

In Silico Identification of Natural Inhibitory Compounds
Against the *Mycobacterium tuberculosis* Enzyme
Pyrazinamidase Using High-throughput Virtual Screening
Techniques

A mini-thesis submitted in partial fulfilment of the requirements for the degree

MASTER OF SCIENCE OF RHODES UNIVERISTY

by

Coursework and Thesis

in

Bioinformatics and Computational Molecular Biology

Research Unit in Bioinformatics (RUBi)

Department of Biochemistry and Microbiology

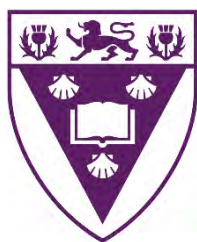
Faculty of Science

by

Thomas Kenyon

15k7151

June 2021



ABSTRACT

Tuberculosis (TB) is most commonly a pulmonary infection caused by the bacterium *Mycobacterium tuberculosis*. With the exception of the COVID-19 pandemic, TB was the most common cause of death due to an infectious disease for a number of years up until 2020. In 2019, 10 million people fell ill with TB worldwide and 1.4 million people died (WHO, 2020a). Additionally, multidrug-resistant TB (MDR-TB) remains a public health crisis and a health security threat. A global total of 206 030 people with multidrug- or rifampicin-resistant TB (MDR/RR-TB) were reported in 2019, a 10% increase from 186 883 in 2018. South Africa is ranked among the 48 high TB burden countries, with an estimated 360 000 people falling ill in 2019, resulting in 58 000 deaths, the majority of which being among people living with HIV. Unlike HIV, however, TB is a curable disease when managed correctly with long durations of antitubercular chemotherapy. Pyrazinamide (PZA) is an important first-line tuberculosis drug unique for its activity against latent TB. PZA is a prodrug, being converted into its active form, pyrazinoic acid (POA) by the Mtb gene *pncA*, coding for the pyrazinamidase enzyme (PZase). TB resistance to first-line drugs such as PZA is commonly associated with mutations in the *pncA*/PZase enzyme. This study aimed to identify potential novel inhibitors that bind to the active site of PZase. By making use of molecular docking studies and molecular dynamics (MD) simulations, high throughput virtual screening was performed on 623 compounds from the South African Natural Compounds database (SANCDDB; <https://sancdb.rubi.ru.ac.za>). Ligands that selectively bound to the PZase active site were identified using docking studies, followed by MD simulations to assess ligand-PZase complex stability. Finally, hit compounds identified from the first round of MD simulations were screened again against PZase structures with high confidence point mutations known to infer PZA resistance in order to identify any novel compounds which had inhibitory potential against both WT and mutant forms of the PZase enzyme.

DECLARATION

I, Thomas Kenyon, hereby declare that this thesis submitted to Rhodes University is my original work and has never been submitted to any institution for a degree or diploma.



.....

Signature

28/09/2021
.....

Date

DEDICATION

This thesis is dedicated to my parents,

Barbara and Michael

For always being there for me, no matter what. Thanks mom and dad, for all the love and support you have shown me every single day of the 25 years I have been on this Earth.

ACKNOWLEDGEMENTS

My sincere gratitude and appreciation to my supervisor Professor Özlem Tastan Bishop for all the advice, insight, and prompt feedback she has given me for the duration of this project.

I am also grateful to Rita Afriyie Boateng for her endless guidance, advice and patience while assisting me with this project, I truly could not have done it without you.

To Prudence Gowu, my project partner, thank you for sticking by me for the past year, I wish you the best of luck for all your future endeavours and have no doubt that you will be immensely successful in whatever you set your mind to.

To Mbunge Mbunge, my partner of heart and mind. There have been many moments of personal regret over the past few years but becoming friends with you has been one of the truly good things that has happened to me since I signed up for that fateful Honours degree in Biochemistry. Who would have thought we would both be alive and kicking more than a year into a global pandemic?

To the rest of my Bioinformatics MSc colleagues, Nabeelah, Sophakama, Amira and Lillian, spending a large part of 2020 together with all of you was truly special to me and I am grateful that every single one of you are some of the most wonderful people I have had the privilege of suffering through a pandemic with.

I am also grateful to all members of RUBi as a whole; I am truly sad to leave. Some of you lectured me and my colleagues while others gave advice and encouragement, thank you to every single one of you.

This work would not have been possible without the financial support from the National Research Foundation (NRF).

TABLE OF CONTENTS

Abstract.....	ii
Declaration.....	iii
Dedication.....	iv
Acknowledgements.....	v
Table of contents.....	vi
List of figures.....	ix
List of tables.....	xi
List of equations.....	xii
Webservers and software tools used.....	xiii
List of abbreviations.....	xiv
Chapter 1.....	1
Literature review.....	1
1.1 Introduction.....	1
1.2 The Burden of TB.....	1
1.3 TB Etiology.....	2
1.4 The genome of MTB.....	3
1.5 Transmission and Pathophysiology.....	3
1.6 Treatment.....	6
1.7 Pyrazinamide.....	7
1.8 TB Drug resistance:.....	9
1.9 Treatment of MDR- and XDR-TB.....	10
1.10 The PncA/Pyrazinamidase Enzyme.....	10
1.11 Problem Statement:.....	12
1.12 Aim of study:.....	12
1.13 Specific Study Objectives:.....	13
Chapter 2.....	14
Molecular Docking.....	14
2.1 Introduction.....	14
2.2 A note on protein-ligand binding.....	14
2.3 Molecular Docking Studies.....	15
2.4 Conformation sampling.....	16
2.5 Scoring functions.....	16

2.6 AutoDock4	17
2.7 Methodology	18
2.7.1 Data preparation.....	18
2.7.2 Protonation.....	19
2.7.3 Autogrid setup and docking preparation.....	19
2.7.4 Docking Simulations.....	19
2.7.5 Docking Validation.....	19
2.7.6 Docking Analysis.....	20
2.8 Results and discussion.....	22
2.8.1 Docking validation.....	22
2.8.2 SANCDB screened compounds.....	23
2.9 Conclusion.....	28
Chapter 3	29
Molecular dynamics simulations for WILD type – LIGAND complexes.....	29
3.1 Introduction	29
3.2 Molecular Dynamics Simulations	29
3.3 Methods:.....	32
3.3.1 Model Preparation.....	32
3.3.1.1 Protein preparation	32
3.3.1.2 Ligand preparation:.....	32
3.3.2 System preparation.....	32
3.3.3.1 Topology generation.....	32
3.3.3 Minimization.....	33
3.3.4 Equilibration	33
3.3.5 Production MD runs.....	33
3.3.6 Trajectory Analysis	34
3.3.6.1 RMSD.....	34
3.3.6.2 RMSF.....	34
3.3.6.3 Radius of gyration:	34
3.3.6.4 Hydrogen bonding analysis:	34
3.4 Results and Discussion.....	36
3.4.1 20 ns pre-runs.....	36
3.4.2 Ligand RMSD.....	37
3.4.3 Hydrogen bonding	38
3.4.4 Protein RMSD.....	40

3.4.5 Radius of Gyration (Rg).....	41
3.4.6 RMSF.....	43
3.4.7 Hit compounds.....	45
7-Angelylplatynecine.....	46
Rubrolide E.....	47
4'-Demethyl-5- <i>O</i> -methyl-3,9-dihydroeucomin.....	47
(<i>Z</i>)-Eucomin.....	47
5-Hydroxy-7-methoxy-3-(3-hydroxy-4-methoxybenzyl)chroman-4-one.....	48
3.5 Conclusion.....	48
Chapter 4.....	50
Mutant PZase Molecular dynamics simulations.....	50
4.1.1 Introduction.....	50
4.1.2 Mutations in the <i>pncA</i> gene and PZA resistance.....	50
4.1.3 Identification of SANCDB compounds that bind to mutant PZase.....	51
4.2 Methodology.....	53
4.2.1 Mutagenesis.....	53
4.3 Results and Discussion.....	54
4.3.1 Ligand RMSD.....	54
4.3.2 Hydrogen bonding analysis.....	55
4.3.3 Mutations and compound stability.....	61
4.3.4 Top Hits.....	64
4.3.5 Drug-likeness of Hits.....	67
4.4 Conclusion.....	68
Chapter 5.....	69
Concluding remarks and future work.....	69
References.....	71
Appendices.....	82

LIST OF FIGURES

- Figure 1:** Global estimated TB incidence rates in 2019.
- Figure 2:** Architecture of the TB granuloma.
- Figure 3:** Classical activation of alveolar macrophage.
- Figure 4:** Proposed mechanisms of action for Pyrazinamide.
- Figure 5:** Hybrid ribbon and stick representation of the structure of Mtb PncA/Pzase protein.
- Figure 6:** NAD⁺ de novo and salvage pathways present in Mtb.
- Figure 7:** Overall methodology used in this project.
- Figure 8:** Overview of docking methodology in this chapter
- Figure 9:** Docking validation.
- Figure 10:** All PZase protein-ligand docking complexes.
- Figure 11:** Selected PZase protein-ligand docking complexes.
- Figure 12:** Hydrogen bond distribution for all selected ligands.
- Figure 13:** Components of a typical MM force field
- Figure 14:** Overview of methodology used for Molecular Dynamics studies.
- Figure 15:** Violin plots of Ligand RMSD results for 20 ns pre-production MD runs.
- Figure 16:** Violin plots of Ligand RMSD results for 150 ns production MD runs.
- Figure 17:** Heatmap of intermolecular hydrogen bond interactions during production runs.
- Figure 18:** Heatmap of intermolecular hydrogen bond counts during production runs.
- Figure 19:** Violin plots of protein backbone RMSD results for 150 ns production MD runs.
- Figure 20:** Violin plots of protein Radius of gyration results for 150 ns production MD runs.
- Figure 21:** Violin plots of Rg of active site residue results for 150 ns production MD runs.
- Figure 22:** RMSF data for 150 ns production MD runs.
- Figure 23:** Point mutations within the PZase protein used during this screening study.
- Figure 24:** Violin plots of Ligand RMSD results for 150 ns production MD runs for the PZase mutant study.
- Figure 25:** Heatmap of intermolecular hydrogen bond interactions during MD runs.
- Figure 26:** Heatmap of intermolecular hydrogen bond counts during production runs.
- Figure 27:** Mutations within the active site pocket resulted in few stable compounds.
- Figure 28:** Locations of two mutations outside of the *pncA* active site.

Figure 29: First pair of analogous hits.

Figure 30: Second pair of analogous hits.

Figure S1: Violin plots of protein backbone RMSD results for 150 mutant PZase production runs

Figure S2: Violin plots of protein Rg) results for 150 ns mutant PZase production MD runs

Figure S3: Violin plots of Rg of active site residue results for 150 ns mutant PZase production MD runs

Figure S4: RMSF data results for 150 ns mutant PZase production MD runs.

LIST OF TABLES

Table 1: Drugs used in the treatment of TB.

Table 2: Selected Ligands that remained bound to PZase during MD simulations.

Table 3: Characteristics of final selected ligands used for the following mutagenesis studies.

Table 4: PZase point mutations chosen for this study.

Table 5: Summary of ligand stability in mutant systems and H-bonding residues.

Table 6: Tabulated results of the QED test for drug desirability.

LIST OF EQUATIONS

Equation 1: Scoring functions used by AutoDock4

WEBSERVERS AND SOFTWARE TOOLS USED

Accelrys Discovery Studio Visualizer:

<https://discover.3ds.com/discovery-studio-visualizer-download>

ACPYPE: <https://github.com/alanwilter/acpype>

AmberTools20: <https://ambermd.org/AmberTools.php>

AutoDock4.2: <http://autodock.scripps.edu/>

AutoDockTools: <http://autodock.scripps.edu/resources/adt>

AutoDockVina: <http://vina.scripps.edu/>

CHPC: <https://www.chpc.ac.za/>

JupyterLab: <https://jupyter.org/>

GROMACS: <https://www.gromacs.org/>

H++ webserver: <http://biophysics.cs.vt.edu/>

OpenBabel: http://openbabel.org/wiki/Main_Page

PyMOL: <https://pymol.org/2/>

RCSB PDB: <https://www.rcsb.org/>

Reduce tool: DOI: 10.1006/jmbi.1998.2401

SANCDDB: <https://sancdb.rubi.ru.ac.za>

LIST OF ABBREVIATIONS

_lc	largest cluster
_le	lowest binding energy
2D	Two-dimensional
3D	Three-dimensional
AIDS	Acquired immunodeficiency syndrome
ALERTS	Structural alerts for fragments with high toxicity potential
AROM	Aromatic rings
CADD	Computer-Aided Drug Design
CHPC	Center for High Performance Computing
COM	Centre of mass
COVID-19	Coronavirus disease 2019
CPU	Central processing unit
Da	Dalton
DCs	Dendritic cells
GPU	Graphics processing unit
H-bond	Hydrogen bond
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
HIV	Human immunodeficiency virus
Kcal	kilocalories
LogP	octanol-water partition coefficient
LTBs	Latent Tuberculosis infections
MBS	Metal binding site

MD	Molecular Dynamics
MIC	Minimum Inhibitory Concentration
MDR-TB	Multidrug-resistant Tuberculosis
MoA	Mechanism of action
Mtb	<i>Mycobacterium Tuberculosis</i>
MW	Molecular weight
NAD ⁺	Nicotinamide Adenine Dinucleotide
NK cells	Natural Killer cells
NMR	Nuclear magnetic resonance
PDB	Protein Data bank
POA	Pyrazinoic acid
PSA	Polar surface area
PZA	Pyrazinamide
PZase	Pyrazinamidase
QED	Quantitative Estimate of Drug-Likeness
Rg	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
ROTB	Rotatable bonds
RR-TB	Rifampicin-resistant Tuberculosis
SANCDB	South African Natural Compounds Database
STATSSA	Statistics South Africa
TB	Tuberculosis
vdW	van der Waals

WHO

World Health Organisation

WT

Wild-type

XDR-TB

Extensively drug resistant tuberculosis

CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION

Mycobacterium tuberculosis (Mtb), the etiological agent of tuberculosis (TB), remains one of the leading causes of death worldwide due to an infectious disease. In 2018 TB infected 10 million people and was responsible for approximately 1.5 million deaths (WHO, 2020a). TB is spread through the air in aerosols that are released when a person with an active TB infection coughs. It primarily (but not exclusively) infects the lungs. Like many other infectious diseases of global importance, the disease burden of TB varies drastically between different countries as shown in **Figure 1**, with nations in the Global South shouldering most of the burden (WHO, 2020a).

1.2 THE BURDEN OF TB

Approximately one quarter of the world's population has been infected with TB. The overwhelming majority of these infections are latent and asymptomatic but 5-10% of latent

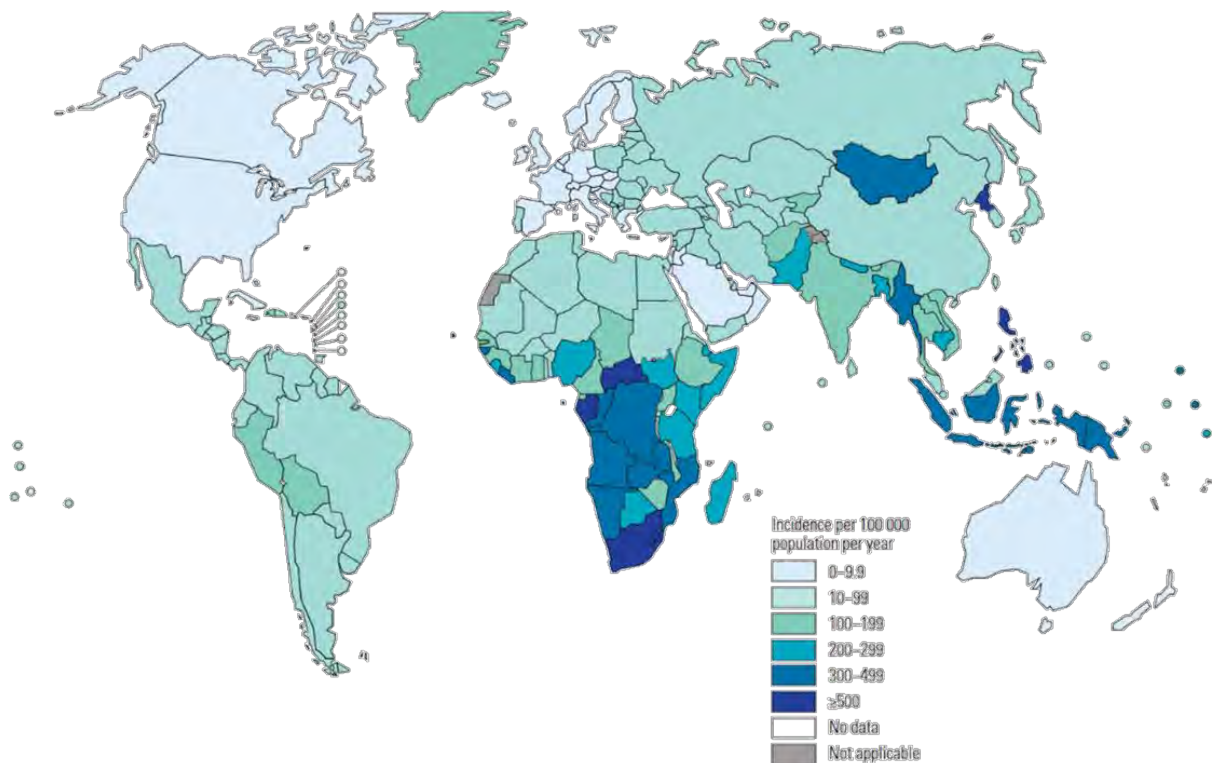


Figure 1: Global estimated TB incidence rates in 2019.
(Adapted from WHO Global Tuberculosis Report 2020)

infections will eventually progress to active TB which, if untreated, will kill approximately half its victims (Falzon *et al.*, 2011). Effective suppression of TB present in the body by the immune system prevents latent infections from becoming active. However immunocompromised individuals such as those living with AIDS are at higher risk of developing active TB infections. The majority of TB cases occur in three regions: South-East Asia (44%), Africa (24%) and the Western Pacific (18%). These three regions combined accounted for 86% of all cases in 2019. The WHO defined 30 countries as having a “high TB burden” in 2019, as they account for 87% of all cases globally (WHO, 2020a) These include India, China, Pakistan, Nigeria and South Africa, amongst others.

In South Africa, TB incidence is among the highest in the world, with approximately 520 active cases per 100 000 population and a total number of cases of 215 – 400 thousand in 2018 (WHO, 2020a). In 2015 TB was responsible for 7.2% of all deaths in South Africa (~65 000) and the leading cause of death due to a communicable disease. South Africa’s TB epidemic is intertwined with the socio-economic failures of post-Apartheid governance: in terms of both wealth and income inequality, South Africa is still one of the most unequal countries in the world. Real unemployment stands at 38.5% and 49.2 % of the adult population lives below the official poverty line (STATSSA, 2015). Inadequate living conditions and poor access to health facilities are widely considered the primary factors behind South Africa’s TB burden (Hartel, Yazbeck and Osewe, 2018). Ultimately these factors result in approximately 100 000 unidentified TB cases each year, and of those that are diagnosed, between 17-25% do not initiate treatment (Skinner and Claassens, 2016).

Incidence of TB amongst South Africans living with HIV/AIDS is significantly higher than the general population (WHO, 2020a) One in six adults in South Africa are HIV positive (WHO, 2020b) and approximately half of all TB cases in 2018 were people living with HIV. HIV-positive mortality was twice the HIV-negative mortality rate (WHO, 2020a).

1.3 TB ETIOLOGY

The causative agent of TB is *Mycobacterium tuberculosis*. Archaeological evidence indicates that tuberculosis has existed since antiquity and has been infecting humans since the Neolithic era approximately 9 000 years ago (Hershkovitz *et al.*, 2008). Mtb is a highly aerobic, non-motile, non-spore forming and non-encapsulated obligate intracellular pathogen belonging to the bacterial family Mycobacteriaceae. It has a unique waxy coating on its surface partially comprised of mycolic acid. This coating is responsible for Mtb responding poorly to gram staining. It is also a key virulence factor (Forrellad *et al.*, 2013) and makes the bacterium

relatively resistant to desiccation compared to many other non-spore-forming bacteria. Mtb is primarily a pathogen of the human respiratory system where it typically (but not exclusively) infects the lungs due to its airborne mode of transmission and preference for highly oxygenated environments (Lawn and Zumla, 2011). Mtb is a fastidious microbe to culture, with complicated nutrient requirements and a division time of 18-24 hours that results in culturing times of weeks instead of hours like a model organism such as *E. coli*. This glacial replication rate and Mtb's ability to persist in a non-replicative latent state in unfavourable conditions mean that successful treatment of tuberculosis infections require extended drug treatment courses that last many months (Lawn and Zumla, 2011).

1.4 THE GENOME OF MTB

The complete genome of the Mtb reference strain H37Rv was published in 1998 by *Cole et al.* This was a landmark achievement that has paved the way for significant advances in understanding the evolution, pathogenesis, and biology of Mtb. Notable is the complete absence of any extrachromosomal elements such as plasmids, with the Mtb nuclear genome containing approximately 4 million base pairs and 4 thousand genes. With a guanine-cytosine content of 65.6%, the Mtb genome is remarkably uniform with respect to other Mycobacterium species such as *M. leprae* (*Cole et al.*, 2001; *Eiglmeier et al.*, 2001) This suggests that the evolution of Mtb was not shaped significantly by horizontal gene transfer events, as is the case with many other bacterial pathogens. One of its most remarkable features is the presence of 250 genes responsible for fatty acid metabolism. This is approximately 5 times the number of such genes found in *E.coli* (*Jeucken et al.*, 2019), pointing to the importance of lipid metabolism for the survival and pathogenesis of Mtb. Indeed Mtb can be cultured in environments where cholesterol is the only source of carbon, and the bacteria have been shown to prefer fatty acids over other source of carbon (*Wipperman et al.*, 2014). Additionally these genes are essential for the synthesis of Mtb's complex and lipid-rich cell wall (*Wilburn et al.*, 2018).

1.5 TRANSMISSION AND PATHOPHYSIOLOGY

Transmission of TB in the overwhelming majority of cases occurs when a person with an active pulmonary TB infection coughs or sneezes and expels thousands of aerosol droplets containing

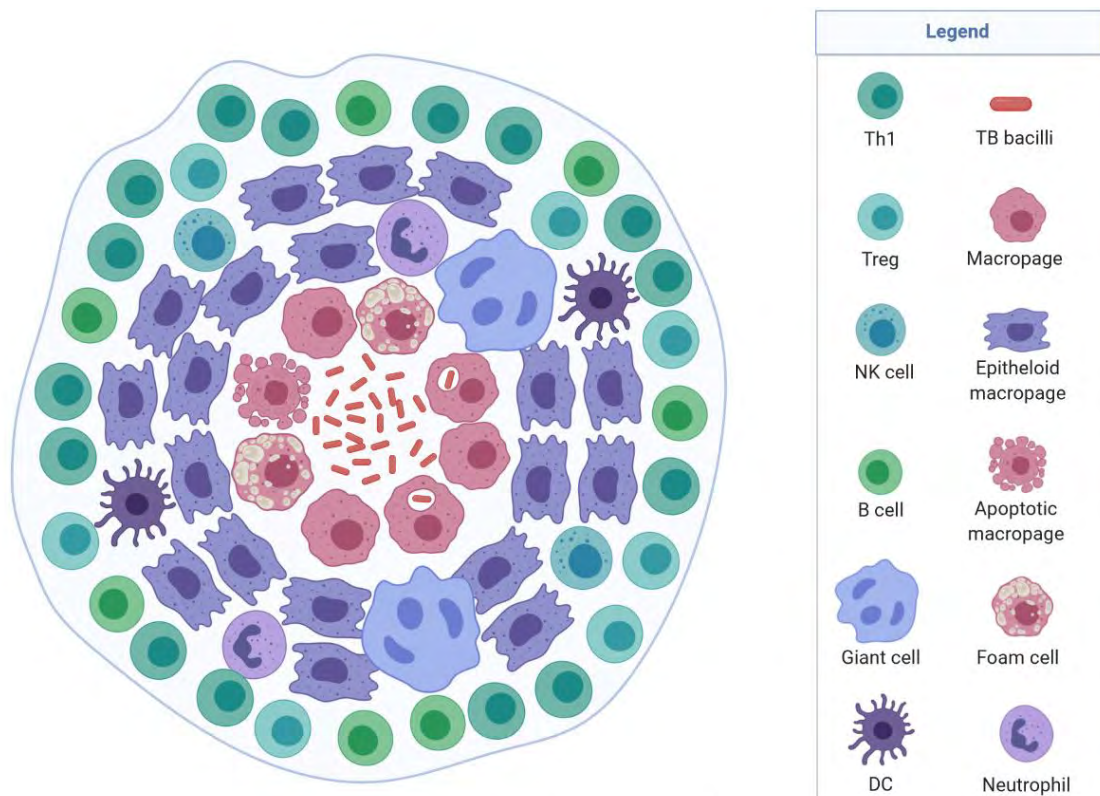


Figure 2: Architecture of the TB granuloma. Located at its centre is the initial site of TB infection, consisting of extracellular TB bacilli, activated alveolar macrophages containing phagocytosed and actively replicating Mtb bacilli, apoptotic macrophages, and foam cells (lipid-laden macrophages). Surrounding this are epithelioid macrophages, dendritic cells (DCs), multinucleated giant cells, neutrophils, and natural killer (NK) cells. The outer layer of a mature granuloma consists of helper and regulatory T-cells, B-cells, and fibroblasts. Adapted from Minton, 2016. Graphic created using BioRender.

Mtb bacteria which are then inhaled by an uninfected person. Inhalation of even one or two droplets may be enough to ultimately cause an infection as less than 10 bacterium are capable of causing an infection (Nicas, *et al.*, 2005). The duration of contact, frequency of contact, and proximity of contact with an infected person are all factors that influence the likelihood of becoming infected (Nicas *et al.*, 2005) Additionally the infectivity of the carrier, degree of ventilation in the immediate environment and susceptibility of the uninfected person also influence the infection probability (Turner *et al.*, 2017).

In a public health setting, early screening, diagnosis and initiation of treatment are essential to disrupt person-to-person transmission as infectivity of individuals with non-resistant infections decreases substantially after approximately two weeks of treatment (Ahmed and Hasnain, 2011). Of those infected with Mtb and classified as asymptomatic latent infections (LTBs), only 5-10% will develop an active TB infection in their lifetimes (Petruccioli *et al.*, 2016). Typically, immunocompetent individuals either eliminate Mtb entirely or contain it in a latent state. However in those with HIV or other immunocompromising conditions the chance of

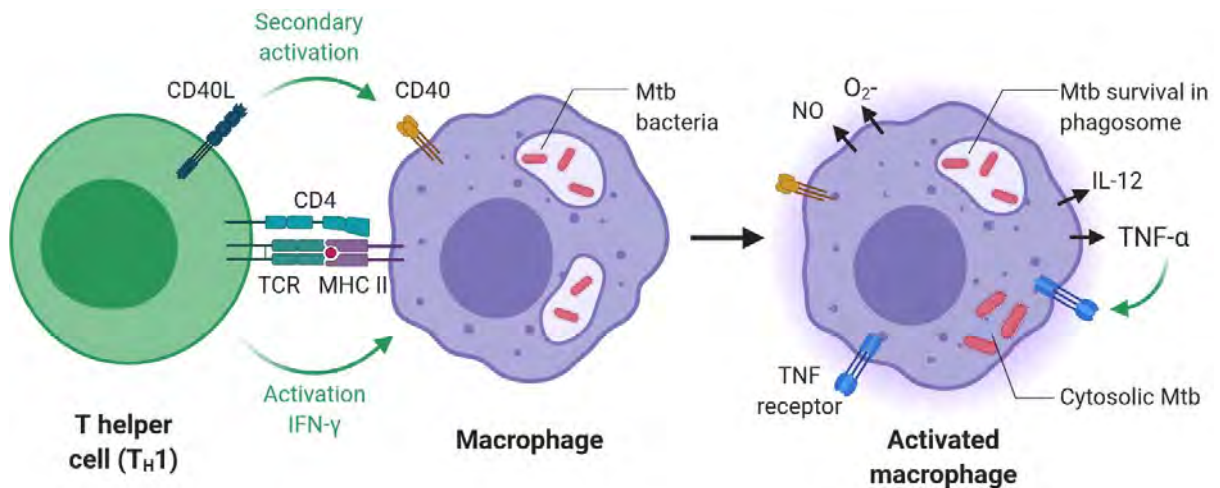


Figure 3: Classical activation of alveolar macrophages does not result in Mtb cell lysis. An armed T_{H1} cell specific to a Mtb peptide makes physical contact with an Mtb-infected macrophage. The T_{H1} cell then secretes IFN-γ and expresses the CD40 ligand (CD40L) on its surface, both of which then activate the macrophage. Activated macrophages are significantly more microbicidal, producing more reactive oxygen species such as nitric oxide (NO) and superoxide (O₂⁻) and promoting fusion of lysosomes and phagosomes. Activated macrophages also amplify the adaptive immune response by expressing more MHC class I and II receptors and secreting cytokines such as TNF-α and IL-12. Mtb has evolved mechanisms to prevent lysosome-phagosome fusion even in activated macrophages, allowing it to avoid cell lysis and escape into the macrophage cytoplasm. Adapted from Murphy and Weaver, 2016. Graphic created using BioRender.

developing an active TB infection can increase as much as 10% a year (Petruccioli *et al.*, 2016). In a country with high burdens of both diseases, this co-morbidity has been the biggest challenge for the South African public health system until the COVID-19 pandemic (Karim *et al.*, 2009).

Both latent and active TB infections ultimately begin once the mycobacteria reach the terminal alveolar air sacs of the lungs (Armstrong and Hart, 1971) where they are identified as foreign and phagocytosed by alveolar macrophages. For most bacteria, phagocytosis results in cell lysis. However Mtb make use of an arsenal of weapons to undermine macrophage defences, block lysosome fusion which prevents cell lysis and allows for replication within the macrophages (Ehrt and Schnappinger, 2009). As shown in **Figure 3**, Mtb, unlike many other pathogens, can survive and replicate within the phagosomes of fully activated macrophages and even escape into the macrophage cytoplasm. Mtb also manipulates intracellular signal transduction pathways within host macrophages which in turn alters the cytokine profile of the body's immune response, attenuating it and promoting intracellular Mtb survival (BoseDasgupta and Pieters, 2014; Orme *et al.*, 2015) The body's immune response is pivotal in determining the outcome of a TB infection: the hallmark of this response is the aggregation of immune cells at the original site of macrophage infection into a structure known as a granuloma, shown in **Figure 2**. This consists of activated infected macrophages at the centre that typically fuse into

a multi-nucleated giant cell, surrounded by T-lymphocytes, recruited macrophages, neutrophils and fibroblasts (de Martino *et al.*, 2019). Fibroblasts surrounding the immune cells lay down connective tissue and deposit calcium, forming the physical granuloma structure that is visible in lung X-rays.

Mtb persistence within granulomas without further spread is what is considered a latent TB infection that is asymptomatic: the immune response has successfully constrained the replication of Mtb and is preventing its spread, and the individual remains free of any tissue damage. If the immune response is disrupted or insufficient, granuloma formation may be impaired, alternatively, a previously adept immune response may become disrupted, resulting in the rupture of existing pulmonary granulomas in a process known as cavitation (Queval *et al.*, 2017). The outcome of either of these two scenarios is the uncontrolled release and replication of Mtb in pulmonary airways, causing an active and symptomatic TB infection. If left untreated, active TB can result in severe lung damage and systemic dissemination of Mtb throughout the body.

1.6 TREATMENT

Treatment of TB infections relies on antibiotics. However unlike many other bacterial pathogens, Mtb is intrinsically more tolerant of antibiotics due to its slow replication rate, persistence in a non-replicative latent state and its unique mycobacterial cell wall that prevents entry of many antibiotics (Brennan and Nikaido, 1995). In 1944 the first effective antibiotic against Mtb was discovered: Streptomycin (Sakula, 1988) and in 1952 isoniazid became the first oral mycobacterial antibiotic to be widely used (Crofto, 1969). Despite the discovery of multiple effective antibiotics, the tendency of mono-therapy to give rise to resistance led to the adoption of combination therapy of at least two antibiotics in the early 1960s (Kerantzas and Jacobs, 2017). More recently numerous drugs with antitubercular activity have been developed including rifampicin, ethambutol, and pyrazinamide.

As of 2011 (Falzon *et al.*, 2011), anti-TB drugs are classified into five main groups as shown in **Table 1**, considering the evidence for their efficacy and their safety/risk of side-effects. As a general rule, the groups start with the most effective drugs available (in group one, **Table 1**), while other drugs are grouped consecutively according to decreasing efficacy and higher risk of serious side-effects.

Table 1: Drugs used in the treatment of TB.

TB Drug Classification (WHO, 2011)	Drug Name
Group 1 First-line oral anti-TB drugs	Rifampicin Isoniazid Pyrazinamide Ethambutol
Group 2 Injectable anti-TB drugs	Streptomycin Kanamycin Amikacin Capreomycin
Group 3 Fluoroquinolones	Levofloxacin Moxifloxacin Gatifloxacin Ofloxacin
Group 4 Oral bacteriostatic second-line anti-TB drugs	Ethionamide/Prothionamide Cycloserine/Terizidone p-aminosalicylic acid
Group 5 Anti-TB drugs with limited data on efficacy and safety. Only used in treatment of drug resistant TB	Linezolid Clofazimine Amoxicillin/Clavulanate Imipenem/Cilastatin Meropenem High-dose isoniazid Thioacetazone Clarithromycin

Group one drugs are all administered orally and are the first drugs to be used when treating non-resistant infections. Rifampicin in group 1 as well as kanamycin and amikacin in group 2 all disrupt RNA synthesis and ultimately inhibit protein synthesis. Isoniazid is a prodrug that is activated by the mycobacterial catalase-peroxidase enzyme *KatG* and ultimately inhibits the formation of the mycobacterial cell wall by preventing the synthesis of mycolic acid, an essential component (Unissa *et al.*, 2016). Ethambutol is another group one drug that disrupts cell wall synthesis. However it inhibits synthesis of arabinogalactan, a disaccharide and essential component of the mycobacterial cell wall (Goude *et al.*, 2009).

1.7 PYRAZINAMIDE

Like isoniazid, pyrazinamide (PZA) is a prodrug that is activated by the mycobacterial enzyme pyrazinamidase *PncA* (PZase) along with numerous host enzymes (Via *et al.*, 2015), forming pyrazinoic acid (POA). Notably PZA, a synthetic nicotinamide analogue, is able to effectively penetrate TB lung granulomas right into their necrotic caseous centres and kill non-replicative latent *Mtb* cells (Gopal *et al.*, 2019) This remarkable characteristic allows for much shorter TB treatment regimens: when PZA was first introduced in the 1970s, uncomplicated TB treatment courses were shortened from 12 to 6 months (Mitchison, 1985). Additionally PZA has desirable pharmacokinetic properties: a high degree of absorption that is not affected by food intake

(Peloquin *et al.*, 1998), with only 3% of absorbed PZA being excreted in urine in its prodrug form (McIlleron *et al.*, 2006). For wild-type Mtb strains the MIC₉₉ values have been reported as between 8 and 64 µg/mL (Werngren *et al.*, 2012).

The underlying mechanism of action (MoA) of PZA is still poorly defined. Furthermore PZA's notably poor *in vitro* potency seemingly conflicts with its ability to kill resistant persister cells that other TB drugs are ineffective against (Blanc *et al.*, 2018). Numerous MoA models have been proposed over the years but virtually all of them now have significant evidence disproving them (Gopal *et al.*, 2019). Four leading models for the mechanism of action of PZA are summarized in **Figure 4** below.

Model 1: PZA is a protonophore. In this model, shown in **Figure 4A**, PZA enters Mtb cells via passive diffusion where it is converted to POA by PZase. The POA anion is then exported out of the cell by a currently unidentified efflux mechanism. Within the acidic environment of an activated phagosome the POA anion becomes protonated and diffuses back into the bacillus where it dissociates into a proton and POA. Over time this results in the acidification of the cytoplasm and the collapse of Mtb's membrane potential resulting in cell death (Zhang *et al.*, 1999).

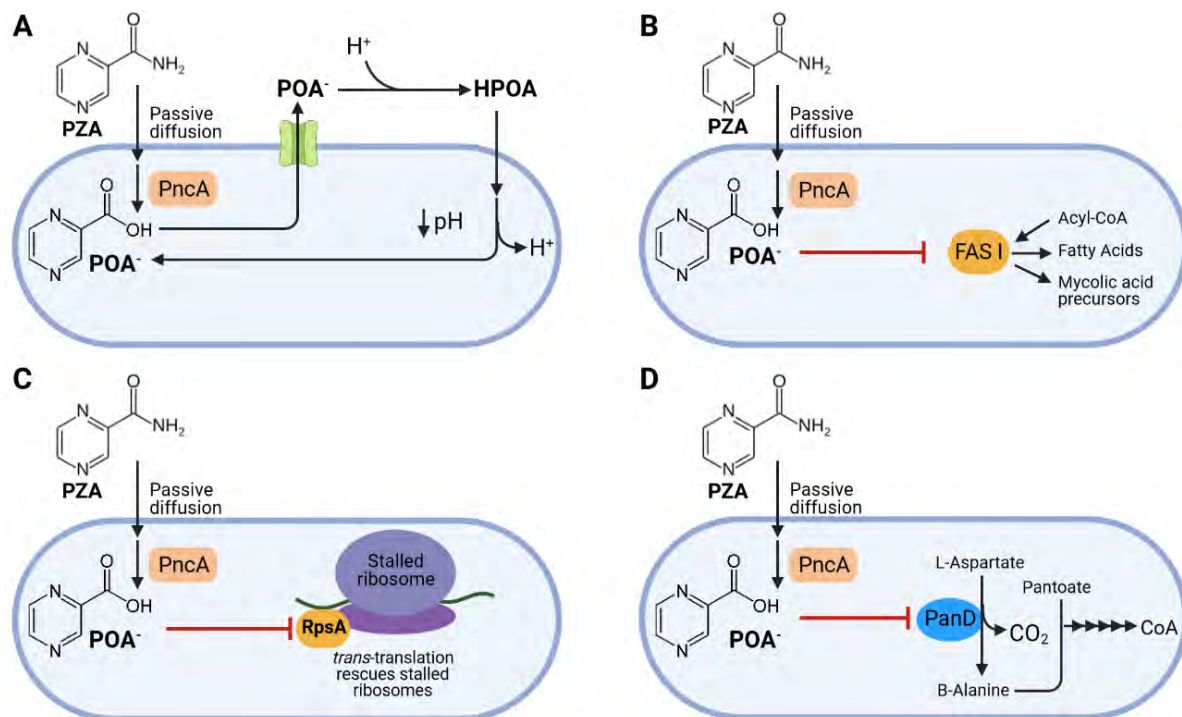


Figure 4: Proposed mechanisms of action for Pyrazinamide. In all cases, PZA enters the Mtb cell by passive diffusion and is converted into the active form (POA) by the nicotinamidase PncA/PZase enzyme. **A:** POA acts as a protonophore leading to the acidification of the bacterial cytoplasm, **B:** POA inhibits the essential fatty acid synthase I (FAS I) enzyme, **C:** POA inhibits *trans*-translation, **D:** POA inhibits biosynthesis of coenzyme A (CoA). Graphic adapted from Lamont *et al.*, 2020 and created using BioRender.

Model 2: Inhibition of Mycobacterial fatty acid synthase I (FAS-I). In this model, shown in **Figure 4B**, POA-driven inhibition of the large multifunctional enzyme stops the synthesis of C16 - C26 fatty acids. Without these fatty acids, numerous cytoplasmic membrane lipids cannot be synthesised, nor can the mycolic acids Mtb uses in its cellular envelope (Zimhony *et al.*, 2000; Schweizer and Hofmann, 2004).

Model 3: PZA inhibits the ribosome salvage pathway. In this model, shown in **Figure 4C**, POA disrupts the process of trans-translation, a process virtually all bacteria use to free ribosomes that cannot disengage from the 3' end of some mRNAs that lack an in-frame stop codon (Keiler, Waller and Sauer, 1996). Ultimately this results in the depletion of free ribosomes in the mycobacterial cytoplasm, preventing protein synthesis and causing cell death. (Shi *et al.*, 2011).

Model 4: PZA inhibits coenzyme A synthesis. In this model, shown in **Figure 4D**, bioactive POA binds to the Mtb aspartate decarboxylase (*PanD*) enzyme and blocks synthesis of coenzyme A (Gopal, *et al.*, 2019), an essential coenzyme in most organisms responsible for the synthesis and oxidation of many fatty acids and the oxidation of pyruvate in the citric acid cycle (Daugherty *et al.*, 2002; Leonardi *et al.*, 2005).

1.8 TB DRUG RESISTANCE:

According to WHO guidelines (Falzon *et al.*, 2011; Kurz *et al.*, 2016) TB drug resistance is classified according to how many drugs a TB strain is resistant to, and which groups those drugs fall into. The four classifications are: Mono-resistant TB, poly-drug resistant TB, multi-drug resistant TB (MDR-TB) and finally Extensively drug resistant TB (XDR-TB). Both mono- and poly-resistant strains are resistant to one or more first-line drugs except for rifampicin and isoniazid. MDR-TB meanwhile is at minimum resistant to both rifampicin and isoniazid, while XDR-TB is further resistant to second-line injectable drugs such as one or more of the fluoroquinolones. While significant progress has been made globally in the management, treatment, and outcome of drug-resistant TB infections, only one in three of the approximately half a million people infected with active MDR-TB or rifampicin-resistant TB (RR-TB) were diagnosed and started on treatment in 2018. Undiagnosed and untreated MDR-/RR-TB is the greatest obstacle for the WHO's global strategy to end the TB epidemic by 2035 (WHO, 2020a), reducing deaths by 95% and cases by 90% between 2015 and 2035 (WHO, 2015).

1.9 TREATMENT OF MDR- AND XDR-TB

Treatment regimens for MDR-TB are complicated, longer, and significantly more expensive than those used to treat drug-susceptible TB. The first-line drugs are used as such because they are all orally administered and have minimal side-effects. Treatment of MDR-TB necessitates the use of drugs with increased toxicity at clinically relevant concentrations and often with reduced efficacy. Initially any first-line drugs that are still effective (as determined by testing) are chosen, usually PZA and isoniazid. However these two drugs are only effective as adjuncts to a treatment regimen due to their limited activity by themselves (Kurz *et al.*, 2016). Therefore one injectable drug from group two and one fluoroquinolone from group three are used to make up a regimen consisting of four drugs and PZA if susceptible (Falzon *et al.*, 2011). Further resistance to injectable group two drugs and one or more fluoroquinolones indicates the presence of an XDR-TB infection. In this case, drugs from group four are substituted into a regimen. These drugs can have very severe side effects and are often toxic at therapeutic concentrations in many individuals (Rendon *et al.*, 2016).

1.10 THE PNCA/PYRAZINAMIDASE ENZYME

PncA (PZase), the enzyme that activates PZA is a monomeric ~20 kDa protein comprised of a six-stranded parallel B-sheet with multiple helices on either side forming a single α/β domain as shown in **Figure 5A** (Petrella *et al.*, 2011b). The metal binding site (MBS) consists of a ferrous ion that is coordinated in a distorted tetragonal bipyramidal arrangement by four histidine residues. Directly adjacent to the MBS are the three substrate-binding residues that form the catalytic triad (**Figure 5B**).

A flexible loop/lid region consisting of residues 52-70 controls access to the active site pocket. Two of the MBS residues, HIS 71 and His51, also function as flexible hinges, attaching the lid/loop region to the rest of the protein. The cavity containing the MBS and catalytic triad shown in **Figure 5B** is relatively small, 10 Å deep and 7 Å wide. On either side of it, the MBS and catalytic triad sit roughly opposite each other.

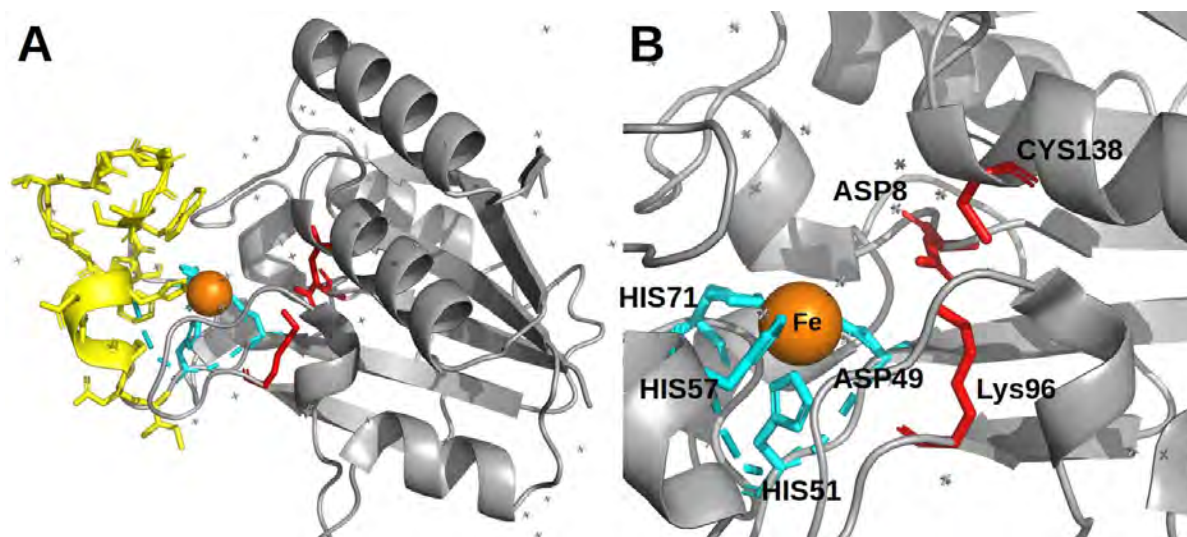


Figure 5: Hybrid ribbon and stick representation of the structure of Mtb PncA/Pzase protein.

A: Overview of important features of *PncA*/PZase. Yellow = Loop or lid region; Orange = iron atom; Blue = Metal binding site (MBS); Red = catalytic triad. **B:** MBS and active site cavity of *PncA*. Obtained from RCSB (PDB ID: 3PL1)

PZase forms a part of Mtb's nicotinamide salvage pathway, summarized in **Figure 6**. This is one of two pathways available to Mtb to obtain NAD^+ , the other pathway being the *de novo* synthesis pathway (Vilchèze et al., 2010) Both pathways are shown in **Figure 6**. These two pathways are both used by Mtb during infection. However Mtb relies on the salvage pathway (and therefore PZase) under certain conditions. Once enclosed within a granuloma, conditions at the necrotic caseous centre can become highly acidic and anoxic (Queval *et al.*, 2017). Under these conditions, Mtb primarily makes use of the salvage pathway to hydrolyse nicotinamide and its analogues in order to regenerate and maintain cytoplasmic NAD^+ concentrations (Zhang *et al.*, 2008, 2014). Since intracellular NAD^+/NADH must be continuously regenerated due to its essential role as a redox agent in all living organisms, disrupting the NAD^+ biosynthesis pathways that are responsible for its regeneration in Mtb is a plausible angle of attack when searching for novel antitubercular compounds. As the Mtb enzymes *NadD* and *NadE* are common to both the *de novo* and salvage pathway, they are each potential drug targets (Gerdes *et al.*, 2002). Blocking an enzyme unique to one pathway however (such as PZase in the salvage pathway) would likely require blocking an additional enzyme unique to the other pathway in order to yield maximum antitubercular effects at a level of therapeutic interest (Boshoff *et al.*, 2008). This is the rationale of this study: although blocking PZase alone may not be sufficient to induce bactericidal effects, there is a high-quality crystal structure of it available, and blocking it as well as an enzyme in Mtb's *de novo* pathway may be an effective therapeutic strategy.

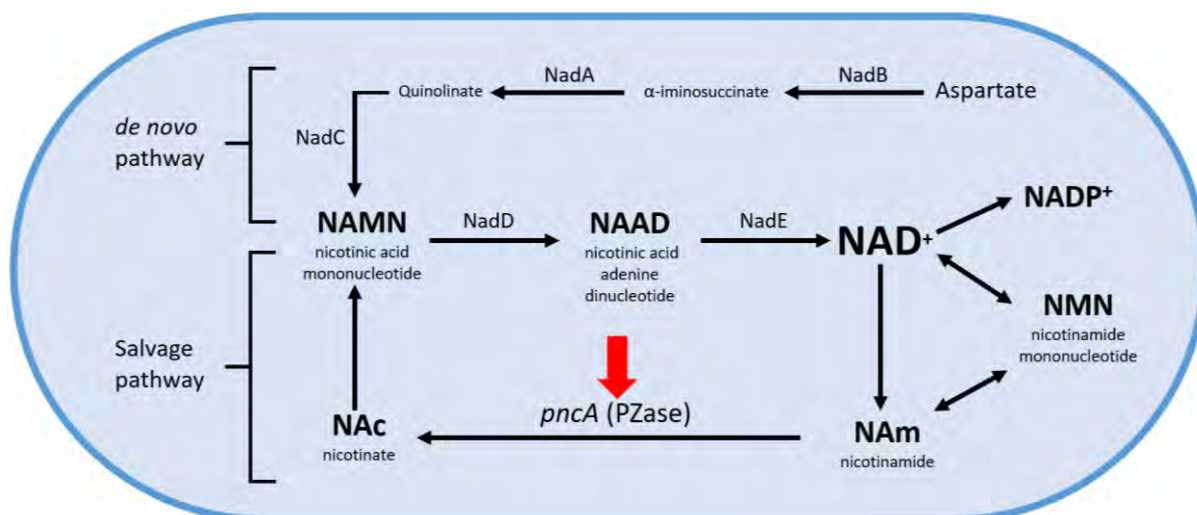


Figure 6: NAD⁺ de novo and salvage pathways present in Mtb. The *de novo* pathway begins with aspartate, followed by intermediate reactions catalysed by 3 enzymes: aspartate oxidase NadB, quinolinic acid synthase NadA and quinolinate phosphoribosyl transferase NadC respectively, all of which are unique to the *de novo* pathway, ultimately producing NAMN. The final two enzymes that convert NAMN to NAD⁺, adenylyl transferase NadD and deamido-NAD ligase NadE, respectively, are common to both the *de novo* and salvage pathways. The salvage pathway begins with NAD⁺ which is either converted to NMN or to NAm. NAm is then converted to NAc by the pyrazinamidase (*pncA*/PZase). Diagram adapted from Vilchèze et al., 2010.

1.11 PROBLEM STATEMENT:

There were between 215 000 and 400 000 cases of TB in South Africa in 2018 and over 10 million cases globally resulting in approximately 1.5 million deaths (WHO, 2020a). Globally 3.4% of new TB cases and 18% of previously treated cases were caused by MDR-/RR-TB. This is major threat to the progress made against TB over the last few decades, both in South Africa and globally. For the WHO's global strategy to end the TB epidemic by reducing deaths by 95% and cases by 90% between 2015 and 2035 to be successful, an enormous amount of research is needed in order to further our knowledge of how TB drug resistance arises, spreads and the mechanisms behind it. Furthermore, new drugs are needed urgently to minimize pressure on existing first- and second-line drugs and reduce the reliance on the current drugs used to treat resistant TB infections.

1.12 AIM OF STUDY:

The principle aim of this study is to identify potential compounds that bind to the active site of PZase via high-throughput virtual screening, using compounds in the South African Natural Compound Database (SANCDDB) (Hatherley *et al.*, 2015). Since PZase functions as part of the essential NAD⁺ biosynthesis pathway, identifying novel compounds that bind to it is of potential scientific and therapeutic value. After molecular dynamics (MD) analysis of interactions between PZase and hit compounds, the effect of mutant-induced structural changes

on binding of hit compounds will be studied in order to investigate some of the underlying mechanisms of resistance used by Mtb that grant phenotypic resistance to PZA. Compounds that bind to PZase mutants will also be identified. An overview of the methodology used during this study is shown in **Figure 7**.

1.13 SPECIFIC STUDY OBJECTIVES:

1. Obtain a high-quality crystal structure of Mtb wild-type (WT) PZase from the Protein Data Bank as well as minimized compound structures from SANCDB.
2. Perform molecular docking of PZA onto PZase using AutoDock4.2 in order to obtain the protein-drug complex.
3. Perform molecular docking of compounds from SANCDB using AutoDock4.2.
4. Perform MD simulations using the GROMACS software package for WT PZase with hit compounds to determine enzyme stability and drug-receptor interactions over relevant timescales.
5. Perform a second round of MD simulations using hit compounds emerging from WT-MD simulations and mutant PZase models.

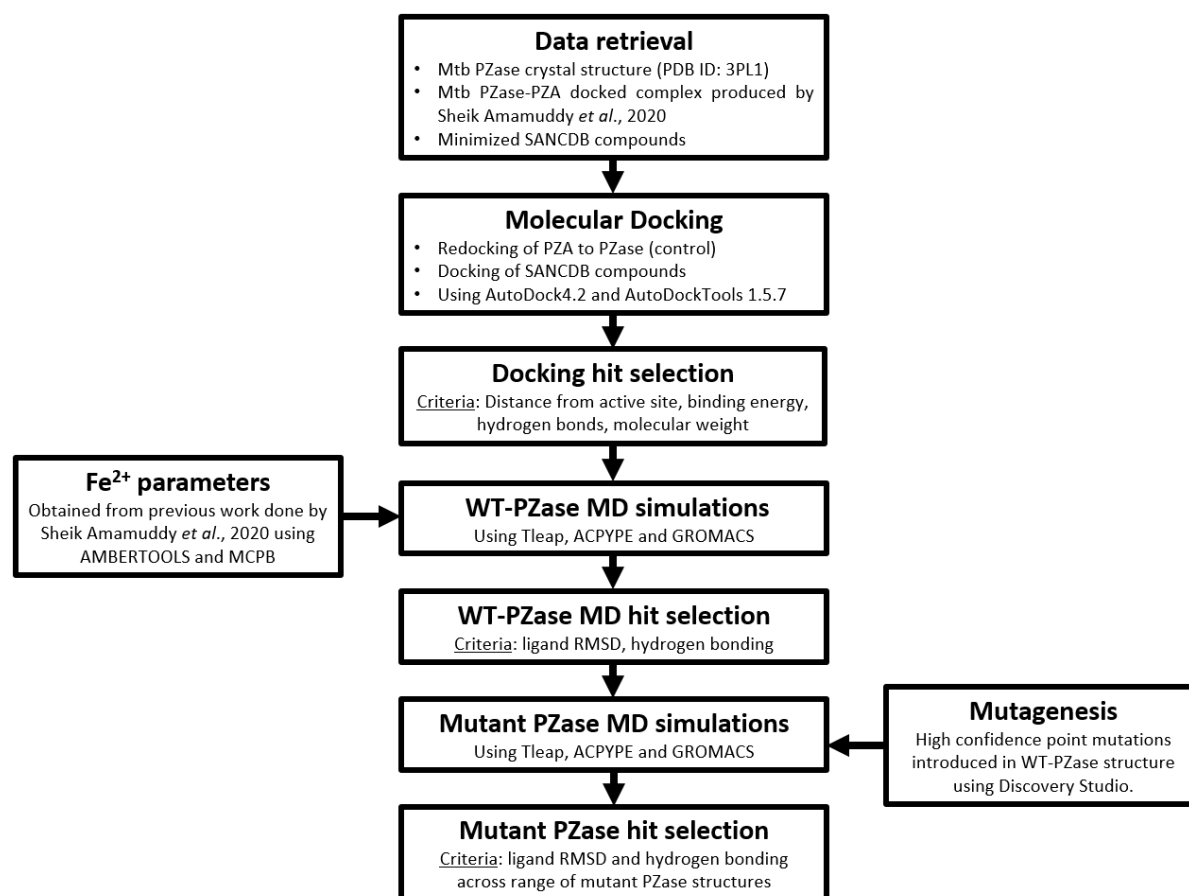


Figure 7: Overall methodology used in this study.

CHAPTER 2

MOLECULAR DOCKING

2.1 INTRODUCTION

Resistance to anti-tuberculosis drugs has become increasingly common worldwide. In order to maintain and improve treatment outcomes, new anti-tuberculosis drugs that act against novel targets within the Mtb proteome are needed. For the past 30 years improvements in computational biology tools and the increasing power of CPUs and GPUs have resulted in Computer-Aided Drug Design (CADD) or rational drug design becoming increasingly common (Abdolmaleki *et al.*, 2017; Yu and Mackerell, 2017; Dar *et al.*, 2019). This chapter describes the identification and selection of compounds that bind favourably to the Mtb protein PZase deposited in the South African Natural Compounds Database (Hatherley *et al.*, 2015). Hit compounds were identified by performing molecular docking studies and systems were evaluated based on binding affinities and intermolecular interactions such as the presence of H-bonds.

2.2 A NOTE ON PROTEIN-LIGAND BINDING

The mechanism of protein-ligand binding has been a topic of study since the late 19th century, with several theories being presented since, each with their own strengths and weaknesses. The first mechanism of protein-ligand binding was proposed by Emil Fischer in 1894, and became known as the ‘lock-and-key’ model wherein both ligand and protein/enzyme are rigid and the ligand shape fits into the active site of its complementary enzyme. While this model can explain how enzymes exhibit specificity for a particular substrate, it cannot explain enzyme promiscuity where a particular enzyme may bind to several different substrates, nor can it explain non-competitive inhibition (Cornish-Bowden, 2013) or allosteric modulation. Daniel E. Koshland Jr. introduced the ‘induced-fit’ theory in 1958, stating that the ligand-binding site on an enzyme undergoes a conformational change in the presence of a ligand, allowing for binding. This theory better explains enzyme promiscuity as multiple ligands can induce unique conformational changes to facilitate binding and further research has generated a volume of evidence that proteins exist in many different conformational states prior to ligand binding, also known as a conformational ensemble (Austin *et al.*, 1975; Foote and Milstein, 1994). The ‘conformational selection and population shift’ model was born out of these observations: the

ligand selects one of the fluctuating protein conformations suitable for binding, shifting the conformational ensemble towards it (Frauenfelder, Sligar and Wolynes, 1991; Ma *et al.*, 1999). This model describes the conformational space accessible to a protein-ligand system as an energy landscape: the system's free energy is a function of its conformations, and the probability of occupying a particular point in the energy landscape depends on the free energy of that state. On the energy landscape are global and local minima which correspond to native and metastable states of the system respectively. These minima are then separated by energy barriers or transition states in the form of local maxima.

2.3 MOLECULAR DOCKING STUDIES

CADD techniques are commonly used at the following stages of the drug discovery process: design of novel drugs based off a pharmacophore or knowledge of target of interest; hit identification by virtual screening; and hit-to-lead optimization to enhance selectivity and/or affinity. CADD techniques can be further classified in ligand-based and structure-based drug design. Ligand-based design does not require a high-quality structure of the target and makes use of existing compounds that are known to bind to a target of interest. This often involves the elucidation of a pharmacophore with the essential sub-structural characteristics required for target binding (Amaro *et al.*, 2008). Structure-based design techniques require accurate knowledge of the structure of the target protein. Data which is typically obtained using techniques such as x-ray crystallography (Petrella *et al.*, 2011b) or NMR spectroscopy (Bezsonova *et al.*, 2008), alternatively homology modelling, has emerged as a viable way to generate accurate 3D structures using experimental structure information of homologous proteins (Hatherley *et al.*, 2016). One popular method of structure-based drug design is virtual screening, where a compound library is searched through using rapid docking software to identify compounds that favourably bind to the target structure.

Docking techniques predict the most favoured orientation or pose of one molecule (a ligand) to another (most commonly a protein) (Huang and Zou, 2010). Typically docking is done in two stages for a given system. First a sampling engine is used to generate numerous ligand poses/conformations, then a scoring function is used to evaluate and rank each pose based on binding energy. To be as accurate as possible, the sampling engine must be able to generate adequate numbers of valid poses, and the scoring function must be able to identify favoured conformations based on them being at or near the global energy minima of the energy landscape of a system.

2.4 CONFORMATION SAMPLING

In principle the search/sampling engine during docking is set an impossible task: generate and search all valid ligand-receptor conformations within a search space. However an exhaustive search is impossible as both the ligand and receptor are dynamic systems that exist in a multitude of states (Austin *et al.*, 1975; Foote and Milstein, 1994). Additionally there are many ways to fit two molecules together in a 3D space: three translational and three rotational degrees of freedom mean that the number of possibilities increases exponentially with increasing size of protein and ligand (Shoichet *et al.*, 1992). In order to reduce the computational cost of searching, most sampling algorithms limit the flexibility of either one or both molecules to varying degrees (Halperin *et al.*, 2002). Docking techniques that limit flexibility in one structure, typically the larger receptor molecule, are known as semi-flexible docking techniques. These techniques assume that a rigid receptor conformation can bind to ligands in a realistic manner, an assumption based on the lock-and-key mechanism for protein-ligand binding proposed by Emil Fischer in 1894. As discussed previously, this mechanism alone is insufficient to explain numerous protein-ligand binding processes as proteins are not static structures but rather dynamic, existing in an ensemble of possible conformations. Numerous methods for flexible docking have also been designed (Alonso *et al.*, 2006). They can be divided between those that consider single protein conformations such as soft docking (Jiang and Kim, 1991) or sidechain flexibility approaches (Leach, 1994) and multiple conformation approaches, where an ensemble of receptor protein conformations is used during docking (Knegtel *et al.*, 1997; Huang and Zou, 2007).

Allowing for increasing flexibility and therefore a greater number of degrees of freedom in a system can increase the accuracy of docking, as this allows the dynamic nature of proteins and ligands to be better simulated, but this also typically comes with a significant increase in computational cost. In order to achieve desired feasibility goals in the context of high-throughput virtual screening, sampling and conformational searching techniques used should be balanced between accuracy and speed.

2.5 SCORING FUNCTIONS

Scoring functions are the second half of any docking program. They are used to identify favourable binding modes from the sample of poses that are generated from the sampling engine. They do this by estimating the binding energy between a ligand and receptor based on many physiochemical properties such as: bonded interactions, non-bonded interactions (van

der Waals forces (vdW), electrostatic potentials) as well as solvation and desolvation effects. They can be separated into 3 categories: force-field, empirical and knowledge-based scoring functions. Force-field based scoring functions estimate the potential energy of a system by summing the strain energy and intermolecular interactions (Morris *et al.*, 2009). Empirical scoring functions count the various types of intermolecular interaction terms in a system and weigh them against coefficients derived using multiple linear regression methods (Böhm, 1998). Knowledge-based scoring functions are derived from statistical observations of the frequency of intermolecular interactions in 3D structure databases such as the Protein Data Bank (PDB), with the assumption that favourable interactions, those that increase binding affinity, are more common than those with a negligible or negative effect on binding affinity (Velec *et al.*, 2005).

2.6 AUTODOCK4

The AutoDock4.2 software (Morris *et al.*, 2009) was used to perform molecular docking in this study. It makes use of a semi-flexible sampling engine and a force field-based scoring function and allows for fully automated docking workflows. An important part of the software is the Autogrid program that calculates grid maps of interaction energies between receptor atoms and any potential ligand atoms prior to any docking. This reduces the time taken during docking as these 3D grid maps are used to determine the net binding energy between a ligand and receptor, and only need to be generated once for a particular ligand. The calculated binding energy of the ligand is therefore the product of corresponding values in the 3D grid maps and the partial charge of ligand and receptor atoms.

AutoDock4 uses a sampling engine based on a Lamarckian genetic algorithm (Morris *et al.*, 1998). These algorithms allow for searching of a large conformational space by representing each receptor-ligand complex as a collection of genes, describing the conformation, translation, and orientation of the ligand. Each of these ‘genes’ are treated as the state variables of the ligand and each state variable therefore corresponds to a gene. As such the ligand’s state

$$\Delta G = \Delta G_{\text{vdW}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \Delta G_{\text{hbond}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \Delta G_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + \Delta G_{\text{tor}} N_{\text{tor}} + \Delta G_{\text{sol}} \sum_{i,j} (S_i V_j - S_j V_i) e^{-(r_{ij}^2 / 2\sigma^2)}$$

Equation 1: Scoring functions used by AutoDock4. The function consists of 5 terms. A Lennard-Jones 12-6 intermolecular pair potential is used to calculate vdW dispersion/repulsion contributions; a directional 12-10 potential hydrogen bonding term for H-bond interactions; a screened coulomb potential term for electrostatic interactions; an unfavorable entropy of ligand binding term to account for restrictions on conformational degrees of freedom and finally a desolvation potential based on the volume of atoms surrounding an atom of interest.

variables correspond to its net genotype while its atomic coordinates in 3D space correspond to its phenotype. A phenotype's 'fitness' is calculated by the scoring function and is therefore the net binding energy between the ligand and protein. Sampling of new poses is done in two ways: by crossover, where a new pose is generated using genes inherited from two parents, or by mutation where a single gene is altered to a random degree. This process is repeated for multiple generations and for each generation selection of the 'offspring' for the next generation is performed using its associated fitness scores.

The scoring function used by AutoDock4 is force-field based and calibrated empirically using 30 validated protein-ligand complexes with experimentally determined binding constants. The function includes five terms as shown in **Equation 1**. The first four terms (vdW forces, hydrogen bonding, electrostatics, and torsional constraints) are standard molecular mechanics terms and are described in more detail in the following chapter.

2.7 METHODOLOGY

2.7.1 Data preparation

Molecular docking was performed using the crystal structure of *Mycobacterium tuberculosis* pyrazinamidase enzyme (PDB ID: 3PL1) and 623 ligands from the South African Natural Compound Database (Hatherley *et al.*, 2015). The 3D structure of PZase used in this study contained the PZA ligand docked to it and was generated during previous research (Sheik Amamuddy *et al.*, 2020). This docked PZA ligand was extracted from the structure and used

to validate the redocking attempt performed in this study as there is no crystal structure of Mtb PZase that has been co-crystalized with PZA

2.7.2 Protonation

All ligand PDB files and the 3PL1 PDB file were converted to PDBQT files suitable for use with AutoDock. The 3PL1 receptor and all ligands were prepared by merging all non-polar hydrogen atoms, adding atom types, and calculating Gasteiger charges. The 3PL1 receptor was also protonated at a pH of 7 that matches the physiological conditions the PZase protein is found in.

2.7.3 Autogrid setup and docking preparation

The Autogrid program was used to calculate grid maps for receptor docking with each ligand. Since this was a blind docking study, the grid was centred at the PZA active site and a 110x110x110 Å box size was used to ensure the entire PZase structure was inside each grid map. Grid maps for all default atom types were generated, in addition to maps for bromine, as some of the SANCDB ligands contained bromide groups.

2.7.4 Docking Simulations

Docking was performed using AutoDock4.2 on the in-house Yoda cluster. As per previous work (Sheik Amamuddy *et al.*, 2020), maximum number of generations per run was set at 27 000 with a maximum number of energy evaluations set at 450 000. RMSD cut-off when performing clustering analysis on poses was left at the default value of 2.0.

2.7.5 Docking Validation

Docking methodology was validated by redocking the PZA ligand with the 3PL1 PZase crystal structure. Initially, docking was going to be performed with AutoDock Vina (Trott and Olson, 2009) due to its performance advantages over AutoDock4.2 (Morris *et al.*, 2009), however AutoDock Vina was unable to replicate the same pose and intermolecular interactions as the original docked PZA ligand (both the water and Asp8 H-bonds between PZA and PZase were absent). Docking with AutoDock4.2 however was able to produce the near identical pose and intermolecular interactions including the water hydrogen bond that is essential for the conversion of PZA to POA (Petrella *et al.*, 2011a). Docking validation was performed with identical parameters as those described above for the docking of SANCDB compounds.

2.7.6 Docking Analysis

Analysis of AutoDock4 docking results is more complicated than analysis of AutoDock Vina results. For each docked system, 100 poses were generated and the clustered according to docking RMSD with a cut-off of 2.0 Å using AutoDock4's internal automated analysis toolset. In order to take into account multiple factors that influence a ligand's affinity and specificity to a particular target such as: binding energy, hydrogen bonds and ligand molecular weight and RMSD clustering; custom scripts were used in conjunction with those part of AutoDock Tools to extract poses from each docked system and then apply criteria to each extracted pose to perform selection. The first script used for this was a slightly customized `process_VSResults.py` script in AutoDockTools that was used to extract 2 poses from each system: **(1)** The global lowest energy pose (LE pose) and **(2)** the lowest energy pose in the largest cluster (LC pose). The output of this would usually be 2 PDBQT files per ligand, 1 for each of the poses, however in instances where the LE and LC poses were the same pose only one PDBQT file was generated. In addition to standard PDBQT file data, the `process_VSResults.py` also generated hydrogen bond information as well as docking clustering data as comments within each PDBQT file. This resulted in 1068 distinct PDBQT files representing 1068 poses. Centre of Mass (COM) distance between ligand and the PZase active site was calculated to determine ligand distance from the active site.

Two rounds of selection with different criteria were performed using a fully custom script on all 1068 PDBQT files:

Selection 1: COM distance ≤ 8 Å; 1 \geq hydrogen bond between ligand and PZase (residue agnostic); binding energy ≤ -6.5 kcal/mol; Molecular mass ≤ 600 Da

Selection 2: 1 \geq hydrogen bond between ligand and PZase (specifically to active site residues); binding energy ≤ -5 kcal/mol; Molecular mass ≤ 600 Da

Selected ligands were converted in PDB file format using Open Babel (O'Boyle *et al.*, 2011). Visualization of docking results and docked ligand-receptor complexes were done using PyMOL (Schrödinger, 2020) and Accelrys Discovery Studio visualizer (BIOVIA, 2021). Visualization of intermolecular interactions was also done with Accelrys Discovery Studio. A graphical overview of the methodology used in this study is shown in **Figure 8**.

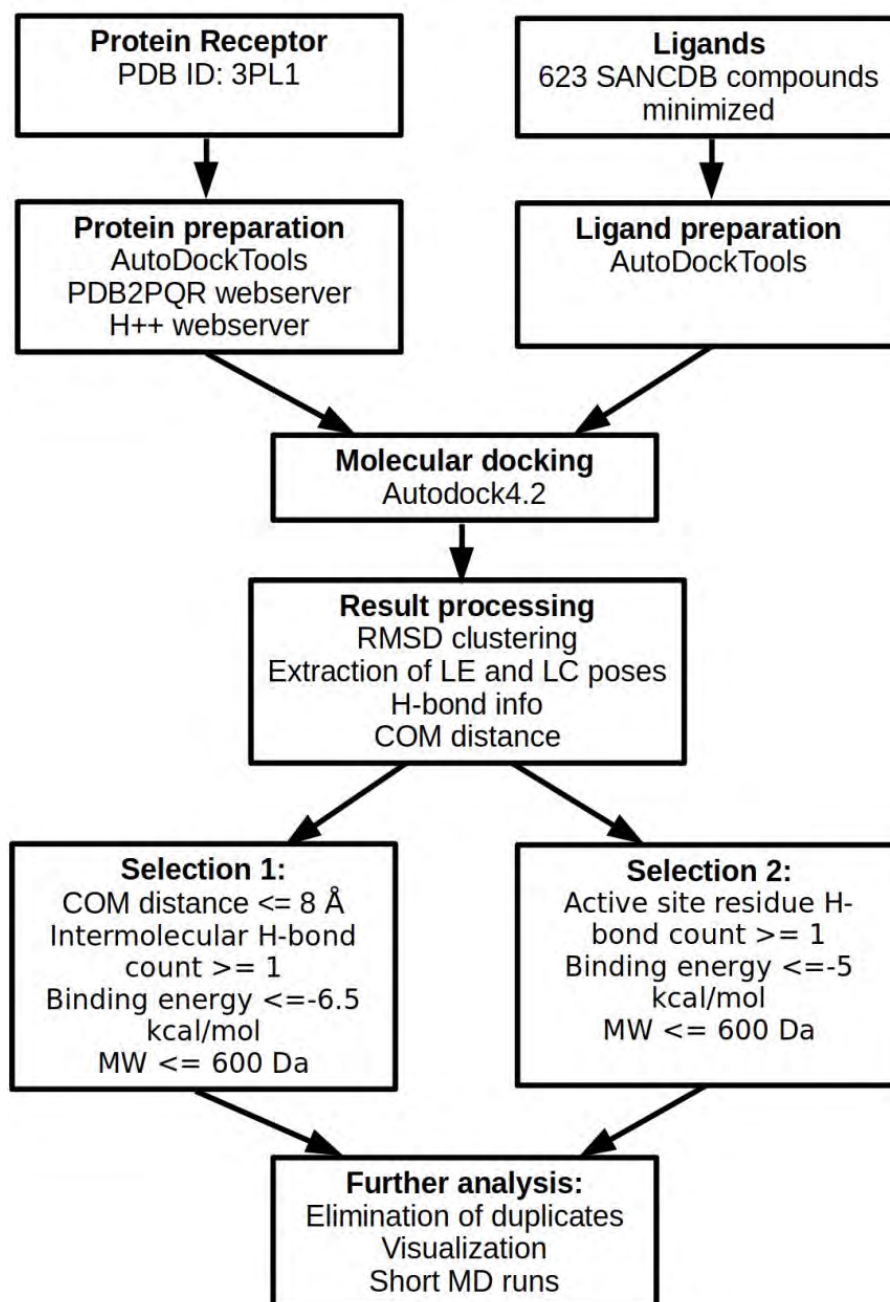


Figure 8: Overview of docking methodology in this study

2.8 RESULTS AND DISCUSSION

2.8.1 Docking validation

The PZA ligand used for re-docking was extracted from the complex produced in previous work. Docking against the Mtb PZase structure is a fairly straightforward process as it is a relatively small monomeric, single chain protein with a well-defined active site. Validation via re-docking of the PZA ligand indicated that the docking protocol used in this study was able to reproduce the correct pose and intermolecular interactions as determined by previous research (Sheik Amamuddy *et al.*, 2020). While the calculated binding energy of the redocking was not

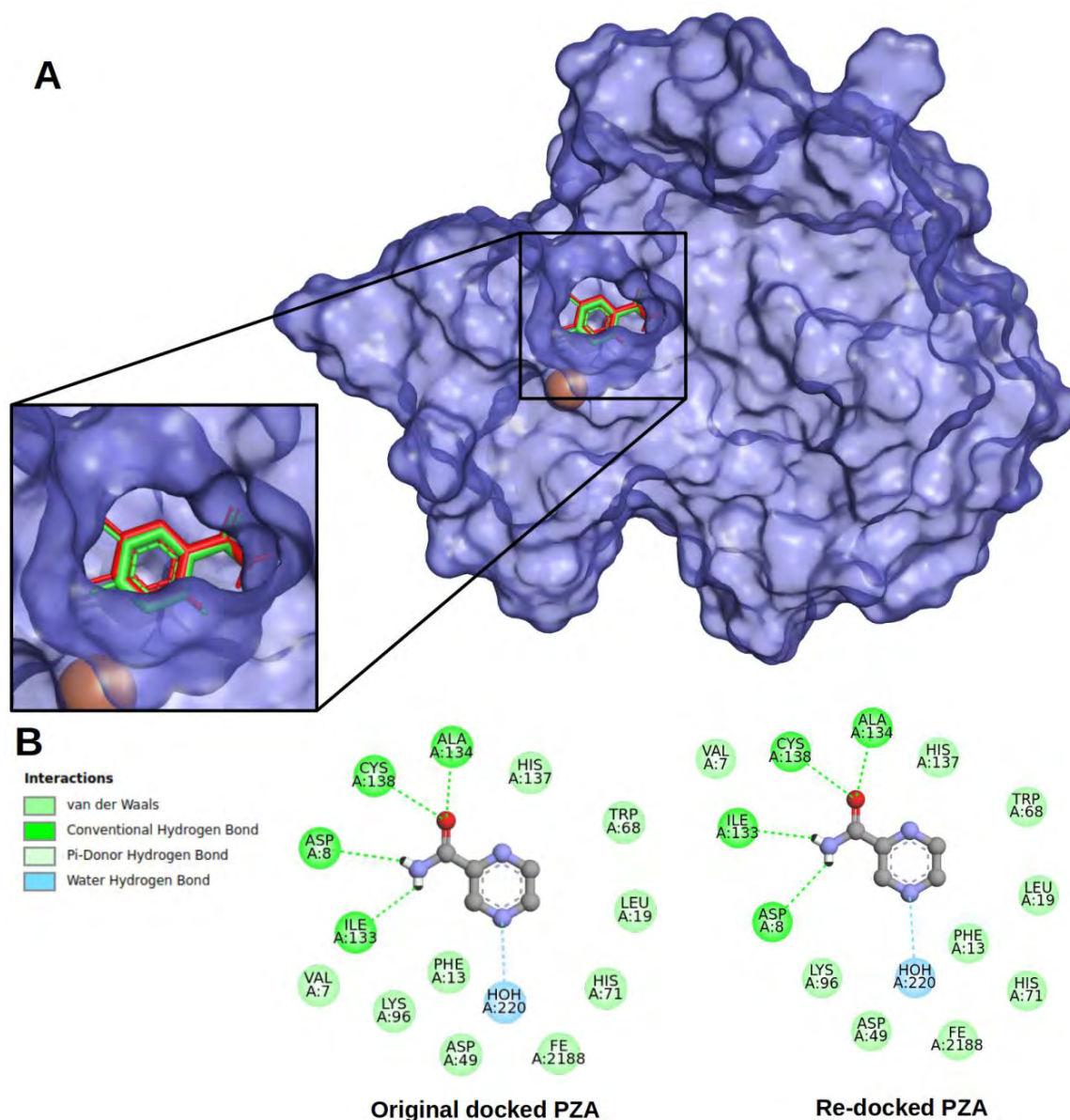


Figure 9: Docking validation. (A) Docking validation results of 3PL1 crystal structure in complex with both the original docked PZA ligand (green) and re-docking attempt (red). Fe^{2+} atom shown in orange. (B) 2D diagram of intermolecular interactions between the 3PL1 PZase receptor and the PZA ligand.

notable at -4.28 kcal/mol, this was in line with previous research which obtained a binding energy of -4.48 kcal/mol. Reproducing the exact pose through docking is essential as specific residues within the PZase active site are required in catalysing the PZA to POA reaction.

Shown in **Figure 9A**, a surface representation of the 3PL1 PZase receptor with both the original and redocked PZA poses overlaid is presented. RMSD between the two poses was calculated at 0.232 using the DockRMSD (Bell and Zhang, 2019) webserver, indicating that the redocked PZA achieved a similar pose and location to the original docked ligand. Discovery studio was used to generate the 2D intermolecular interaction diagrams shown in **Figure 9B** between 3PL1 receptor and the PZA ligands. Redocking was successful in reproducing the correct pose as well as all the essential non-bonding interactions. Notable is the hydrogen bonds between ligand and residues Cys138 and Asp8, both of which are part of the Asp8-Lys96-Cys138 catalytic triad that is conserved amongst numerous homologous proteins (Du *et al.*, 2001) Additionally the essential water hydrogen bond to the aromatic ring nitrogen of PZA is also present. These interactions are essential to PZase activity on PZA: H₂O initiates the reaction, Asp8 acts as a general base and converts Cys138 to its thiolate form which initiates a nucleophilic attack on PZA producing the acyl enzyme intermediate (Du *et al.*, 2001)

2.8.2 SANCDB screened compounds

Virtual screening was performed with 623 compounds from the SANCDB database. All compound structures were pre-optimised and in their lowest energy conformations. Virtually all of these compounds have not been manufactured and are unavailable for purchase, therefore they have not been tested for anti-tuberculosis activity. All protein-ligand complexes produced from docking were viewed in either PyMOL or Discovery Studio Visualizer. Since the original intention of this study was to identify compounds binding to both the active site and any potential allosteric sites, blind docking was performed over the entire surface of the PZase structure and ligands that preferentially bound to within the active site pocket were identified.

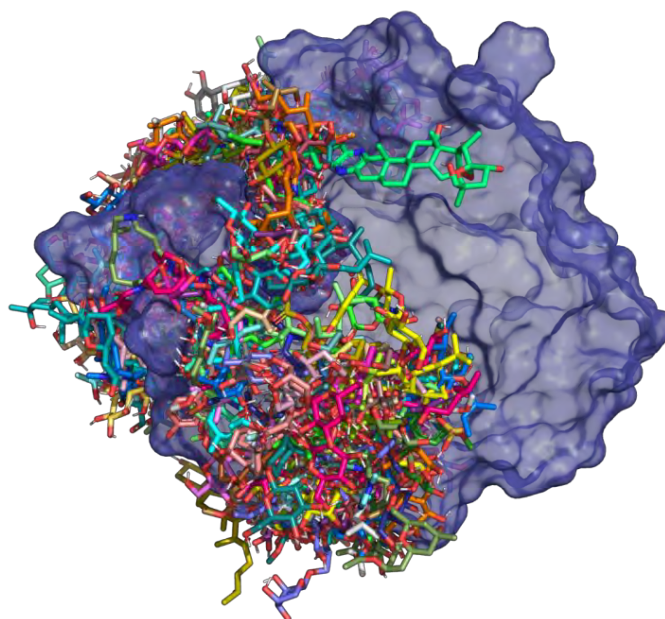


Figure 10: All PZase protein-ligand docking complexes. The 1068 ligand poses are shown as colored sticks while the PZase receptor protein is shown as a semi-transparent blue surface.

The most common method of determining the binding affinity between a ligand and protein after docking uses a calculated value known as binding free energy. AutoDock4 uses a semi-empirical free energy force-field algorithm to determine whether a ligand-protein complex is energetically favourable. Therefore, a lower energy value corresponds to a more favourable binding while a higher value means it is less favourable. However, this calculated binding energy is by no means the only method to evaluate ligand suitability as intermolecular interactions such as the presence of hydrogen bonds have been shown to be important factors to consider during screening (Zhao and Huang, 2011; Chen *et al.*, 2016). Compounding this are the output results of AutoDock4.2, which as per the default parameters used in this study produces 100 of the most favourable poses for each docking system. As per methods described above, multiple poses were extracted from each docking result and filtered using 2 sets of criteria.

Initially selection with the 1st criteria resulted in 107 systems. Given time considerations, this was deemed an adequate number of ligands that would then be analysed and submitted for MD simulations. However, a second round of selection was performed to ensure as many SANCDB ligands that bound preferentially to PZase active sites were selected. The selection criteria differed firstly due to further work being made on the custom script used for selection and secondly because the presence of a hydrogen bond between ligand and an active site residue such as those forming hydrogen bonds in the PZA control was judged an alternate method to select ligands that a) were binding near active site and b) were interacting with an active site

residue.. The 8 Å COM distance criterion used for the first selection was also very strict and resulted in many ligands being excluded from selection despite being present in the active site. The second selection yielded 115 systems after removal of duplicates. In total 221 ligands binding within the active site pocket were selected as shown in **Figure 11**. In a few instances, both of those extracted poses for a SANCDB compound passed one or both of the selections. They are denoted with the suffix “_le” and “_lc” to distinguish between them when relevant. As will become clear in the following chapter, the two different poses can exhibit very different trajectories during MD simulations despite being the same ligand. The introduction of H-bond requirements in the two selection criteria were due to two reasons: Firstly, during docking validation, as well as in previous work (Sheik Amamuddy *et al.*, 2020), binding energy of the redocked PZA-PZase system was only -4.28 kcal/mol, this is a very poor score and would certainly not be indicative of a favourable binding. However, it is important to note that the binding energy reported by docking programs is a synthetic value that is determined by the scoring function used, and these scoring functions are not always accurate when calculating binding energies across a diverse set of ligands such as those in the SANCDB database. Indeed, several studies have cautioned against relying solely on binding energy to determine promising compounds (Ramírez and Caballero, 2016, 2018; Pagadala *et al.*, 2017; Pantzar and Poso, 2018). This is especially true in cases of cross-docking where a ligand is docked to a receptor structure that it was not co-crystallized with. Cross-docking is the most popular type of docking since it is used to identify novel drug candidates but because receptor binding sites are often

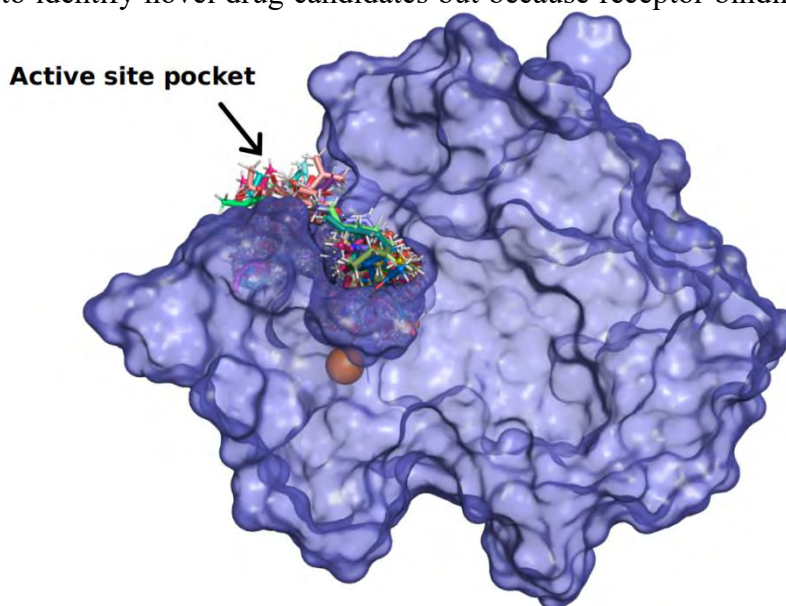


Figure 11: Selected PZase protein-ligand docking complexes. The 221 ligand poses are shown as colored sticks while the PZase receptor protein is shown as a semi-transparent blue surface. Ferrous ion shown in orange.

fitted to the ligand it was co-crystallized with (Skjærven *et al.*, 2011; Ramírez and Caballero, 2018), scoring functions may report inaccurate binding energies. The ‘induced-fit’ model of protein-ligand binding was borne out of observations such as this (Koshland, 1958; Austin *et al.*, 1975; Foote and Milstein, 1994), and docking programs have traditionally struggled to accurately account for this model due to the computational challenges associated with introducing residue flexibility into the receptor structure. One avenue to evaluate docking poses is by the presence of hydrogen bonds between polar groups in the ligand and on the protein surface. Complementary hydrogen bonds are essential to a significant number of ligand-protein binding events (Fersht *et al.*, 1985; Chen *et al.*, 2016), the breaking of water-ligand hydrogen bonds and subsequent formation of protein-ligand hydrogen bonds are an important part of this process (Zhao and Huang, 2011). Hydrogen bonding can also further modulate ligand affinity and specificity when the ligand-receptor complex has significantly higher or significantly lower hydrogen bonding potential than the hydrogen and oxygen atoms in the surrounding bulk water (Chen *et al.*, 2016). It was therefore judged that the presence of a hydrogen bond between a ligand and the PZase protein during docking would be a positive indicator of a favourable binding. Results of this investigation are shown in **Figure 12**. Shown in this figure are the 238 ligand poses that passed one or both of the selection criteria and were submitted for short MD simulations. Ligands with no hydrogen bonds shown were selected with criteria 1 and contained hydrogen bonds with residues outside the active site.

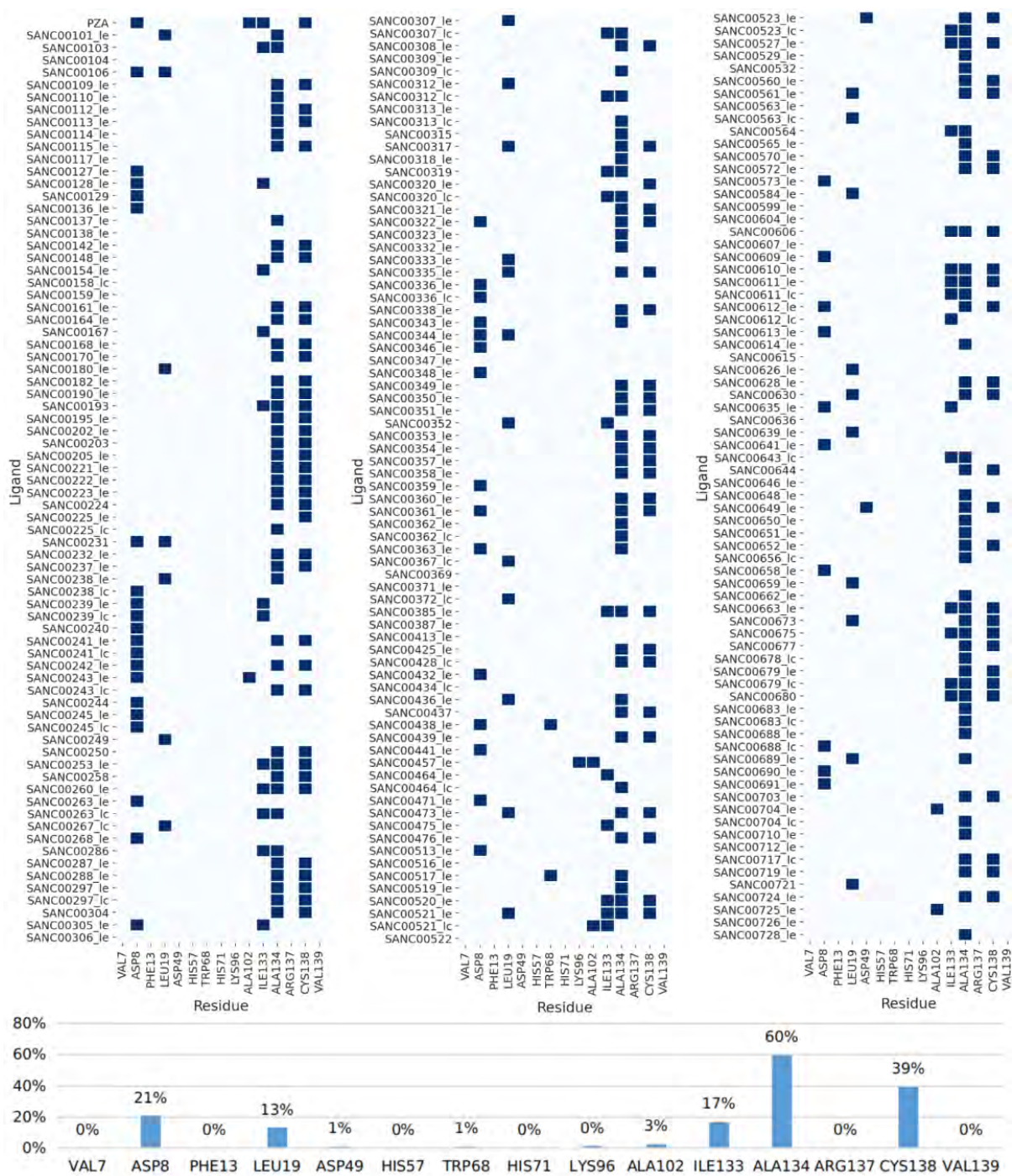


Figure 12: Hydrogen bond distribution for all selected ligands. Y axis represents each ligand-PZase complex. On X-axis all active site residues within 5 Å of original docked PZA location. The presence of an H-bond between a ligand and specific residue is shown by a navy square, a white square denotes the absence of an H bond.

2.9 CONCLUSION

The development of novel drugs is challenging, time-consuming and expensive. Computer assisted drug design (CADD) can speed up this process and reduce costs by using computational tools in the early stages of the drug design and screening process to rapidly identify large numbers of potential drug candidates that can then be further investigated. In this chapter, 623 small molecules from the SANCDB were screened against the crystal structure of Mtb PZase (PDB ID: 3PL1). Compounds that bound at or near the active site, exhibited relatively low binding energies and had hydrogen bonds with the PZase receptor protein were preferentially selected. In total 238 ligand poses were identified and selected for further MD analysis in the following chapter.

CHAPTER 3

MOLECULAR DYNAMICS SIMULATIONS FOR WILD TYPE – LIGAND COMPLEXES

3.1 INTRODUCTION

Proteins are dynamic structures that move and change conformation over time based on the physiological environment they exist in. *In silico* studies on protein dynamics often make use of molecular dynamics simulations to predict and analyse protein behaviour as well as protein-protein and protein-ligand interactions within a system. This differs from the approach used during standard docking studies, during docking protein receptors are modelled as rigid structures with certain isolated residues being allowed to flex. Ligand conformation and orientation sampling followed by scoring are then done in the absence of a fully physically relevant environment. In this chapter, the stability and dynamics of selected SANCDB compounds in complex with the WT-PZase protein were investigated using MD simulations.

3.2 MOLECULAR DYNAMICS SIMULATIONS

With an emphasis on dynamics, MD simulations involve the simulation of molecular systems as a function of time. The movement and trajectories of all atoms in a system is calculated by integrating Isaac Newton's equations governing objects in motion, otherwise known as classical mechanics. MD as a proof-of-concept as technique was first demonstrated in 1957 and was used to simulate collisions between hard spheres (Alder and Wainwright, 1957; Wood and Jacobson, 1957), in 1977 the first simulation on a protein was achieved: a 9.2 ps trajectory of bovine pancreatic trypsin inhibitor (McCammon *et al.*, 1977). Today, MD simulations are a powerful tool used for studying protein folding and unfolding, enzymatic reaction mechanisms, protein dynamics, rational drug design as well as elucidation of drug and protein mechanisms of action (Karplus and Petsko, 1990; Karplus and McCammon, 2002; Durrant and McCammon, 2011). During a simulation, trajectories of atoms are determined by solving Newton's equations governing objects in motion, inter- and intramolecular interactions, forces and the kinetic energies of atoms are evaluated using empirically derived force fields based on classical mechanics. This means each atom in a system is simulated as a particle with a radius and net charge, while bonded and many non-bonded interactions between particles are simulated as springs. A force field therefore is in effect a set of equations and constants that can be used to

reproduce the properties and geometry of atoms and all interactions within a system. Since protein structure files such as the PDB format do not contain any of this information, atom coordinates are derived from the PDB structure, then a velocity for each atom is assigned randomly from a Maxwellian distribution that is centred on the desired temperature. The velocities of each atom are then adjusted until a net angular momentum of zero is achieved.

Bond length, bond angle and angle torsion are factors that influence the force exerted during bonded interactions, as shown in **Figure 13**. Hooke's Law ($F = -kx$) is typically used to approximate this force. F being the force that is applied to the atom, $-k$ is the spring constant and x is the extension distance. For non-bonded interactions, Lennard-Jones equations are used to calculate both attractive and repulsive interactions for electronically neutral atoms.

MD simulations are frequently coupled in various manners with docking studies. The trajectories produced by MD simulations are essentially a collection of conformations for any protein(s) and ligand(s) included, as such they have been used in docking studies that use multiple receptor protein conformations instead of a common single-conformation receptor (Lin *et al.*, 2002; Amaro *et al.*, 2008). Aside from novel MD-docking hybrid techniques MD simulations can complement standard ligand-protein docking studies by using trajectory data to determine ligand (or protein) stability for the simulation duration. Due to receptor rigidity and lack of explicit solvent simulation, poses that appear favourable during docking may be unstable or even dissociate from the protein receptor at adequate timescales in an MD simulation.

For the purposes of simulations modelling biological macromolecules, MD simulations that rely on molecular mechanics and empirically derived force fields have one notable shortcoming: they are unable to describe the interactions between a protein and any coordinating metal atom(s). The metal atoms in many proteins play essential roles in enzymatic catalysis as well as protein dynamics and folding (Sousa *et al.*, 2010), therefore they cannot be omitted during MD simulations. However commonly used force fields such as CHARRM (Vanommeslaeghe *et al.*, 2010), GROMOS (Oostenbrink *et al.*, 2004) and AMBER (Maier *et al.*, 2015) typically contain molecular mechanical parameters for an expanded set of standard and non-standard amino acid residues and common post-translational modifications, but little to no metal atom parameters, and since molecular mechanic methods bypass the concept of electrons as well as

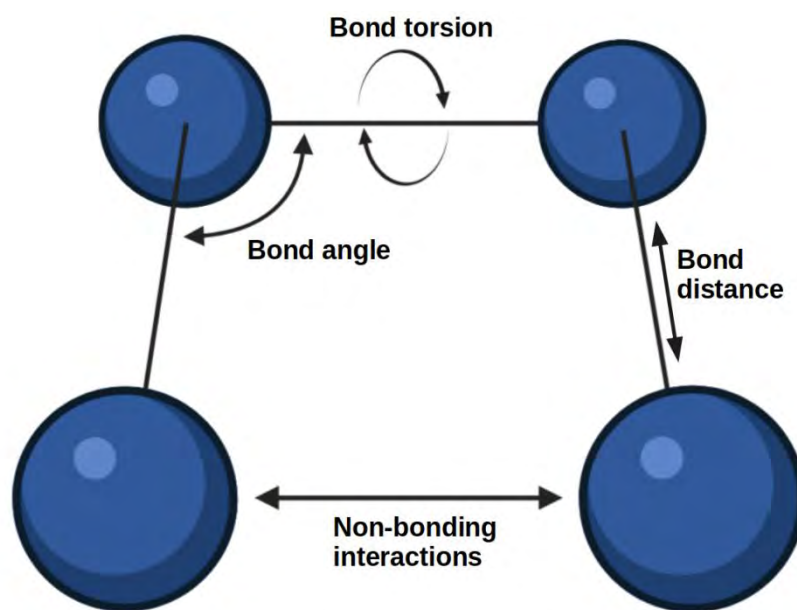


Figure 13: Components of a typical MM force field. The total energy of each atom in a system is the summation of all potential energy terms (Bond distance, bond angle, bond torsion, bond angle and non-bonding interactions). Graphic created in Biorender.

the quantum mechanical nature of molecules, these force fields cannot hope to simulate interactions between a protein and a metal atom in an accurate manner in the absence of empirically derived parameters.

3.3 METHODS:

Molecular dynamics simulations for screening hit compounds and wild-type PZase were performed in a standardized step-by-step protocol as illustrated in **Figure 14** at the end of this section.

3.3.1 Model Preparation

3.3.1.1 Protein preparation

The *Reduce* tool (Word *et al.*, 1999) was used to remove all hydrogens from the PZase protein structure (PDB ID: 3PL1). The de-protonated structure was then protonated by the H++ webserver (Anandakrishnan *et al.*, 2012) at a pH of 7 with a salinity of 0.15 M and default dielectric parameters. Generated AMBER topology and coordinate files (.top and .crd) were downloaded and used to generate correctly protonated protein structure files using the *amb2PDB* tool. Catalytic site residues were edited and re-labelled in order to work with inferred Fe²⁺ parameters and the Fe²⁺ atom was manually assigned a correct charge of +2.

3.3.1.2 Ligand preparation:

Ligand PDBQT files of hits from the docking study were converted to a PDB file with all implicit hydrogens added using open Babel (O'Boyle *et al.*, 2011). A custom python script was used to automate this and to correctly renumber atoms as open Babel did not update atom numbers when adding hydrogens. AmberTools' antechamber tool was then used to generate coordinate and parameter files for each ligand using the AM1-BCC charge model (Jakalian *et al.*, 2002) to calculate atomic point charges. Potential non-parameterized atom combinations were then checked for using AmberTools' *parmchk* tool.

3.3.2 System preparation

3.3.3.1 Topology generation

AmberTools' *LEaP* program was used to generate Amber topology and coordinate files of the PZase-ligand complexes with the AMBER ff14SB (Maier *et al.*, 2015) force field and generated Fe²⁺ force field parameters. Systems were solvated using the TIP3P water model (Mark and Nilsson, 2001) in a cubic box with a padding distance of 10 nm and all systems were neutralized by addition of NaCl. The AnteChamber Python Parser Interface (ACPYPE) tool (Da Silva and Vranken, 2012) was used to convert finalized AMBER topology and coordinate

files to GROMACS-compatible structure (.gro) and topology (.top) files with corrected atom types and charges.

3.3.3 Minimization

Energy minimization of systems prior to production runs is essential in order to remove any unfavourable geometries and steric clashes in a system, particularly after water molecules and neutralization ions are added to the system. Systems were minimized using the GROMACS (Abraham *et al.*, 2015) preparatory *grompp* command followed by the *mdrun* command. Minimization was done using the steepest descent algorithm for 50 000 steps and was halted when a maximum force < 1000 kJ/mol was achieved for every system. Successful minimization was verified by generating an energy minimization curve with the GROMACS energy command. The resulting .xvg files were viewed in xmgrace.

3.3.4 Equilibration

The equilibration step ensures that a solvated system is at a correct temperature and pressure. Equilibration was done by taking the system to the desired simulation temperature of 300 K with an NVT ensemble (constant number of particles, volume and temperature) for 200 ps using a Brendsen thermostat. Following this pressure equilibration was performed with an NPT ensemble (constant number of particles, constant pressure and constant temperature) for 200 ps in order to achieve a stable pressure of approximately 1 bar using the Parrinello-Rahman barostat. Successful equilibration was verified by generating pressure and temperature curves with the GROMACS energy command. The resulting .xvg files were viewed in xmgrace.

3.3.5 Production MD runs

Following minimization and equilibration, system position restraints were relaxed, and 150 ns production runs were conducted using the GROMACS *mdrun* command. Trajectory and coordinate information was saved every 10 ps, yielding 15 000 states for every system simulated. All production runs were done on the CHPC (Center for High Performance Computing) in Cape Town.

3.3.6 Trajectory Analysis

3.3.6.1 RMSD

RMSD (Root mean square deviation) in the context of MD simulations is a measure of the average distance between atoms (usually backbone atoms or ligand atoms in the case of protein MD). Here it is a measure of distance between atoms during simulation and the initial simulation state after equilibration and minimization. It is used to evaluate the stability of a simulated structure through time and for this study RMSD of protein backbone atoms and RMSD of ligand atoms was analysed to evaluate protein and ligand stability, respectively. RMSD data was generated using the GROMACS *rms* command, processed in R studio and then visualized using the Pandas (McKinney, 2010; Reback *et al.*, 2021) and Seaborn toolkits (Waskom *et al.*, 2020).

3.3.6.2 RMSF

While RMSD provides a global measure of deviation of a system through time, RMSF (Root means square fluctuation) provides an average deviation over time at the individual residue level and is therefore a measure of residue rigidity over the course of the simulation. RMSF data was generated using the GROMACS *rmsf* command, processed in R studio and then visualized using the Pandas and Seaborn toolkits

3.3.6.3 Radius of gyration:

The Rg (Radius of gyration) of protein is a measure of its ‘compactness’. When a protein is stable and in its native fold it is most likely to maintain a constant value gyration value over time. For this study Rg was used to determine the overall distribution of a system. The GROMACS command *gyrate* was used to calculate Rg for the PZase protein as a whole as well as active site residues within 8 angstroms of the binding site of PZA. Data was processed in R studio and then visualized using the Pandas and Seaborn toolkits.

3.3.6.4 Hydrogen bonding analysis:

Hydrogen bonding was evaluated in a two-fold manner. Firstly, the *hbond* tool built into GROMACS was used to generate data on the total number of intermolecular H-bonds between ligand and PZase throughout the 150 ns simulations. The second tool used was AmberTool20’s CPPTRAJ program (Case *et al.*, 2020). This tool was used to generate specific hydrogen bonding data and occupancy values for every residue in the PZase active site pocket. Both tools

were supplied identical parameters (max H-bond distance: 3.5 Å; angle cut-off: $\pm 35^\circ$) to ensure equivalence between the two sets of data.

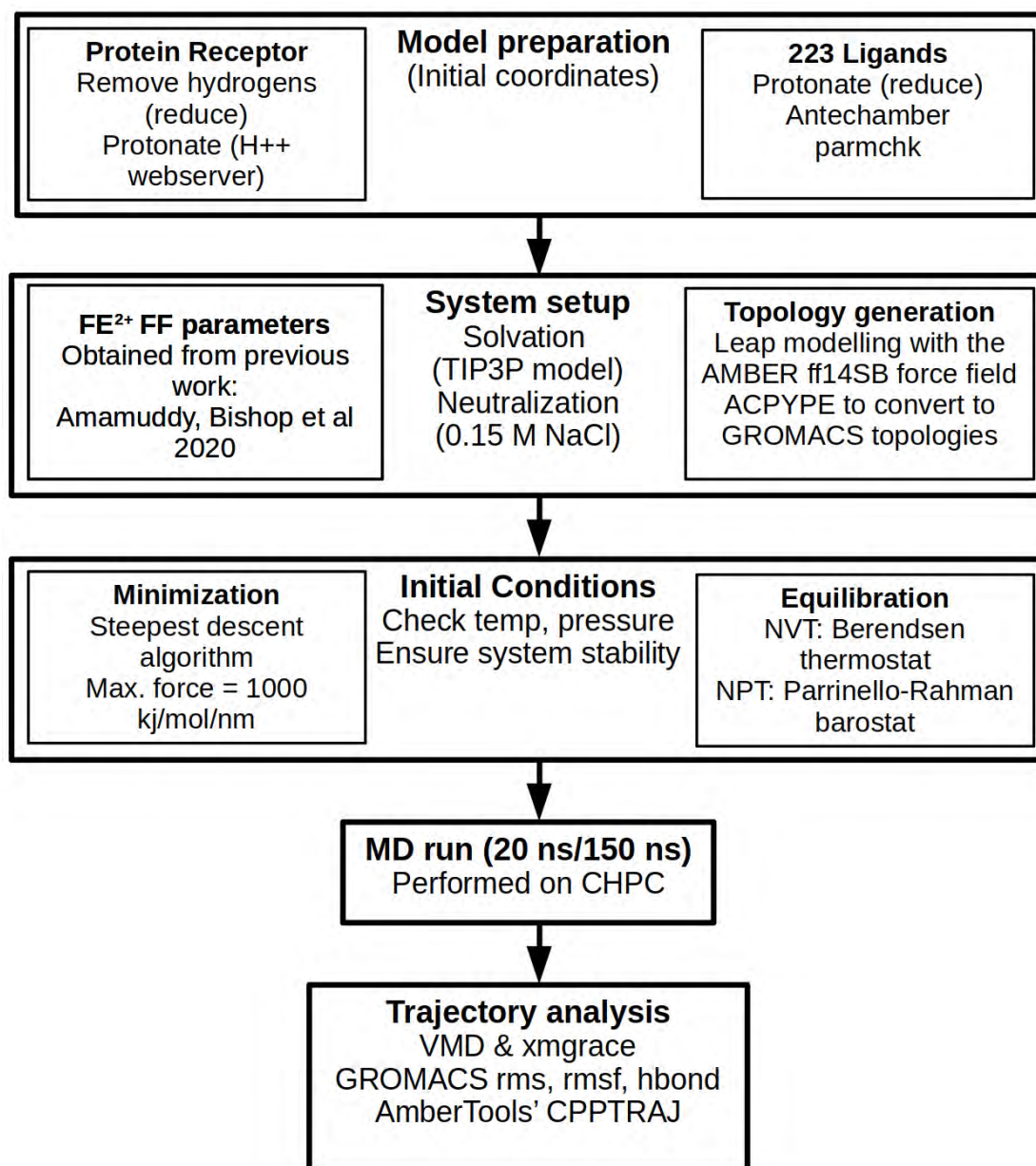


Figure 14: Overview of methodology used for Molecular Dynamics studies.

3.4 RESULTS AND DISCUSSION

3.4.1 20 ns pre-runs

As a first step, 20 ns MD simulations were performed on all ligands selected from the docking study. Due to the computational cost of these simulations, it was decided that pre-production 20 ns runs would be used initially to eliminate ligands that were immediately unstable during MD simulations. To determine this, ligand-PZase complexes were simulated for 20 ns MD runs, then ligand RMSD for each system at 10 ps intervals was determined to evaluate stability -whether each ligand underwent significant changes in conformation or location. Only the last 10 ns of these simulations were used to calculate ligand RMSD in order to allow the ligands to settle into a favourable conformation, if one exists. Ligands with RMSD clustering that indicated initial stable binding to PZase were then submitted for 150 ns production MD runs, as shown in **Figure 15**. RMSD data is present here using violin plots, which are similar to box plots but with the added benefit that they show probability density values. Ligands that exhibited unimodal RMSD clustering with a median value at or below 0.2 nm were selected

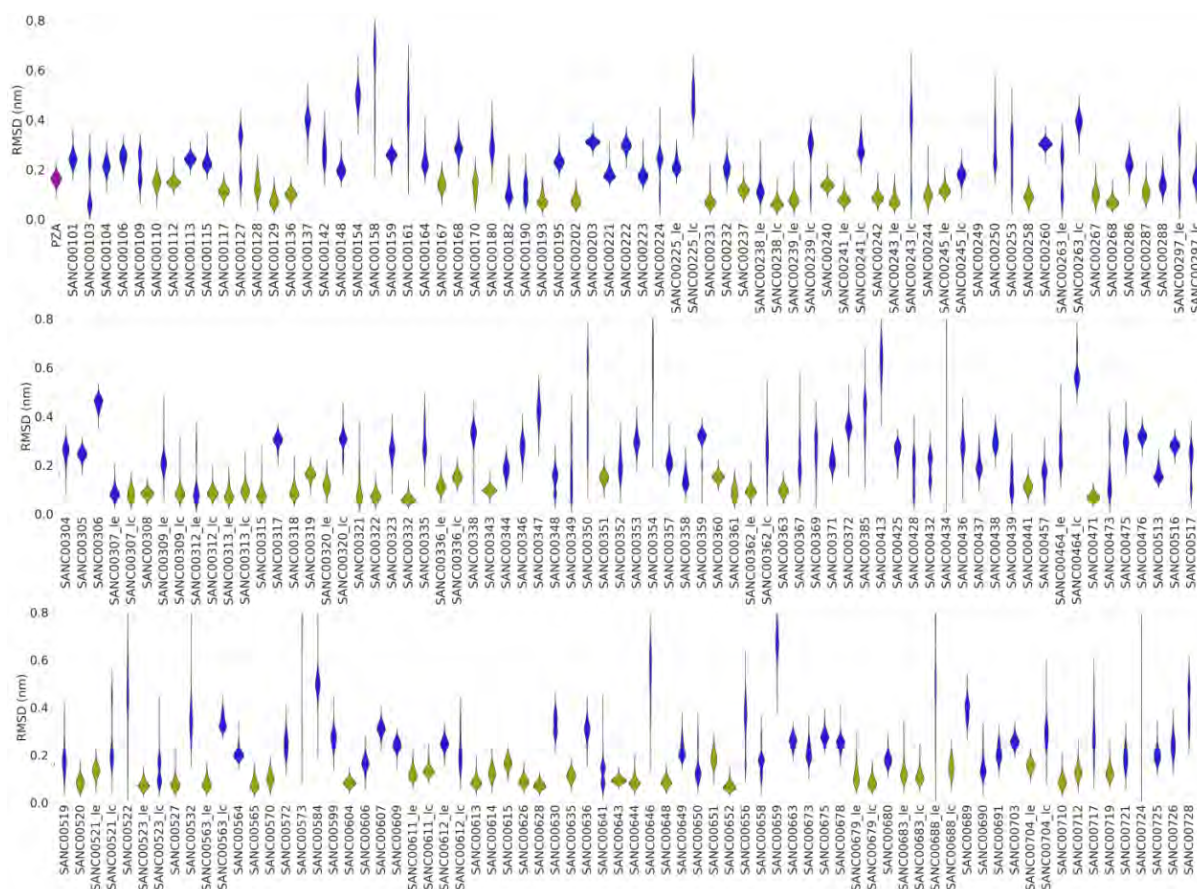


Figure 15: Violin plots of Ligand RMSD results for 20 ns pre-production MD runs. In purple: PZA-PZase control; blue: ligands judged to be unstable; yellow: ligands appearing stable.

based on this data, while all others were discarded. By this metric the PZA control was also considered sufficiently stable and submitted for a 150 ns MD run. Unimodal clustering at or below 0.2 nm indicates that the ligand is very stable throughout the simulation and was likely only in one pose. Multimodal clustering and clustering at values much higher than 0.2 nm indicates that a ligand is fluctuating substantially and is alternating between different poses during simulation, suggesting instability. Finally, ligands that dissociate from the protein and leave the system entirely are visible as violin plots stretched out over large RMSD ranges with no apparent clustering. Since the plots in **Figure 15** only represent the last 10 ns of a 20 ns simulation, analysis of any multi-modal clustering was not performed as the simulation timescale was too short.

3.4.2 Ligand RMSD

Extended 150 ns production MD runs revealed the limitations of the shorter MD runs: <100 ns timescales are not sufficient to simulate ligand-binding and ligand-unbinding events and therefore cannot reliably predict ligand stability. Many ligands that appeared stable during 20 ns simulation were shown to be unstable in the 150 ns simulations. Still, since it was unfeasible to perform 150 ns MD simulations on all 238 systems, pre-screening using 20 ns MD runs was an effective method to reduce the number of 150 ns simulations needed. Selection of hit ligands from MD production runs was based primarily on ligand stability observed with ligand RMSD

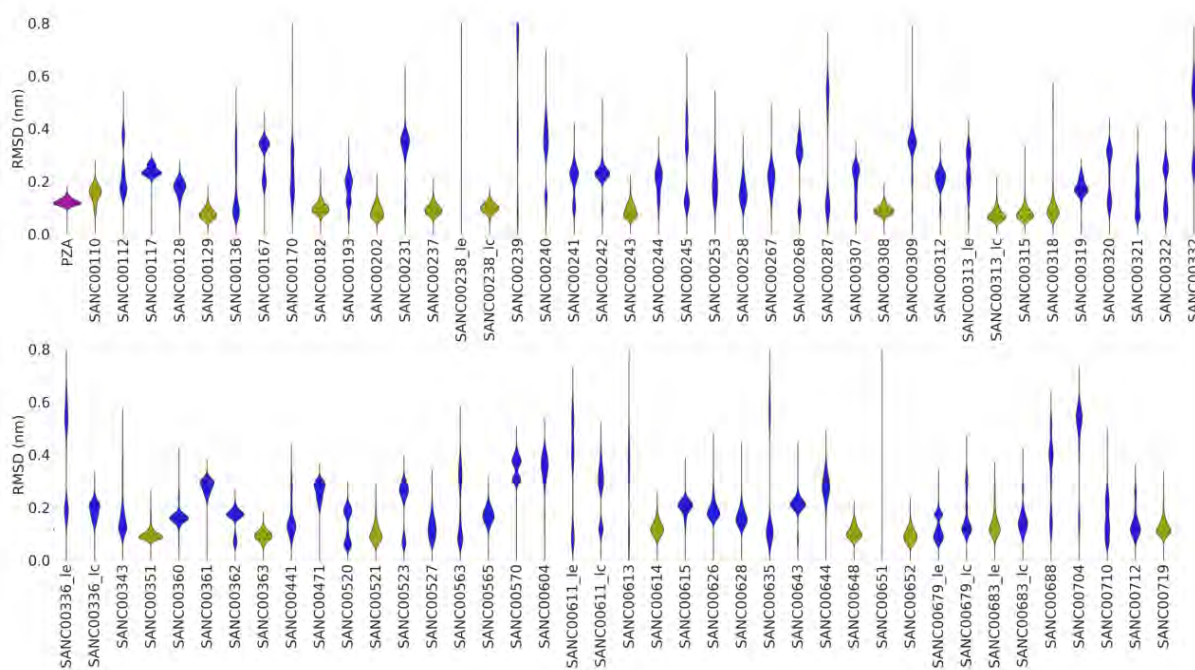


Figure 16: Violin plots of Ligand RMSD results for 150 ns production MD runs. In purple: PZA-PZase control; blue: ligands judged to be unstable; yellow: ligands appearing stable.

data shown in **Figure 16**, additionally, hydrogen bond data was taken into account too, as the presence of persistent hydrogen bonds is an important indicator of ligand stability.

3.4.3 Hydrogen bonding

In **Figure 17**, systems are ranked according to the sum of the occupancy values of H-bonds that are present between ligand and the active site residues shown on the x-axis of Figure 17A. Most notable is that a significant number of stable systems exhibit weak and even non-existent hydrogen bonding between ligand and protein. This is also apparent in **Figure 18**, showing non-specific hydrogen bond counts. Take for example the top 4 systems in Figure 18: all four

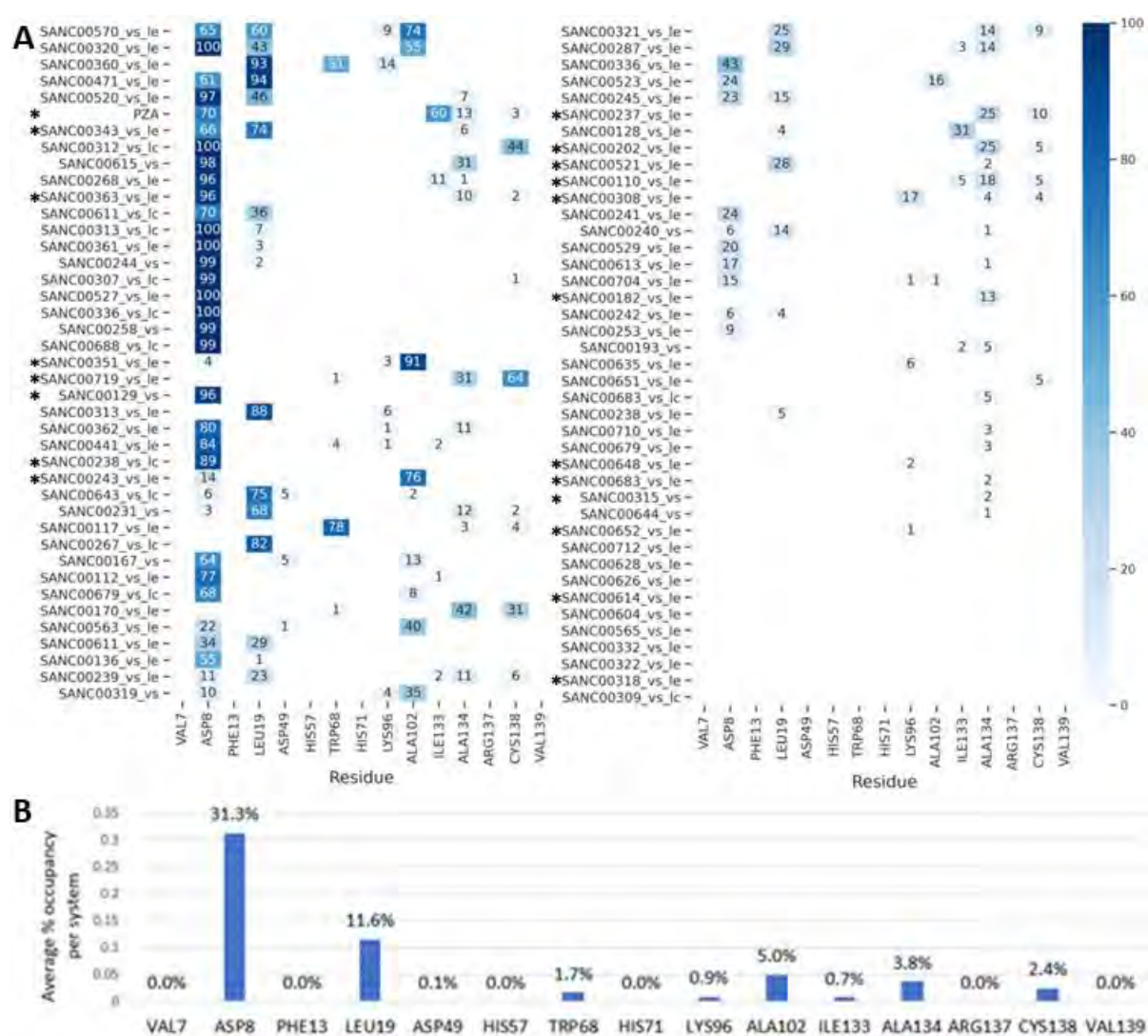


Figure 17: Heatmap of intermolecular hydrogen bond interactions during production runs. (A) Each system is denoted by the ligand name on the y-axis, residues making up the PZase active site pocket are listed on the x-axis. The colour intensity of coloured squares and the numbers in each denote hydrogen bond occupancy, or the percentage of time during the simulation where the hydrogen bond was present. The absence of a coloured square and a percentage therefore denotes the absence of a hydrogen bond. Systems identified as stable in **Figure 16** are marked with an asterisk (B) Average occupancy per residue.

exhibit hydrogen bonding throughout the simulation duration with the exception of SANC00679_lc. Despite this, none of the ligands in these systems are stable, as per ligand RMSD data shown in **Figure 16**. This suggests that hydrogen bonding alone is not a reliable way to predict ligand binding potential.

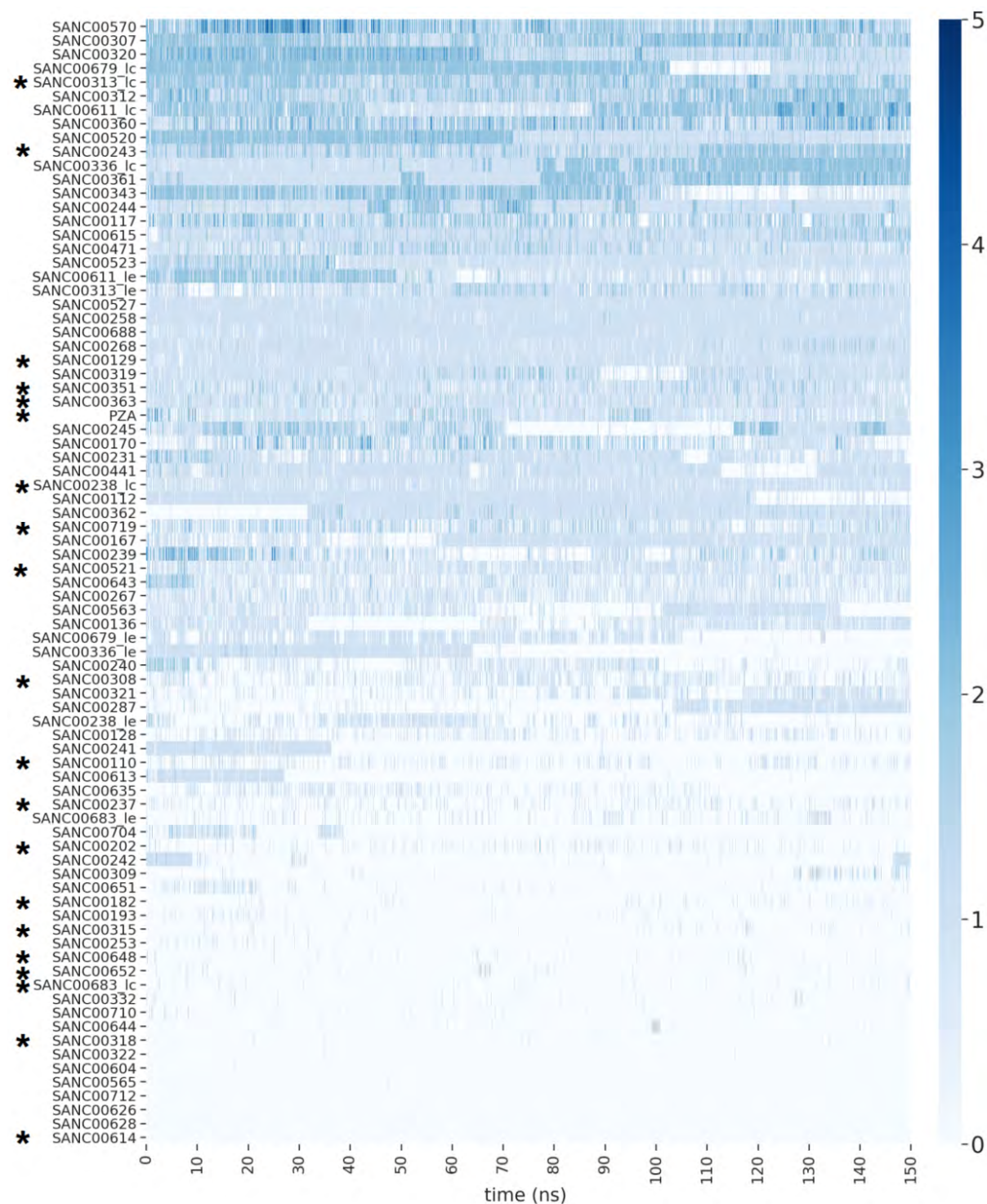


Figure 18: Heatmap of intermolecular hydrogen bond counts during production runs. Each system is denoted by the ligand name on the y-axis, with simulation time on the x-axis. Increasing colour intensity corresponds to a higher number of hydrogens bonds between ligand and PZase.

Also notable, is the large difference in common hydrogen bonding residues between the docking results and this MD study. During docking, as shown in **Figure 12**, the most common residues that participated in hydrogen bonding were ALA134, CYS138 and ASP8, present in 60%, 39% and 21% of systems respectively, while during MD simulations as shown in **Figure 17** the most common residues were ASP8, LEU19 and ALA102 at 31.3%, 11.6% and 5% respectively. The comparison is not perfect as overall residue occupancy for this MD study was determined by calculating the average % occupancy for each residue by summing the per-system occupancies over the 150 ns simulation duration, whereas the docking study produced binary data: either a hydrogen bond was present, or it was not. What is clear is that residue ALA134 and CYS138 displayed significant hydrogen bonding potential during the docking study, but did not do so during this MD simulation, meanwhile ASP8 and LEU19, the residues exhibiting the most H-bond potential during MD simulations had significantly less potential during the docking study. Part of the reason for this is the fundamental difference between molecular dynamics simulations and molecular docking performed in this study: during docking, a single protein conformation is used, in this case, the pose of the crystal structure, the orientation of the ligand is the factor that determines what hydrogen bonds will be present. However, during MD simulations everything in a system is in motion, ligand, water, ions, and protein; and only a fraction of the conformational ensembles produced during MD simulations will make certain hydrogen bonds favourable.

3.4.4 Protein RMSD

Shown in **Figure 19**, protein backbone RMSD was one metric used to evaluate PZase stability in each system, showing average deviation of protein structure from initial coordinates. RMSD values for all systems clustered between 0.10 and 0.20 nm over the 150 ns simulation time. This indicates that all systems, including the PZA control were stable for the duration of the 150 ns simulation time. Only one system, SANC00244, exhibited a slightly higher average RMSD at 0.25 nm. Additionally, most systems and a majority of stable systems (**Figure 19**, yellow) exhibit unimodal clustering, suggesting that the PZase structure converged on a single stable state for the simulation duration. However, a minority of systems such as SANC00170, SANC00231, SANC00343 and SANC00648 exhibited multimodal clustering, usually bimodal, which indicates the PZase structure may have converged into multiple stable states.

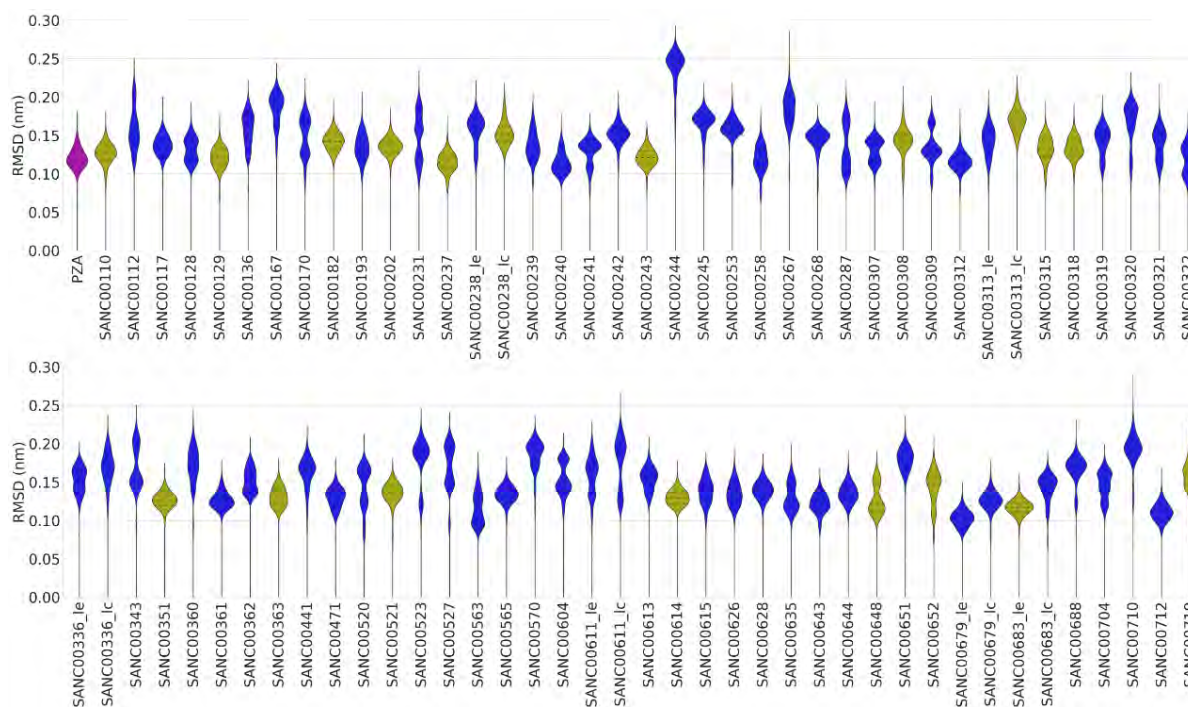


Figure 19: Violin plots of protein backbone RMSD results for 150 ns production MD runs. In purple: PZA-PZase control; blue: systems judged to be unstable based on ligand RMSD data; yellow: systems judged stable based on ligand RMSD data.

3.4.5 Radius of Gyration (Rg)

Net distribution and density of PZase structures was monitored by calculating Rg (radius of gyration) data for each 150 ns MD run. Since Rg is a measure of root mean distance of a collection of atoms from a common centre of mass, Rg describes the ‘compactness’ of the protein structure and therefore a protein’s stability in terms of whether it maintains its native structure or partially or even fully denatures during the simulation duration. A protein that denatures during an MD simulation would result in steadily increasing Rg value throughout the simulation, on a violin plot, this would result in the absence of clustering, with violin plots being stretched out. Significant changes in protein conformation can also typically be detected in Rg data as movement of a large number of atoms towards or away from the protein center of mass will alter Rg values. As shown in **Figure 20**, all PZase structures maintained Rg values of 2.54 ± 0.02 nm, indicating that all structures maintained a stable state for the duration of simulation time, with insignificant changes in overall Rg states.

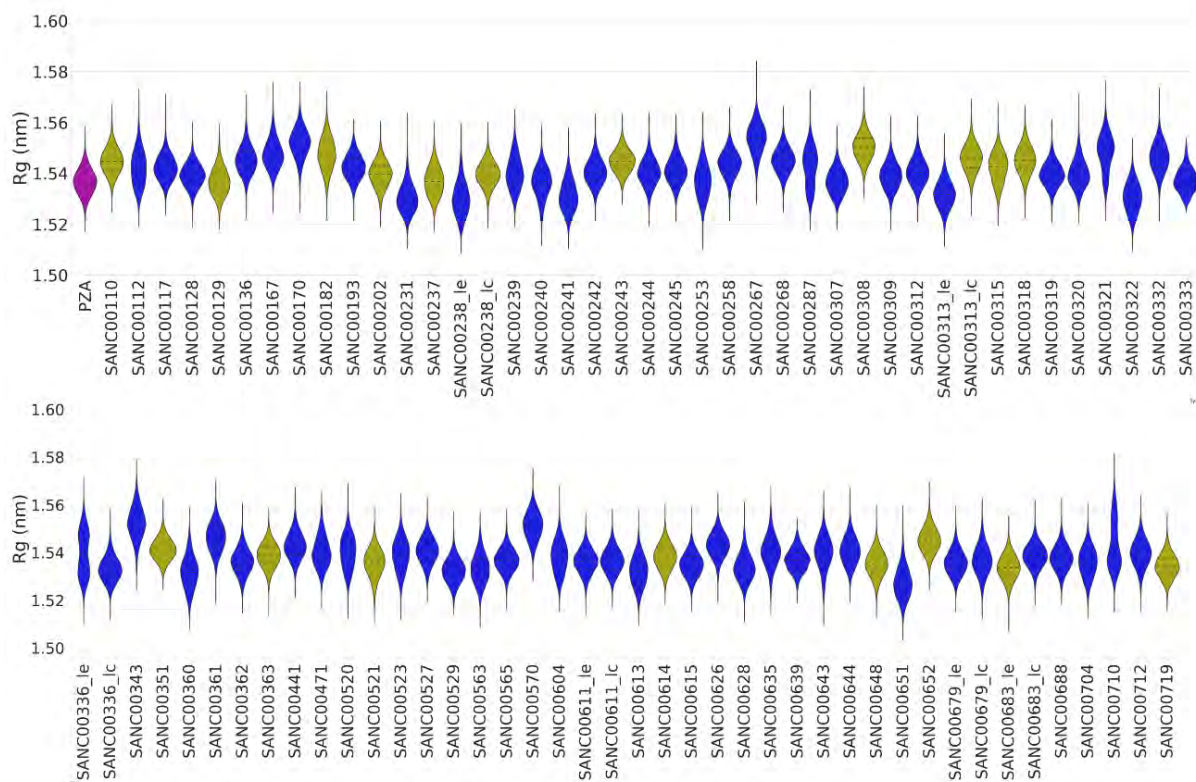


Figure 20: Violin plots of protein Radius of gyration (Rg) results for 150 ns production MD runs. In purple: PZA-PZase control; blue: systems judged to be unstable based on ligand RMSD data; yellow: systems judged stable based on ligand RMSD data.

Radius of gyration of the active site residues was also evaluated by selection of residues within 8 Å of the binding site of PZA. This was used to determine whether there were significant variations in active site conformation or topology between systems. As shown in **Figure 21**, absolute Rg values were approximately 0.5 nm lower than the Rg values shown in **Figure 20**. This is expected as the active site residues are all in the active site pocket buried within the PZase surface and are therefore closer to the enzyme's centre of mass than an average residue. Overall active site residues exhibited average Rg values of 0.975 ± 0.08 nm. This variation is larger than the variation observed for whole protein Rg shown in Figure 20, however this is largely due to outlier systems that show increased Rg variation over the majority of systems. This includes SANC00170, SANC00267, SANC00343 and SANC00710.

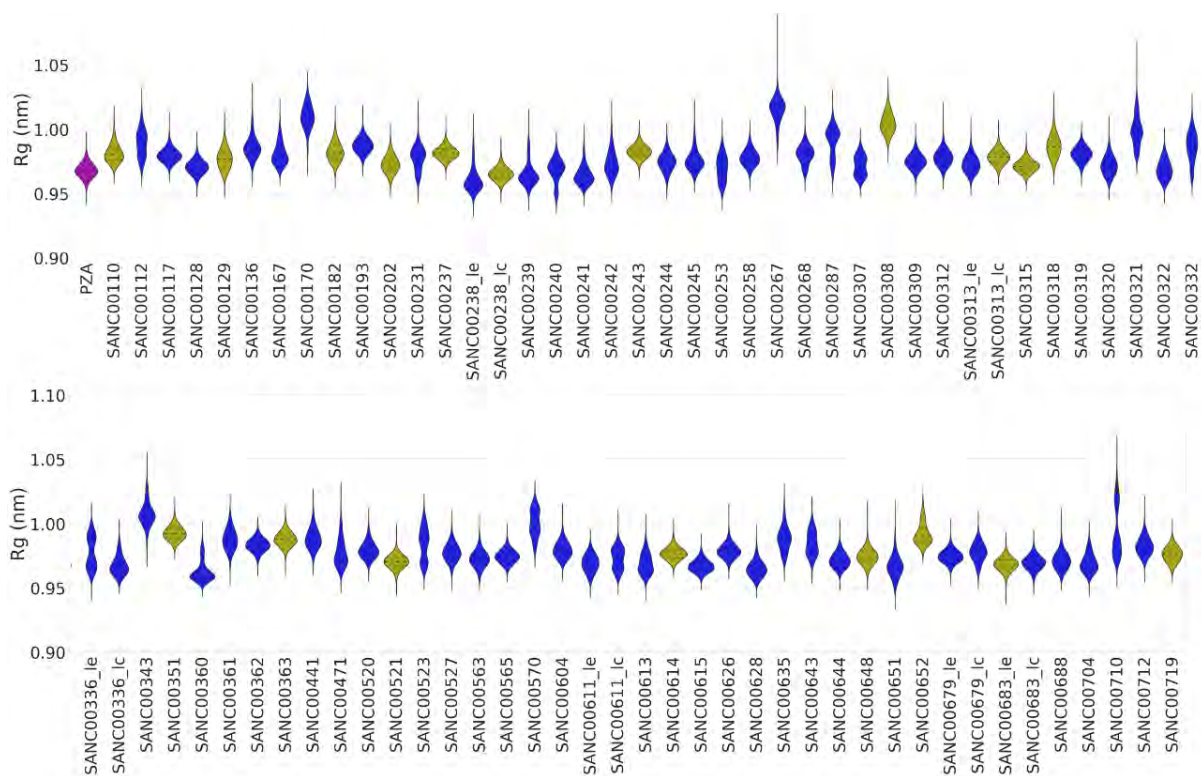


Figure 21: Violin plots of Rg of active site residue results for 150 ns production MD runs. In purple: PZA-PZase control; blue: systems judged to be unstable based on ligand RMSD data; yellow: systems judged stable based on ligand RMSD data.

3.4.6 RMSF

RMSF data was analysed to determine protein stability at the per-residue level. For this, RMSD data for the duration of the 150 ns simulation was averaged at the per-residue level to identify any regions of fluctuating residues (**Figure 22**). Regions that consistently exhibited increased fluctuations across most systems include several loop regions, as shown in **Figure 22C**. This is to be expected as structured regions such as α -helices and β -sheets typically exhibit more restricted motion than loops when in their native folded states. The most consistently flexible regions across all systems were residues 14-17 and 35-41, both of which are loop regions. Semi-flexible residues 59 – 65 and residue 53 are part of the loop region that forms the ‘lid’ controlling access to the PZase active site pocket, and accordingly exhibit a mild to moderate increase in RMSF.

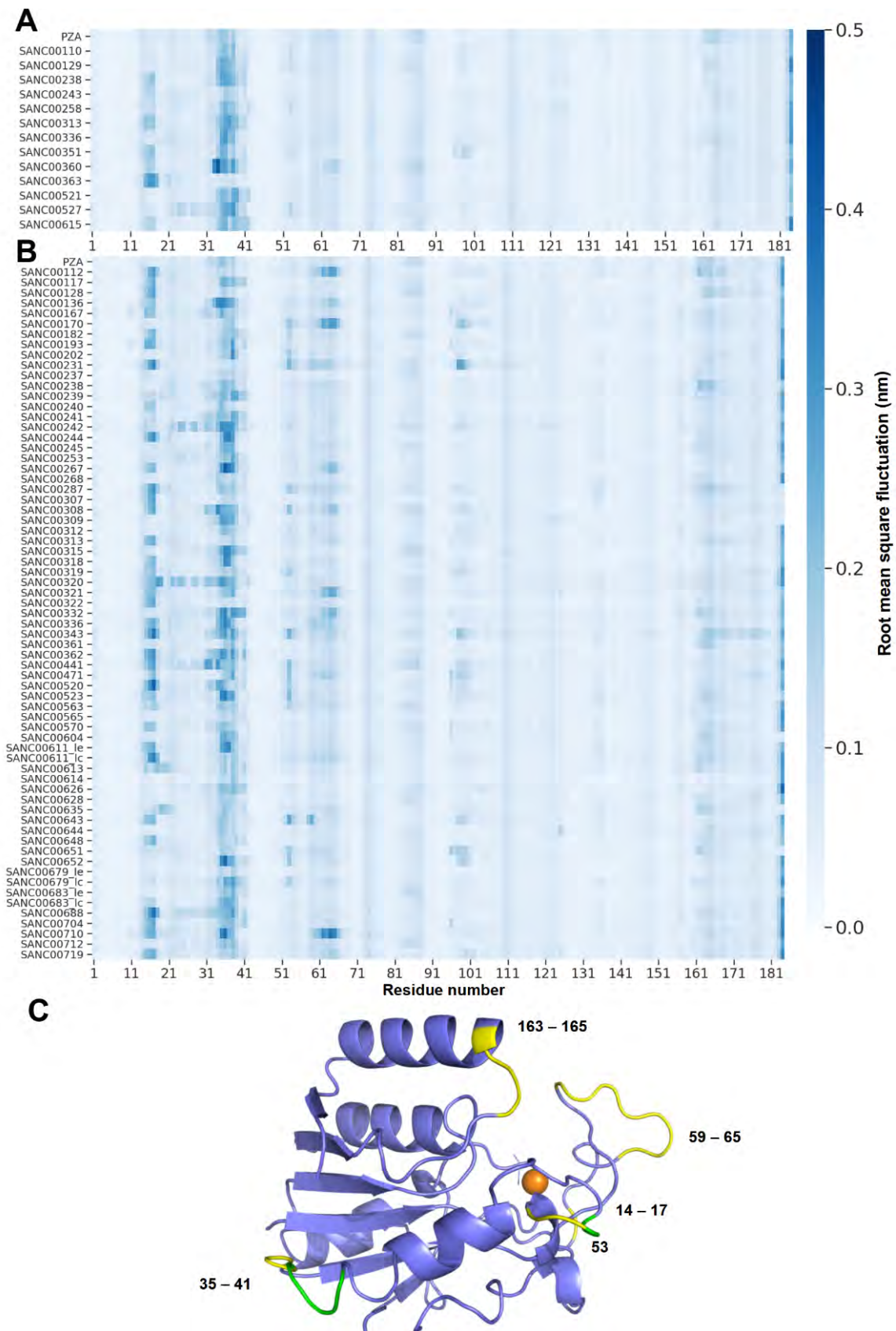


Figure 22: RMSF data for 150 ns production MD runs. (A) RMSF profiles of all systems selected for further screening during mutagenesis studies. **(B)** RMSF profiles of all other systems. **(C)** Crystal structure of PZase with regions of consistent higher RMSF values colored in yellow and green, adjacent labels indicate residue numbers.

3.4.7 Hit compounds

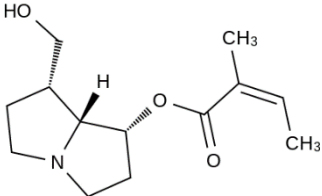
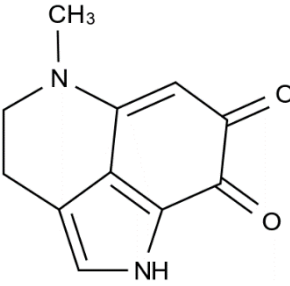
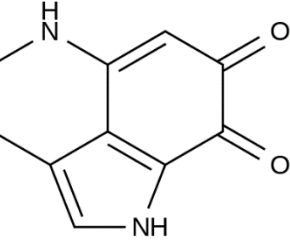
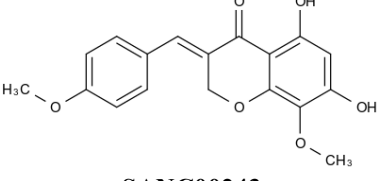
Ligand RMSD data generated from the 150 ns production MD runs indicated that there were approximately 20 ligands that remained stable and bound to PZase for the duration of the simulation timescale (**Figure 16**, yellow). This selection was based primarily on ligand RMSD as ligand stability and potential for sustained binding at the PZase active site is essential for any potential drug candidate.

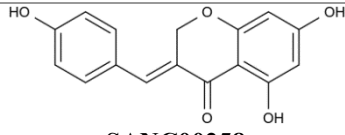
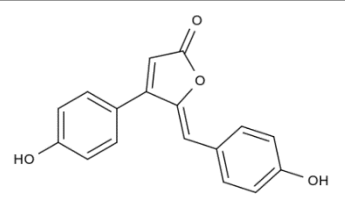
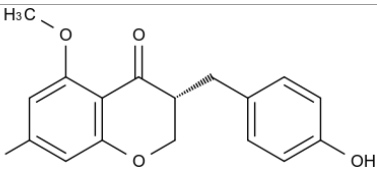
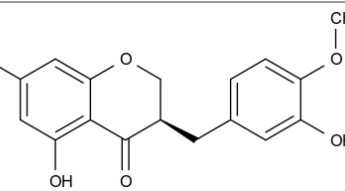
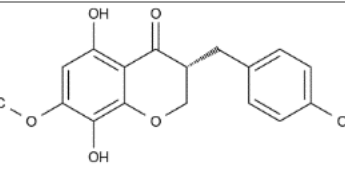
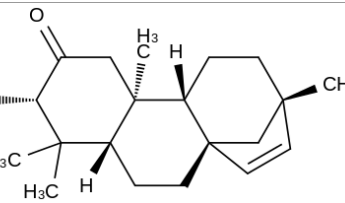
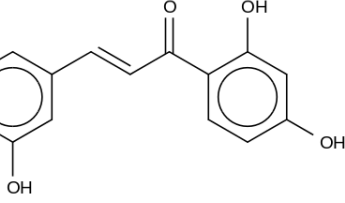
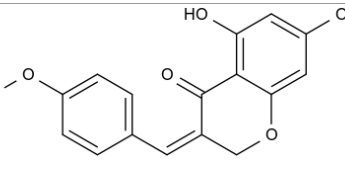
Table 2: Selected Ligands that remained bound to PZase during MD simulations. A particular H-bond was considered present if it had a >10% occupancy value.

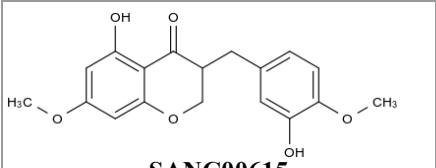
Compound	Docking binding Energy (kcal/mol)	Docking H-bonds	Mean MD ligand RMSD (nm)	MD H-bonds
PZA	-4.48	ASP8, ALA102, ILE133, CYS138	0.120	ASP8, ILE133, ALA134
SANC00110	-5.46	ALA134	0.153	ALA134
SANC00129	-5.8	ASP8	0.0763	ASP8
SANC00182	-6.75	ALA134, CYS138	0.099	ALA134
SANC00202	-5.53	ALA134, CYS138	0.085	ALA134
SANC00237	-9.31	ALA134, CYS138	0.093	ALA134, CYS138
SANC00238	-5.42	ASP8	0.099	ASP8
SANC00258	-7.45	ASP8	0.159	ASP8
SANC00308	-7.59	ALA134, CYS138	0.090	LYS96
SANC00313	-7.02	ALA20, ALA134	0.072	ASP8
SANC00315	-7.28	ALA20, ALA134	0.078	--
SANC00318	-7.12	ALA134	0.114	--
SANC00336	-6.70	ASP8, TYR64	0.190	ASP8
SANC00351	-6.82	ALA134, CYS138	0.097	ALA102
SANC00360	-7.21	ALA134, CYS138, ALA20	0.165	LEU19, TRP68, LYS96
SANC00363	-8.16	ASP8, ALA134	0.093	ASP8, ALA134
SANC00521	-6.55	ALA102, ILE133, ARG140	0.103	LEU19
SANC00527	-7.00	ILE133, ALA134, CYS138	0.122	ASP8
SANC00614	-7.50	ALA134	0.122	--
SANC00615	-6.66	ALA20	0.202	ASP8, ALA134
SANC00648	-7.13	ALA134	0.106	--
SANC00652	-5.89	ALA134, CYS138	0.094	--
SANC00683	-5.99	ALA134	0.152	--
SANC00719	-5.71	ALA134, CYS138	0.124	ALA134, CYS138

Given time considerations, compounds with no significant H-bonding during MD simulations were removed from the selection. Notable in **Table 2** is the wide range of binding energy values during docking simulations, many of these compounds would not have been selected for MD simulations if the criteria for binding energy had been a more conventional -8 kcal/mol instead of -5 kcal/mol, despite being stable during MD simulations. As stated previously the H-bonding interactions between docking and MD simulations are also different for many compounds such as SANC00308, SANC00313, SANC00351 and SANC00652, among others.

Table 3: Characteristics of final selected ligands used for the following mutagenesis studies.

Compound	Compound Name Classification	Source Organism	Uses	References
 SANC00110	7-Angelylplatynecine Alkaloid Pyrrolizodine MW: 239 Da		N/A	(Logie, 1996)
 SANC00129	Damirone B Alkaloid Damirone Pyrroloiminoquinone MW: 202 Da	Latrunculia lorii	N/A	(Antunes, 2003)
 SANC00238	Damirone C Damirone Pyrroloiminoquinone MW: 188 Da	Strongylodesm a aliwaliensis	Anticancer activity	(Whibley <i>et al.</i> , 2005)
 SANC00243	4'-O-Methyl-punctatin Flavonoid Homoisoflavonoid MW: 328 Da	Eucomis autumnalis	N/A	(Sidwell and Tamm, 1970)

 <p>SANC00258</p>	<p>4'-Demethyl-eucomin Homoisoflavonoid MW: 284 Da</p>	<p>Eucomis punctata</p>	<p>N/A</p>	<p>(Finckh and Tamm, 1970)</p>
 <p>SANC00313</p>	<p>Rubrolide E Furanone MW: 280 Da</p>	<p>Synoicum globosum</p>	<p>N/A</p>	<p>(Sikorska <i>et al.</i>, 2012)</p>
 <p>SANC00336</p>	<p>4'-Demethyl-5-O-methyl-3,9-dihydroeucomin Homoisoflavonoid MW: 300 Da</p>	<p>Eucomis punctata Eucomis montana Eucomis comosa</p>	<p>N/A</p>	<p>(Finckh and Tamm, 1970)</p>
 <p>SANC00351</p>	<p>5,7-Dihydroxy-3-(3'-methoxybenzyl)-4'-chromanone Homoisoflavonoid MW: 316 Da</p>	<p>Resnova humifusa</p>	<p>N/A</p>	<p>(N. A. Koorbanall <i>et al.</i>, 2006)</p>
 <p>SANC00360</p>	<p>8-O-Demethyl-7-O-methyl-3,9-dihydropunctatin Homoisoflavonoid MW: 316 Da</p>	<p>Eucomis punctata Eucomis montana</p>	<p>N/A</p>	<p>(N. A. Koorbanall <i>et al.</i>, 2006)</p>
 <p>SANC00363</p>	<p>Ent-3β-hydroxy-beyer-15-ene-2-one Diterpene Terpenoid MW: 302 Da</p>	<p>Spirostachys africana</p>	<p>Antibacteri al activity</p>	<p>(Mathabe <i>et al.</i>, 2008)</p>
 <p>SANC00521</p>	<p>(E)-3,2',4'-Trihydroxychalcone Flavonoid MW: 256 Da</p>	<p>Galenia africana</p>	<p>N/A</p>	<p>(Mativandl ela <i>et al.</i>, 2009)</p>
	<p>(Z)-Eucomin Homoisoflavonoid MW: 298 Da</p>	<p>Eucomis comosa</p>	<p>N/A</p>	<p>(C. Koorbanall <i>et al.</i>, 2006)</p>

SANC00527				
 <p data-bbox="336 436 480 465">SANC00615</p>	<p data-bbox="639 286 927 416">5-Hydroxy-7-methoxy-3-(3-hydroxy-4-methoxybenzyl)chroman-4-one</p> <p data-bbox="639 450 863 479">Homoisoflavonoid</p> <p data-bbox="639 517 799 546">MW: 330 Da</p>	Scilla nervosa	N/A	(Bangani, Crouch and Mulholland, 1999)

Shown in table 3 are the properties of the 13 final compounds selected for further analysis in the following PZase mutant study. Nearly all of them except for SANC00238 and SANC00363 have no known uses or specific activities. Also notable is that many of them have significant structural similarities, with SANC00129 and SANC00238 sharing a common structure and SANC00243, SANC00258, SANC00336, SANC00351, SANC00360 and SANC00527 also share a common structure. Finally, PZA is a small molecule with a MW of 123 Da, while the hit compounds identified in this chapter have MW values 2-3 times larger, as shown in Table 3. Given the dimensions of the PZase active site pocket (approximately 10 Å deep and 7 Å wide), the increased size of these compounds may hinder access to the active site in vitro to a greater extent than is evident during this in silico study.

3.5 CONCLUSION

As described previously, prediction of protein-ligand for the purposes of rational drug design is challenging and computationally expensive, many docking protocols rely on shortcuts that come at the expense of accuracy, such as simulating the protein receptor as a rigid structure. MD simulations assist with this shortcoming by allowing for the precise simulation of the conformational dynamics between protein and ligand over physically relevant timescales. In this chapter, 238 ligand poses were subjected to 20 ns MD simulations with the PZase enzyme in order to discard immediately unstable ligands, this was followed by 83 150 ns simulations to identify stable ligands. Regrettably, system SANC00307_le was erroneously not selected for further 150 ns MD simulations despite it passing the 0.2 nm median ligand RMSD stability criterion. Ligand potential was evaluated using ligand RMSD data as well as H-bonding interactions between ligand and PZase. Protein RMSD, Rg and RMSF were used to evaluate the PZase protein stability and dynamics over the simulation period and to identify any local points of increased fluctuation. In total there were 13 ligands identified that showed significant potential primarily due to ligand RMSD data, many of which fell into two distinct common

structures. The next chapter will be focused on investigating whether these ligands exhibit similar behaviour against mutant PZase enzymes, that induce PZA resistance in Mtb.

CHAPTER 4

MUTANT PZASE MOLECULAR DYNAMICS SIMULATIONS

4.1.1 INTRODUCTION

Worldwide, an estimated 500 000 cases of MDR/RR-TB were reported in 2019, a 10% increase compared to 2018 (WHO, 2020a), this is alarming as there is likely a significant number of unreported MDR/RR-TB cases worldwide, and only 57% of MDR-TB infected people are successfully treated. The increasing incidence of MDR/RR-TB, as well as XDR/TDR-TB poses a significant threat to the control of TB globally, but particularly in the Global South, where TB burden is significantly higher. Since PZA is such a vital drug for the treatment of both drug-susceptible and drug-resistant TB, it is imperative that insights into its MoA are uncovered, and novel potential drug candidates are discovered.

4.1.2 Mutations in the *pncA* gene and PZA resistance

The gene encoding the Mtb PZase enzyme, *pncA*, was first associated with PZA resistance by Scorpio and Zhang in 1996, and shortly after that mutations in the *pncA* were found to be responsible for 70-97% of cases of PZA-resistant Mtb infections (Hirano *et al.*, 1998). Although not all cases of PZA-resistance have been attributed to mutated *pncA*, recombinant studies using mutated *pncA* have shown that mutated enzymes can reduce PZA to POA enzymatic activity up to 10 times depending on the type of mutation and its location(s) in the protein (Cheng *et al.*, 2000; Sheen *et al.*, 2009). Notable residues include those at the active site: Asp8, Ile133, Ala134 and Cys138; all of which, when mutated, have significant effects on PZase activity (Sheen *et al.*, 2009; Petrella *et al.*, 2011c). Although not all PZA resistance can be attributed to mutations in *pncA*, there is a very high correlation between PZA resistance and *pncA* mutations in South Africa (Louw *et al.*, 2006; Mphahlele *et al.*, 2008).

4.1.3 Identification of SANCDB compounds that bind to mutant PZase

In the previous chapter, SANCDB compounds that favourably bind to PZase during MD simulations were identified and described. In this chapter, these identified compounds will be screened against mutant PZase structures in order to identify any compounds that are less affected by PZase mutations than PZA is. A total of 13 high confidence missense mutations (**Figure 23**) were retrieved from previous work done by Sheik Amamuddy *et al.*, 2020, who originally obtained mutation data from the TB Drug Resistance Mutation Database (Sandgren *et al.*, 2009). Sheik Amamuddy *et al.* classified the mutations according to their distance from the PZA ligand's COM: Group 1, green spheres are ≤ 6.7 Å from COM; Group 2, red spheres are ≤ 11 Å from COM; Group 3, yellow spheres are ≥ 11 Å from COM and finally Group 4, purple spheres formed part of the MBS of PZase. This was done to determine whether distance from active site was a factor that affected a mutation's effect. For this study, 13 point mutations were chosen with a few from each group, listed below in **Table 4**

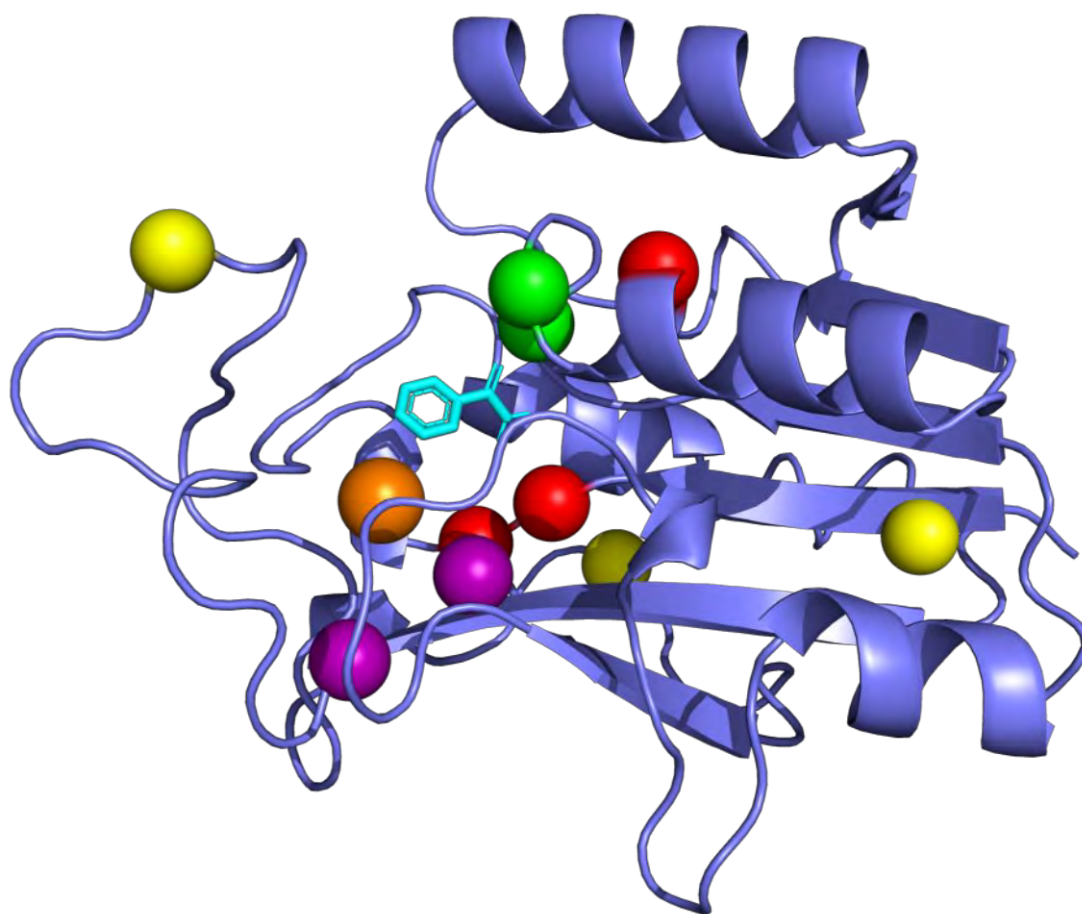


Figure 23: Point mutations within the PZase protein used during this screening study. PZA ligand shown in cyan within the active site pocket; FE^{2+} atom shown in orange. Group 1: green; Group 2: red; Group 3: yellow; Group 4: purple.

Table 4: PZase point mutations chosen for this study. Mutations originally obtained from the TB drug resistance database (Sandgren *et al.*, 2009) and studied by Sheik Amamuddy *et al.*, 2020

Mutant	Group
A134V	1 (≤ 6.7 Å from active site)
H137R	
D8G	2 (≤ 11 Å from active site)
Q10P	
Y103S	
V139M	
R140S	
A3P	3 are ≥ 11 Å from active site)
T61P	
D63G	
L85R	
L116R	
A146V	
D49A	4 (MBS residues)
D49G	
H51P	

4.2 METHODOLOGY

The overall methodology for this study was very similar to the methodology used for the previous wild-type MD study. The only significant differences were the added step of mutagenesis where PZase structures were mutated and modifications to Fe^{2+} force field parameters where a mutation changed the coordination environment surrounding the Fe^{2+} atom.

4.2.1 Mutagenesis

Point mutations were introduced into the PZase crystal structure (PDB ID: 3PL1) using the residue editing tool built into Discovery Studio Visualizer (BIOVIA, 2021). Mutated structures were then deprotonated using the *Reduce* tool (Word *et al.*, 1999). The de-protonated structures were then protonated by the H++ webserver (Anandakrishnan, Aguilar and Onufriev, 2012) at a pH of 7 with a salinity of 0.15 M and default dielectric parameters, from here on the methodology used was identical to that of the previous chapter. Energies of mutated structures were minimized during the energy minimization step of the MD simulations, this was essential to eliminate any unfavourable geometries and steric clashes that may have been introduced during mutagenesis,

4.3 RESULTS AND DISCUSSION

4.3.1 Ligand RMSD

Like the previous chapter, ligand RMSD was the primary method used to determine ligand stability, as shown in **Figure 24**, unlike chapter 3 however, all plots in this chapter have been split so that each ligand tested has its own subplot, and each subplot characterizes a ligand across all mutant PZase structures tested.

Criteria for determining stability via ligand RMSD data was identical to that used in chapter 3, only systems that exhibited unimodal clustering below 0.2 nm RMSD are considered stable, with small exceptions made if clustering occurred on the 0.2 nm boundary with significant levels of hydrogen bonding. As expected, PZA performed poorly, only being stable in 2 out of

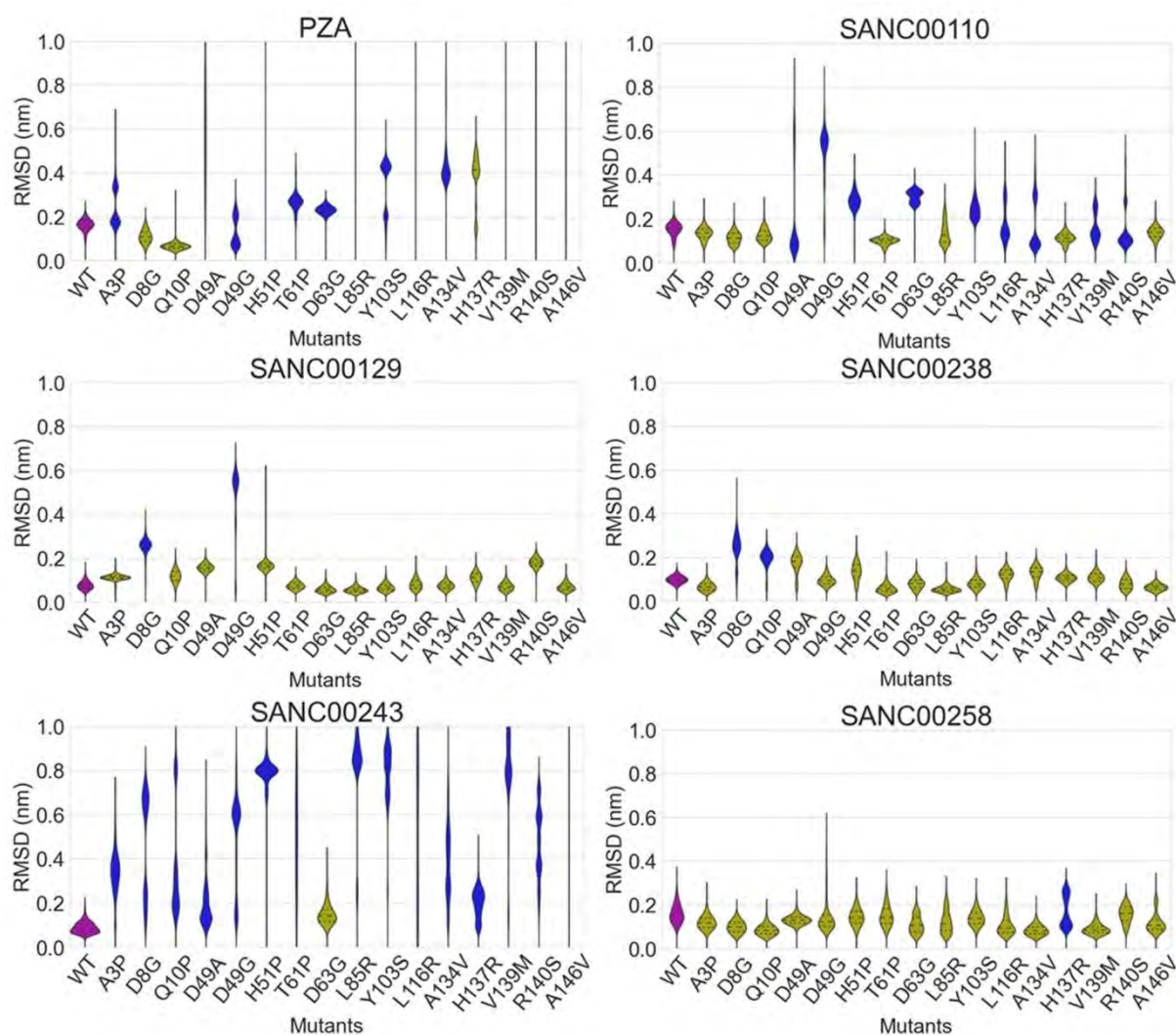


Figure 24: Violin plots of Ligand RMSD results for 150 ns production MD runs for the PZase mutant study. In purple: WT control; blue: ligands judged to be unstable; yellow: ligands appearing stable.

the 13 mutant systems, this is somewhat expected, as the majority of these mutations are detected in PZA resistant isolates of Mtb (Sandgren *et al.*, 2009). D8G and Q10P, the two systems where PZA was stable are both group 2 mutations. Ligands SANC00243, SANC00313, SANC00360, SANC00521, SANC00615 performed similarly poorly, only being stable in 1-4 mutant systems, suggesting that they would not be promising candidates against PZase mutants. SANC00110 and SANC00336 showed slightly more promise being stable in 7 and 6 systems, respectively. The 6 ligands that were stable in more than half of systems were SANC00129, SANC00238, SANC00258, SANC00351, SANC00363, SANC00527, with the most promising systems: SANC00129, SANC00238 and SANC00258 being stable in 14,14 and 15 systems, respectively. Supplementary Figures S1-S4 show protein backbone RMSD, protein Rg, active site Rg and protein RMSF respectively for all mutant system MD simulations. Protein RMSD data (**Figure S1**) showed that there were no systems rendered significantly more unstable due to introduced point mutations. Protein Rg data, shown in **Figure S2** indicates a similar situation, with the exception of D49A and D49G systems, Rg values remained unimodally clustered and very similar to those of WT PZase. Active site Rg data, shown in **Figure S3** is largely the same: active site residues in a majority of systems were comparable to WT systems. The exception to this were systems with the D49A and D49G mutations, since Asp49 is part of the protein MBS, it is likely a point mutation of this residue significantly alters the stability of the MBS as well as the active site as a whole. In vitro investigations of mutations of PZase MBS residues show that they often abolish nicotinamidase activity and protein stability (Petrella *et al.*, 2011a). A similar pattern is observed in protein RMSF data, shown in **Figure S4**. RMSF patterns largely correspond to that of the WT systems, however systems with the D49G and D49A residues.

4.3.2 Hydrogen bonding analysis

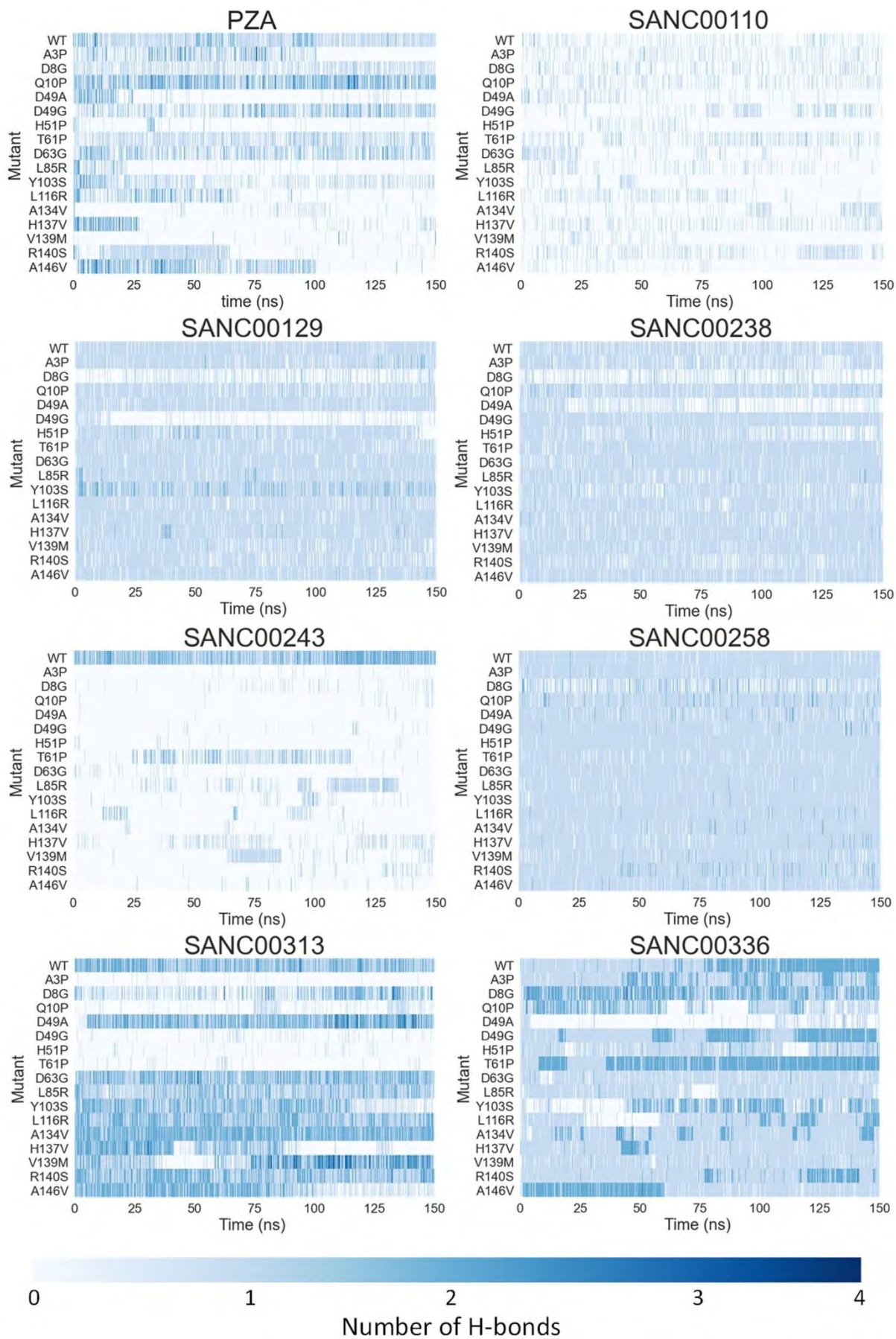
Hydrogen bonding data, shown in **Figures 25 and 26**, reveals that the most common residue participating in hydrogen bonding with ligands is the active site residue Asp8. In approximately eight ligands, Asp8 shows significant hydrogen bonding in the majority of mutants, with consistently high occupancy ranges, often in the 80-100% range. Being part of the catalytic triad, Asp8, is hypothesized to facilitate the nucleophilic attack of Cys138 on the carbonyl atom of PZA, forming the intermediate acyl-enzyme complex. Asp8 also facilitates water-mediated hydrolysis of the acyl-intermediate by acting as a general base, completing the conversion of PZA into POA (Petrella *et al.*, 2011c). Asp8 is an even more important hydrogen-bonding residue for SANCDB compounds that exhibit the most favorable binding across a wide range

of mutants: SANC00129, SANC00238, SANC000258 and SANC00527 all show significant hydrogen bonding with most mutants. Occupancy values are almost always > 90% and there is very little hydrogen bonding to other residues in any of the systems.

Table 5: Summary of ligand stability in mutant systems and H-bonding residues.

Compounds shaded in grey showed the most favorable binding pattern in a majority of mutant systems.

Compound	No. stable systems	H-bonding residues
PZA	2	Ile133, Ala134, Cys138
SANC00110	7	Ile133, Ala134, Cys138
SANC00129	14	Asp8
SANC00238	14	Asp8
SANC00243	1	--
SANC00258	15	Asp8
SANC00313	4	Asp8
SANC00336	6	Asp8, Ala102
SANC00351	9	Asp8, Ala102
SANC00360	1	Leu19
SANC00363	10	Asp8, Ala134
SANC00521	4	Asp8, Leu19
SANC00527	12	Asp8
SANC00615	3	Asp8, Leu19



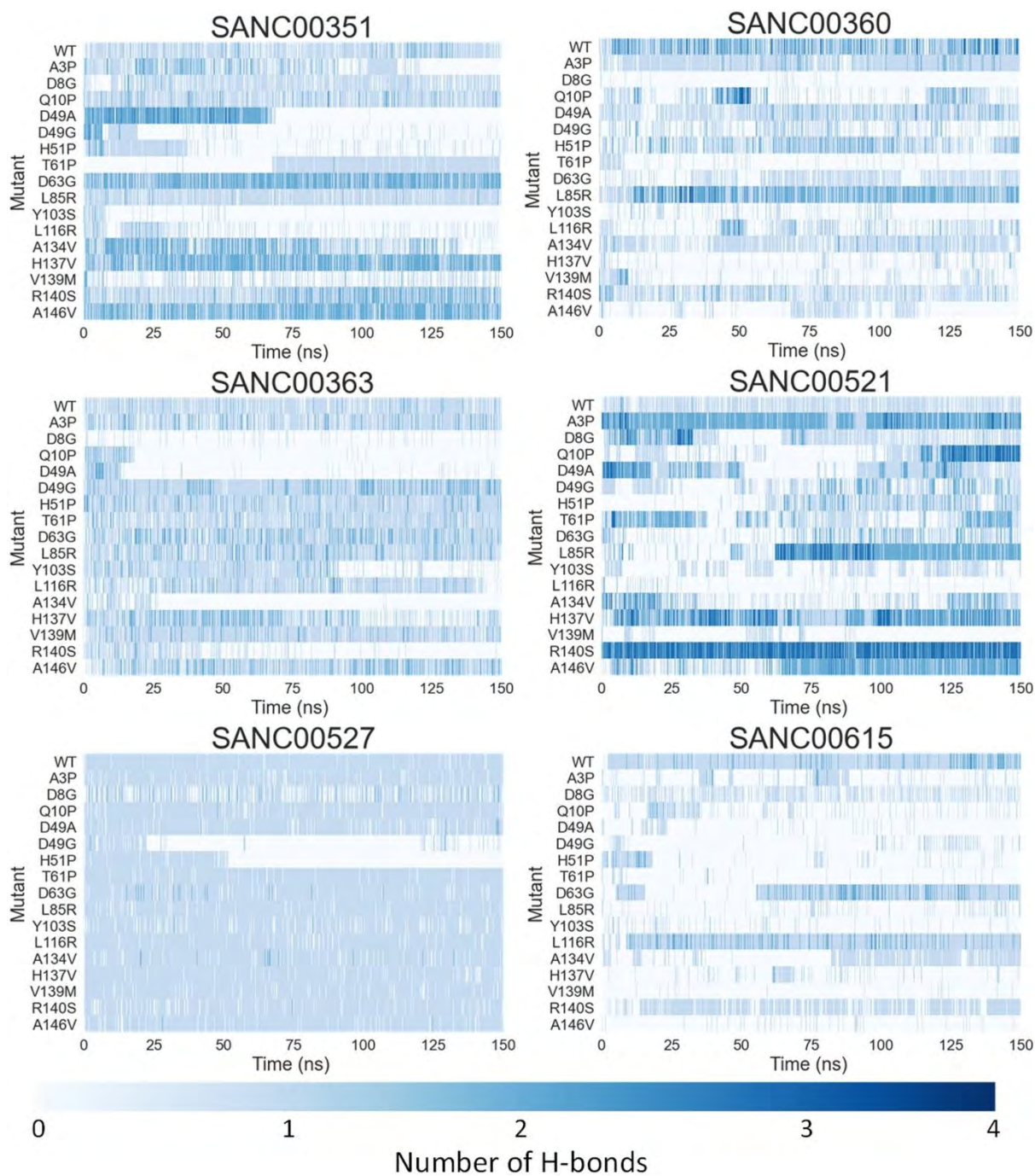


Figure 26: Heatmap of intermolecular hydrogen bond counts during production runs. Each system is denoted by the mutant on the y-axis, with simulation time on the x-axis. Increasing colour intensity corresponds to a higher number of hydrogens bonds between ligand and PZase.

4.3.3 Mutations and compound stability

Point mutations can have significant effects on protein structure, substrate specificity and enzymatic activity, (Rodríguez-Zavala, 2008; Ishida, 2010) and therefore play roles in the development of drug resistance in many diseases (van Doorn *et al.*, 2003; Bohnert *et al.*, 2007).

In this study, 3 residues in the active site pocket were mutated: 2 MBS residues (mutations D49A, D49G and H51P) and 1 residue that was part of the catalytic triad (D8G). The other 12 mutations are located outside of the active site cavity, and so any effects they may have on protein activity and function are not due to any direct alteration of the active site –although they may very well still influence the active site via allosteric mechanisms. The 3 residues in the active site that were mutated (D8G, D49A, D49G, H51P) all resulted in the most significant reductions in the number of stable compounds compared to the other mutations, as shown in **Figure 27C**. In addition to being missense mutations, the substituted amino acids are all

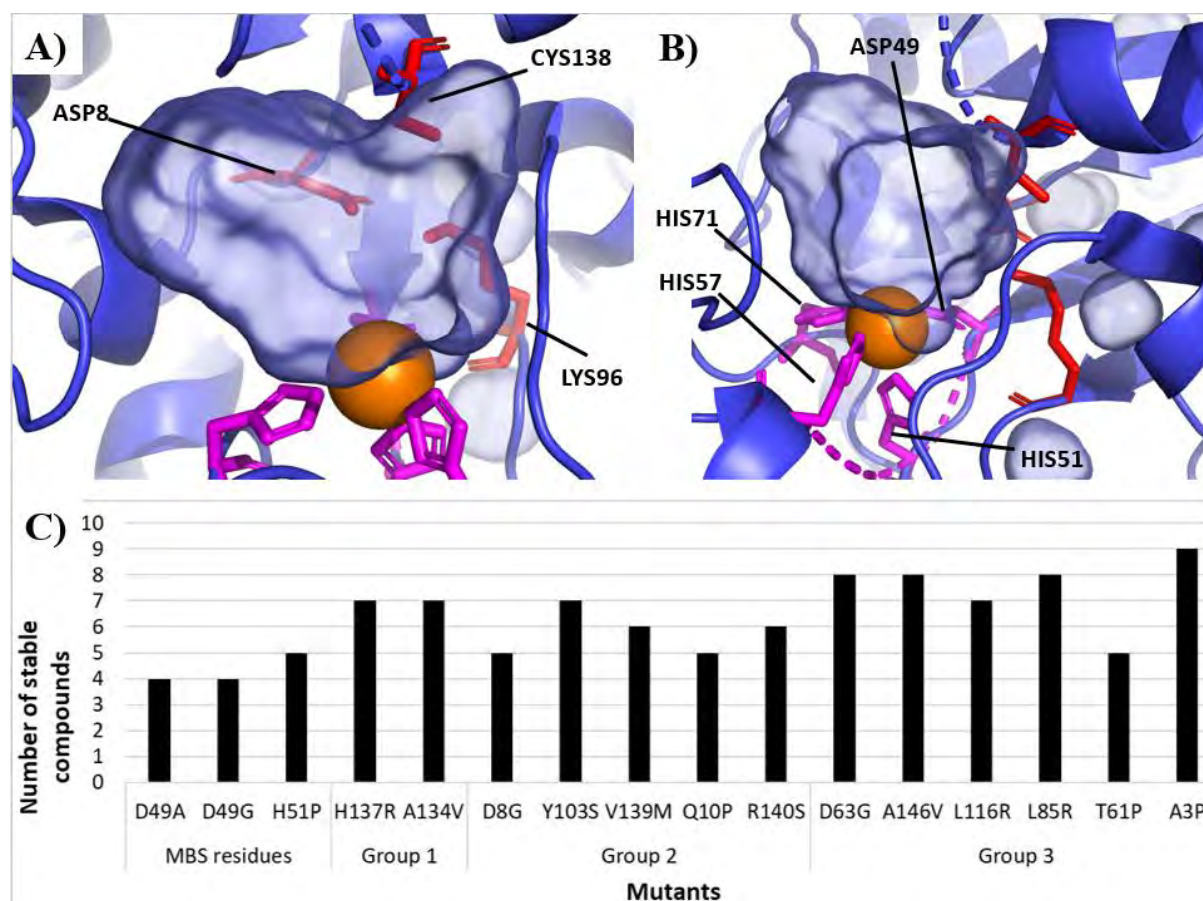


Figure 27: Mutations within the active site pocket resulted in few stable compounds. A) and B) are hybrid representations of the active site cavity of PZase/pncA, rotated 90°. In red are the active site residues forming the catalytic triad, in purple are the MBS residues that coordinate with the Fe²⁺ atom, shown in orange. (C) Number of stable compounds for each set of mutant systems, grouped according to approximate distance to the binding site of PZA within the active site cavity. 3D diagrams generated using PyMOL.

biochemically distinct from the WT residues they replace. Asp8, an acidic residue essential for catalysis in numerous enzymes due to its ability to provide a carboxylate group (Buller and Townsend, 2013), is replaced with glycine, the smallest amino acid with no side chain and therefore no polar groups. Asp49 is replaced with glycine or proline, both small, aliphatic, and non-polar residues. His51, a basic residue with an imidazole group is replaced with proline, a residue with neither of these features. Both Asp49 and His51 are coordinating residues of the Fe^{2+} atom and are essential for the proper coordination geometry of the metal ion. Therefore, these mutations will lead to alterations of the ion's coordination geometry. As all these mutations perturb either the MBS or catalytic triad, it is expected that they would result in significant structural alterations to the active site itself (**Figure 27A & 27B**) and therefore binding affinity. In vitro assays on heterologously expressed PZase with the D8G, D49G and H51P mutations have shown that mutations at these positions abolish PZase activity or, in the case of the D8G mutation, cause the protein to precipitate out of solution (Petrella *et al.*, 2011b), likely due to protein denaturation and exposure of hydrophobic residues to the bulk solvent.

Mutations outside of the active site can influence protein function and enzymatic activity in two ways: via allosteric effects as mentioned above where a mutation occurs at an allosteric site (Qiang *et al.*, 2017; Mishra *et al.*, 2018), thereby altering binding behaviour of any effector molecules (or drugs that target the allosteric site); or by disrupting protein structure and dynamics to such an extent that the active site of a protein is functionally altered (Shakhnovich

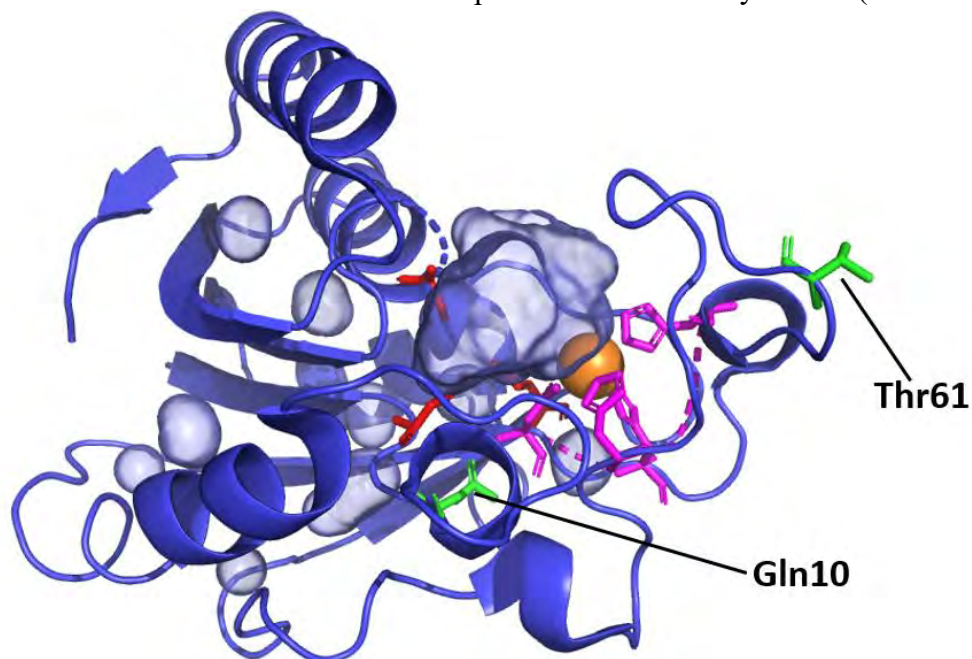


Figure 28: Locations of two mutations outside of the *pncA* active site. Despite being outside the active site cavity, both mutations introduced here resulted in a significant reduction in compound stability. Diagram generated using PyMOL.

and Gutin, 1991; Shanthirabalan *et al.*, 2018). A notable characteristic of the *pncA* gene is the diversity in mutations that confer PZA resistance (Petrella *et al.*, 2011b), which occur along the length of the *pncA* gene. The mutations investigated in this study are all point mutations, but there are numerous insertions, deletions and frameshift mutations that have all been documented to confer PZA resistance as well (Aono *et al.*, 2014; Khan *et al.*, 2019). As shown in **Figure 27C**, mutations in the active site pocket (specifically residues part of the catalytic triad and MBS coordinating residues) appear to have the most significant effect on the number of stable compounds at the active site. However, mutations outside of the active site cavity also reduced the number of stable compounds at the active site. Two mutations, Q10P and T61P, shown in **Figure 28**, resulted in only five of the 14 compounds tested being stable in the active site. Gln10 in the WT protein, while still being relatively close to the active site cavity is situated beneath it, within the interior of the protein. Thr61 on the other hand is located further from the active site cavity being one of the outermost residues on the flexible loop/lid region of the protein. Once again, both of these mutations, result in the substitution of a biochemically distinct amino acid, proline. In the case of the Q10P mutation, tertiary protein structure is formed by networks of disulphide bonds, hydrogen bonds, ionic bonds, and non-polar hydrophobic interactions and such a mutation may disrupt this network. Gln10, with its amide functional group, has a high propensity to form H-bonds, important for both protein secondary and tertiary structure, a property that proline does not share. The Thr61 residue on the other hand is located on the exterior, solvent-facing surface of the protein and as such likely interacts with water molecules via its polar hydroxyl group. Proline, being an aliphatic residue has no such polar group and is therefore a hydrophobic residue, exposure of a hydrophobic residue to the water molecules may alter the behaviour of the flexible loop region that controls access to the active site cavity. Other mutations introduced have a smaller but still considerable effect on compound stability, since they do not form part of the active site cavity, they likely influence compound stability indirectly, by altering the networks of intramolecular interactions responsible for protein secondary and tertiary structure.

4.3.4 Top Hits

Four clear hit compounds, that bound favourably to both WT PZase and the majority of PZase mutants were identified, as shown in **Figures 27 and 28**. However these 4 hits are actually just two sets of analogues: SANC00129 and SANC00238 share an identical chemical structure except for SANC00129's addition of a methyl group onto the nitrogen atom of the pyridine ring. The near identical structures also produce very similar hydrogen bonding patterns, with both compounds exhibiting consistent hydrogen bonding with Asp8, in the D8G mutated system, this hydrogen bond is no longer possible and both compounds become unstable in these systems. SANC00129 also exhibits significantly reduced hydrogen bonding to Asp8 in the D49G mutant system and subsequently is also unstable here, this is however not the case with

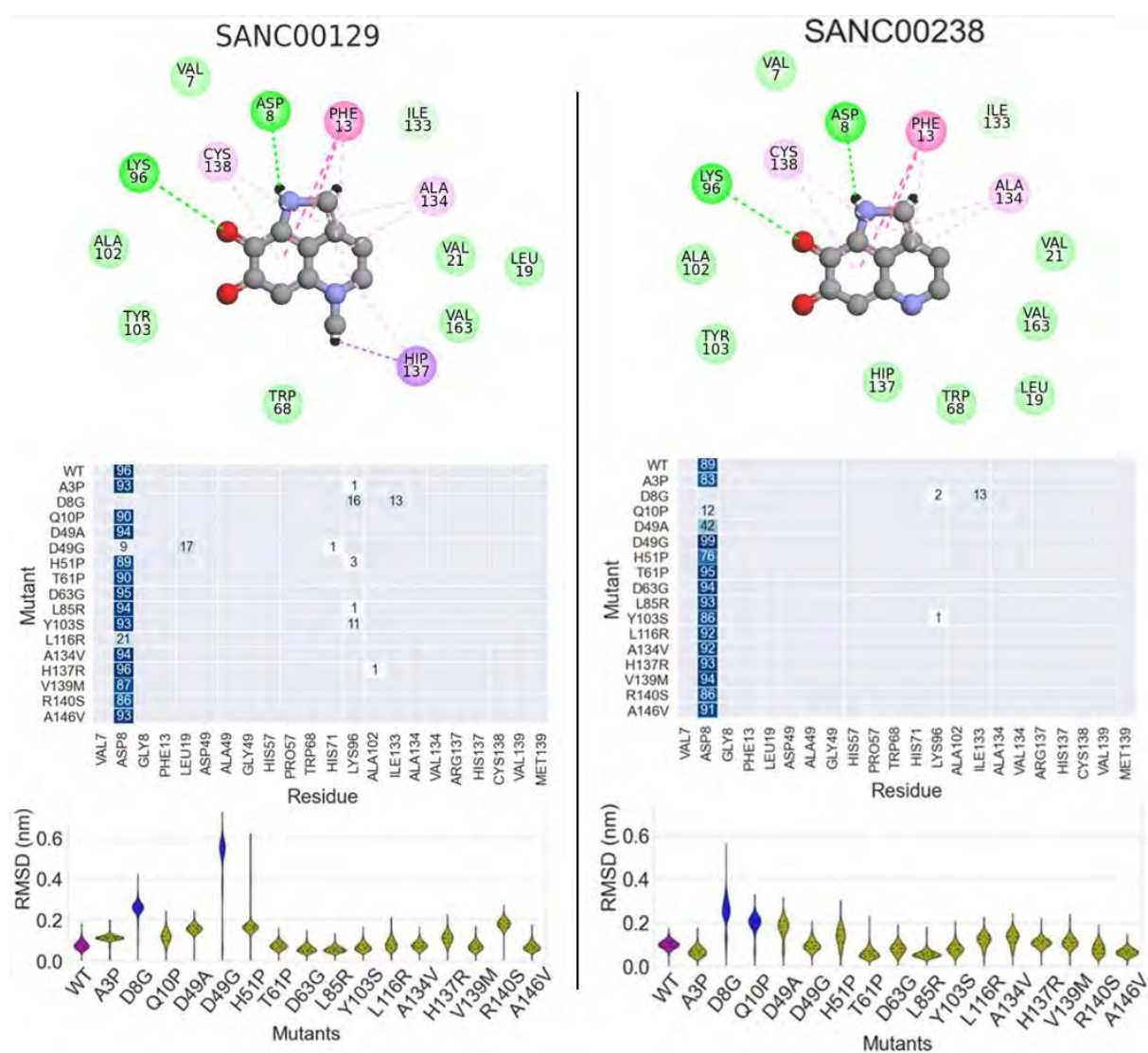


Figure 29: First pair of analogous hits. SANC00129 and SANC00238 performed favourably against WT PZase as well as the majority of PZase mutants. Sub-figures for each ligand correspond to 2D structure and docking interactions, hydrogen bonding during MD simulations and ligand RMSD, respectively.

SANC00238 which maintains the Asp8 hydrogen bond in the D49G system but loses the majority of this interaction in the Q10P system and therefore becomes unstable. Interestingly, SANC00238 does exhibit significantly reduced hydrogen bonding to Asp8 in the D49A mutant system, unlike SANC00129. A possible explanation for this, as highlighted previously is that Asp49 is one of the residues forming the Fe²⁺ MBS and as such a mutation in this residue likely destabilizes the Fe²⁺ atom or at the very least alters its coordination geometry, both scenarios could lead to changes in the topology of the active site, disrupting ligand stability. Curiously, SANC00129 exhibited significantly reduced Asp8 hydrogen bonding in the L116R system, but this did not have a detrimental effect on ligand stability as indicated by ligand RMSD data. To conclude, the Asp8 residue plays an important role in stabilizing both compounds within the active site pocket.

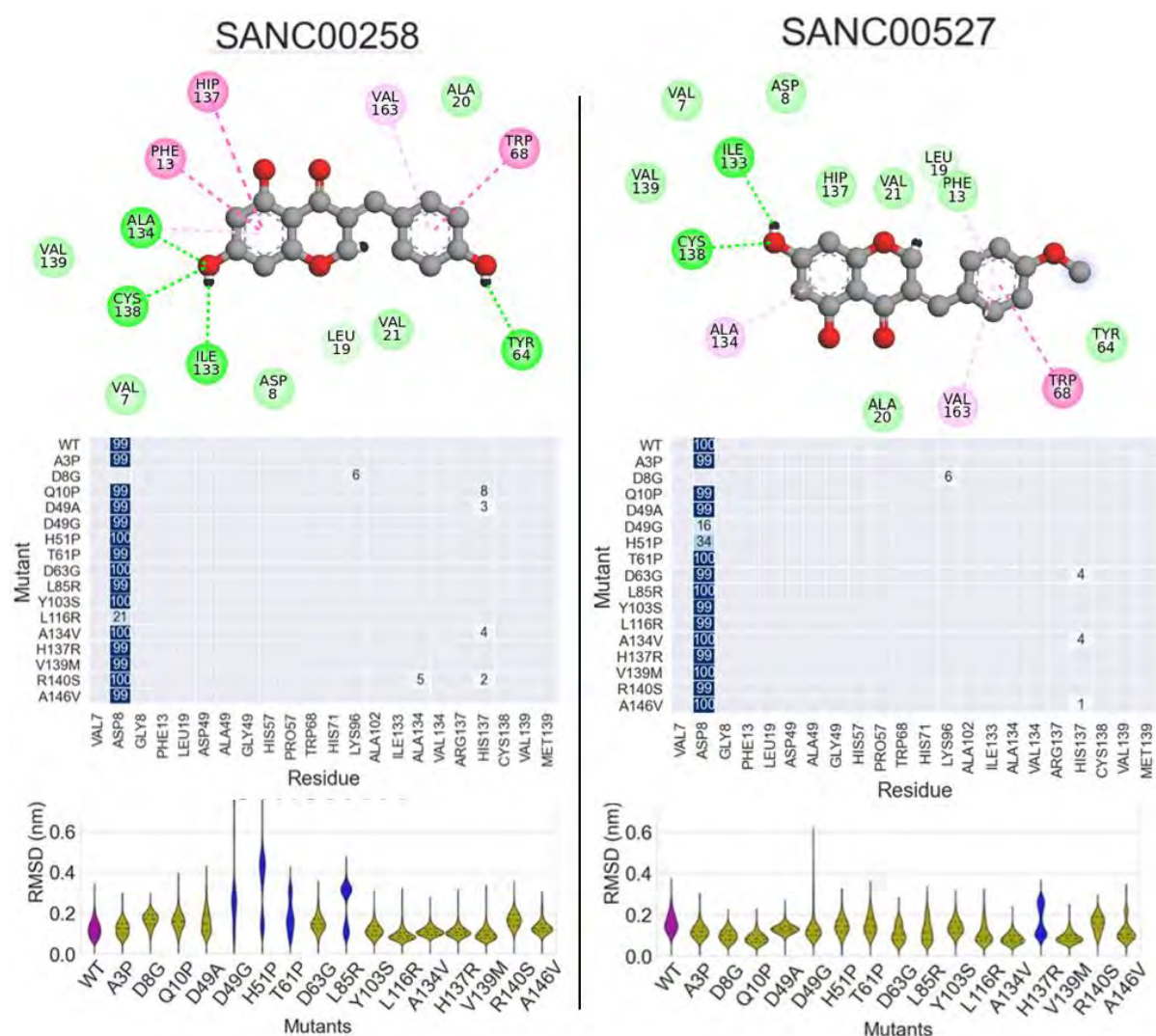


Figure 30: Second pair of analogous hits. SANC00258 and SANC00527 performed favourably against WT PZase as well as the majority of PZase mutants. Sub figures for each ligand correspond to 2D structure and docking interactions, hydrogen bonding during MD simulations and ligand RMSD, respectively.

The second set of analogues, SANC00258 and SANC00527 behave somewhat differently, both structures share an identical structure except for the oxygen bonded to the terminal benzene ring. In SANC00258 this oxygen is just bonded to a hydrogen, forming a polar hydroxylic group ideal for potential hydrogen bonding, this oxygen in SANC00527 on the other hand is bonded to a methyl carbon, reducing this oxygen's potential to form hydrogen bonds. This difference does not result in any significant changes in hydrogen bonding patterns during MD simulations between the two compounds however, just like the previous pair, SANC00258 and SANC00527 hydrogen bond almost exclusively with Asp8. Again, the first exception to this is in the D8G mutant system, where the absence of Asp8 results in the absence of this hydrogen bond, however unlike the previous pair its absence does not disrupt the stability of either compound, which remain stable in the active site pocket throughout the duration of MD simulations and instead exhibit very intermittent hydrogen bonding with Lys96, another residue in the catalytic triad. Important to note in both these compounds is that in mutant systems where Asp8 hydrogen bonding is disrupted and virtually no other hydrogen bonding occurs, ligand stability does not appear to be affected in any significant manner: SANC00258 does not hydrogen bond with Asp8 in the D8G and L116R mutants but is still stable in both, and SANC00527 does not hydrogen bond with Asp8 in the D8G, D49G and H51P systems but is still stable in all 3. This may suggest that for these two compounds, H-bonding is not a major factor keeping these compounds stable at the active site and that hydrophobic interactions may instead be responsible.

It is important to note just how different hydrogen bonding patterns are between the docking study systems and MD systems. In **Figures 29 and 30** H-bonds are shown in bright green, in **Figure 29**, both SANC00129 and SANC00238 are shown to be H-bonding to Asp8 which is then also present during MD simulations, however this is not the case for Lys96, which is present in the docking simulations but then almost completely absent during the MD simulations. In **Figure 30**, this difference is even more stark, the Asp8 was the only residue showing significant levels of H-bonding during MD simulations for both SANC00258 and SANC00527. However, Asp8 is not considered to be H-bonding to either compound during the docking study. One possible explanation for this is that the criteria used by Discovery Studio Visualizer to identify H-bonds during the docking study may be different to the criteria that was used with CPPTRAJ for the MD simulations (3.5 Å distance, 35° H-bond angle tolerance). This is unlikely to be the only factor however as the difference in H-bonding patterns between docking and MD is significant, with H-bonds during docking being

completely absent during MD or vice-versa. Another possibility is that the hydrogen bonds detected after docking are not actually favourable or possible when the system is simulated in a dynamic state such as that in an MD simulation. This may be due to the rigidity of the protein structure during docking: when the protein is placed into the MD environment it undergoes any necessary conformational changes to enter its lowest energy state(s) because it is no longer a rigid structure. This may lead to alterations in the topology of the active site and therefore potentially significant changes in H-bonding.

4.3.5 Drug-likeness of Hits

The concept of drug-likeness allows for more efficient screening during the early stages of drug discovery. For example, The Lipinski Rule of five is one of the most commonly used guidelines to evaluate whether potential drug candidates are suitable for oral intake (Lipinski *et al.*, 2001; Lipinski, 2004). Lipinski's rules (Ro5) state that a suitable candidate passes a minimum of 3 of the 4 criteria for drug-likeness: >5 H-bond donors, >10 H-bond acceptors, >500 Da molecular weight and a >5 LogP value. While the Ro5 is certainly a powerful tool to predict oral bioavailability, 16% of current oral drugs violate one of these criteria, and 6% violate 2 or more, the Ro5 also does not apply to natural products which very often violate the Ro5 criteria but are still bioavailable (Doak and Kihlberg, 2017). For this reason, applying the Ro5 criteria alone to SANCDB compounds, all of which are natural products is likely to yield incorrect results.

Bickerton *et al.*, 2012, proposed a different, more quantitative metric for assessing drug-likeness based on the concept of desirability. This metric, termed the Quantitative Estimate of Drug-Likeness (QED) produces a single value between zero and one. Zero indicates that all desirability properties are unfavorable, while one indicates that all desirability properties are favorable for oral bioavailability. This single dimensionless score is based on 8 commonly used drug properties used to determine drug likeness: Molecular weight, octanol-water partition coefficient (LogP), number of H-bond donors (HBDs), number of H-bond acceptors (HBAs), molecular polar surface area (PSA), number of rotatable bonds (ROTBs), number of aromatic rings (AROMs) and the number of structural alerts for fragments with high toxicity potential (ALERTS) (Limban *et al.*, 2018). All these numerical properties are then used to produce the final QED value, calculated based on a curated collection of 771 orally dosed and approved

Table 6: Tabulated results of the QED test for drug desirability. Properties highlighted in red indicate a value that contributes negatively to the compound's QED score.

Compound	MW (Da)	LogP	HBD	HBA	PSA	ROTB	AROM	ALERTS	QED
SANC00129	212	-0.979	4	3	55.73	0	0	0	0.5
SANC00238	198	-1.322	4	4	64.52	0	0	0	0.4
SANC00258	300	0.435	5	4	90.15	2	0	0	0.6
SANC00527	314	1.089	5	3	79.15	3	0	0	0.7

drugs that were used to generate the asymmetric double sigmoidal functions used to calculate a final QED score.

SANC00129 and SANC00238, the first pair of analogous hits (**Figure 27**) received slightly below average QED scores, brought down by their low MW, low lipophilicity (LogP) and absence of any rotatable bonds. SANC00258 and SANC00527, the second pair of analogous hits received improved scores due to the increase MW and lipophilicity and the presence of 2 and 3 rotatable bonds, respectively. Based on the QED scores of these compounds, all of them are possible candidates for further screening and testing.

4.4 CONCLUSION

Given the increasing prevalence of PZA resistant Mtb infections worldwide, it is important to evaluate any potential drug candidates for potential against PZase mutant enzymes as they are the primary mechanism by which PZA-resistance arises in Mtb. In this chapter, 13 hit compounds arising from WT-PZase screening were screened against 16 mutant PZase structures in a total of 224 MD simulations. Post-MD analysis of H-bonding patterns and ligand RMSD data revealed that of the 13 hit compounds, 4 behaved favourably with the majority of mutant PZase structures. Protein RMSD, RMSF, and Rg data indicated that all mutant structures remained stable despite introduced mutations, albeit with some localized residue-level increases in fluctuations as indicated by RMSF data. All 4 hit compounds exhibited consistent hydrogen bonding with Asp8 in virtually all MD simulations indicating that this residue is important for localizing potential ligands at the PZase active site.

CHAPTER 5

CONCLUDING REMARKS AND FUTURE WORK

The crystal structure of Mtb PZase (PDB ID: 3PL1) and natural compounds from the SANCDB were retrieved and used in a molecular docking study using Autdock4.2. Blind molecular docking allowed for the identification of compounds that selectively bound to the active site of PZase with binding energies below that of the PZA control complex and hydrogen bonding to active site residues.

After docking evaluation and selection based on multiple criteria, 211 SANCDB compounds were then tested in 20ns MD pre-runs to evaluate whether the hit docking compounds were stable when in complex with PZase in an MD environment. Ligand stability was evaluated using ligand RMSD data in the last 10ns of each 20ns simulation. After selection, approximately 80 systems were identified as being stable and submitted for extended, 150ns production MD runs to evaluate system stability and behaviour over more physiologically relevant timescales. After evaluating system stability using RMSF, Rg and protein RMSD, ligand stability was evaluated using ligand RMSD data and hydrogen bonding interactions were assessed. Both ligand RMSD data and hydrogen bonding data was used to select the 13 compounds that behaved favourably enough to justify MD screening with mutant PZase structures

The 13 selected ligands and 16 mutant PZase structures each containing a single, high confidence point mutation were then screened in 224 150 ns MD simulations, with identical parameters to those used in the WT PZase MD simulations. Systems were evaluated for stability and PZase structure compactness using protein RMSD, Rg and RMSF data, following this ligand stability was evaluated using ligand RMSD data and hydrogen bonding interactions were again assessed. Of the 13 compounds screened against mutants, 4 compounds remained bound and stable in complex with the majority of mutant PZase structures, SANC00129, SANC00238, SANC00258 and SANC00527. Further analysis of these compounds revealed that they were 2 pairs of analogous structures that differed only by the presence or absence of a single methyl group. Drug-likeness of these compounds was assessed using the QED metric and revealed that the first pair of analogues, SANC00129 and SANC00238 would likely need to undergo optimization before any further screening, while the other two analogues, SANC00258 and SANC00527 achieved promising QED scores.

In this study, the first major compromise made was the use of semi-flexible docking with AutoDock4.2, only the ligands were allowed to be flexible while the PZase protein was simulated as a rigid structure, in the future it may be beneficial to perform molecular docking with multiple active site residues being made flexible in order to better simulate the active site pocket of PZase. Additionally, since blind docking was performed in this study, another avenue to pursue would be to identify potential allosteric sites where disproportionately large numbers of ligands have preferentially bound to sites other than the active site pocket of PZase. This could lead to the discovery of potential PZase allosteric sites of therapeutic interest.

Additionally, given more time, it may be prudent to attempt further short MD simulations using ligands emerging from the docking study, as the criteria for selection of ligands was primarily based on time constraints. For example, given that the docking validation with PZA and PZase in complex yielded a poor binding energy of -4.28 kcal/mol, lowering selection criteria for binding energy to at least this value would yield many more compounds requiring further screening with at least 20 ns pre-run MD simulations.

For the screening study performed on mutant PZase structures, one of the compromises made due to time limits was that point mutations were introduced using Discovery Studio visualizer. Ideally point mutations would be introduced in the protein sequence and homology modelling using a tool like MODELLER (Webb and Sali, 2016) would be used to re-generate mutant structures based on the WT 3PL1 PZase structure, which would then be validated. This is what was done by Sheik Amamuddy *et al.* in 2020, and would result in more accurate mutant structures as even point mutations can induce changes in global protein structure and dynamics (Petrella *et al.*, 2011c).

REFERENCES

- Abdolmaleki, A., Ghasemi, J. and Ghasemi, F. (2017) 'Computer Aided Drug Design for Multi-Target Drug Design: SAR /QSAR, Molecular Docking and Pharmacophore Methods', *Current Drug Targets*, 18(5), pp. 556–575. doi: 10.2174/1389450117666160101120822.
- Abraham, M. J. *et al.* (2015) 'Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*, 1–2, pp. 19–25. doi: 10.1016/j.softx.2015.06.001.
- Ahmed, N. and Hasnain, S. E. (2011) 'Molecular epidemiology of tuberculosis in India: Moving forward with a systems biology approach', *Tuberculosis*. Churchill Livingstone, pp. 407–413. doi: 10.1016/j.tube.2011.03.006.
- Alder, B. J. and Wainwright, T. E. (1957) 'Phase transition for a hard sphere system', *The Journal of Chemical Physics*. American Institute of PhysicsAIP, pp. 1208–1209. doi: 10.1063/1.1743957.
- Alonso, H., Bliznyuk, A. A. and Gready, J. E. (2006) 'Combining docking and molecular dynamic simulations in drug design', *Medicinal Research Reviews*. John Wiley & Sons, Ltd, pp. 531–568. doi: 10.1002/med.20067.
- Amaro, R. E., Baron, R. and McCammon, J. A. (2008) 'An improved relaxed complex scheme for receptor flexibility in computer-aided drug design', *Journal of Computer-Aided Molecular Design*, 22(9), pp. 693–705. doi: 10.1007/s10822-007-9159-2.
- Anandakrishnan, R., Aguilar, B. and Onufriev, A. V. (2012) 'H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations', *Nucleic Acids Research*, 40(W1). doi: 10.1093/nar/gks375.
- Antunes, E. M. (2003) *Pyrrroloiminoquinone metabolites from South African Latrunculid sponges*. Rhodes University; Faculty of Science, Chemistry. Available at: <https://vital.seals.ac.za/vital/access/manager/Repository/vital:4340> (Accessed: 1 March 2021).
- Aono, A. *et al.* (2014) 'Association between pncA gene mutations, pyrazinamidase activity, and pyrazinamide susceptibility testing in Mycobacterium tuberculosis', *Antimicrobial Agents and Chemotherapy*, 58(8), pp. 4928–4930. doi: 10.1128/AAC.02394-14.
- Armstrong, J. A. and Hart, D. (1971) 'Response of cultured macrophages to Mycobacterium Tuberculosis, with observations on fusion of lysosomes with phagosomes', *Journal of Experimental Medicine*, 134(3), pp. 713–740. doi: 10.1084/jem.134.3.713.
- Austin, R. H. *et al.* (1975) 'Dynamics of Ligand Binding to Myoglobin', *Biochemistry*, 14(24), pp. 5355–5373. doi: 10.1021/bi00695a021.
- Bangani, V., Crouch, N. R. and Mulholland, D. A. (1999) 'Homoisoflavanones and stilbenoids from Scilla nervosa', *Phytochemistry*, 51(7), pp. 947–951. doi: 10.1016/S0031-9422(99)00155-7.
- Bell, E. W. and Zhang, Y. (2019) 'DockRMSD: An open-source tool for atom mapping and RMSD calculation of symmetric molecules through graph isomorphism', *Journal of Cheminformatics*, 11(1), p. 40. doi: 10.1186/s13321-019-0362-7.

- Bezsonova, I. *et al.* (2008) ‘Interactions between the three CIN85 SH3 domains and ubiquitin: Implications for CIN85 ubiquitination’, *Biochemistry*, 47(34), pp. 8937–8949. doi: 10.1021/bi800439t.
- Bickerton, G. R. *et al.* (2012) ‘Quantifying the chemical beauty of drugs’, *Nature Chemistry*, 4(2), pp. 90–98. doi: 10.1038/nchem.1243.
- BIOVIA, D. S. (2021) ‘Discovery Studio Visualizer’. San Diego.
- Blanc, L. *et al.* (2018) ‘Impact of immunopathology on the antituberculous activity of pyrazinamide’, *Journal of Experimental Medicine*, 215(8), pp. 1975–1986. doi: 10.1084/jem.20180518.
- Böhm, H. J. (1998) ‘Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs’, *Journal of Computer-Aided Molecular Design*, 12(4), p. 309. doi: 10.1023/a:1007999920146.
- Bohnert, J. A. *et al.* (2007) ‘Altered spectrum of multidrug resistance associated with a single point mutation in the Escherichia coli RND-type MDR efflux pump YhiV (MdtF)’, *Journal of Antimicrobial Chemotherapy*, 59(6), pp. 1216–1222. doi: 10.1093/jac/dkl426.
- BoseDasgupta, S. and Pieters, J. (2014) ‘Striking the right balance determines TB or not TB’, *Frontiers in Immunology*. Frontiers Media S.A., p. 455. doi: 10.3389/fimmu.2014.00455.
- Boshoff, H. I. M. *et al.* (2008) ‘Biosynthesis and recycling of nicotinamide cofactors in Mycobacterium tuberculosis: An essential role for NAD in nonreplicating bacilli’, *Journal of Biological Chemistry*, 283(28), pp. 19329–19341. doi: 10.1074/jbc.M800694200.
- Brennan, P. J. and Nikaido, H. (1995) ‘The Envelope of Mycobacteria’, *Annual Review of Biochemistry*, 64(1), pp. 29–63. doi: 10.1146/annurev.bi.64.070195.000333.
- Buller, A. R. and Townsend, C. A. (2013) ‘Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(8), pp. E653–E661. doi: 10.1073/pnas.1221050110.
- Case, D. A. *et al.* (2020) ‘AmberTools20’, *Journal of Computational Chemistry*. San Francisco: University of California, San Francisco. doi: 10.1002/jcc.20290.
- Chen, D. *et al.* (2016) ‘Regulation of protein-ligand binding affinity by hydrogen bond pairing’, *Science Advances*, 2(3), p. e1501240. doi: 10.1126/sciadv.1501240.
- Cheng, S. J. *et al.* (2000) ‘pncA mutations as a major mechanism of pyrazinamide resistance in Mycobacterium tuberculosis: Spread of a monoresistant strain in Quebec, Canada’, *Antimicrobial Agents and Chemotherapy*, 44(3), pp. 528–532. doi: 10.1128/AAC.44.3.528-532.2000.
- Cole, S. T. *et al.* (1998) ‘Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence’, *Nature*. Nature Publishing Group, pp. 537–544. doi: 10.1038/31159.
- Cole, S. T. *et al.* (2001) ‘Massive gene decay in the leprosy bacillus’, *Nature*, 409(6823), pp. 1007–1011. doi: 10.1038/35059006.
- Cornish-Bowden, A. (2013) ‘The origins of enzyme kinetics’, *FEBS Letters*, 587(17), pp. 2725–2730. doi: 10.1016/j.febslet.2013.06.009.

- Crofto, J. (1969) ‘Some Principles in the Chemotherapy of Bacterial Infections’, *British Medical Journal*, 2(5650), pp. 137–141. doi: 10.1136/bmj.2.5650.137.
- Dar, K. B. *et al.* (2019) ‘Modern Computational Strategies for Designing Drugs to Curb Human Diseases: A Prospect’, *Current Topics in Medicinal Chemistry*, 18(31), pp. 2702–2719. doi: 10.2174/1568026619666190119150741.
- Daugherty, M. *et al.* (2002) ‘Complete reconstitution of the human coenzyme A biosynthetic pathway via comparative genomics’, *Journal of Biological Chemistry*, 277(24), pp. 21431–21439. doi: 10.1074/jbc.M201708200.
- Doak, B. C. and Kihlberg, J. (2017) ‘Drug discovery beyond the rule of 5 - Opportunities and challenges’, *Expert Opinion on Drug Discovery*. Taylor and Francis Ltd, pp. 115–119. doi: 10.1080/17460441.2017.1264385.
- van Doorn, H. R. *et al.* (2003) ‘Detection of a point mutation associated with high-level isoniazid resistance in Mycobacterium tuberculosis by using real-time PCR technology with 3'-minor groove binder-DNA probes’, *Journal of Clinical Microbiology*, 41(10), pp. 4630–4635. doi: 10.1128/JCM.41.10.4630-4635.2003.
- Du, X. *et al.* (2001) ‘Crystal structure and mechanism of catalysis of a pyrazinamidase from *Pyrococcus horikoshii*’, *Biochemistry*, 40(47), pp. 14166–14172. doi: 10.1021/bi0115479.
- Durrant, J. D. and McCammon, J. A. (2011) ‘Molecular dynamics simulations and drug discovery’, *BMC Biology*. BioMed Central, p. 71. doi: 10.1186/1741-7007-9-71.
- Ehrt, S. and Schnappinger, D. (2009) ‘Mycobacterial survival strategies in the phagosome: Defence against host stresses’, *Cellular Microbiology*. Cell Microbiol, pp. 1170–1178. doi: 10.1111/j.1462-5822.2009.01335.x.
- Eiglmeier, K. *et al.* (2001) ‘The decaying genome of mycobacterium leprae’, *Leprosy Review*, 72(4), pp. 387–398. doi: 10.5935/0305-7518.20010047.
- Falzon, D. *et al.* (2011) ‘WHO guidelines for the programmatic management of drug-resistant tuberculosis: 2011 update’, *European Respiratory Journal*, 38(3), pp. 516–528. doi: 10.1183/09031936.00073611.
- Fersht, A. R. *et al.* (1985) ‘Hydrogen bonding and biological specificity analysed by protein engineering’, *Nature*, 314(6008), pp. 235–238. doi: 10.1038/314235a0.
- Finckh, R. E. and Tamm, C. (1970) ‘The homo-isoflavones III. isolation and structure of punctatin, 3,9-dihydro-punctatin, 4'-O-methyl-3,9-dihydro-punctatin, 4'-demethyl-eucomin and 4'-demethyl-5-O-methyl-3,9-dihydro-eucomin’, *Experientia*, 26(5), pp. 472–473. doi: 10.1007/BF01898447.
- Fischer, E. (1894) ‘Einfluss der Configuration auf die Wirkung der Enzyme’, *Berichte der deutschen chemischen Gesellschaft*, 27(3), pp. 2985–2993. doi: 10.1002/cber.18940270364.
- Foote, J. and Milstein, C. (1994) ‘Conformational isomerism and the diversity of antibodies’, *Proceedings of the National Academy of Sciences of the United States of America*, 91(22), pp. 10370–10374. doi: 10.1073/pnas.91.22.10370.
- Forrellad, M. A. *et al.* (2013) ‘Virulence factors of the mycobacterium tuberculosis complex’, *Virulence*. Taylor and Francis Inc., pp. 3–66. doi: 10.4161/viru.22329.
- Frauenfelder, H., Sligar, S. G. and Wolynes, P. G. (1991) ‘The energy landscapes and

- motions of proteins', *Science*, 254(5038), pp. 1598–1603. doi: 10.1126/science.1749933.
- Gerdes, S. Y. *et al.* (2002) 'From genetic footprinting to antimicrobial drug targets: Examples in cofactor biosynthetic pathways', *Journal of Bacteriology*, 184(16), pp. 4555–4572. doi: 10.1128/JB.184.16.4555-4572.2002.
- Gopal, P., Grüber, G., *et al.* (2019) 'Pharmacological and Molecular Mechanisms Behind the Sterilizing Activity of Pyrazinamide', *Trends in Pharmacological Sciences*. Elsevier Ltd, pp. 930–940. doi: 10.1016/j.tips.2019.10.005.
- Gopal, P., Sarathy, J., *et al.* (2019) 'Pyrazinamide triggers degradation of its target aspartate decarboxylase', *bioRxiv*, p. 674416. doi: 10.1101/674416.
- Goude, R. *et al.* (2009) 'The arabinosyltransferase EmbC is inhibited by ethambutol in *Mycobacterium tuberculosis*', *Antimicrobial Agents and Chemotherapy*, 53(10), pp. 4138–4146. doi: 10.1128/AAC.00162-09.
- Halperin, I. *et al.* (2002) 'Principles of docking: An overview of search algorithms and a guide to scoring functions', *Proteins: Structure, Function and Genetics*. Proteins, pp. 409–443. doi: 10.1002/prot.10115.
- Hartel, L. A., Yazbeck, A. S. and Osewe, P. L. (2018) 'Responding to Health System Failure on Tuberculosis in Southern Africa', *Health Systems & Reform*, 4(2), pp. 93–100. doi: 10.1080/23288604.2018.1441621.
- Hatherley, R. *et al.* (2015) 'SANCDDB: a South African natural compound database', *Journal of Cheminformatics*, 7(1), p. 29. doi: 10.1186/s13321-015-0080-8.
- Hatherley, R. *et al.* (2016) 'PRIMO: An Interactive Homology Modeling Pipeline', *PLOS ONE*. Edited by S. C. E. Tosatto, 11(11), p. e0166698. doi: 10.1371/journal.pone.0166698.
- Hershkovitz, I. *et al.* (2008) 'Detection and Molecular Characterization of 9000-Year-Old *Mycobacterium tuberculosis* from a Neolithic Settlement in the Eastern Mediterranean', *PLoS ONE*. Edited by N. Ahmed, 3(10), p. e3426. doi: 10.1371/journal.pone.0003426.
- Hirano, K. *et al.* (1998) 'Mutation in *pncA* is a major mechanism of pyrazinamide resistance in *Mycobacterium tuberculosis*', *Tubercle and Lung Disease*, 78(2), pp. 117–122. doi: 10.1016/S0962-8479(98)80004-X.
- Huang, S.-Y. and Zou, X. (2010) 'Advances and Challenges in Protein-Ligand Docking', *International Journal of Molecular Sciences*, 11(8), pp. 3016–3034. doi: 10.3390/ijms11083016.
- Huang, S. Y. and Zou, X. (2007) 'Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking', *Proteins: Structure, Function and Genetics*, 66(2), pp. 399–421. doi: 10.1002/prot.21214.
- Ishida, T. (2010) 'Effects of point mutation on enzymatic activity: Correlation between protein electronic structure and motion in chorismate mutase reaction', *Journal of the American Chemical Society*, 132(20), pp. 7104–7118. doi: 10.1021/ja100744h.
- Jakalian, A., Jack, D. B. and Bayly, C. I. (2002) 'Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation', *Journal of Computational Chemistry*, 23(16), pp. 1623–1641. doi: 10.1002/jcc.10128.
- Jeucken, A. *et al.* (2019) 'A Comprehensive Functional Characterization of *Escherichia coli*

- Lipid Genes', *Cell Reports*, 27(5), pp. 1597-1606.e2. doi: 10.1016/j.celrep.2019.04.018.
- Jiang, F. and Kim, S. H. (1991) "'Soft docking": Matching of molecular surface cubes', *Journal of Molecular Biology*, 219(1), pp. 79–102. doi: 10.1016/0022-2836(91)90859-5.
- Karim, S. S. A. *et al.* (2009) 'HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response', *The Lancet*. NIH Public Access, pp. 921–933. doi: 10.1016/S0140-6736(09)60916-8.
- Karplus, M. and McCammon, J. A. (2002) 'Molecular dynamics simulations of biomolecules', *Nature Structural Biology*. Nature Publishing Group, pp. 646–652. doi: 10.1038/nsb0902-646.
- Karplus, M. and Petsko, G. A. (1990) 'Molecular dynamics simulations in biology', *Nature*. Nature Publishing Group, pp. 631–639. doi: 10.1038/347631a0.
- Keiler, K. C., Waller, P. R. H. and Sauer, R. T. (1996) 'Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA', *Science*, 271(5251), pp. 990–993. doi: 10.1126/science.271.5251.990.
- Kerantzas, C. A. and Jacobs, W. R. (2017) 'Origins of combination therapy for tuberculosis: Lessons for future antimicrobial development and application', *mBio*. American Society for Microbiology. doi: 10.1128/mBio.01586-16.
- Khan, M. T. *et al.* (2019) 'Pyrazinamide resistance and mutations in *pncA* among isolates of *Mycobacterium tuberculosis* from Khyber Pakhtunkhwa, Pakistan', *BMC Infectious Diseases*, 19(1). doi: 10.1186/s12879-019-3764-2.
- Knegtel, R. M. A., Kuntz, I. D. and Oshiro, C. M. (1997) 'Molecular docking to ensembles of protein structures', *Journal of Molecular Biology*, 266(2), pp. 424–440. doi: 10.1006/jmbi.1996.0776.
- Koorbanally, C. *et al.* (2006) 'Homoisoflavanones and spirocyclic nortriterpenoids from three *Eucomis* species: *E. comosa*, *E. schijffii* and *E. pallidiflora* subsp. *pole-evansii* (Hyacinthaceae)', *South African Journal of Botany*, 72(3), pp. 428–433. doi: 10.1016/j.sajb.2005.12.006.
- Koorbanally, N. A. *et al.* (2006) 'Coincident isolation of a novel homoisoflavonoid from *Resnova humifusa* and *Eucomis montana* (Hyacinthoideae: Hyacinthaceae)', *Biochemical Systematics and Ecology*, 34(2), pp. 114–118. doi: 10.1016/j.bse.2005.08.003.
- Koshland, D. E. (1958) 'Application of a Theory of Enzyme Specificity to Protein Synthesis', *Proceedings of the National Academy of Sciences*, 44(2), pp. 98–104. doi: 10.1073/pnas.44.2.98.
- Kurz, S. G., Furin, J. J. and Bark, C. M. (2016) 'Drug-Resistant Tuberculosis: Challenges and Progress', *Infectious Disease Clinics of North America*. W.B. Saunders, pp. 509–522. doi: 10.1016/j.idc.2016.02.010.
- Lamont, E. A., Dillon, N. A. and Baughn, A. D. (2020) 'The Bewildering Antitubercular Action of Pyrazinamide', *Microbiology and Molecular Biology Reviews*, 84(2). doi: 10.1128/mmbr.00070-19.
- Lawn, S. D. and Zumla, A. I. (2011) 'Tuberculosis', in *The Lancet*. Elsevier, pp. 57–72. doi: 10.1016/S0140-6736(10)62173-3.

- Leach, A. R. (1994) 'Ligand docking to proteins with discrete side-chain flexibility', *Journal of Molecular Biology*, 235(1), pp. 345–356. doi: 10.1016/S0022-2836(05)80038-5.
- Leonardi, R. *et al.* (2005) 'Coenzyme A: Back in action', *Progress in Lipid Research*. Pergamon, pp. 125–153. doi: 10.1016/j.plipres.2005.04.001.
- Limban, C. *et al.* (2018) 'The use of structural alerts to avoid the toxicity of pharmaceuticals', *Toxicology Reports*. Elsevier Inc., pp. 943–953. doi: 10.1016/j.toxrep.2018.08.017.
- Lin, J. H. *et al.* (2002) 'Computational drug design accommodating receptor flexibility: The relaxed complex scheme', *Journal of the American Chemical Society*, 124(20), pp. 5632–5633. doi: 10.1021/ja0260162.
- Lipinski, C. A. *et al.* (2001) 'Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings', *Advanced Drug Delivery Reviews*, 46(1–3), pp. 3–26. doi: 10.1016/S0169-409X(00)00129-0.
- Lipinski, C. A. (2004) 'Lead- and drug-like compounds: The rule-of-five revolution', *Drug Discovery Today: Technologies*. Elsevier, pp. 337–341. doi: 10.1016/j.ddtec.2004.11.007.
- Logie, C. G. (1996) *The pyrrolizidine alkaloids of Senecio chrysocoma and Senecio paniculatus*. Rhodes University; Faculty of Science, Chemistry. Available at: <http://vital.seals.ac.za:8080/vital/access/manager/Repository/vital:4339> (Accessed: 1 March 2021).
- Louw, G. E. *et al.* (2006) 'Frequency and implications of pyrazinamide resistance in managing previously treated tuberculosis patients', *International Journal of Tuberculosis and Lung Disease*, 10(7), pp. 802–807.
- Ma, B. *et al.* (1999) 'Folding funnels and binding mechanisms', *Protein Engineering, Design and Selection*, 12(9), pp. 713–720. doi: 10.1093/protein/12.9.713.
- Maier, J. A. *et al.* (2015) 'ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB', *Journal of Chemical Theory and Computation*, 11(8), pp. 3696–3713. doi: 10.1021/acs.jctc.5b00255.
- Mark, P. and Nilsson, L. (2001) 'Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K', *Journal of Physical Chemistry A*, 105(43), pp. 9954–9960. doi: 10.1021/jp003020w.
- de Martino, M. *et al.* (2019) 'Immune Response to Mycobacterium tuberculosis: A Narrative Review', *Frontiers in Pediatrics*, 7. doi: 10.3389/fped.2019.00350.
- Mathabe, M. C. *et al.* (2008) 'Antibacterial activities and cytotoxicity of terpenoids isolated from *Spirostachys africana*', *Journal of Ethnopharmacology*, 116(1), pp. 194–197. doi: 10.1016/j.jep.2007.11.017.
- Mativandlela, S. P. N. *et al.* (2009) 'Antimycobacterial flavonoids from the leaf extract of *Galenia africana*', *Journal of Natural Products*, 72(12), pp. 2169–2171. doi: 10.1021/np800778b.
- McCammon, J. A., Gelin, B. R. and Karplus, M. (1977) 'Dynamics of folded proteins', *Nature*, 267(5612), pp. 585–590. doi: 10.1038/267585a0.
- McIlleron, H. *et al.* (2006) 'Determinants of rifampin, isoniazid, pyrazinamide, and

- ethambutol pharmacokinetics in a cohort of tuberculosis patients’, *Antimicrobial Agents and Chemotherapy*, 50(4), pp. 1170–1177. doi: 10.1128/AAC.50.4.1170-1177.2006.
- McKinney, W. (2010) ‘Data Structures for Statistical Computing in Python’, in *Proceedings of the 9th Python in Science Conference*. SciPy, pp. 56–61. doi: 10.25080/majora-92bf1922-00a.
- Minton, K. (2016) ‘Macrophages: Granuloma macrophage differentiation’, *Nature Reviews Immunology*. Nature Publishing Group, pp. 718–719. doi: 10.1038/nri.2016.132.
- Mishra, A. *et al.* (2018) ‘An allosteric inhibitor of Mycobacterium tuberculosis ArgJ: Implications to a novel combinatorial therapy’, *EMBO Molecular Medicine*, 10(4), p. e8038. doi: 10.15252/emmm.201708038.
- Mitchison, D. A. (1985) ‘The action of antituberculosis drugs in short-course chemotherapy’, *Tubercle*, 66(3), pp. 219–225. doi: 10.1016/0041-3879(85)90040-6.
- Morris, G. M. *et al.* (1998) ‘Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function’, *Journal of Computational Chemistry*, 19(14), pp. 1639–1662. Available at: https://www.academia.edu/3182339/Automated_docking_using_a_Lamarckian_genetic_algorithm_and_an_empirical_binding_free_energy_function (Accessed: 24 May 2021).
- Morris, G. M. *et al.* (2009) ‘AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility’, *Journal of Computational Chemistry*, 30(16), pp. 2785–2791. doi: 10.1002/jcc.21256.
- Mphahlele, M. *et al.* (2008) ‘Pyrazinamide resistance among South African multidrug-resistant Mycobacterium tuberculosis isolates’, *Journal of Clinical Microbiology*, 46(10), pp. 3459–3464. doi: 10.1128/JCM.00973-08.
- Murphy, K. M. and Weaver, C. (2016) *Janeway’s Immunobiology: Ninth International Student Edition*. W.W. Norton & Company. Available at: <https://wwnorton.com/books/9780815345053>.
- Nicas, M., Nazaroff, W. W. and Hubbard, A. (2005) ‘Toward understanding the risk of secondary airborne infection: Emission of respirable pathogens’, *Journal of Occupational and Environmental Hygiene*, 2(3), pp. 143–154. doi: 10.1080/15459620590918466.
- O’Boyle, N. M. *et al.* (2011) ‘Open Babel: An Open chemical toolbox’, *Journal of Cheminformatics*, 3(10), p. 33. doi: 10.1186/1758-2946-3-33.
- Oostenbrink, C. *et al.* (2004) ‘A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6’, *Journal of Computational Chemistry*, 25(13), pp. 1656–1676. doi: 10.1002/jcc.20090.
- Orme, I. M., Robinson, R. T. and Cooper, A. M. (2015) ‘The balance between protective and pathogenic immune responses in the TB-infected lung’, *Nature Immunology*. Nature Publishing Group, pp. 57–63. doi: 10.1038/ni.3048.
- Pagadala, N. S., Syed, K. and Tuszynski, J. (2017) ‘Software for molecular docking: a review’, *Biophysical Reviews*, 9(2), pp. 91–102. doi: 10.1007/s12551-016-0247-1.
- Pantsar, T. and Poso, A. (2018) ‘Binding affinity via docking: Fact and fiction’, *Molecules*, 23(8). doi: 10.3390/molecules23081899.

- Peloquin, C. A. *et al.* (1998) ‘Pharmacokinetics of pyrazinamide under fasting conditions, with food, and with antacids.’, *Pharmacotherapy*, 18(6), pp. 1205–11. doi: 10.1002/j.1875-9114.1998.tb03138.x.
- Petrella, S. *et al.* (2011a) ‘Crystal structure of the pyrazinamidase of mycobacterium tuberculosis: Insights into natural and acquired resistance to pyrazinamide’, *PLoS ONE*, 6(1). doi: 10.1371/journal.pone.0015785.
- Petrella, S. *et al.* (2011b) ‘Crystal Structure of the Pyrazinamidase of Mycobacterium tuberculosis: Insights into Natural and Acquired Resistance to Pyrazinamide’, *PLoS ONE*. Edited by H. W. van Veen, 6(1), p. e15785. doi: 10.1371/journal.pone.0015785.
- Petrella, S. *et al.* (2011c) ‘Crystal Structure of the Pyrazinamidase of Mycobacterium tuberculosis: Insights into Natural and Acquired Resistance to Pyrazinamide’, *PLoS ONE*. Edited by H. W. van Veen, 6(1), p. e15785. doi: 10.1371/journal.pone.0015785.
- Petruccioli, E. *et al.* (2016) ‘Correlates of tuberculosis risk: Predictive biomarkers for progression to active tuberculosis’, *European Respiratory Journal*. European Respiratory Society, pp. 1751–1763. doi: 10.1183/13993003.01012-2016.
- Qiang, W. *et al.* (2017) ‘Mechanisms of resistance to the BCR-ABL1 allosteric inhibitor asciminib’, *Leukemia*. Nature Publishing Group, pp. 2844–2847. doi: 10.1038/leu.2017.264.
- Queval, C. J., Brosch, R. and Simeone, R. (2017) ‘The macrophage: A disputed fortress in the battle against Mycobacterium tuberculosis’, *Frontiers in Microbiology*. Frontiers Media S.A., p. 23. doi: 10.3389/fmicb.2017.02284.
- Ramírez, D. and Caballero, J. (2016) ‘Is It Reliable to Use Common Molecular Docking Methods for Comparing the Binding Affinities of Enantiomer Pairs for Their Protein Target?’, *International Journal of Molecular Sciences*, 17(4), p. 525. doi: 10.3390/ijms17040525.
- Ramírez, D. and Caballero, J. (2018) ‘Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data?’, *Molecules*, 23(5), p. 1038. doi: 10.3390/molecules23051038.
- Reback, J. *et al.* (2021) ‘pandas-dev/pandas: Pandas’. Zenodo. doi: 10.5281/ZENODO.4524629.
- Rendon, A. *et al.* (2016) ‘Classification of drugs to treat multidrug-resistant tuberculosis (MDR-TB): Evidence and perspectives’, *Journal of Thoracic Disease*. AME Publishing Company, pp. 2666–2671. doi: 10.21037/jtd.2016.10.14.
- Rodríguez-Zavala, J. S. (2008) ‘Enhancement of coenzyme binding by a single point mutation at the coenzyme binding domain of E. coli lactaldehyde dehydrogenase’, *Protein Science*, 17(3), pp. 563–570. doi: 10.1110/ps.073277108.
- Sakula, A. (1988) ‘Selman Waksman (1888-1973), discoverer of streptomycin: a centenary review.’, *British journal of diseases of the chest*, 82(1), pp. 23–31. doi: 10.1016/0007-0971(88)90005-8.
- Sandgren, A. *et al.* (2009) ‘Tuberculosis Drug Resistance Mutation Database’, *PLoS Medicine*, 6(2), p. e1000002. doi: 10.1371/journal.pmed.1000002.
- Schrödinger, L. (2020) ‘The PyMOL Molecular Graphics System, Version 2.0’. Available at: <https://pymol.org>.

Schweizer, E. and Hofmann, J. (2004) 'Microbial Type I Fatty Acid Synthases (FAS): Major Players in a Network of Cellular FAS Systems', *Microbiology and Molecular Biology Reviews*, 68(3), pp. 501–517. doi: 10.1128/mnbr.68.3.501-517.2004.

Scorpio, A. and Zhang, Y. (1996) 'Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus', *Nature Medicine*, 2(6), pp. 662–667. doi: 10.1038/nm0696-662.

Shakhnovich, E. I. and Gutin, A. M. (1991) 'Influence of point mutations on protein structure: Probability of a neutral mutation', *Journal of Theoretical Biology*, 149(4), pp. 537–546. doi: 10.1016/S0022-5193(05)80097-9.

Shanthirabalan, S., Chomilier, J. and Carpentier, M. (2018) 'Structural effects of point mutations in proteins', *Proteins: Structure, Function and Bioinformatics*, 86(8), pp. 853–867. doi: 10.1002/prot.25499.

Sheen, P. *et al.* (2009) 'Effect of pyrazinamidase activity on pyrazinamide resistance in *Mycobacterium tuberculosis*', *Tuberculosis*, 89(2), pp. 109–113. doi: 10.1016/j.tube.2009.01.004.

Sheik Amamuddy, O. *et al.* (2020) 'Determining the unbinding events and conserved motions associated with the pyrazinamide release due to resistance mutations of *Mycobacterium tuberculosis* pyrazinamidase', *Computational and Structural Biotechnology Journal*, 18, pp. 1103–1120. doi: 10.1016/j.csbj.2020.05.009.

Shi, W. *et al.* (2011) 'Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*', *Science*, 333(6049), pp. 1630–1632. doi: 10.1126/science.1208813.

Shoichet, B. K., Kuntz, I. D. and Bodian, D. L. (1992) 'Molecular docking using shape descriptors', *Journal of Computational Chemistry*, 13(3), pp. 380–397. doi: 10.1002/jcc.540130311.

Sidwell, W. T. L. and Tamm, C. (1970) 'The homo-isoflavones III). Isolation and structure of 4'-o-methyl-punctatin, autumnalin and 3,9-dihydro-autumnalin', *Tetrahedron Letters*, 11(7), pp. 475–478. doi: 10.1016/0040-4039(70)89003-7.

Sikorska, J. *et al.* (2012) 'Antimicrobial rubrolides from a South African species of *synoicum* tunicate', *Journal of Natural Products*, 75(10), pp. 1824–1827. doi: 10.1021/np300580z.

Skinner, D. and Claassens, M. (2016) 'It's complicated: Why do tuberculosis patients not initiate or stay adherent to treatment? A qualitative study from South Africa', *BMC Infectious Diseases*, 16(1). doi: 10.1186/s12879-016-2054-5.

Skjærven, L., Reuter, N. and Martinez, A. (2011) 'Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: a computational approach', *Future Medicinal Chemistry*, 3(16), pp. 2079–2100. doi: 10.4155/fmc.11.159.

Sousa Da Silva, A. W. and Vranken, W. F. (2012) 'ACPYPE - AnteChamber PYthon Parser interface', *BMC Research Notes*, 5, p. 367. doi: 10.1186/1756-0500-5-367.

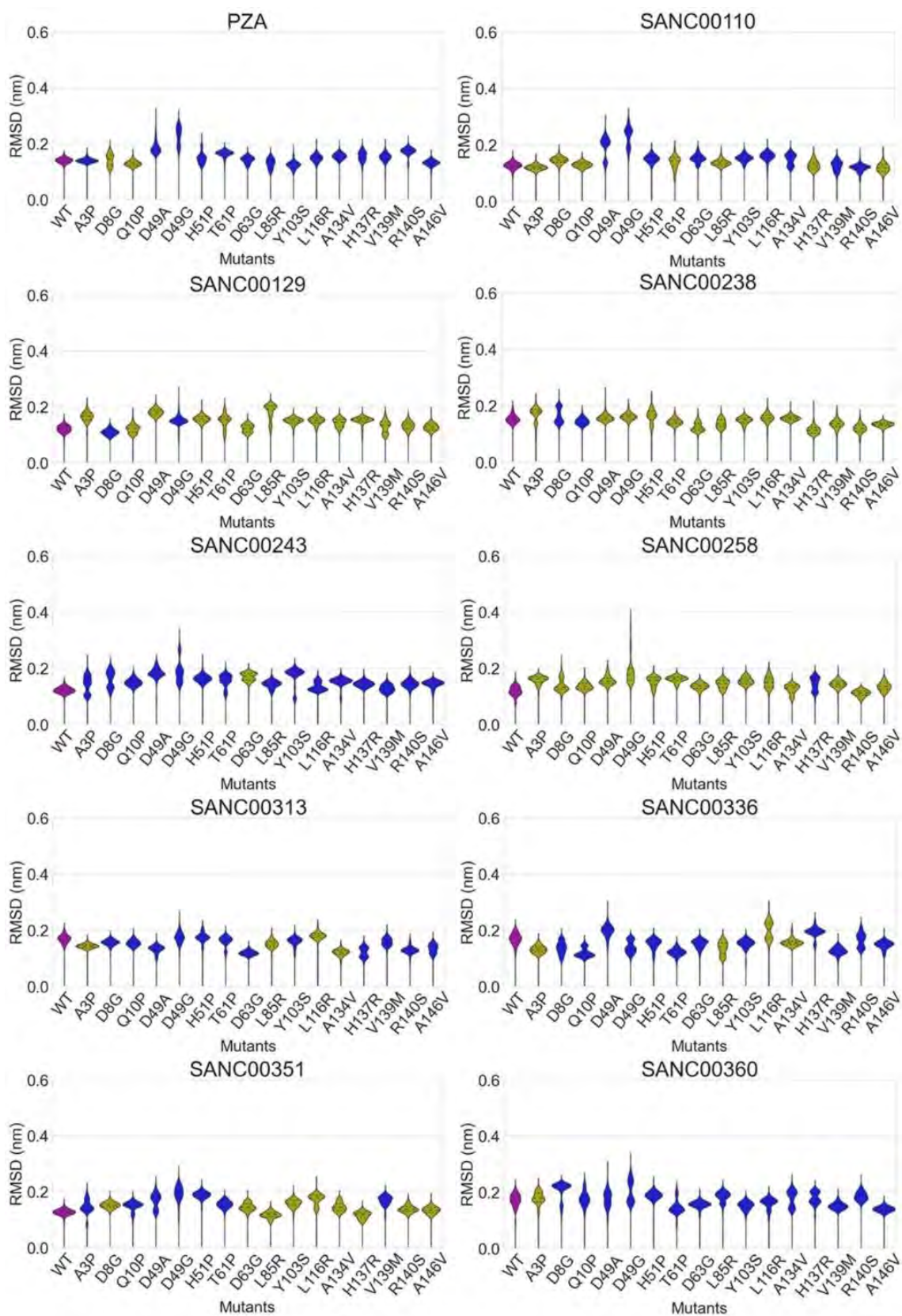
Sousa, S. F., Fernandes, P. A. and Ramos, M. J. (2010) 'Molecular Dynamics Simulations: Difficulties, Solutions and Strategies for Treating Metalloenzymes', in *Challenges and Advances in Computational Chemistry and Physics: From Nano- to Bio-Scale*. Springer, pp. 299–330. doi: 10.1007/978-90-481-3034-4_11.

STATSSA (no date) *South Africa - Living Conditions Survey 2014-2015, 2015*. Available at:

- <https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/608> (Accessed: 28 July 2020).
- Trott, O. and Olson, A. J. (2009) ‘AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading’, *Journal of Computational Chemistry*, 31(2), p. NA-NA. doi: 10.1002/jcc.21334.
- Turner, R. D. *et al.* (2017) ‘Tuberculosis Infectiousness and Host Susceptibility’, *The Journal of Infectious Diseases*, 636(6), p. 216. doi: 10.1093/infdis/jix361.
- Unissa, A. N. *et al.* (2016) ‘Overview on mechanisms of isoniazid action and resistance in Mycobacterium tuberculosis’, *Infection, Genetics and Evolution*. Elsevier B.V., pp. 474–492. doi: 10.1016/j.meegid.2016.09.004.
- Vanommeslaeghe, K. *et al.* (2010) ‘CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields’, *Journal of Computational Chemistry*, 31(4), pp. 671–690. doi: 10.1002/jcc.21367.
- Veleg, H. F. G., Gohlke, H. and Klebe, G. (2005) ‘DrugScoreCSD-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction’, *Journal of Medicinal Chemistry*, 48(20), pp. 6296–6303. doi: 10.1021/jm050436v.
- Via, L. E. *et al.* (2015) ‘Host-Mediated Bioactivation of Pyrazinamide: Implications for Efficacy, Resistance, and Therapeutic Alternatives’, *ACS Infectious Diseases*, 1(5), pp. 203–214. doi: 10.1021/id500028m.
- Vilchèze, C. *et al.* (2010) ‘NAD⁺ auxotrophy is bacteriocidal for the tubercle bacilli’, *Molecular Microbiology*, 76(2), pp. 365–377. doi: 10.1111/j.1365-2958.2010.07099.x.
- Waskom, M. *et al.* (2020) ‘mwaskom/seaborn’. Zenodo. doi: 10.5281/ZENODO.4379347.
- Webb, B. and Sali, A. (2016) ‘Comparative protein structure modeling using MODELLER’, *Current Protocols in Bioinformatics*, 2016, pp. 5.6.1-5.6.37. doi: 10.1002/cpbi.3.
- Werngren, J. *et al.* (2012) ‘Reevaluation of the critical concentration for drug susceptibility testing of Mycobacterium tuberculosis against pyrazinamide using wild-type MIC distributions and pncA gene sequencing’, *Antimicrobial Agents and Chemotherapy*, 56(3), pp. 1253–1257. doi: 10.1128/AAC.05894-11.
- Whibley, C. E. *et al.* (2005) ‘Antiesophageal cancer activity from Southern African marine organisms’, in *Annals of the New York Academy of Sciences*. Blackwell Publishing Inc., pp. 405–412. doi: 10.1196/annals.1352.031.
- WHO (2020a) *Global tuberculosis report 2019*, WHO. World Health Organization. Available at: http://www.who.int/tb/publications/global_report/en/ (Accessed: 6 July 2020).
- WHO (2020b) ‘Progress report on HIV, viral hepatitis and sexually transmitted infections 2019’, WHO. Available at: <http://www.who.int/hiv/strategy2016-2021/progress-report-2019/en/> (Accessed: 28 July 2020).
- WHO (no date) *WHO | WHO End TB Strategy*. Available at: https://www.who.int/tb/post2015_strategy/en/ (Accessed: 27 July 2020).
- Wilburn, K. M., Fieweger, R. A. and VanderVen, B. C. (2018) ‘Cholesterol and fatty acids grease the wheels of Mycobacterium tuberculosis pathogenesis’, *Pathogens and disease*. Oxford Academic, p. 21. doi: 10.1093/femspd/fty021.

- Wipperman, M. F., Sampson, N. S. and Thomas, S. T. (2014) 'Pathogen roid rage: Cholesterol utilization by Mycobacterium tuberculosis', *Critical Reviews in Biochemistry and Molecular Biology*. Informa Healthcare, pp. 269–293. doi: 10.3109/10409238.2014.895700.
- Wood, W. W. and Jacobson, J. D. (1957) 'Preliminary results from a recalculation of the Monte Carlo equation of state of hard spheres', *The Journal of Chemical Physics*, pp. 1207–1208. doi: 10.1063/1.1743956.
- Word, J. M. *et al.* (1999) 'Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation', *Journal of Molecular Biology*, 285(4), pp. 1735–1747. doi: 10.1006/jmbi.1998.2401.
- Yu, W. and Mackerell, A. D. (2017) 'Computer-aided drug design methods', in *Methods in Molecular Biology*. Humana Press Inc., pp. 85–106. doi: 10.1007/978-1-4939-6634-9_5.
- Zhang, H. *et al.* (2008) 'Characterization of Mycobacterium tuberculosis nicotinamidase/pyrazinamidase', *FEBS Journal*, 275(4), pp. 753–762. doi: 10.1111/j.1742-4658.2007.06241.x.
- Zhang, Y. *et al.* (1999) 'Role of acid pH and deficient efflux of pyrazinoic acid in unique susceptibility of Mycobacterium tuberculosis to pyrazinamide', *Journal of Bacteriology*, 181(7), pp. 2044–2049. doi: 10.1128/jb.181.7.2044-2049.1999.
- Zhang, Y. *et al.* (2014) 'Mechanisms of Pyrazinamide Action and Resistance', *Microbiology Spectrum*, 2(4). doi: 10.1128/microbiolspec.mgm2-0023-2013.
- Zhao, H. and Huang, D. (2011) 'Hydrogen Bonding Penalty upon Ligand Binding', *PLoS ONE*. Edited by P. Butko, 6(6), p. e19923. doi: 10.1371/journal.pone.0019923.
- Zimhony, O. *et al.* (2000) 'Pyrazinamide inhibits the eukaryotic-like fatty acid synthetase I (FASI) of Mycobacterium tuberculosis', *Nature Medicine*, 6(9), pp. 1043–1047. doi: 10.1038/79558.

APPENDICES



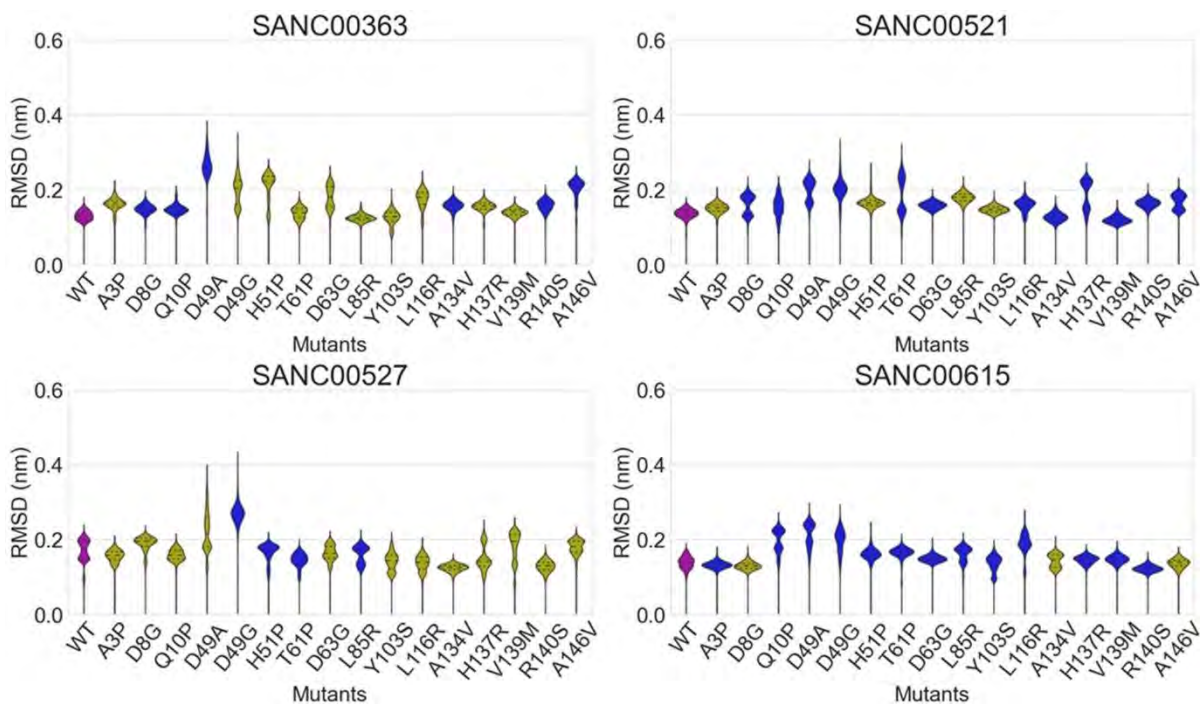
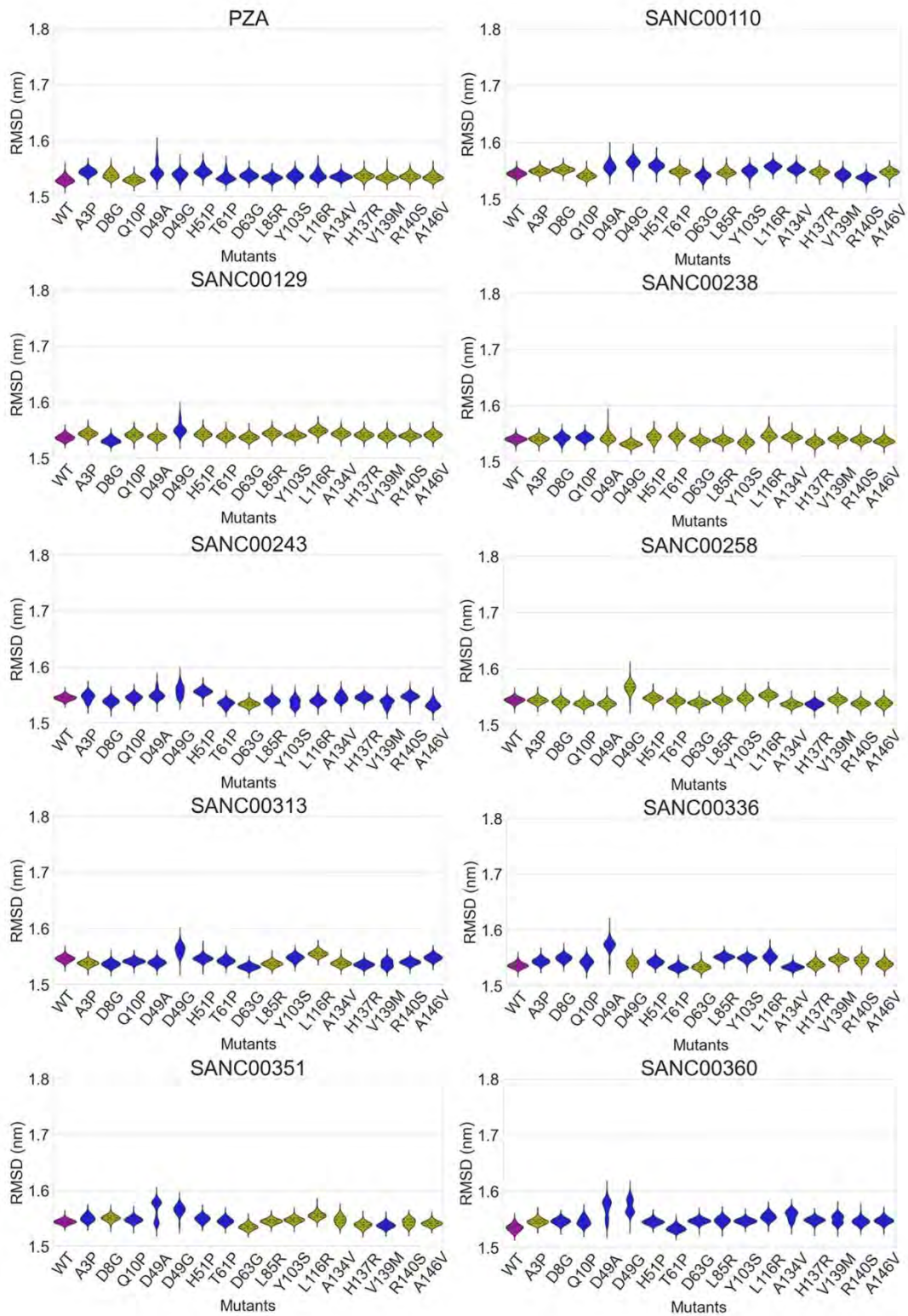


Figure S1: Violin plots of protein backbone RMSD results for 150 mutant PZase production runs. In purple: PZA-PZase control; blue: systems judged to be unstable based on ligand RMSD data; yellow: systems judged stable based on ligand RMSD data.



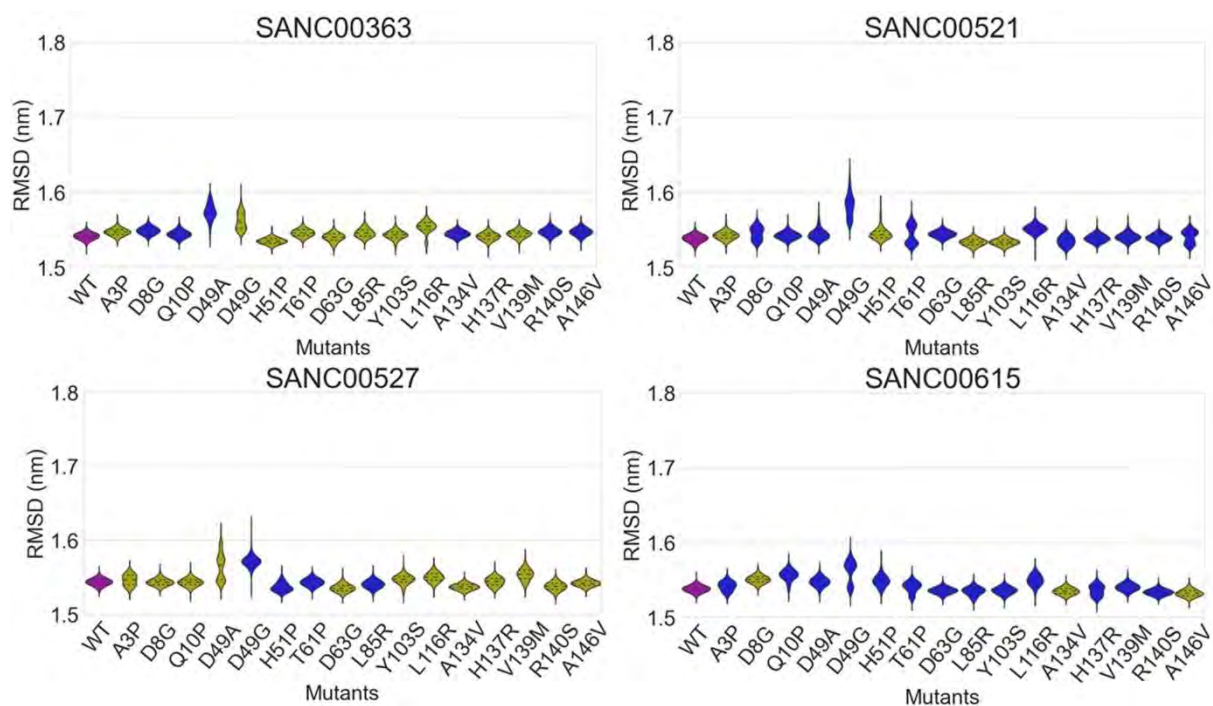
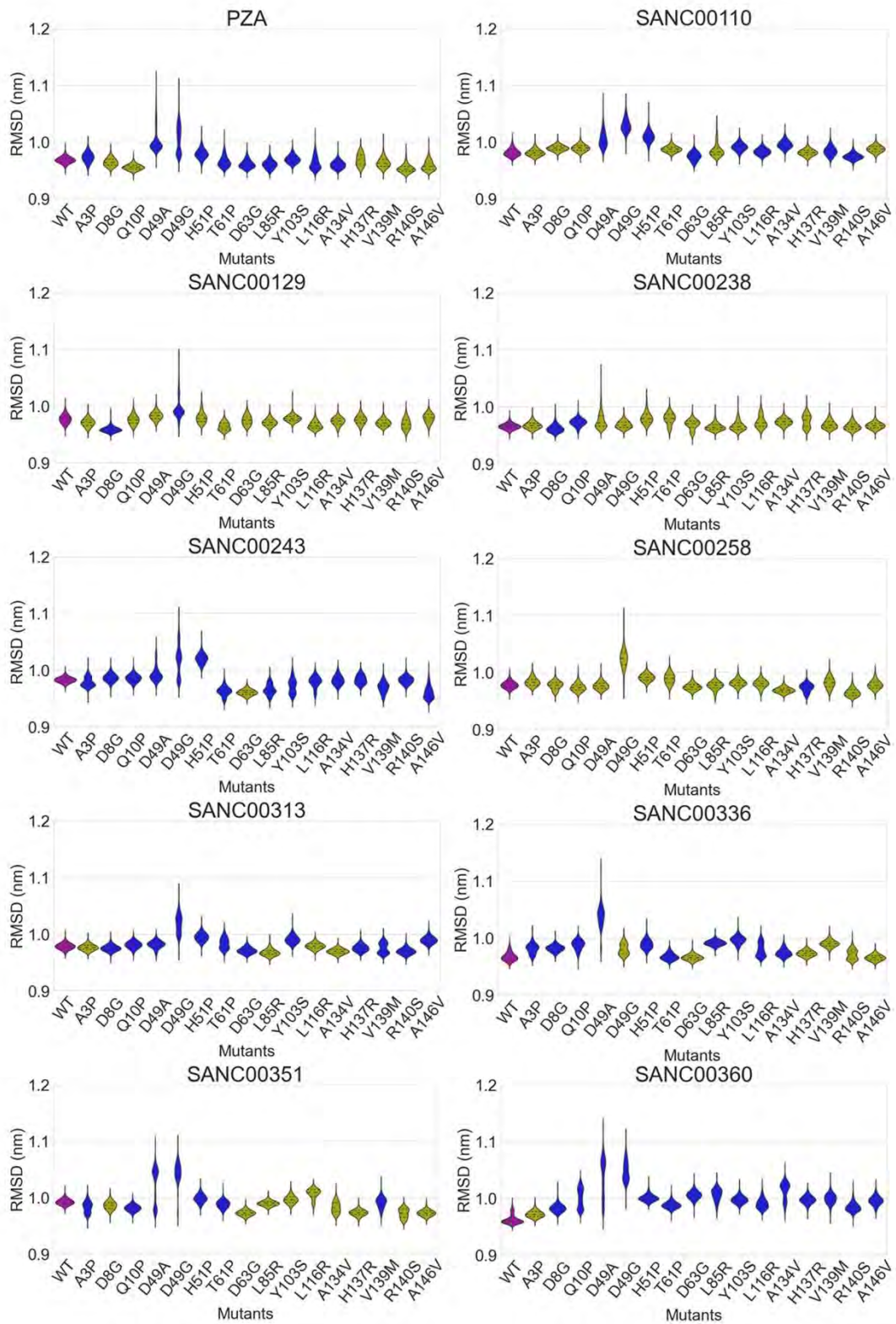


Figure S2: Violin plots of protein Radius of gyration (Rg) results for 150 ns mutant PZase production MD runs. In purple: PZA-PZase control; blue: systems judged to be unstable based on ligand RMSD data; yellow: systems judged stable based on ligand RMSD data.



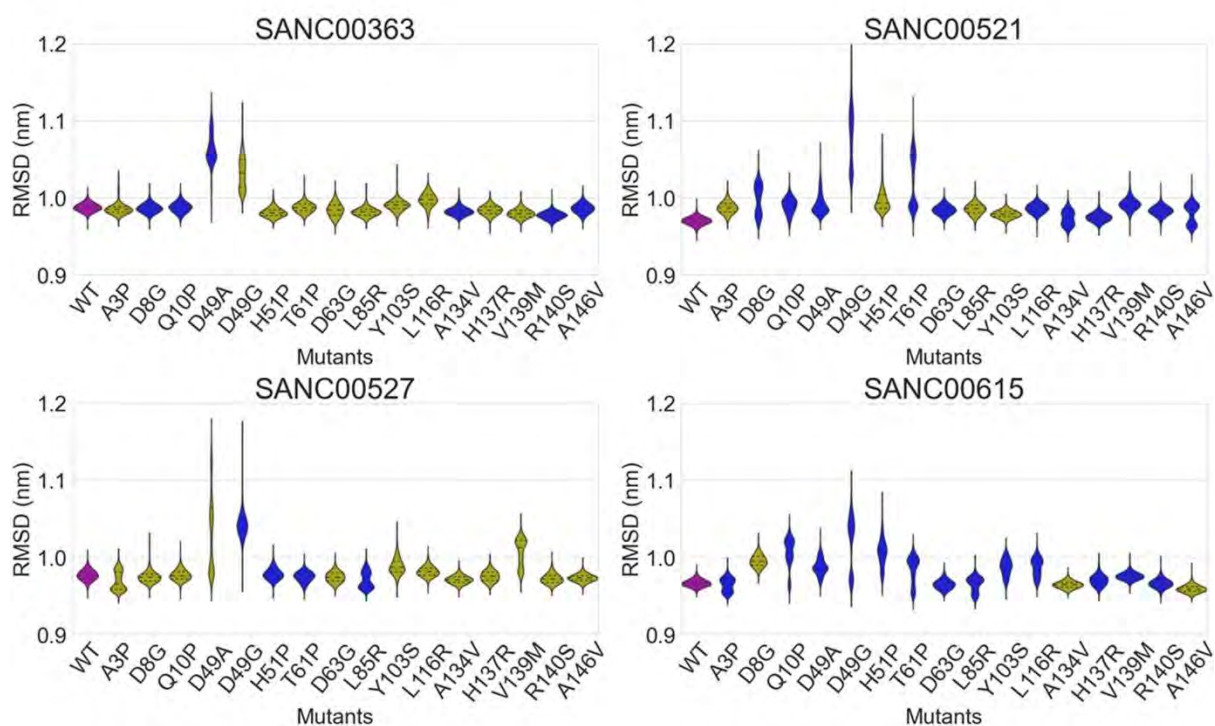
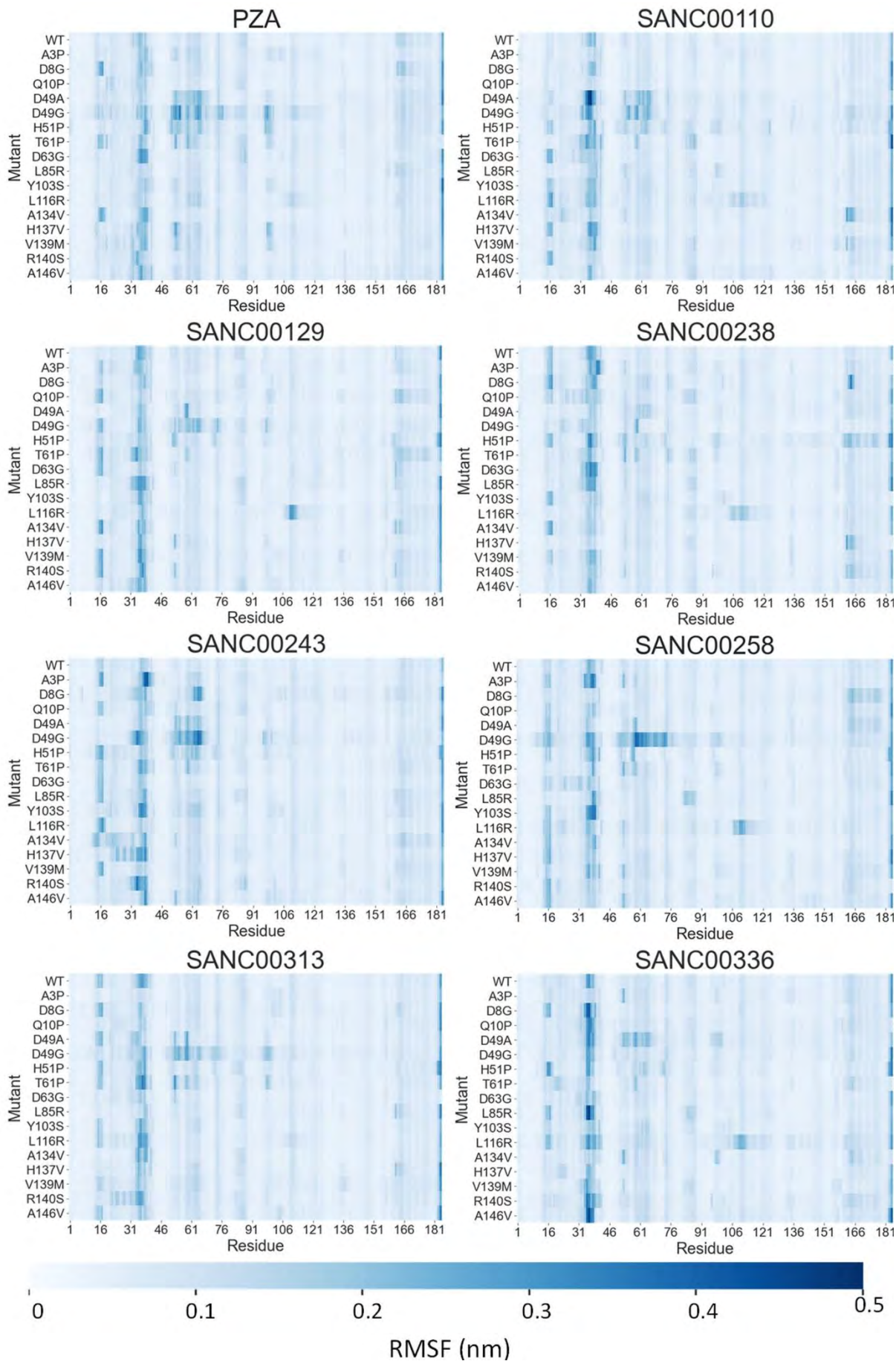


Figure S3: Violin plots of Rg of active site residue results for 150 ns mutant PZase production MD runs. In purple: PZA-PZase control; blue: systems judged to be unstable based on ligand RMSD data; yellow: systems judged stable based on ligand RMSD data.



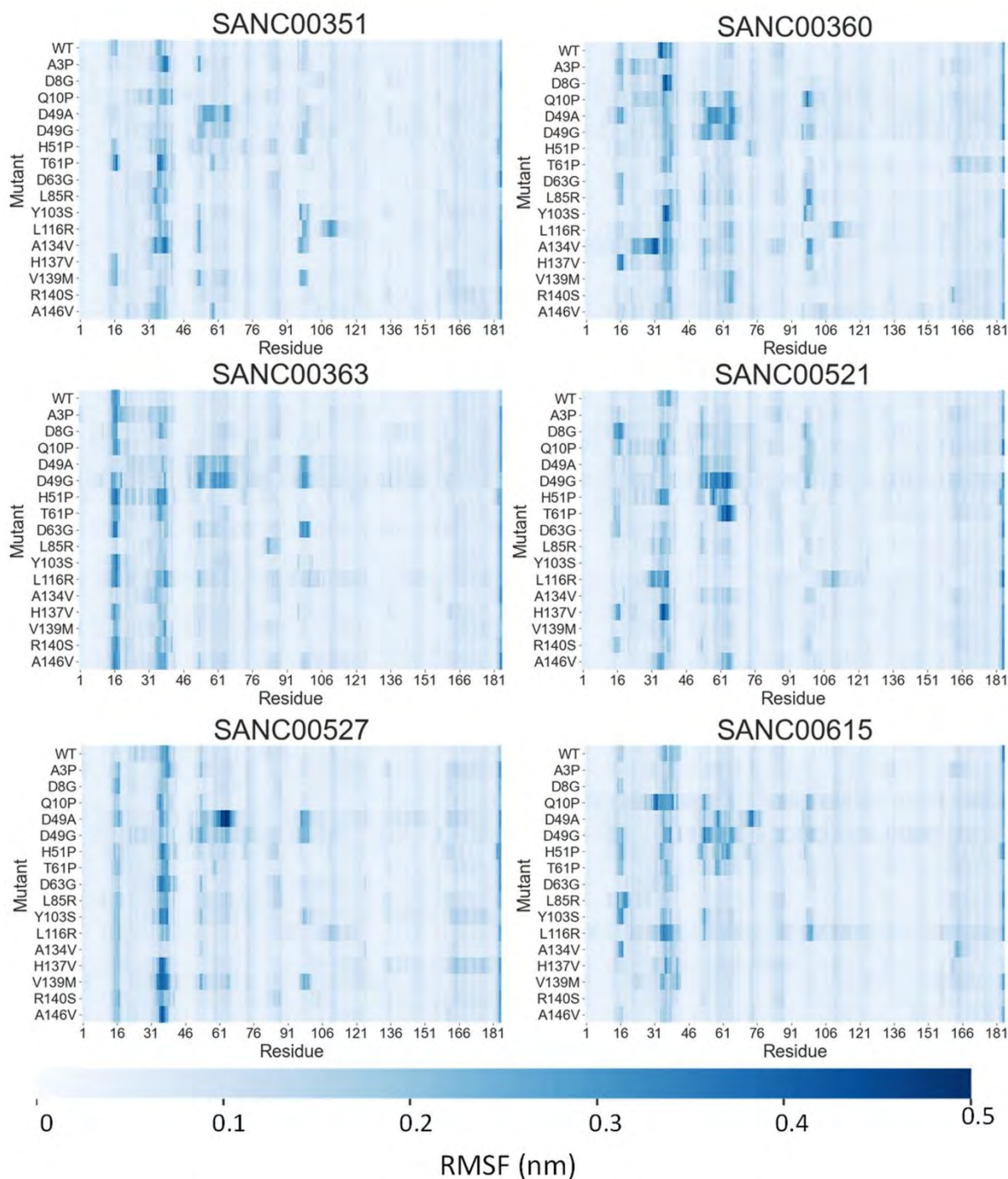


Figure S4: RMSF data results for 150 ns mutant PZase production MD runs. (A) RMSF profiles of all systems selected for further screening during mutagenesis studies. **(B)** RMSF profiles of all other systems. **(C)** Crystal structure of PZase with regions of consistent higher RMSF values colored in yellow and green, adjacent labels indicate residue numbers.