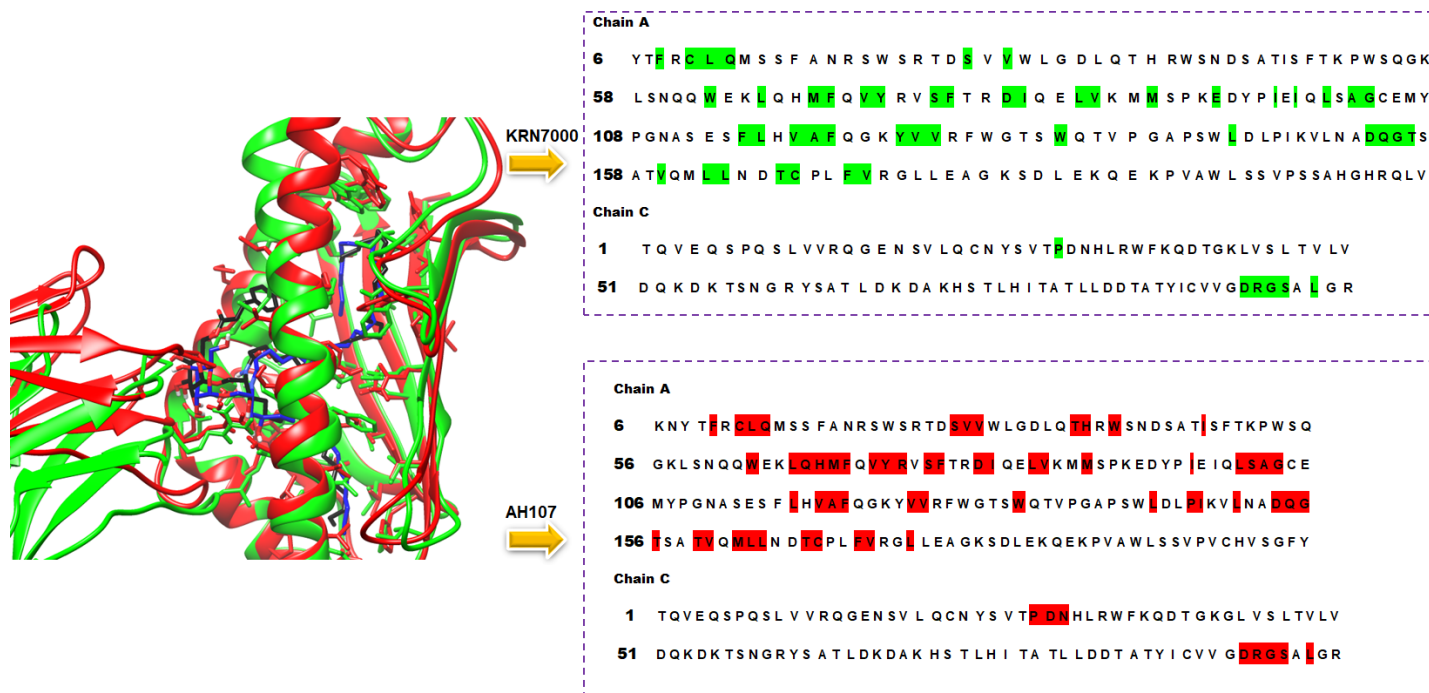


Figure 12.2: Comparative sequence analysis of AH10-7 and KRN7000-2C12 TCR-*m*CD1d complexes with corresponding highlights of interacting residues at the hydrophobic pockets. Green inserts indicate residues that directly interacts with KRN7000 while red highlights indicate residues that directly interact with AH10-7

3.1 Enhanced binding affinity and mechanistic electivity mediated by the incorporated hydrocinnamoyl ester on the C6'' of the sugar

The ability of α -GalCer to potentially activate in human *i*NKT cells is largely dependent on their respective interactions at the hydrophobic pocket on TCR-CD1d in which they bind, and the corresponding binding energies elicited. The respective interactions between the α -GalCer derivatives and amino acids at the binding pocket of TCR-CD1d are crucial to the overall stability, selectivity and total binding affinity. With a highly favorable binding energy of AH10-7 over KRN7000 from QM/MM calculations as estimated in earlier reports, we again calculated their total binding free energy (ΔG) over the course of the MD simulation period, in an attempt to corroborate earlier reports using the extensively explored MM/PBSA approach [49,51,52,55]. This procedure was performed to also understand the molecular basis of the reported favorable binding of AH10-7 over KRN7000 from experimental reports. The calculated binding free energy also provided a deeper quantitative understanding of the binding of both KRN7000 and AH10-7 while also providing a direct insight into driving forces that support their binding. As shown in Table 12.1, AH10-7 exhibited a total binding free energy of -30.12 kcal/mol while KRN7000 exhibited a total binding free energy of -28.2 kcal/mol. As estimated, AH10-7 exhibited a higher binding affinity (ΔG) relative to KRN7000 which corroborates earlier reports in which AH10-7 exhibited superiority in binding over KRN7000 [34]. As presented in Table 12.1, the intermolecular van der Waals and electrostatic interactions provide the major contribution for the binding of both KRN7000 and AH10-7 even though the van der Waals forces were more prominent towards the binding of AH10-7 than KRN7000. The van der Waals and electrostatic energies in the TCR-CD1d-AH10-7 were estimated as -28.11 kcal/mol and -11.81 kcal/mol, respectively as shown in Table 12.1. Also, a positive polar solvation energy estimation indicates the binding interactions

were more favorable in the hydrophobic (non-polar) regions of the ligand-protein complex. Likewise, we observe that van der Waals interactions further contributed to complex formation in both ligands. These interactions altogether, could account for the binding affinities exhibited by both ligands when bound to *h*CD1D as previously reported [34].

Table 12.1: MM/PBSA-based binding free energy profile of KRN7000 and AH10-7

	ΔE_{vdw}	ΔE_{ele}	ΔG_{gas}	ΔG_{sol}	ΔG_{bind}
AH10-7	-28.11±0.13	-11.81±0.16	-39.92±0.19	9.80±0.08	-30.12±0.15
KRN7000	-24.75± 0.13	-13.98±0.17	-38.73±0.17	10.54±0.09	-28.2±0.14

ΔE_{ele} = electrostatic energy; ΔE_{vdw} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{sol} = solvation free energy ΔG = gas phase free energy.

3.2 Energy-based decomposition of binding site residues and corresponding contributions towards ligand binding

To further understand the favorable and preferential binding mechanisms of AH10-7 over KRN7000, we performed the decomposition of total ΔG involved with respect to the contributions of individual ligand binding site residues. This operation was carried out using the MM/PBSA-incorporated per-residue energy decomposition method. Estimated per-residue energies allowed a more thorough quantitative insight into the differential binding mechanisms of both KRN7000 and AH10-7 while showcasing the superiority of AH10-7 based on the energy contribution of each binding site residue. An interesting observation was the obvious variation in major energy contributing residues when comparing both the AH10-7- and KRN7000-bound systems. Basically,

the involvement of certain residues that interacted with the hydrocinnamoyl ester group on C6'' of AH10-7 could account for enhanced binding when compared to KRN7000. Figure 12.3 showcases the energy contributions of residues unique to the binding of both AH10-7 and KRN7000. The introduction of the hydrocinnamoyl ester in AH10-7 visibly increased the number of interactions as compared to KRN7000 with relatively fewer interactions. As shown, many of the residues with higher energy contributions interacted are those involved in complimentary interactions with the hydrocinnamoyl ester moiety. Cumulatively, these unique interactions with the hydrocinnamoyl ester group in AH10-7 could have contributed to a higher affinity binding which favored its potent therapeutic activity in human variants.

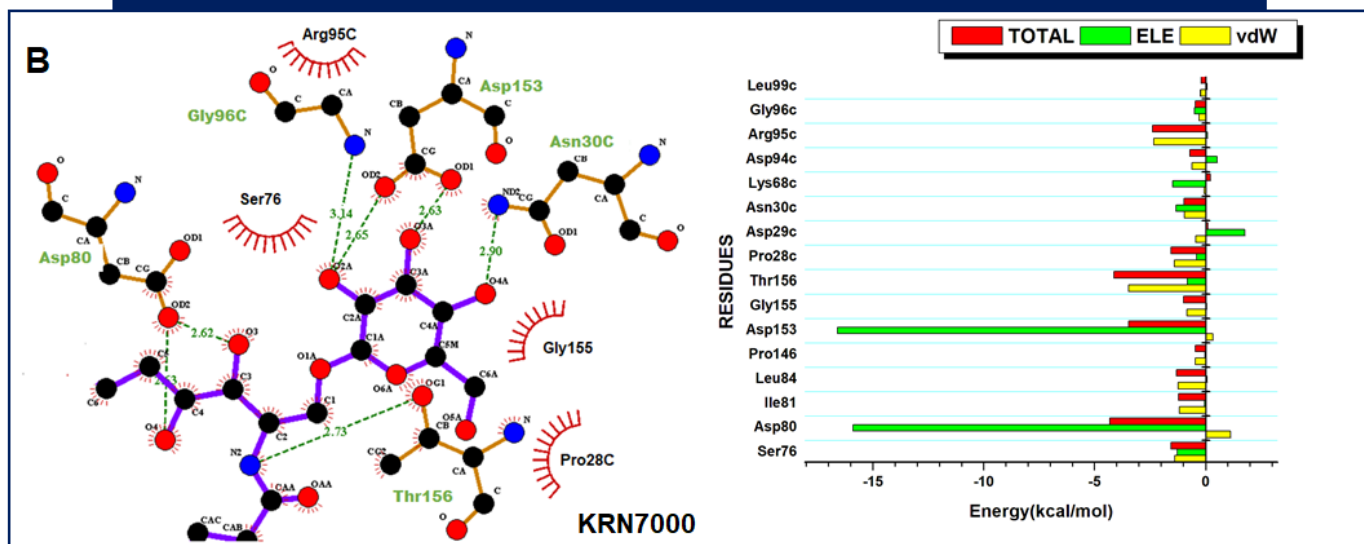
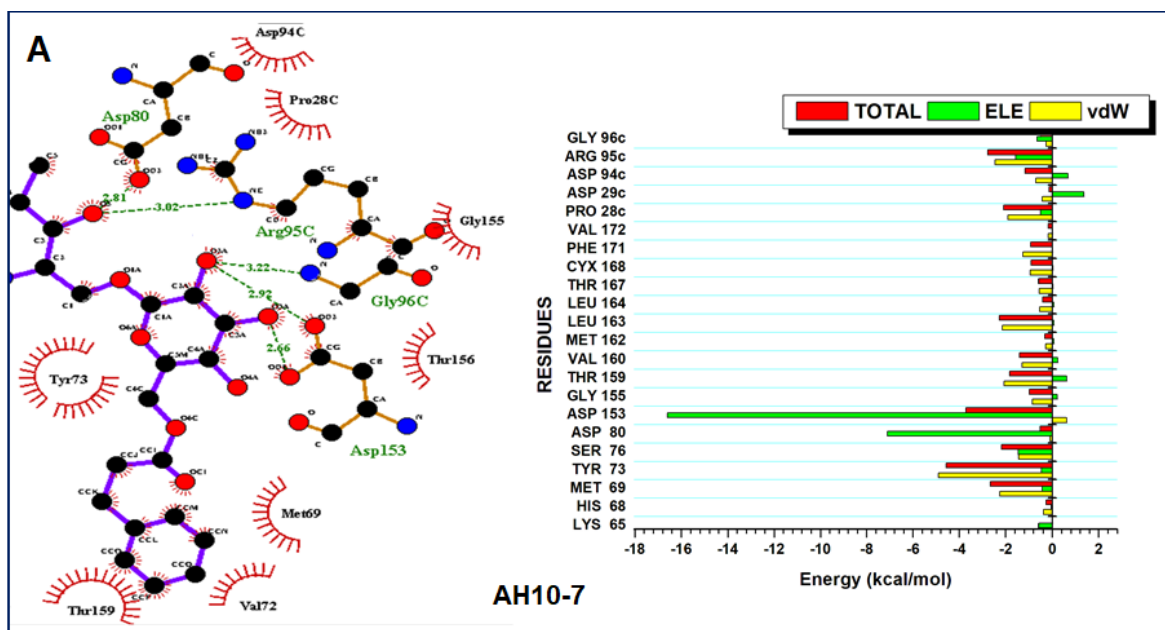


Figure 12.3: Per-residue energy decomposition plot of uncommon binding site residues and their differential interactions with AH10-7 and KRN7000, with corresponding ligand interaction plots of average structures.

As shown in Figure 12.4A, the modification in AH10-7 accounted for a higher number of conventional hydrogen bonds with shorter distances relative to those elicited by KRN7000 as presented in Figure 12.4B using average structures obtained from the MD trajectories. The higher number of conventional hydrogen interactions amidst other notable interactions such as van der

Waals and alkyl interactions with surrounding hydrophobic residues could have equally contributed to the overall high affinity binding of AH10-7 relative to KRN7000 as estimated.

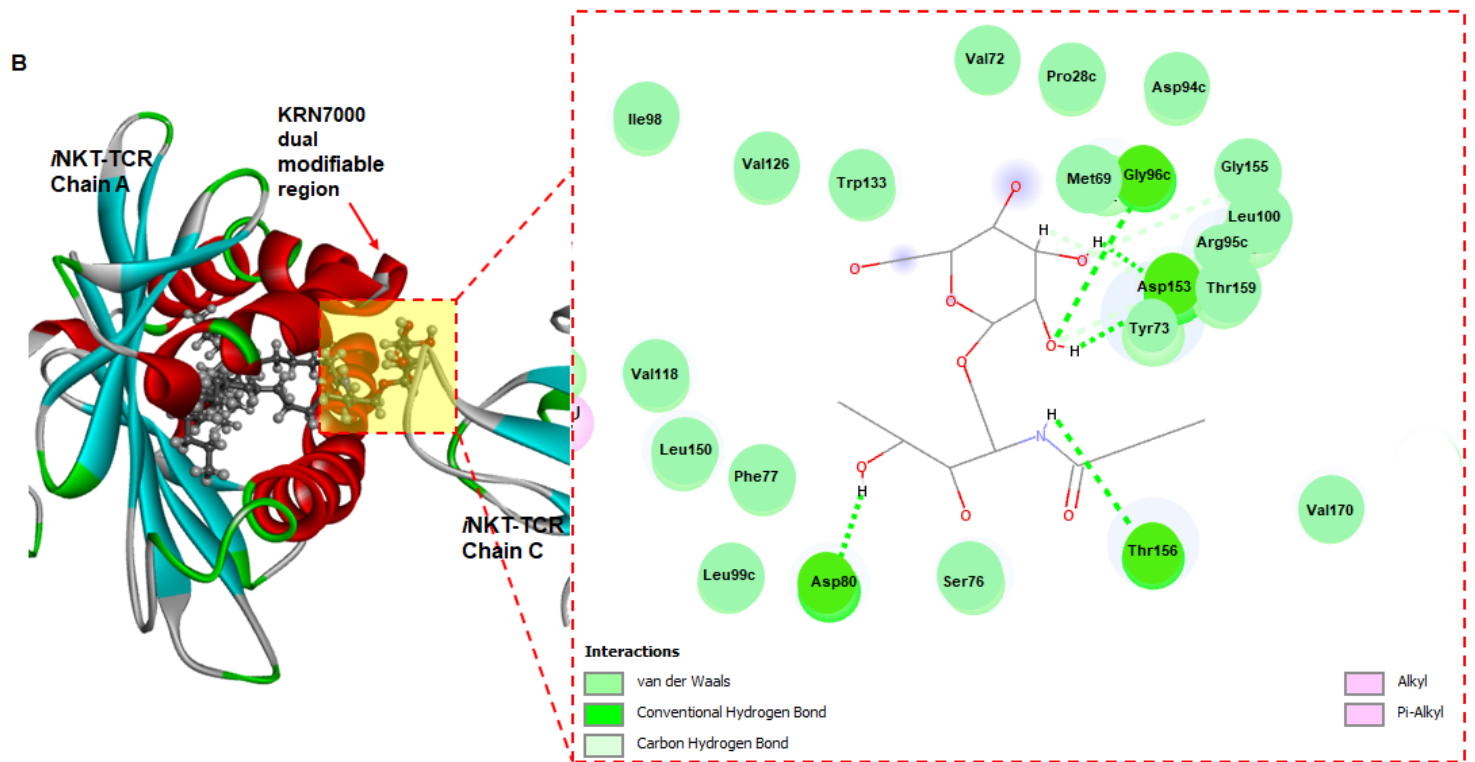
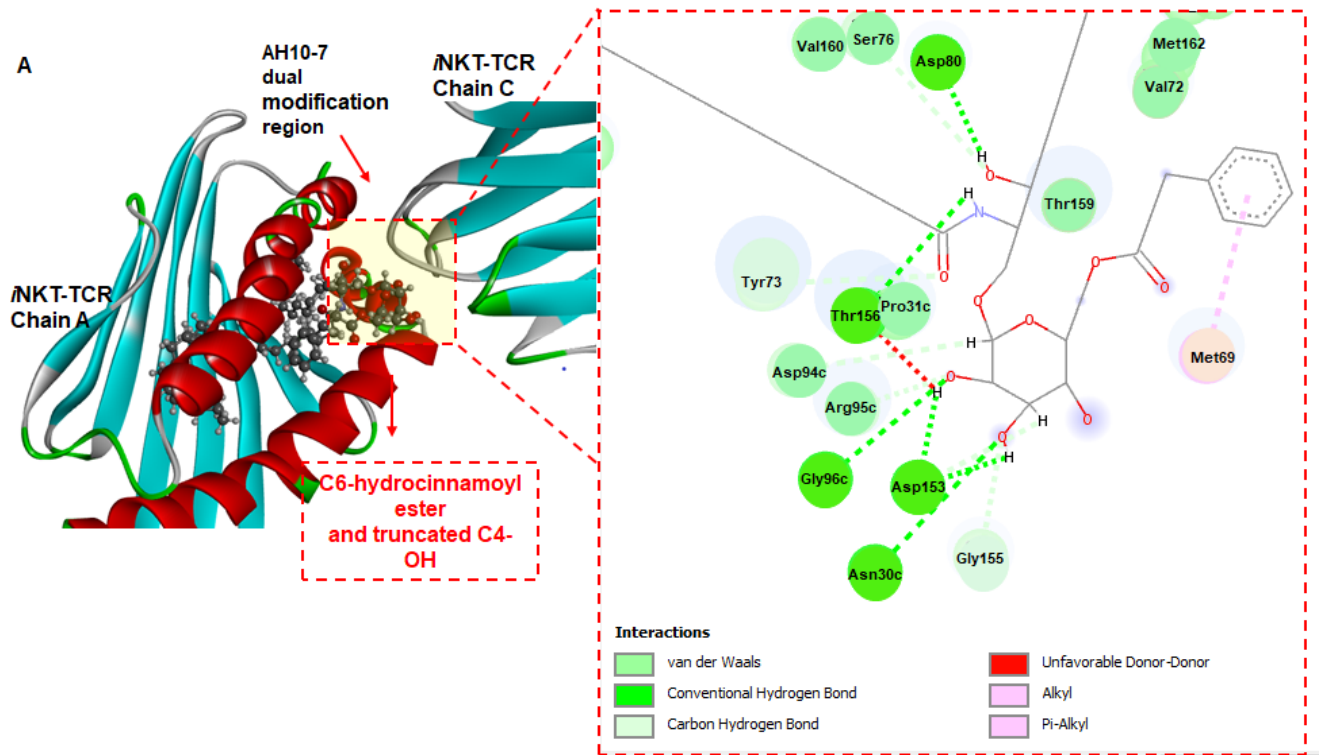


Figure 12.4: Complementary binding site-surface interactions of [A] C6'' modified hydrocinnamoyl ester in AH10-7 and [B] "non-modified" KRN7000 galactosyl moiety. Conventional (OH--O) and non-conventional/carbon-hydrogen (CH--O) bonds are also indicated.

A broader analysis of the differential activities of AH10-7 and KRN7000 focused on binding pocket residues within 5Å radius from them, which revealed striking insights that could further support the binding superiority of AH10-7 to 2C12 TCR-*m*CD1d over KRN7000 as previously established. Overall, residues with energy contributions less than -1.0 kcal/mol were considered very crucial to the interaction of the α -GalCers within the hydrophobic pocket and could consequently influence their overall activity [56]. From the calculations, residues within 5Å radius which contributed prominently to KRN7000 binding included Cys12, Val30, Ile47, Phe70, Tyr73, Ser76, Phe77, Asp80, Ile81, Leu84, Leu100, Leu116, Val118, Leu151, Asp153, Thr156 and Pro28, all located on chain C. Amongst these, residues Thr156, Asp80 and Tyr73 exhibited the highest energies of -4.11kcal/mol, -4.31kcal/mol and -3.88kcal/mol, respectively as shown in TableS1. On the other hand, residues that notably contributed to the binding of AH10-7 included Cys12, Val30, Lue66, Met69, Phe70, Val72, Tyr73, Ser76, Phe77, Leu79, Leu100, Leu116, Val118, Leu150, Asp153, Thr156, Thr159, Val160, Leu163, and Ile76, Pro28, Asn30, Asp94, and Arg95 on chain C. Moreover, residues Asp153, Thr156 and Tyr73 contributed the highest energies towards the binding and stabilization of AH10-7 with estimated total energies of -3.72 kcal/mol, -3.52 kcal/mol and -4.56 kcal/mol, respectively (Table S2). As observed, many of the high energy contributing residues in both AH10-7 and KRN7000 complexes were both similar suggesting an analogous binding of both AH10-7 and KRN7000 as shown in the superimposed structures in Figure 12.5.

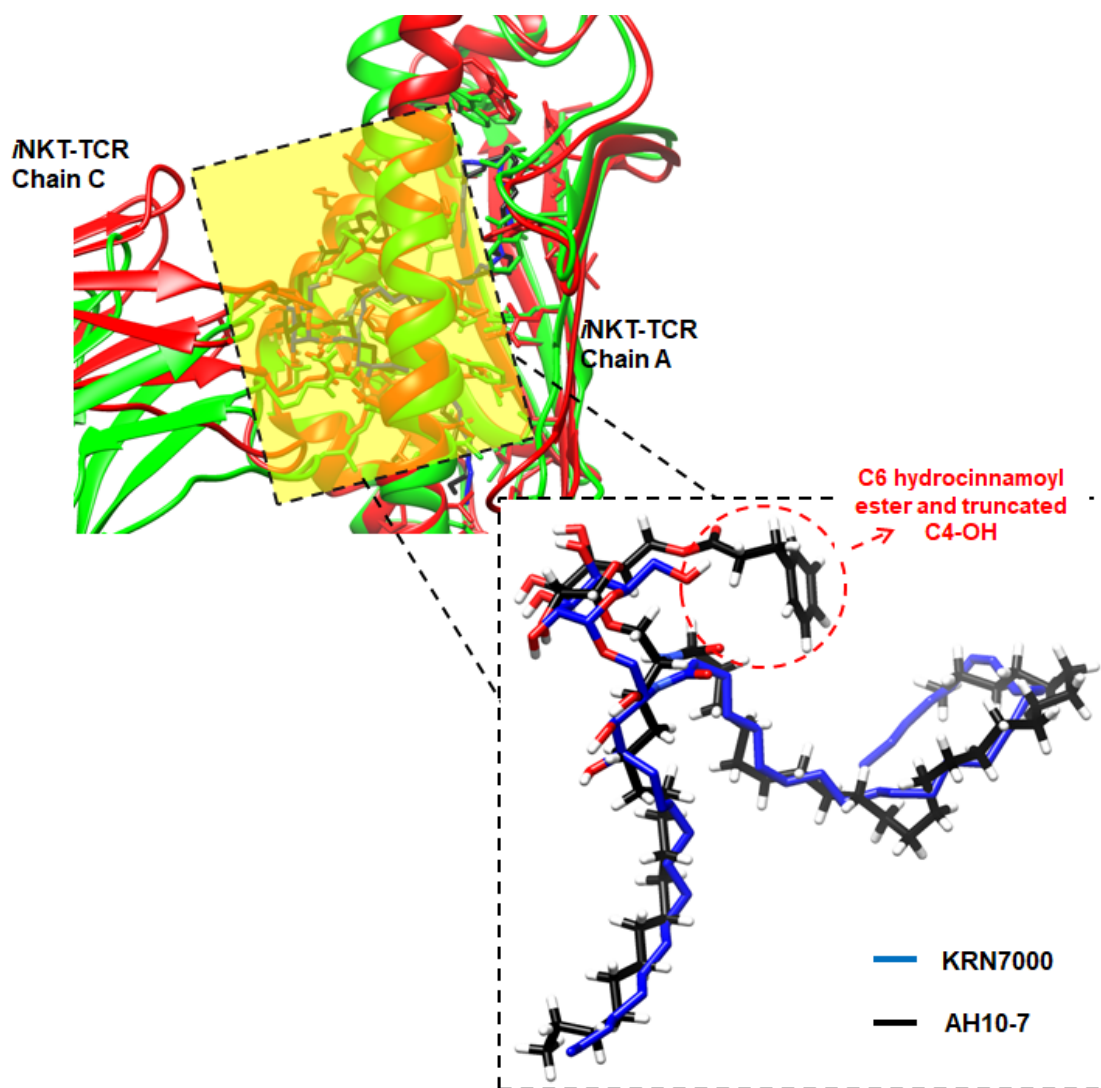


Figure 12.5: Structural superposition of AH10-7 and KRN7000 at the binding pocket of human *i*NKT-TCR to depict their analogous binding modes.

3.3 Dual sphinganine base and hydrocinnamoyl ester modification of α GalCer enhanced residue interactions with TCR-CD1d.

To get a further molecular grasp of the effects of hydrocinnamoyl ester incorporation on the therapeutic potential of AH10-7 towards *iNKT* cellular activation in humans, we performed a time-scale molecular visualization of resulting snapshots (10 ns, 40 ns, 70 ns and 100 ns) taken across the MD simulation period as shown Figure 12.6A and 12.6B. This procedure permitted a thorough and dynamical exploration of the nature of interactions that the incorporated moieties elicited during at the hydrophobic pocket region which could have accounted for its favorable interaction relative to KRN7000 as evidenced by the differential ΔG s earlier estimated. As shown in Figure 12.6A and 12.6B, certain residues maintained steady “strong intermolecular” interactions with the incorporated moiety of AH10-7 at the hydrophobic pocket across the 100 ns simulation time. At the beginning of the simulation at around 10 ns, the residues: Ser30, Asp153, Thr156 and Asp80 on chain A and Gly96 on chain C formed strong hydrogen interactions with the modified moieties as shown in Figure 12.6A. At about 40 ns, these same interactions were preserved as shown in Figure 12.6A. Towards the end of the simulation at about 70 ns as shown in Figure 12.6B, two new hydrogen bond interactions were formed between the hydrocinnamoyl ester group, Asp29 and Asn30 on chain C in addition to Asp153 and Thr156. At the end of the simulation at 100 ns as shown in Figure 12.6B, strong hydrogen bonds were formed between the modified moieties and Asp80, Arg153 on chain A and Gly96 and Arg95 on chain C. It was observed, that although the interacting residues varied over the course of the simulation, most of the interactions were

steadily maintained with Thr153. These same set of residues were amongst the highest energy contributing residues in the per-residue energy decomposition calculations shown in Table S1. These strong interactions could have contributed to the stronger binding of AH10-7 over KRN7000 as estimated. These could also contribute to the enhanced therapeutic activity in human cells.

Likewise, we monitored the interactions of KRN7000 at the hydrophobic pocket at the same time-scales of 10 ns, 40 ns, 70 ns and 100 ns, with keen interest on the “non-modified” region, allowing for comparative analysis with AH10-7 so as to properly comprehend the impact of the dual modification. As shown in Figure 12.6A, at 10 ns, although KRN7000 almost engaged in similar number of hydrogen bond interactions as AH10-7, KRN7000 had fewer hydrophobic contacts. Also, KRN7000 did not interact with Ser30 which was shown to form a strong hydrogen bond with AH10-7 at 10 ns. Similarly, at 40 ns, as shown in Figure 12.6A, whilst AH10-7 formed an even stronger hydrogen bond with Ser30, KRN7000 did not still engage in any interaction with Ser30 with few hydrophobic contacts as well. Towards the end of the simulation at 70 ns as shown in Figure 12.6B, although AH10-7 formed two strong hydrogen bond interactions with Asp153, there was no interaction of KRN7000 with Asp153. At the end of our simulation at 100 ns, AH10-7 again formed strong hydrogen bond interaction with Arg153, an interaction that did not occur again with KRN7000. Therefore, taken together, the strong hydrogen bond interactions engaged between AH10-7 and Ser30 at 10 ns and 40 ns as well as the two hydrogen bond interactions with Asp153 at 70 ns and the single hydrogen bond with Arg95 on chain C, coupled with other hydrophobic contacts,

could have accounted for the more favorable activity of AH10-7 towards human TCR-CD1d variant; hence, the stronger binding free energy estimated compared to KRN7000.

Previous studies have revealed that the modified sphinganine base together with the stabilized surface binding mediated by the C6'' substitution synergistically play crucial roles in *h*CD1d binding affinity and Th1 biasing. From our findings (Figure 12.4A and B), we observed that the modified sphinganine base was positioned in the deep hydrophobic pockets of the receptor. However, the substituted hydrocinnamoyl ester group was protruded and made stabilizing contacts with surface residues which was essential for stabilizing the polar galactosyl head group for TCR recognition, affinity and potency (Figure 12.5). Additional interactions elicited by the C6'' substituents presumably compensated for the sphinganine base modification. Hence, strong affinity and interaction compensation synergistically mediated by the C6'' substituent and sphinganine base (4'OH) modification could suggest the Th1 biasing activity and high *h*CD1d binding affinity/potency of AH10-7. These are in line with previous reports [34].

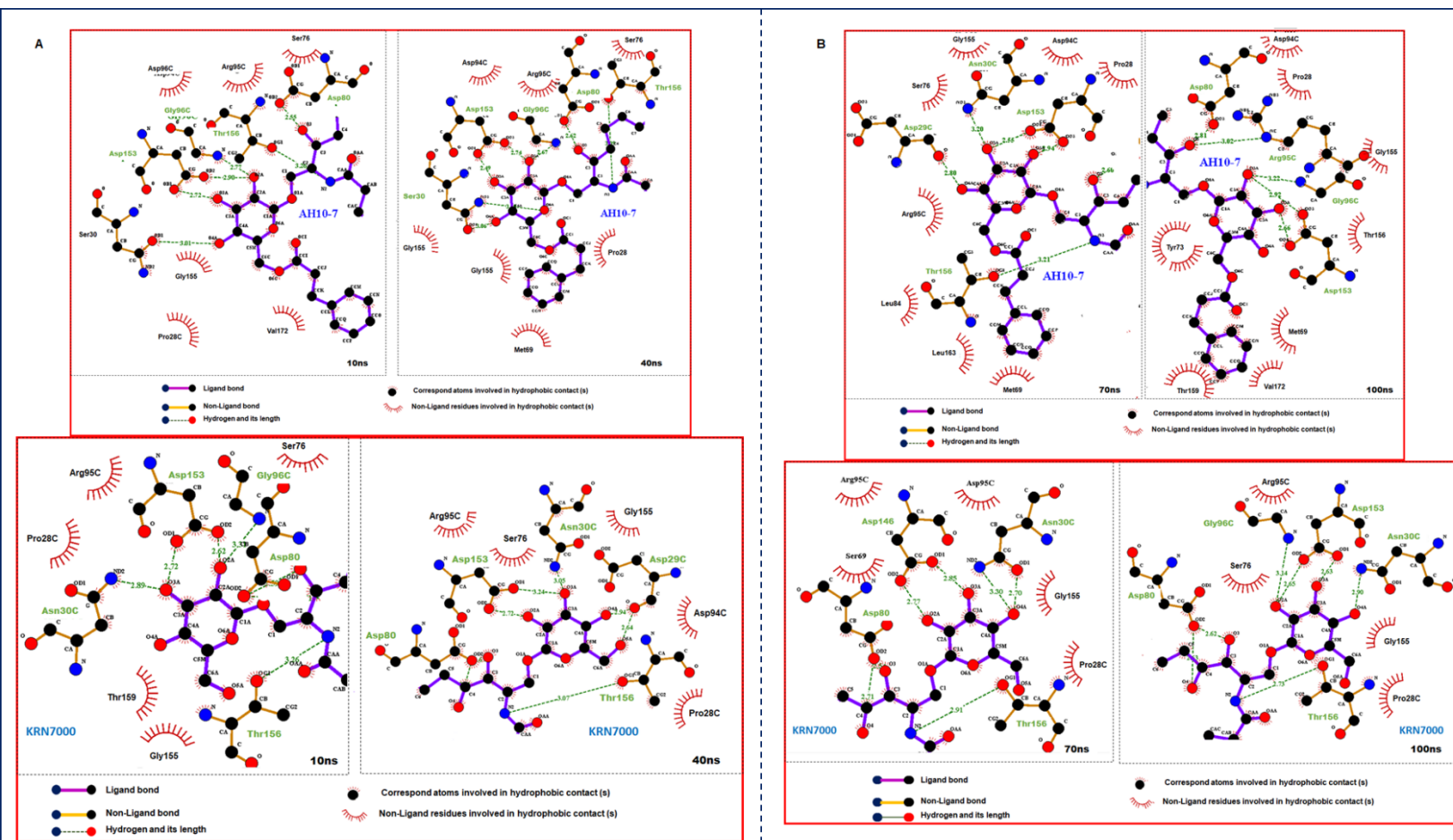


Figure 12.6: [A] Comparative time-based monitoring of complimentary interactions of “modified” and “non-modified” regions of AH10-7 and KRN7000, respectively at 10 ns and 40 ns. [B] Comparative time-based monitoring of complimentary interactions of “modified” and “non-modified” regions of AH10-7 and KRN7000, respectively at 70 ns and 100 ns.

3.4 Conformational effects of AH10-7 and KRN7000 and differential binding on hTCR-CD1d

Since proteins are basically made up of different amino acids [57], structural changes and alterations among these components, particularly those involved in ligand binding and stabilization could greatly influence overall protein function [58,59]. Moreover, the binding of small molecule inhibitors greatly influences the integrity of the protein active site and

overall conformation with regards to structure-functional inactivation or activation [59]. We, therefore, performed a conformational analysis of *h*TCR-CD1d binding site residues over the course of simulation comparing their dynamics in the unbound state and when distinctly bound by both AH10-7 and KRN7000. This procedure also enabled a thorough insight into the impact of the incorporated groups on AH10-7 relative to its superior therapeutic efficacy. Prior to the analysis of the structural dynamics, we monitored the validity of our simulated models which was ascertained by calculating the root mean square deviations (RMSD) of the C α atoms of *h*TCR-CD1d across the 100 ns simulation as presented in Figure S1 [60–62]. It was observed from the plot that all systems achieved convergence after about 20 ns following an initial gradual rise in structural deviation due to motions of constituent C α atoms from the beginning of the simulation. Overall unbound *h*TCR-CD1d exhibited the highest deviation of C α atoms with an average RMSD of 8.84Å while the KRN7000 bound system showcased the lowest deviation with an average RMSD of 6.70Å while the AH10-7-bound *h*TCR-CD1d exhibited an average RMSD of 7.14Å. Taken together, the bound systems were characterized by comparatively lower deviations of C α atoms relative to the unbound *h*TCR-CD1d, suggesting that the binding of AH10-7 and KRN7000 stabilized *h*TCR-CD1d structure. With all studied systems having achieved convergence and subsequent stability after about 20 ns, it is suggestive that subsequent deductions and assumptions drawn from the generated MD trajectories are trustworthy [63]. Moreover, the binding of AH10-7 and KRN7000 to *h*TCR-CD1d induced a consistent pattern of structural alteration, which was characterized by a decrease in deviation among the C α backbone atoms in contrast to the unbound *h*TCR-CD1d conformation that relatively exhibited higher structural instability. Suggestively, the minimal

disruption of the backbone atoms in *h*TCR-CD1d upon binding of AH10-7 and KRN7000 could highlight the mechanistic therapeutic potencies of these α -GalCer derivatives.

We then proceeded to calculate the root mean square fluctuation (RMSF) of the C α atoms as well as the solvent accessible surface area (SASA) of each of the residues on both chains over the course of the 100 ns simulation as shown in Figure 12.7. With RMSF calculations as shown in Figures 12.7C and 12.7D, we were able to assess the flexibility of each of the residues while SASA (Figures 12.7A and 12.7B) calculations showcased the exposure and unfolding mechanism of buried residues within the hydrophobic pocket which could have engaged in essential interactions with bound ligands. From the RMSF plots it was observed that the ligand bound conformations of the hydrophobic pocket residues were less flexible relative to the unbound conformation, possibly implying that the binding of both AH10-7 and KRN7000 restricted the movement of these residues and subsequently imposed some level of rigidity. However, relative to KRN7000, hydrophobic residues in the AH10-7-bound system were more flexible with an average RMSF of 7.53Å while KRN7000 exhibited an average RMSF of 5.8Å. The higher flexibility could be attributed to the structural destabilization and distortion induced by the structural modification incorporated in AH10-7. The SASA plots also revealed that the binding of both AH10-7 and KRN7000 induced lower average SASA values relative to the unbound *h*TCR-CD1d which was consistent with protein folding and an expulsion of the hydrophobic residues toward the protein core. However, among AH10-7 and KRN7000, AH10-7 showcased higher average SASA which further implied that the hydrophobic residues were more exposed to the aqueous environment hereby exposing buried surfaces, which justifies an increase in SASA value. The exposure

of the buried surfaces could have potentially favored the suitable intermolecular interactions that contributed to the enhanced binding free energy estimated.

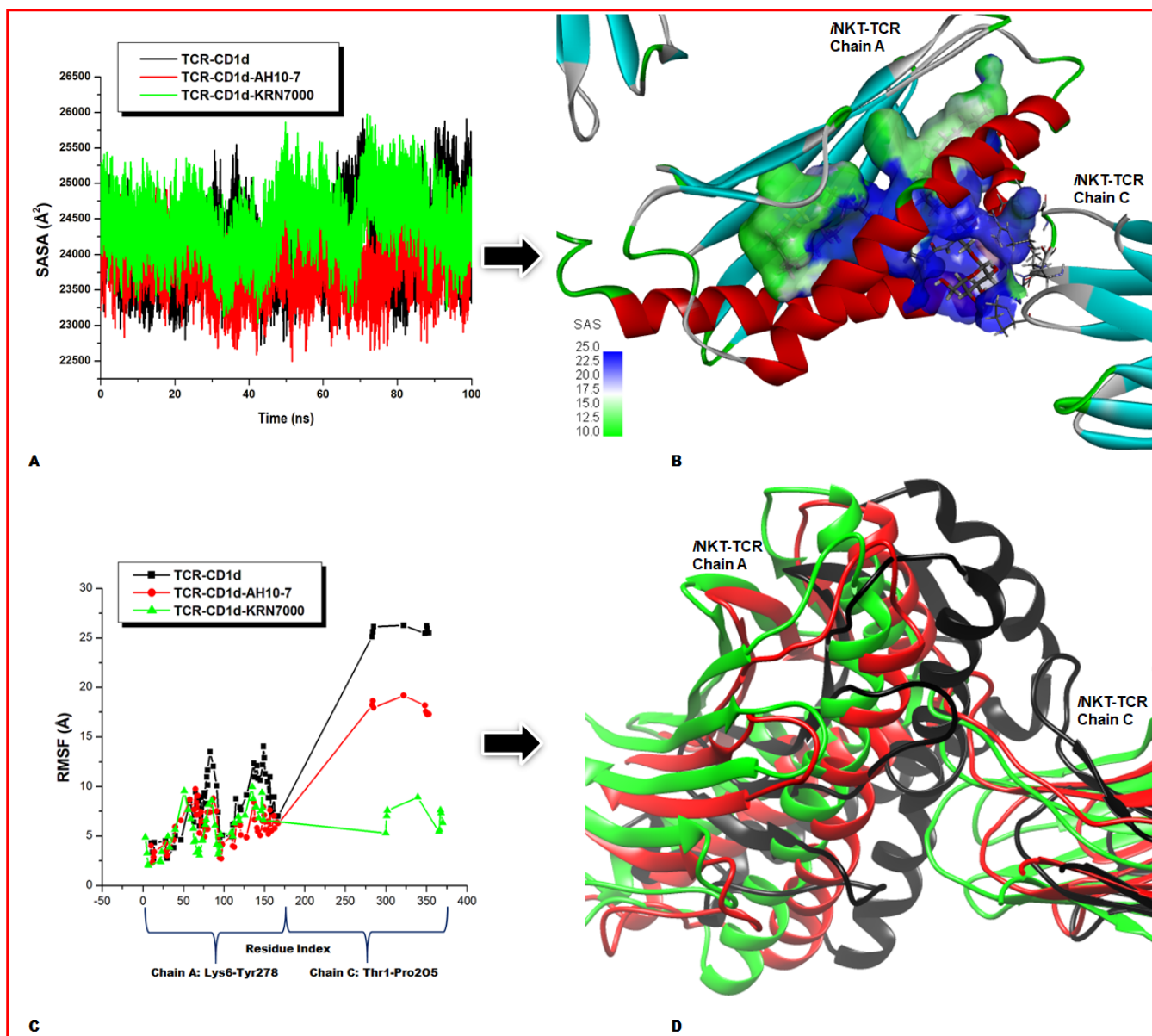


Figure 12.7: A) Comparative SASA plots of residues that make up the binding pocket of both AH10-7 and KRN7000 as well as the unbound conformation. B) 3D surface representation of the higher SASA of the AH10-7 complex with hydrocinnamoyl ester region showcasing the highest SAS (deep blue). C) Comparative RMSF plots of binding site residues with the unbound *i*NKT-TCR showing the highest residue flexibility. D) 3D representation of superimposed X-ray crystal structures of unbound *i*NKT-TCR (black) and the AH10-7 (red) and KRN7000 (green) bound complexes.

4.0 Conclusion

The plethora of scientific reports that have been conducted over the years towards enhancing the therapeutic competence of earlier α -GalCer derivatives towards the design of improved molecules such as AH10-7. The recent discovery of AH10-7, which exhibited impressive therapeutic efficacy in human *i*NKT-TCR variant typified a significant breakthrough in the search for an alternative approach to physiological immune surveillance. The uniqueness of AH10-7 over previous α -GalCer derivatives is attributed to its sphingoid base and hydrocinnamoyl ester on its C6". However, regardless of the exciting experimental studies, there is little to no reports on the structural dynamics, molecular mechanisms and binding profile of AH10-7 in support of its therapeutic prowess. As such, this study sought to provide essential and augmentative details using MD simulation techniques coupled with post-MD analytical methods such as RMSD, RMSF, SASA and MM/PBSA. Dual modifications of AH10-7 enhanced its binding superiority over the earlier α -GalCer KRN7000 by forming stronger intermolecular and hydrophobic interactions with some specific binding site residues of TCR-CD1d. Unique hydrogen bonds interactions of AH10-7 with Ser30, Asp153 on chain A and Arg95 on the chain C could have equally accounted for its outstanding binding superiority as estimated in the binding free energy analysis. Although there exists similarity in ligand-interacting residues due to structural similarities between AH10-7 and KRN7000, the introduced hydrocinnamoyl ester in AH10-7 potentiated favorable intermolecular interactions which probably contributed to enhanced binding. Also, the dual modification on AH10-7 induced a higher structural flexibility and increased the exposure of buried hydrophobic residues which could have favored their respective interactions when

compared with KRN7000. Findings unveiled atomistic insights into the binding profile of AH10-7 and could serve as a basis for further research into sphinganine derivatives towards the design of novel immunotherapeutic agents that could augment current cancer therapies.

5.0 Conflicts of Interest

Authors declare no financial and intellectual conflict of interests.

6.0 Ethics approval and consent to participate

Not applicable

7.0 Human and Animal Rights

No Animals/Humans were used in the course of this study.

8.0 Consent for Publication

Not Applicable

9.0 Funding

No funding was obtained for this study

10.0 Acknowledgements

We also acknowledge the Center for High Performance Computing (CHPC, www.chpc.ac.za), Cape Town, for computational resources.

11.0 Availability of Data And Materials

Not applicable

ORCID ID: Houda Washah: <https://orcid.org/0000-0003-2572-6839>

Clement Agoni: <https://orcid.org/0000-0001-6130-8031>

Fisayo A. Olotu: <https://orcid.org/0000-0003-3604-5983>

Geraldene Munsamy: <https://orcid.org/0000-0002-0619-2109>

Mahmoud Soliman: <https://orcid.org/0000-0002-8711-7783>

11.0 References

- [1] Ribatti, D. The Concept of Immune Surveillance against Tumors: The First Theories. *Oncotarget*, **2017**, 8, 7175–7180.
- [2] Dellabona, P.; Padovan, E.; Casorati, G.; Brockhaus, M.; Lanzavecchia, A. An Invariant V α 24-J α Q/V β 11 T Cell Receptor Is Expressed in All Individuals by Clonally Expanded CD4–8– T Cells. *J. Exp. Med.*, **1994**, 180, 1171–1176.
- [3] Lantz, O.; Bendelac, A. An Invariant T Cell Receptor α Chain Is Used by a Unique Subset of Major Histocompatibility Complex Class I- Specific CD4+ and CD4–8– T Cells in Mice and Humans. *J. Exp. Med.*, **1994**, 180, 1097–1106.
- [4] McEwen-Smith, R.M.; Salio, M.; Cerundolo, V. The Regulatory Role of Invariant NKT Cells in Tumor Immunity. *Cancer Immunol. Res.*, **2015**, 3, 425–435.
- [5] Lam, P.Y.; Nissen, M.D.; Mattarollo, S.R. Invariant Natural Killer T Cells in Immune Regulation of Blood Cancers: Harnessing Their Potential in Immunotherapies. *Front. Immunol.*, **2017**, 8, 1355.
- [6] Brennan, P.J.; Brigl, M.; Brenner, M.B. Invariant Natural Killer T Cells: An Innate Activation Scheme Linked to Diverse Effector Functions. *Nat. Rev. Immunol.*, **2013**, 13, 101–117.
- [7] Bendelac, A.; Savage, P.B.; Teyton, L. The Biology of NKT Cells. *Annu. Rev. Immunol.*, **2007**, 25, 297–336.
- [8] Macho-Fernandez, E.; Brigl, M. The Extended Family of CD1d-Restricted NKT Cells: Sifting through a Mixed Bag of TCRs, Antigens, and Functions. *Front. Immunol.*, **2015**, 6, 362.
- [9] Crosby, C.M.; Kronenberg, M. Tissue-Specific Functions of Invariant Natural Killer T Cells. *Nat. Rev. Immunol.*, **2018**, 18, 559–574.
- [10] Rossjohn, J.; Pellicci, D.G.; Patel, O.; Gapin, L.; Godfrey, D.I. Recognition of CD1d-Restricted Antigens by Natural Killer T Cells. *Nat. Rev. Immunol.*, **2012**, 12, 845–857.

- [11] Burdin, N. et al. Selective Ability of Mouse CD1 to Present Glycolipids: α -Galactosylceramide Specifically Stimulates V α 14⁺ NK T Lymphocytes. *J. Immunol.*, **1998**, *161*, 3271–3281.
- [12] Jones, E.Y.; Salio, M.; Cerundolo, V. T Cell Receptors Get Back to Basics. *Nat. Immunol.*, **2007**, *8*, 1033–1035.
- [13] Dougan, S.K.; Kaser, A.; Blumberg, R.S. CD1 Expression on Antigen-Presenting Cells. In: Moody DB, Editor. T Cell Activation by CD1 and Lipid Antigens. Berlin, Heidelberg. *Springer*, **2007**, 113–41. ISBN:978-3-540-65511-0.
- [14] Griewank, K.; Borowski, C.; Rietdijk, S.; Wang, N.; Julien, A.; Wei, D.G. et. al. Homotypic Interactions Mediated by Slamf1 and Slamf6 Receptors Control NKT Cell Lineage Development. *Immunity*, **2007**, *27*, 751–762.
- [15] Coquet, J.M.; Chakravarti, S.; Kyparissoudis, K.; McNab, F.W.; Pitt, L.A.; McKenzie, B.S. et al. Diverse Cytokine Production by NKT Cell Subsets and Identification of an IL-17-Producing CD4-NK1.1- NKT Cell Population. *Proc. Natl. Acad. Sci. U S A*, **2008**, *105*, 11287–11292.
- [16] Lynch, L.; Michelet, X.; Zhang, S.; Brennan, P.J.; Moseman, A.; Lester, C. et. al. Regulatory INKT Cells Lack Expression of the Transcription Factor PLZF and Control the Homeostasis of T(Reg) Cells and Macrophages in Adipose Tissue. *Nat Immunol*, **2015**, *16*, 85–96.
- [17] Lynch, L.; Nowak, M.; Varghese, B.; Clark, J.; Hogan, A.E.; Toxavidis, V. et. al. Adipose Tissue Invariant NKT Cells Protect against Diet-Induced Obesity and Metabolic Disorder through Regulatory Cytokine Production. *Immunity*, **2012**, *37*, 574–587.
- [18] Stetson, D.B.; Mohrs, M.; Reinhardt, R.L.; Baron, J.L.; Wang, Z.E.; Gapin, L. et al. Constitutive Cytokine MRNAs Mark Natural Killer (NK) and NK T Cells Poised for Rapid Effector Function. *J. Exp. Med.*, **2003**, *198*, 1069–1076.
- [19] Beyaz, S.; Kim, J.H.; Pinello, L.; Xifaras, M.E.; Hu, Y.; Huang, J. et al. The Histone Demethylase UTX Regulates the Lineage-Specific Epigenetic Program of Invariant Natural

Killer T Cells. *Nat. Immunol.*, **2017**, *18*, 184–195.

[20] Wesley, J.D.; Tessmer, M.S.; Chaukos, D.; Brossay, L. NK Cell-like Behavior of Valpha14i NK T Cells during MCMV Infection. *PLoS Pathog.*, **2008**, *4*, e1000106.

[21] Nieuwenhuis, E.E.S.; Matsumoto, T.; Exley, M.; Schleipman, R.A.; Glickman, J.; Bailey, D.T. et al. CD1d-Dependent Macrophage-Mediated Clearance of *Pseudomonas aeruginosa* from Lung. *Nat. Med.*, **2002**, *8*, 588–593.

[22] Schmieg, J.; Yang, G.; Franck, R.W.; Van Rooijen, N.; Tsuji, M. Glycolipid Presentation to Natural Killer T Cells Differs in an Organ-Dependent Fashion. *Proc. Natl. Acad. Sci. U S A*, **2005**, *102*, 1127–1132.

[23] Monteiro, M.; Graca, L. *i*NKT Cells: Innate Lymphocytes with a Diverse Response. *Crit. Rev. Immunol.*, **2014**, *34*, 81–90.

[24] Yamashita, K.; Arimoto, A.; Nishi, M.; Tanaka, T.; Fujita, M.; Fukuoka, E. et al. Application of *i*NKT Cell-Targeted Active Immunotherapy in Cancer Treatment. *Anticancer Res.*, **2018**, *38*, 4233–4239.

[25] Crowe, N.Y.; Smyth, M.J.; Godfrey, D.I. A Critical Role for Natural Killer T Cells in Immunosurveillance of Methylcholanthrene-Induced Sarcomas. *J. Exp. Med.*, **2002**, *196*, 119–127.

[26] Cerundolo, V.; Salio, M. Harnessing NKT Cells for Therapeutic Applications. *Curr. Top. Microbiol. Immunol.*, **2007**, *314*, 325–340.

[27] Bedard, M.; Salio, M.; Cerundolo, V. Harnessing the Power of Invariant Natural Killer T Cells in Cancer Immunotherapy. *Front. Immunol.*, **2017**, *8*, 1829.

[28] Morita, M.; Motoki, K.; Akimoto, K.; Natori, T.; Sakai, T.; Sawa, E.; Yamaji, K.; Koezuka, Y.; Kobayashi, E.; Fukushima, H. Structure-Activity Relationship of Alpha-Galactosylceramides against B16-Bearing Mice. *J. Med. Chem.*, **1995**, *38*, 2176–2187.

[29] Kawano, T.; Cui, J.; Koezuka, Y.; Taura, I.; Kaneko, Y.; Motoki, K.; Ueno, H.; Nakagawa, R.; Sato, H.; Kondo, E. et al. CD1d-Restricted and TCR-mediated Activation of

Valpha14 NKT Cells by Glycosylceramides. *Science*, **1997**, *278*, 1626–1629.

[30] Natori, T.; Morita, M.; Akimoto, K.; Koezuka, Y. Agelasphins, Novel Antitumor and Immunostimulatory Cerebrosides from the Marine Sponge *Agelas-mauritianus*. *Tetrahedron*, **1994**, *50*, 2771–2784.

[31] Kharkwal, S.S.; Arora, P.; Porcelli, S.A. Glycolipid Activators of Invariant NKT Cells as Vaccine Adjuvants. *Immunogenetics*, **2016**, *68*, 597–610.

[32] Kobayashi, E.; Motoki, K.; Uchida, T.; Fukushima, H.; Koezuka, Y. KRN7000, a Novel Immunomodulator, and Its Antitumor Activities. *Oncol. Res.*, **1995**, *7*, 529–534.

[33] Kronenberg, M.; Gapin, L. The Unconventional Lifestyle of NKT Cells. *Nat. Rev. Immunol.*, **2002**, *2*, 557–568.

[34] Chennamadhavuni, D.; Saavedra-Avila, N.A.; Carreño, L.J.; Guberman-Pfeffer, M.J.; Arora, P.; Yongqing, T. et. al. Dual Modifications of α -Galactosylceramide Synergize to Promote Activation of Human Invariant Natural Killer T Cells and Stimulate Anti-Tumor Immunity. *Cell Chem. Biol.*, **2018**, *25*, P571–584.e8.

[35] Nair, S.; Dhodapkar, M.V. Natural Killer T Cells in Cancer Immunotherapy. *Front. Immunol.*, **2017**, *8*, 1178.

[36] Yu, K.O.; Porcelli, S.A. The Diverse Functions of CD1d-Restricted NKT Cells and Their Potential for Immunotherapy. *Immunol. Lett.*, **2005**, *100*, 42–55.

[37] Laurent, X.; Bertin, B.; Renault, N.; Farce, A.; Speca, S.; Milhomme, O.; Millet, R.; Desreumaux, P.; Henon, E.; Chavatte, P. Switching Invariant Natural Killer T (*i*NKT) Cell Response from Anticancerous to Anti-Inflammatory Effect: Molecular Bases. *J. Med. Chem.*, **2014**, *57*, 5489–5508.

[38] Kopecky-Bromberg, S.A.; Fraser, K.A.; Pica, N.; Carnero, E.; Moran, T.M.; Franck, R.W.; Tsuji, M.; Palese, P. Alpha-C-Galactosylceramide as an Adjuvant for a Live Attenuated Influenza Virus Vaccine. *Vaccine*, **2009**, *27*, 3766.

[39] Salmaso, V.; Moro, S. Bridging Molecular Docking to Molecular Dynamics in

Exploring Ligand-Protein Recognition Process: An Overview. *Front. Pharmacol.*, **2018**, *9*, 923.

[40] Salmaso, V. Exploring Protein Flexibility During Docking to Investigate Ligand-Target Recognition., University of Padova, Padova., **2018**.<http://paduaresearch.cab.unipd.it/10780/>

[41] Berman, H.M.; Battistuz, T.; Bhat, T.N.; Bluhm, W.F. et. al. The Protein Data Bank. *Biol. Crystallogr.*, **2002**, 899–907.

[42] Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.*, **2004**, *25*, 1605–1612.

[43] Case, D. a.; T.E. Cheatham, I.; Darden, T.; Gohlke, H.; Luo, R.; K.M. Merz, J.; Onufriev, a.; Simmerling, C.; Wang, B.; Woods., R. The Amber Biomelecular Simulation Programs. *J. Comput. Chem.*, **2005**, *26*, 1668–1688.

[44] Grest, G.S.; Kremer, K. Molecular Dynamics Simulation for Polymers in the Presence of a Heat Bath. *Phys. Rev. A*, **1986**, *33*, 3628–3631.

[45] Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A; Haak, J.R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, **1984**, *81*, 3684–3690.

[46] 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.*, **1977**, *23*, 327–341.

[47] Roe, D.R.; Cheatham III, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.*, **2013**, *9*, 3084–3095.

[48] Seifert, E. OriginPro 9.1: Scientific Data Analysis and Graphing Software-Software Review. *J. Chem. Inf. Model.*, **2014**, *54*, 1552.

- [49] Hou, T.; Wang, J.; Li, Y.; Wang, W.; Houa, T.; Wangb, J.; Lia, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Comput. Sci.*, **2011**, *51*, 69–82.
- [50] Agoni, C.; Ramharack, P.; Soliman, M.E.S. Synergistic Interplay of The Co-Administration of Rifampin And Newly Developed Anti-TB Drug : Could It Be a Promising New Line of TB Therapy ? *Comb. Chem. High Throughput Screen.*, **2018**, *8*, 40187–40197. <https://doi.org/10.2174/1386207321666180716093617>
- [51] Agoni, C.; Ramharack, P.; Soliman, M.E.S. Allosteric Inhibition Induces an Open WPD-Loop: A New Avenue towards Glioblastoma Therapy. *RSC Adv.*, **2018**, *8*, 40187–40197.
- [52] Agoni, C.; Ramharack, P.; Soliman, M. Co-Inhibition as a Strategic Therapeutic Approach to Overcome Rifampin Resistance in Tuberculosis Therapy: Atomistic Insights. *Future Med. Chem.*, **2018**, *10*, 1665–1675.
- [53] Munsamy, G.; Agoni, C.; Soliman, M.E.S. A Dual Target of Plasmepsin IX and X: Unveiling the Atomistic Superiority of a Core Chemical Scaffold in Malaria Therapy. *J. Cell Biol.*, **2018**, 1–12.
- [54] Olotu, F.A.; Agoni, C.; Adeniji, E.; Abdullahi, M.; Soliman, M.E. Probing Gallate-Mediated Selectivity and High-Affinity Binding of Epigallocatechin Gallate: A Way-Forward in the Design of Selective Inhibitors for Anti-Apoptotic Bcl-2 Proteins. *Appl. Biochem. Biotechnol.*, **2019**, *187*, 1061-1080.
- [55] Miller, B.R.; Mcgee, T.D.; Swails, J.M.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA . Py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.*, **2012**, *8*, 3314-3321.
- [56] Akher, F.B.; Farrokhzadeh, A.; Olotu, F.A.; Agoni, C.; Soliman, M.E.S. The irony of chirality – unveiling the distinct mechanistic binding and activities of 1-(3-(4-amino-5-(7-methoxy-5-methylbenzo[b]thiophen-2-yl)-7H-pyrrolo[2,3-d]pyrimidin-7-yl)pyrrolidin-1-

yl)prop-2-en-1-one enantiomers as irreversible covalent FGFR4 inhibitors. *Org. Biomol. Chem.*, **2019**, *17*, 1176-1190. DOI: 10.1039/C8OB02811G

[57] Wu, G. Amino Acids: Metabolism, Functions, and Nutrition. *Amino Acids*, **2009**, *37*, 1–17.

[58] Mukherjee, J.; Gupta, M.N. Increasing Importance of Protein Flexibility in Designing Biocatalytic Processes. *Biotechnol. Reports*, **2015**, *6*, 119–123.

[59] Olotu, F.A.; Soliman, M.E.S. Mutational Inactivation to Aberrant Gain-of-Function: Unraveling the Structural Basis of Mutant P53 Oncogenic Transition. *J. Cell. Biochem.*, **2018**, *119*, 2646–2652.

[60] Maiorov, V.N.; Crippen, G.M. Significance of Root-Mean-Square Deviation in Comparing Three-Dimensional Structures of Globular Proteins. *J. Mol. Biol.*, **1994**, *235*, 625–634.

[61] Knapp, B.; Frantal, S.; Cibena, M.; Schreiner, W.; Bauer, P. Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible? *J. Comput. Biol.*, **2011**, *18*, 997–1005.

[62] Pitera, J.W. Expected Distributions of Root-Mean-Square Positional Deviations in Proteins. *J. Phys. Chem. B*, **2014**, *118*, 6526–6530.

[63] Mhlongo, N.N.; Soliman, M.E.S. Single H5N1 Influenza A Neuraminidase Mutation Develops Resistance to Oseltamivir Due to Distorted Conformational and Drug Binding Landscape: Multiple Molecular Dynamics Analyses. *RSC Adv.*, **2015**, *5*, 10849–10861.

CHAPTER 13

13.0 Conclusion and Future Perspectives

13.1 Conclusions

Studies in this thesis employed CADD approaches in unraveling the mechanism of inhibition of some known therapeutic targets in treating viral infections, cancer, and tuberculosis. Drug-target interactions and mechanisms of actions elucidated in this thesis are relevant since the interaction dynamics of drugs at binding sites play a vital role in the drug discovery process by influencing new drugs' overall therapeutic success. These interactions dynamics also provide essential insights that could inform the identification of new drug compounds for biological targets or improving the therapeutic activity of existing compounds. Therapeutic targets investigated included; SARS-CoV-2 RNA dependent RNA polymerase (SARS-CoV-2 RdRp), Human Rhinovirus B14 (HRV-B14) and human *N*-myristoyltransferases in viral infections, Dihydrofolate reductase (DHFR) and Flavin-dependent thymidylate synthase (FDTS) in tuberculosis, human variants of TCR-CD1d and Protein Tyrosine Phosphatase Receptor Zeta (PTPRZ). As such, the focus of this research was divided into three domains. The research that culminated in this thesis began with a thorough review of the concept of druggability and drug-likeness since the crux of the following research objectives revolved around therapeutics and their inhibitions by small molecule inhibitors. The review highlighted druggability and drug-likeness principles and their recent advancements in the drug discovery field. Subsequently, the different computational tools, their predictive analysis and reliability in drug discovery were presented.

In the first domain of the research, studies presented sought to unravel some small molecule inhibitors' mechanism against some therapeutic targets in viral infections by explicitly focusing on the therapeutic targets; SARS-CoV-2 RdRp, HRV-B14, and *N*-myristoyltransferase.

Studies in the first domain of the research began with an exploration of the binding mechanism of Remdesivir by unraveling the structural and conformational implications on SARS-CoV-2 RdRp towards identifying crucial pharmacophoric moieties of Remdesivir. Our analysis showed that the modulatory activity of Remdesivir is characterized by an extensive array of high-affinity and consistent molecular interactions with specific active site residues that anchor Remdesivir within the binding pocket for efficient binding. Results also showed that Remdesivir binding induces minimal individual amino acid perturbations, subtly interferes with deviations of C- α atoms, and restricts the systematic transition of SARS-CoV-2 RdRp from the “buried” hydrophobic region to the “surface-exposed” hydrophilic region. We also mapped a pharmacophore model based on observed high-affinity interactions with SARS-CoV-2 RdRp, which showcased the crucial functional moieties of Remdesivir and was subsequently employed for virtual screening. The structural insights and the optimized pharmacophoric model provided would augment the design of improved analogs of Remdesivir that could expand treatment options for COVID-19 the molecular mechanism of action of a novel benzothiophene derivative against HRV-B14 was investigated. Residue interaction network analysis revealed that the binding of the benzothiophene derivative upon binding into the “canyon” region of the active site of HRV-B14 distorts its initially extensively networked and compact residue profile with conformational changes

characterized by fewer inter-residue hydrogen bonds, reduced van der Waals interactions, and increased residue flexibility. Also, the binding of this benzothiophene derivative decreased the north-south wall's flexibility around the canyon region, possibly impeding the “breathing motion” of HRV-B14.

In addition, the structural and molecular mechanism of action associated with the dual inhibitory activity of IMP-1088, a novel compound that reportedly inhibits both HsNMT1 and HsNMT2 towards common cold therapy was also investigated. An augmentative computational and nanosecond-based analyses reveal that the steady and consistent interactions of IMP-1088 with specific residues; Tyr296, Phe190, Tyr420, Leu453, Gln496, Val181, Leu474, Glu182, and Asn246, shared within the binding pockets of both HNMT subtypes, in addition to peculiar structural changes account for its dual inhibitory potency. Findings unveil atomistic and structural perspectives that could form the basis for the design of novel dual-acting inhibitors of *N*-myristoyltransferase towards common cold therapy.

In the second domain of the research, studies presented attempted to unravel the mechanism of actions of some small molecule inhibitors against some therapeutic targets in tuberculosis therapy by unraveling the molecular mechanisms and structural dynamics associated with dual inhibitory activity of PAS-M against both DHFR and FDTS towards tuberculosis therapy. MD simulations revealed that PAS-M binding towards DHFR and FDTS is characterized by a recurrence of strong conventional hydrogen bond interactions between a peculiar site residue the 2-amino-decahydropteridin-4-ol group of PAS-M.

Using an atomistic model of mycobacterial c-ring the structural dynamics that explicate the experimentally reported antagonistic features of BDQ in halting ion shuttling by the

mycobacterial c-ring towards the treatment of TB was also described. Results showed that BDQ exhibited a considerably high ΔG while it specifically maintained high-affinity interactions with Glu65B -and Asp32B, blocking their crucial roles in proton binding and shuttling required for ATP synthesis. Moreover, the bulky nature of BDQ induced a rigid and compact conformation of the rotor c-ring, which impedes the essential rotatory motion that drives ion exchange and shuttling. The binding affinity of a BDQ molecule was considerably increased by the complementary binding of another BDQ molecule, which indicates that an increase in BDQ molecule enhances inhibitory potency against *Mtb* ATP synthase.

Furthermore, the effects of triple mutations (L59V, E61D, and I66M) on BDQ-*Mtb* F1F0 ATP-synthase binding were also investigated. Findings revealed a drastic reduction in BDQ binding affinity (ΔG) in the triple mutant protein, which was caused by a systemic loss in high-affinity interactions primarily mediated by L59, E61, and I66. Likewise, these mutations distorted the binding site and overall structural architecture of F1F0 ATP-synthase in the presence of BDQ, as revealed by the RIN and conformational analyses.

The last domain of the research in this thesis unraveled some small molecule inhibitors' mechanism against some therapeutic targets in cancer, specifically PTPRZ and *h*TCR-CD1d. In studying the possible therapeutic inhibition of PTPRZ towards cancer treatment, we explored the impact of NAZ2329, a recently identified allosteric inhibitor of PTPRZ, on the atomic flexibility of the WPD-loop. Structural analysis revealed that NAZ2329 induced an open conformation of the crucial WPD-loop, consequently impeding enzyme activity even upon substrate binding. Based on the molecular interactions between of NAZ2329 and

tyrosine-protein phosphatase zeta, a pharmacophore model was generated to exhibit the essential functional moieties of NAZ2329.

Lastly, the mechanism behind the selective inhibition *h*TCR-CD1d by AH10-7 towards the activation of *i*NKT cells was also investigated. Findings revealed that AH10-7 exhibited higher-affinity binding and structural effects on *h*TCR-CD1d, mediated by the incorporated hydrocinnamoyl ester moiety, which accounted for stronger intermolecular interactions with 'non-common' binding site residues.

Overall, this study has provided essential conformational and structural molecular insights into the design and development of new therapeutic approaches to treat cancer, viral infections, and tuberculosis through molecular modeling and CADD.

13.2 Future Scope and beyond

As evidenced in this research, CADD is vital in drug design research and has gained prominence in the pharmaceutical industry and academic research, particularly towards scouting and identifying potential drug candidates. The atomistic insights and molecular mechanisms unraveled in the research conducted in this thesis could inform future drug design processes, including but not limited to improving the potency of the studied inhibitors, repurposing the inhibitors' activity, and redesigning the compounds towards dual-targeting propensities. Also, the structural dynamics and conformational changes showcased upon inhibitor binding towards the studied inhibitors could also form the basis for the design of novel inhibitors to overcome resistance and increased selectivity. However, as technology advances, there is a need to continuously improve CADD methods towards enhancing their computational efficiency. Also, as the challenges of drug resistance and adverse side effects of drugs persist, there is a need to continue drug discovery efforts considering different perspectives. The concept of multi-targeting drugs should be extensively explored, considering its advantages in minimizing the amounts of drugs consumed for therapy and reducing the risk of side effects. Finally, the *in silico* insights provided in the reports could be augmented by experimental investigations while essential scaffolds highlighted could be optimized towards the development of novel inhibitors with improved therapeutic activity.

APPENDIX

APPENDIX A

Clement Agoni and Mahmoud Soliman (2020). “The Binding of Remdesivir to SARS-CoV-2 RNA-Dependent Polymerase May Pave the Way Towards the Design of Potential Drugs for COVID-19 Treatment”.

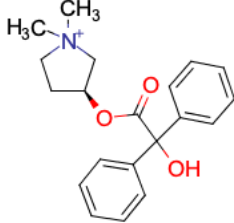
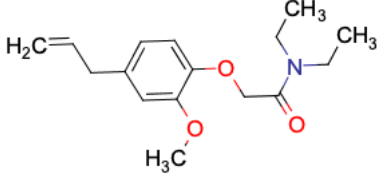
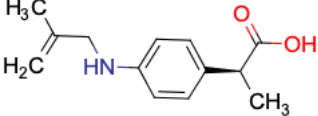
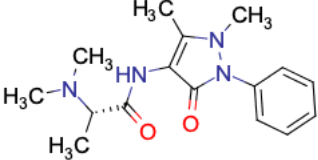
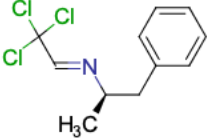
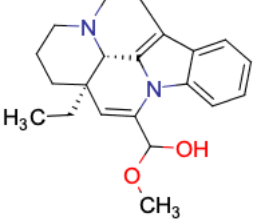
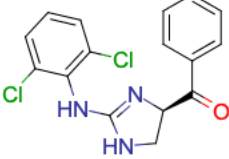
Current Pharmaceutical Biotechnology, **(in press)**

Supplementary Documents

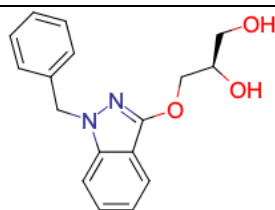
RMSD: ca	7bif.pdb, chain A	1	SAD	AQSFLNR	VCGVSAARLT	PCGTGTSTDV	VYRAFDIYND	KVAGFAKFLK
	model.pdb	117
RMSD: ca	7bif.pdb, chain A	51	TNCCRFQEKD	EDDNLIDSYF	VVKRHTFSNY	QHEETIYNLL	KDCPAVAKHD	
	model.pdb	117	
RMSD: ca	7bif.pdb, chain A	101	FFKFRIDGDM	VPHISRQR.L	TKYTMADLVY	ALRHFD.EGN.	CDTLKEILVT	
	model.pdb	117QR...LT	.KYTMADLVY	ALRHFD.EGN	CDTLKEILVT	
RMSD: ca	7bif.pdb, chain A	149	YNCCD.DD.Y	FNKCDWYDFV	ENPDILRVYA	NLGERVROAL	LKTQVQCD.A	
	model.pdb	149	YNCCD.DD.Y	FNKCDWYDFV	ENPDILRVYA	NLGERVROAL	LKTQVQCD.AM	
RMSD: ca	7bif.pdb, chain A	196	MREN.AGIVG	V.LTLDNQDLN	GNW.YDFGD	FI...QT.TP	.GSGVPVVD	
	model.pdb	197	RNA.GIVGV	.LTLDNQDLN	GNWYDF	..GDFI.QTT	PGSGVPVVD.	
RMSD: ca	7bif.pdb, chain A	236	SYYSLLMPIL	TLTRALTAES	HVD.TD.LTK	P.YI.KWDLK	YDFTEERLKL	
	model.pdb	236	SYYSLLMPIL	TLTRALTAES	HVD.TD.LTK	P.YI.KWDLK	YDFTEERLKL	
RMSD: ca	7bif.pdb, chain A	283	FDRYFKYWDQ	TYHPNCVNCL	DDRCILHCAN	FNVLFSTVFP	PTSFGPLVRK	
	model.pdb	283	FDRYFKYWDQ	TYHPNCVNCL	DDRCILHCAN	FNVLFSTVFP	PTSFGPLVRK	
RMSD: ca	7bif.pdb, chain A	333	IFVDGVPFV	STGYHFRELG	VVHNQDVN.LH	.S.SRLSFKE	LLVYAADPAM	
	model.pdb	333	IFVDGVPFV	STGYHFRELG	VVHNQDVN..	LHS.SRLSFKE	LLVYAADPAM	
RMSD: ca	7bif.pdb, chain A	381	HAASGN.L.LL	DKRTTCFSVA	ALTNN.VAFQ	TVKPGNFNKD	FYDFAVSKGF	
	model.pdb	381	HAASGN.L.LL	DKRTTCFSVA	ALTNN.VAFQ	TVKPGNFNKD	FYDFAVSKGF	
RMSD: ca	7bif.pdb, chain A	429	F.KEGSS.V	ELKHFFFAQD	GNAAISDYDY	YRYNLPMTCD	IRQLLFVVEV	
	model.pdb	429	FKEG...SSV	ELKHFFFAQD	GNAAISDYDY	YRYNLPMTCD	IRQLLFVVEV	
RMSD: ca	7bif.pdb, chain A	476	VDKYFDCYDG	GCINANQVIV	NNLDKSAGFP	FNKWGKARLY	YDSMSYEDDD	
	model.pdb	476	VDKYFDCYDG	GCINANQVIV	NNLDKSAGFP	FNKWGKARLY	YDSMSYEDDD	
RMSD: ca	7bif.pdb, chain A	526	ALFAYTKRNV	IPTITQMNLK	YAISAKNRAR	TVAGVSIKST	MTNRQFHQKL	
	model.pdb	526	ALFAYTKRNV	IPTITQMNLK	YAISAKNRAR	TVAGVSIKST	MTNRQFHQKL	
						Active site		
RMSD: ca	7bif.pdb, chain A	576	LKSTAAATRGA	TVVIGTSKFY	GGWHNMLKTV	YSDVENPHLM	GWDYPKCDRA	
	model.pdb	576	LKSTAAATRGA	TVVIGTSKFY	GGWHNMLKTV	YSDVENPHLM	GWDYPKCDRA	
RMSD: ca	7bif.pdb, chain A	626	MPNMLRIMAS	LVLARKHTTC	CSLSHRFYRL	ANECAQVLE	MVMCGGSLYV	
	model.pdb	626	MPNMLRIMAS	LVLARKHTTC	CSLSHRFYRL	ANECAQVLE	MVMCGGSLYV	
RMSD: ca	7bif.pdb, chain A	676	KPGGTSSGDA	TTAYANSVFN	ICQAVTANVN	ALLSTDGNKI	ADKYVRNLQH	
	model.pdb	676	KPGGTSSGDA	TTAYANSVFN	ICQAVTANVN	ALLSTDGNKI	ADKYVRNLQH	
RMSD: ca	7bif.pdb, chain A	726	RLYECLYRNR	DVDTDFVNEF	YAYLRKHFSM	MILSDDAVVC	FNSTYASQGL	
	model.pdb	726	RLYECLYRNR	DVDTDFVNEF	YAYLRKHFSM	MILSDDAVVC	FNSTYASQGL	
RMSD: ca	7bif.pdb, chain A	776	VASIKNFKSV	LYYQNVFMS	EAKCWTETDL	TKGPHEFCSQ	HTMLVK.QGD	
	model.pdb	776	VASIKNFKSV	LYYQNVFMS	EAKCWTETDL	TKGPHEFCSQ	HTMLVK.QGD	
RMSD: ca	7bif.pdb, chain A	825	D.VVYLPYPDP	SRILGAGCFV	D.D.I.VKI.T.DG	TLMIERFVSL	AIDAYPLTKH	
	model.pdb	826	.VVYLPYPDP	SRILGAGCFV	D.D.I.VKI.T.DG	TLMIERFVSL	AIDAYPLTKH	
RMSD: ca	7bif.pdb, chain A	873	PNQEYADV FH	LYLQYIRKLLH	D.E.L.T.GHMLD	MYS.....	..V.ML.TNDN	
	model.pdb	873	PNQEYADV FH	LYLQYIRKLLH	.D.E.L.T.GHMLD	...TGHMLDM	YSV.ML.TNDN	
RMSD: ca	7bif.pdb, chain A	912	TSRYWEP.EFY	EAMYTPHTVL	QHHHHHHHHH.H			
	model.pdb	912	TSRYWEP.EFY	

Figure S1: Comparative sequence alignment of the constructed homology model of SARS-CoV-2 RNA-dependent polymerase and the experimentally resolved X-ray crystal structure of SARS-CoV-2 RNA-dependent polymerase (PDB ID: 7BTF) . Insert highlights the active site residue, VAL557 as previously reported

Table S1: Hits compounds identified from the ZINC database using the developed pharmacophore model

ZINC ID OF HITS	2D STRUCTURE
ZINC77	
ZINC7	
ZINC22	
ZINC31	
ZINC30690712	
ZINC32275064	
ZINC72266866	

ZINC73



APPENDIX B

Clement Agoni, Pritika Ramharack, Geraldene Munsamy and Mahmoud Soliman (2020),
"Human Rhinovirus Inhibition through Capsid "Canyon" Perturbation: Structural Insights into
the Role of a Novel Benzothiophene Derivative".

Cell Biochemistry and Biophysics 78:3–13.

Supplementary Documents

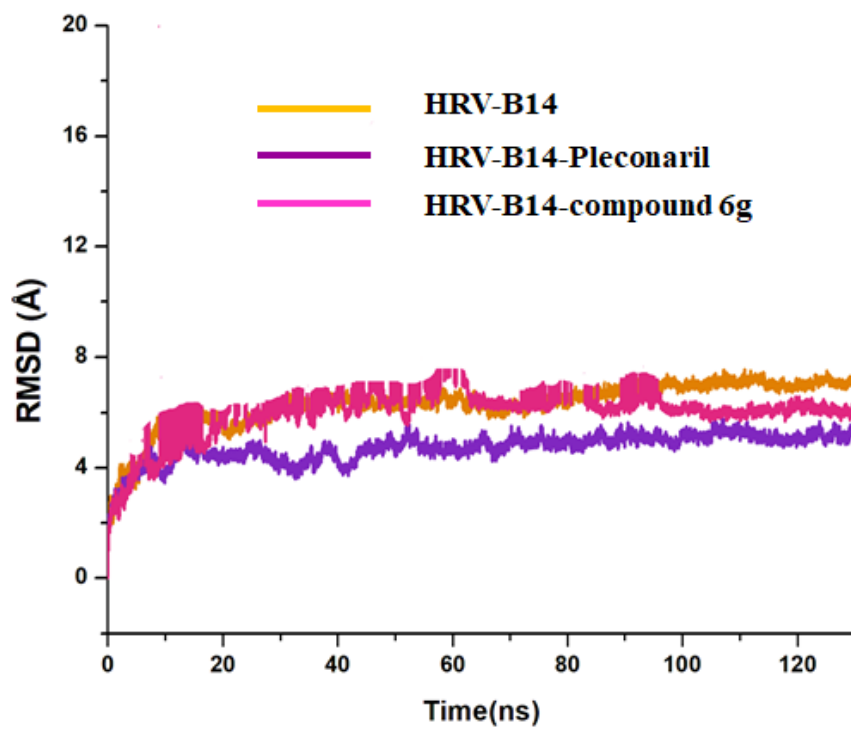


Figure S1: The RMSD plot depicting stability of each of the three systems throughout the 130ns simulation.

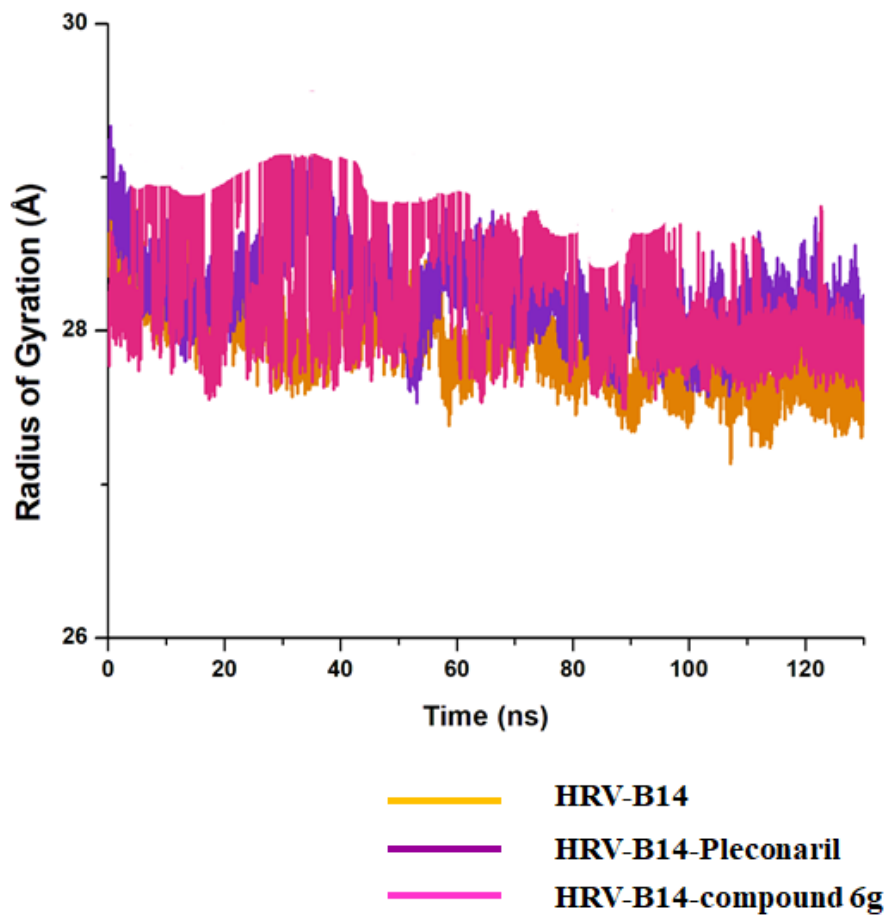


Figure S2: The RG plot illustrating the difference in compactness and surface area of all four relative to the starting minimized structure over 130 ns simulation.

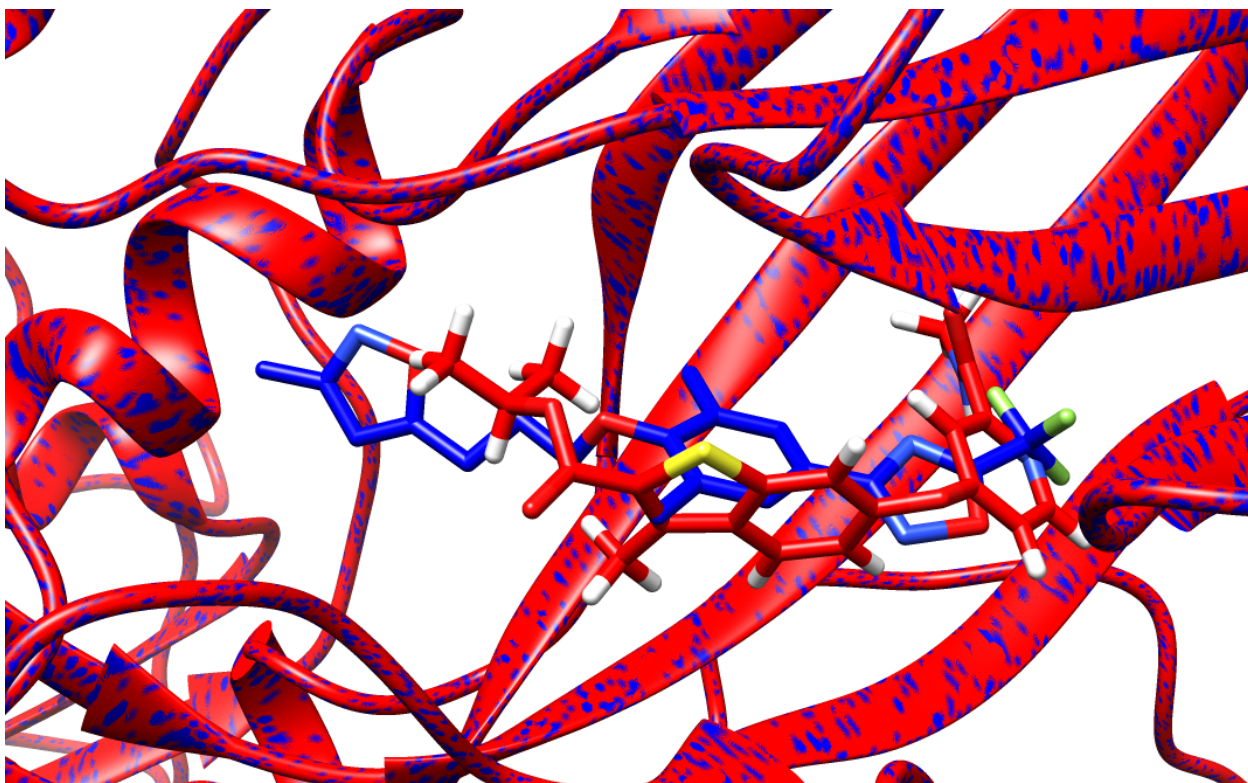


Figure S3: Superimposed representation of the Pleconaril bound complex (PDB:1NCQ) (blue) and the docked compound 6g bound complex (red). This validates that compound 6g was successfully docked into the Pleconaril binding site of HRV-B14

APPENDIX C

Clement Agoni, Eliasu Y. Salifu and Mahmoud Soliman (2020). “Dual-Targeting of Human *N*-Myristoyltransferase Subtype 1/2 by IMP-1088 Halts Common Cold Pathogenesis: Atomistic insight”.

Future Virology (Submitted)

Supplementary Documents

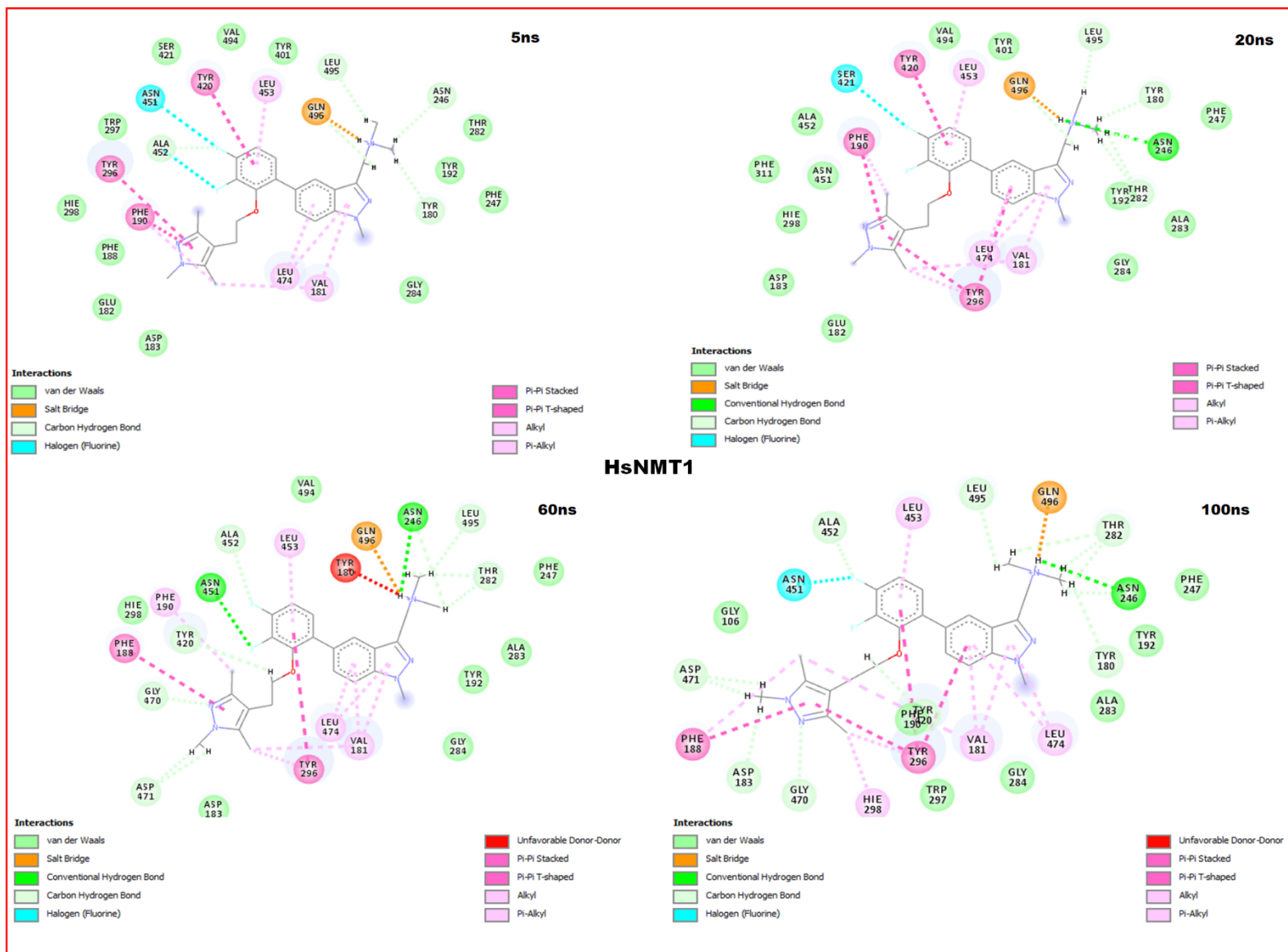


Figure S1: Showing nanosecond-based variation in residue-ligand interaction between compound IMP-1088 and binding pocket residues of HsNMT1

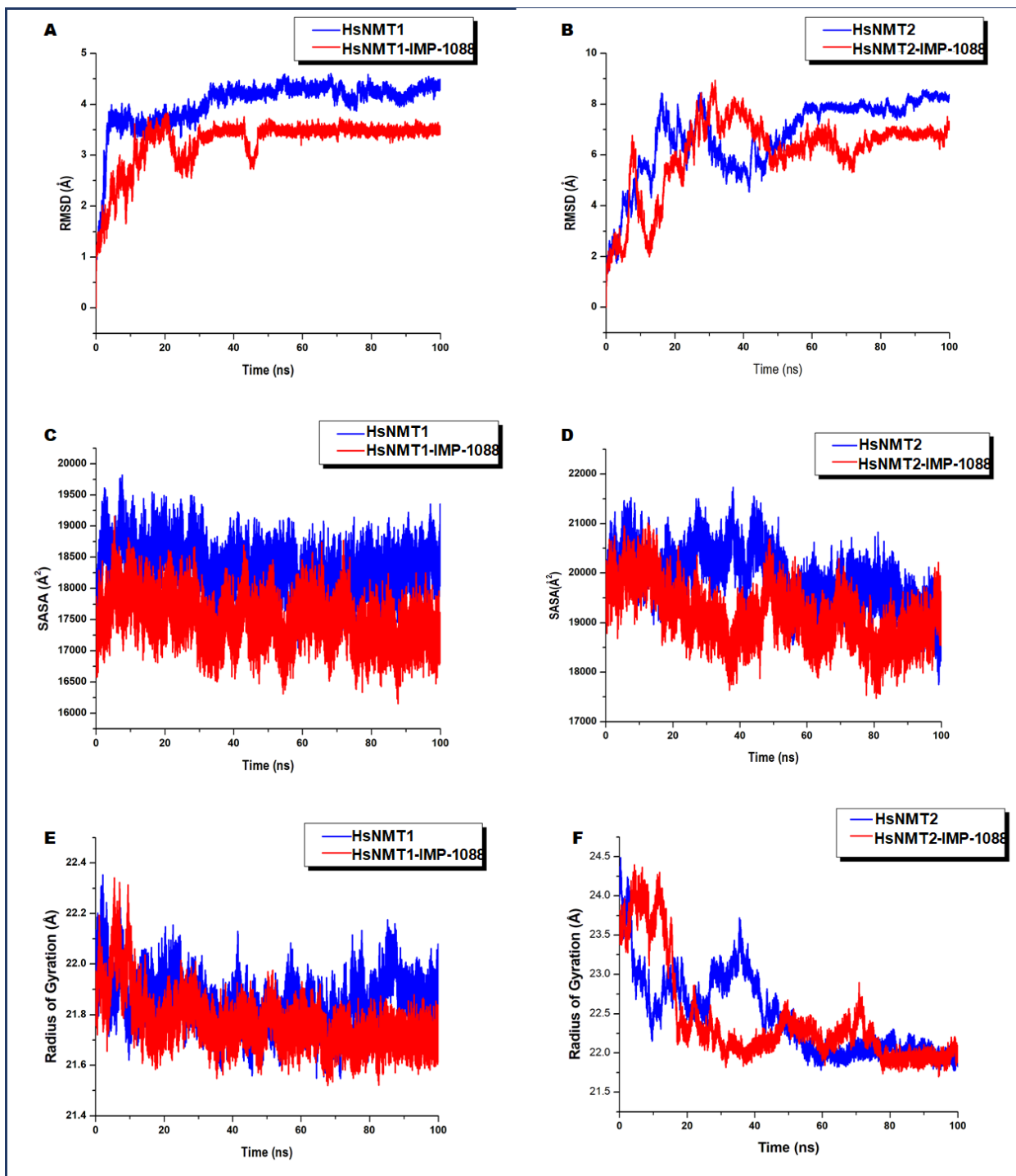


Figure S3: Showing the comparative RMSD plots (A,B), SASA plots (C,D) and radius of gyration plots (E,F) for the bound and unbound conformation of HsNMT1 and HsNMT2

APPENDIX D

Clement Agoni, Pritika Ramharack, Elliasu Y. Salifu and Mahmoud Soliman (2020). The Dual-Targeting Activity of The Metabolite Substrate of Para-Amino Salicylic Acid In The Mycobacterial Folate Pathway: Atomistic And Structural Perspectives.

The Protein Journal, 39:106–117.

Supplementary Documents

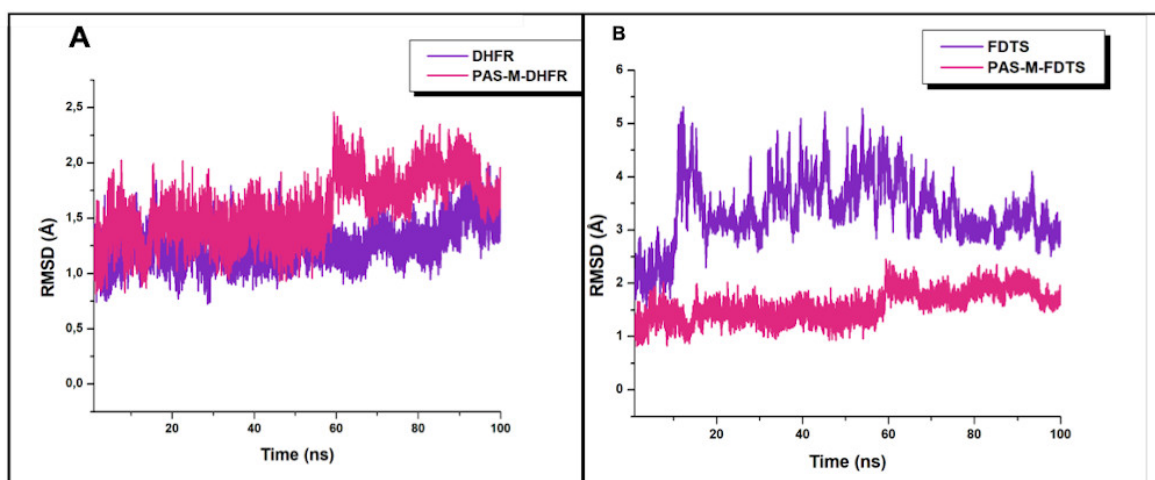


Figure S1: Comparative RMSD plot of unbound DHFR and FDTS(**magenta**) and when they are complexed with PAS-M (**purple**)

APPENDIX E

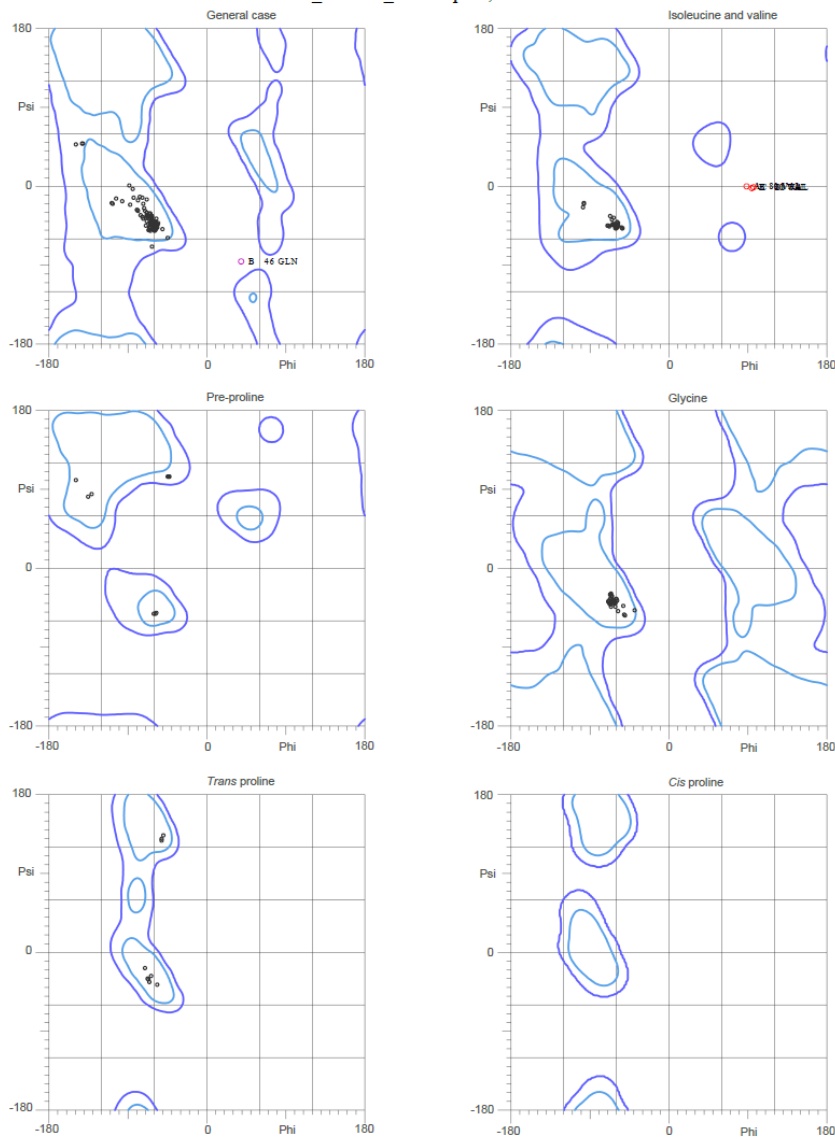
Elliasu Y. Salifu, **Clement Agoni**, Fisayo A. Olotu, Yussif M. Dokurugu, Mahmoud E. S. Soliman (2020), Triple *Mycobacterial* ATP-synthase mutations impede Bedaquiline binding: Atomistic and structural perspectives.

Computational Biology and Chemistry, 85:107204.

Supplementary Documents

MolProbity Ramachandran analysis

Mtb_Model_clean.pdb, model 1



95.2% (220/231) of all residues were in favored (98%) regions.
98.3% (227/231) of all residues were in allowed (>99.8%) regions.

There were 4 outliers (phi, psi):

- A 80 VAL (88.8, 0.1)
- B 46 GLN (39.6, -86.6)
- B 80 VAL (94.5, -2.0)
- C 80 VAL (96.9, -1.6)

<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure S1: A ramachandran plot of *Mtb* F₁F₀ ATP synthase protein.

Overall model quality

[HELP](#)

Z-Score: **-1.88**

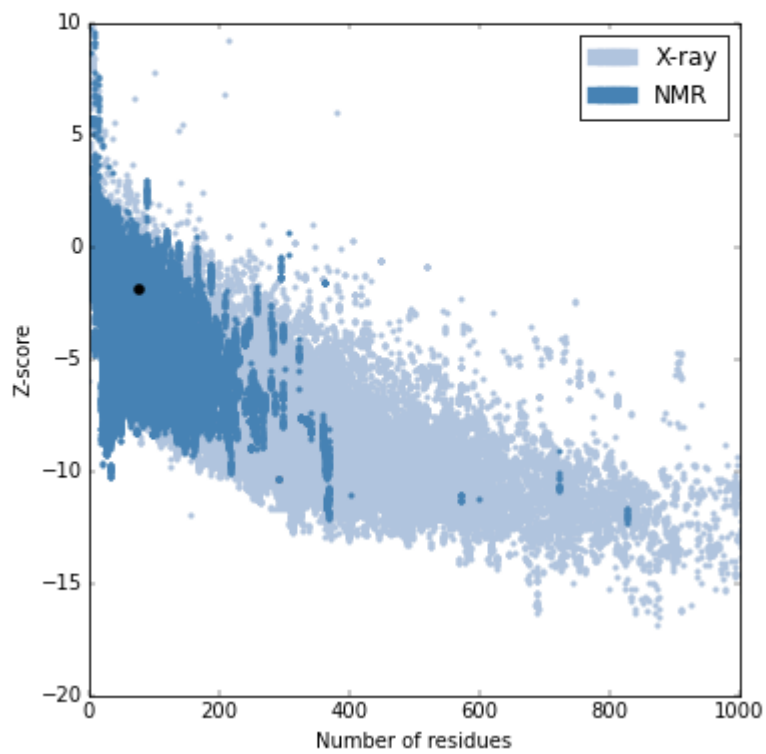


Figure S2: PROSA plot showing the model quality.

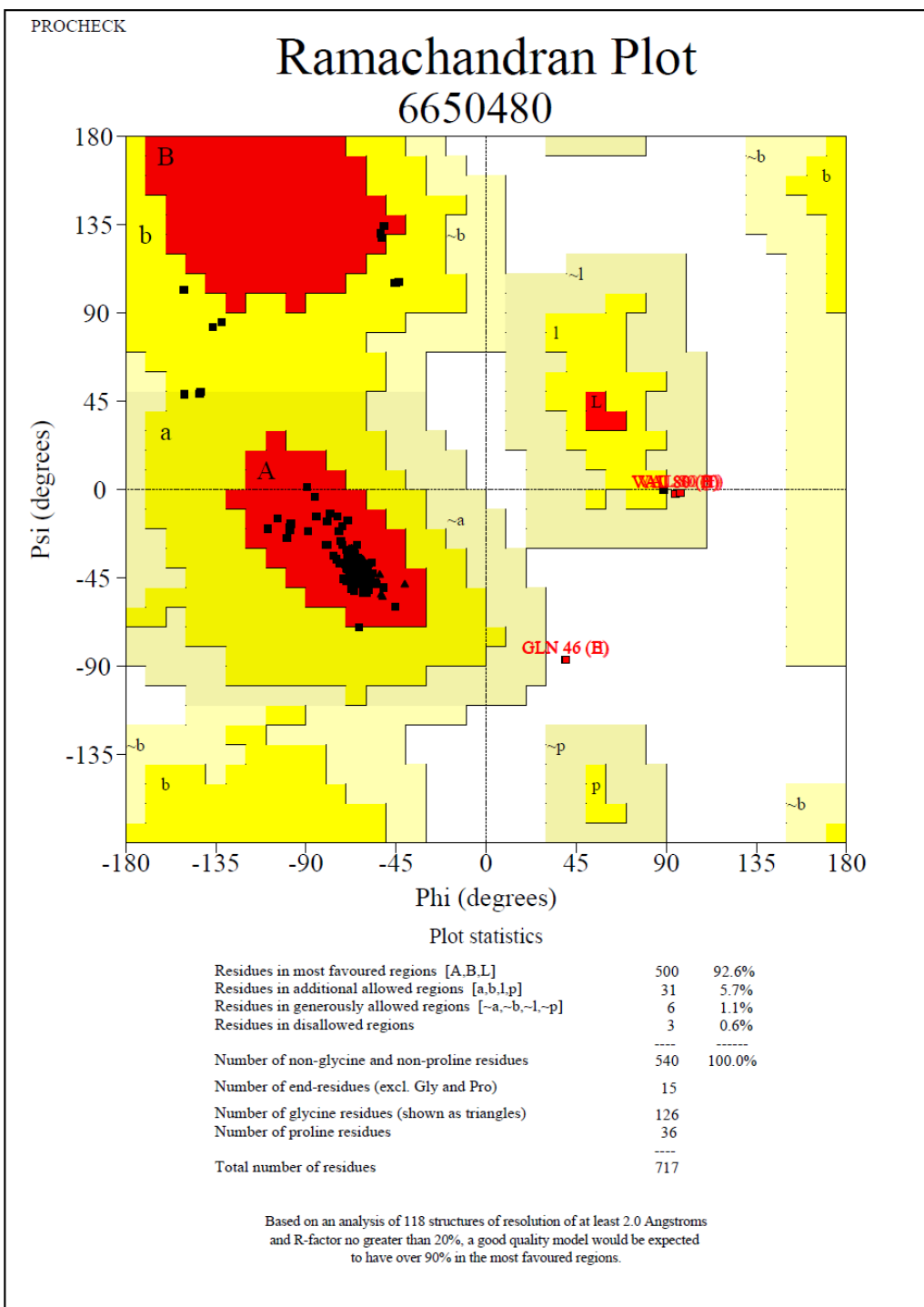


Figure S3: PROCHECK analysis Mtb F₁F₀ ATP-synthase modelled structure.

APPENDIX F

Clement Agoni, Pritika Ramharack and Mahmoud Soliman (2018), Allosteric Inhibition
Induces an Open WPD-Loop: A New Avenue Towards Glioblastoma Therapy.

RSC Advances, 8:40187-40197.

Supplementary Documents

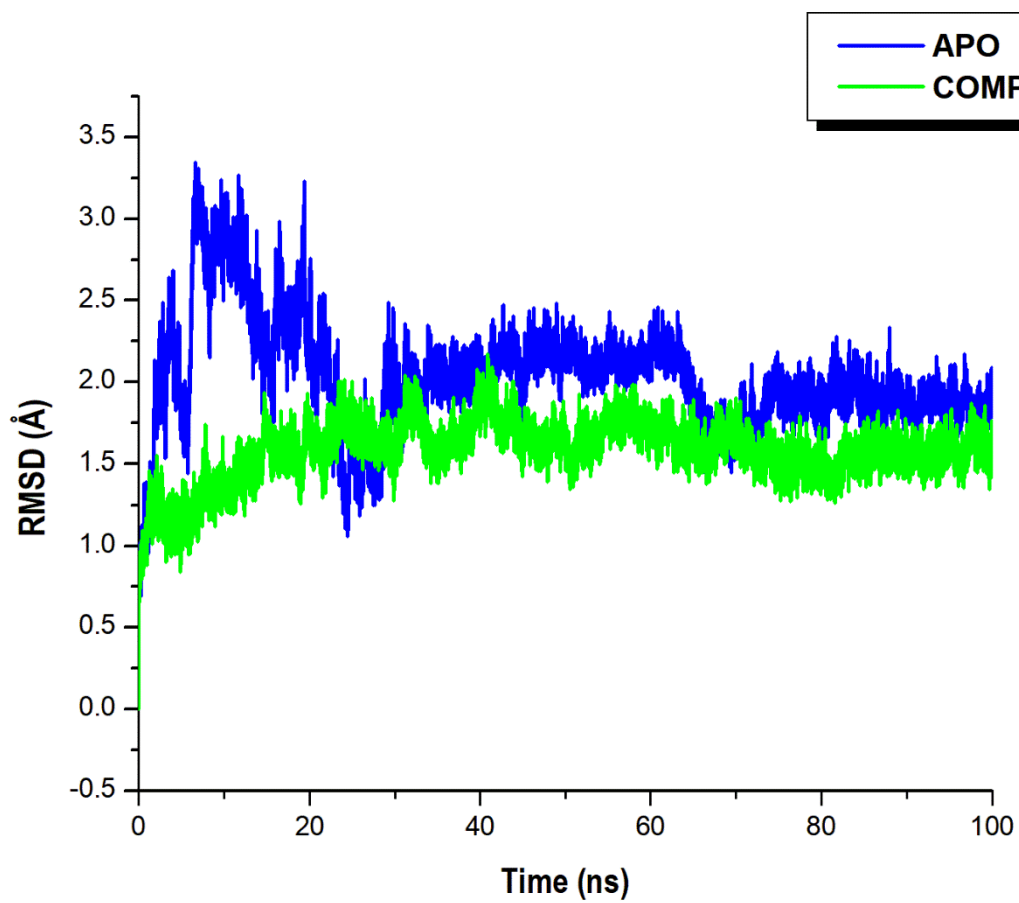


Figure S1: Comparative RMSD of PTPRZ when bound to the allosteric inhibitor NAZ2329 (COMPLEX) and when unbound to any inhibitor (APO).

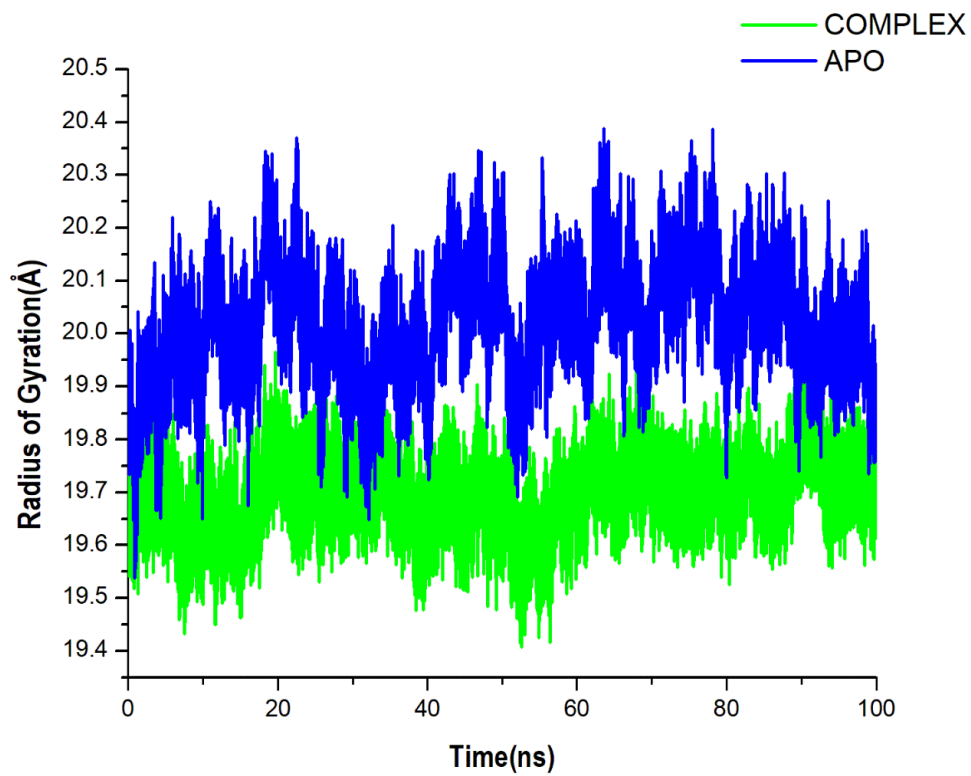


Figure S2: Comparative compactness of PTPRZ when bound to NAZ2329 (COMPLEX) and when unbound to any inhibitor (APO).

APPENDIX G

Houda Washah, **Clement Agoni**, Fisayo A. Olotu, Geraldene Munsamy, and Mahmoud E. S.

Soliman (2020), "Tweaking α -galactoceramides: Probing the dynamical mechanisms of improved recognition for invariant natural killer T-cell receptor in cancer immunotherapeutics".

Current Pharmaceutical Biotechnology (in press).

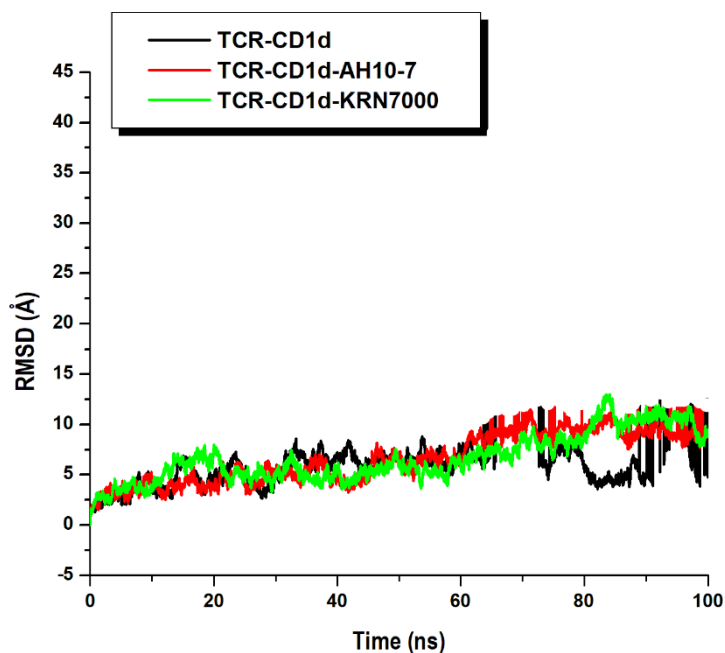


Figure S1: Comparative RMSD plot of unbound human TCR-CD1d (**black**) and complexed with both AH10-7 (**red**) and KRN7000 (**green**) showing stability of simulated systems

Table S1: Hydrophobic site residues and corresponding per-residue energies in the presence of KRN7000. Residues with energies < -1kcal/mol were considered prominent.

Residue	vdW (kcal/mol)	ELE (kcal/mol)	TOTAL (kcal/mol)
CYS 12	-1.683	-0.022	-1.43
VAL 30	-0.903	-0.003	-1.124
ILE 47	-0.92	0.005	-1.066
PHE 70	-3.079	0.171	-2.08
TYR 73	-4.265	-0.749	-3.879
SER 76	-1.375	-1.264	-1.551
PHE 77	-1.843	-0.245	-1.553
ASP 80	1.121	-15.883	-4.315
ILE 81	-1.163	-0.042	-1.2
LEU 84	-1.218	0.071	-1.305
VAL 85	-0.368	0.022	-0.332
LEU 100	-1.416	0.041	-1.581
LEU 116	-0.869	0.002	-1.04
VAL 118	-0.915	-0.005	-1.074
LEU 143	-1.525	-0.21	-1.669
ASP 153	0.333	-16.588	-3.463
THR 156	-3.466	-0.81	-4.114
PRO 28c	-1.384	-0.382	-1.535
ARG 95c	-2.328	0.091	-2.378

Table S2: Hydrophobic site residues and corresponding per-residue energies in the presence of AH10-7. Residues with energies < -1kcal/mol were considered prominent.

Residue	VDW (kcal/mol)	ELE (kcal/mol)	TOTAL (kcal/mol)
CYS 12	-1.902	-0.043	-1.629
VAL 30	-0.908	-0.003	-1.129
LEU 66	-0.999	0.149	-1.109
MET 69	-2.247	-0.435	-2.65
PHE 70	-3.597	0.161	-2.374
VAL 72	-1.296	0.248	-1.565
TYR 73	-4.895	-0.459	-4.56
SER 76	-1.437	-1.451	-2.169
PHE 77	-1.891	-0.126	-1.543
ILE 76	-1.26	-0.094	-1.292
LEU 79	-1.346	0.046	-1.415
LEU 100	-1.773	0.049	-1.846
LEU 116	-0.908	-0.002	-1.07
VAL 118	-0.893	-0.005	-1.052
LEU 150	-1.397	-0.27	-1.524
ASP 153	0.625	-16.595	-3.716
THR 156	-3.391	0.014	-3.521
THR 159	-2.06	0.628	-1.816
VAL 160	-1.276	0.229	-1.394
LEU 163	-2.144	0.073	-2.277
PRO 28c	-1.892	-0.512	-2.09
ASN 30c	-1.099	-2.189	-1.495
ASP 94c	-0.688	0.683	-1.15
ARG 95c	-2.454	-1.559	-2.763

