![Rhodes University logo](Where leaders learn)

RHODES UNIVERSITY
Where leaders learn

IN SILICO IDENTIFICATION OF SELECTIVE NOVEL HITS AGAINST THE ACTIVE SITE OF WILD TYPE *MYCOBACTERIUM TUBERCULOSIS* PYRAZINAMIDASE AND ITS MUTANTS

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

of

Master of Science in Bioinformatics and Computational Molecular Biology

(Coursework and Thesis)

of

RHODES UNIVERSITY, SOUTH AFRICA

Research Unit in Bioinformatics (RUBi)

DEPARTMENT OF BIOCHEMISTRY AND MICROBIOLOGY

Faculty of Science

by

GOWO PRUDENCE

G20G2285

FEBRUARY 2021

DECLARATION

I, Prudence Gowo, declare that this mini thesis entitled, *In silico* identification of selective novel hits against the active site of wild type *Mycobacterium tuberculosis* pyrazinamidase and its mutants, submitted to Rhodes University is solely my own research work. I have acknowledged all authors' concepts and referenced direct quotations from their works. I also declare that this thesis has never been submitted to any different institution for whatever degree.

Signature ....

Date...

DEDICATION

To my best friend Loveness, you taught me everything, including achieving the impossible. I love

you mom!

ACKNOWLEDGEMENTS

For its completion and success, I would like to acknowledge the supremacy of God, my ultimate Enabler; The One who started this journey and completed it. His love and grace never fails me !

For her unrivalled love and support in every way, my heartfelt gratitude goes to my mom, Mrs Loveness Muisa. You made sure I went through, you are irreplaceable!

I also want to acknowledge the unpararelled work and guidance of my supervisor Professor Özlem Tastan Bishop, may God bless you. I would like to recognise the work of my project mentor Mrs Rita Boateng; her patience, guidance and encouragement was second to none. You were the power.

I will also want to offer my invaluable gratitude to the NRF bursary for the financial support, all my lecturers for the coursework and guidance. I am deeply grateful.

To this clique;

    Dr. "bro" Kenny Chiwarawara,

    Loving sis Chipo Magwaba,

    Dr. "sis" Nomzamo Dube,

    Project partner Thomas Kenyon,

    Marvel, Lillian Mbaisi, Mingi and all my classmates;

You were my social and academic pillars, the fire that kept pushing me and my absolute cheerleaders. I appreciate you, your presence, encouragement, love and dedication to my success.

Thank you.

ABSTRACT

The World Health Organization declared Tuberculosis a global health emergency and has set a goal to eradicate it by 2035. However, effective treatment and control of the disease is being hindered by the emerging Multi-Drug Resistant and Extensively Drug Resistant strains on the most effective first line prodrug, Pyrazinamide (PZA). Studies have shown that the main cause of PZA resistance is due to mutations in the *pncA* gene that codes for the target protein Pyrazinamidase (PZase). Therefore, this study aimed to identify novel drug compounds that bind to the active site of wild type PZase and study the dynamics of these potential anti-TB drugs in the mutant systems of PZase. This approach will aid in identifying drugs that may be repurposed for TB therapy and/or designed to counteract PZA resistance. This was achieved by screening 2089 DrugBank compounds against the whole wild type (WT) PZase protein in molecular docking using AutoDOCK4.2. Compound screening based on docking binding energy, hydrogen bonds, molecular weight and active site proximity identified 47 compounds meeting all the set selection criteria. The stability of these compounds were analysed in Molecular Dynamic (MD) simulations and were further studied in PZase mutant systems of A3P, A134V, A146V, D8G, D49A, D49G, D63G, H51P, H137R, L85R, L116R, Q10P, R140S, T61P, V139M and Y103S. Generally, mutant-ligand systems displayed little deviation from the WT systems. The compound systems remained compact, with less fluctuations and more hydrogen bond interactions throughout the simulation (DB00255, DB00655, DB00672, DB00782, DB00977, DB01196, DB04573, DB06414, DB08981, DB11181, DB11760, DB13867, DB13952). From this research study, potential drugs that may be repurposed for TB therapy were identified. Majority of these drugs are currently used in the treatment of hypertension, menopause disorders and inflammation. To further understand the mutant-ligand dynamic systems, calculations such as Dynamic Residue Network (DRN) may be done. Also, the bioactivity of these drugs on *Mycobacterium tuberculosis* may be studied in wet laboratory, to understand their clinical impart *in vivo* experiments.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIDS   Acquired Immune Deficiency Syndrome

BCG   Baccille Calmette-Guerin

CHPC   Center for High Performance Computing

FAS1   Fatty Acid Synthase1

GROMACS   GROningen MAchine for Chemical Simulations

HIV   Human Immunodeficiency Virus

ID   Identification

LAM   Lipoarabinomannan

LB   Ligand-based

MBS   Metal Binding Site

MD   Molecular Dynamics

MDR-TB   Multidrug Resistant TB

M *tb*   Mycobacterium tuberculosis

M.tbC   Mycobacterium tuberculosis Complex

PDB   Protein Data Bank

POA   Pyrazinoic Acid

PROSA   Protein Structure Analysis

PZA   Pyrazinamide

PZase   Pyrazinamidase

Rg   Radius of Gyration

RMSD   Root Mean Square Deviation

RMSF   Root Mean Square Fluctuation

rpsA   30S Ribosomal Protein S1

SB   Structure-based

TB   Tuberculosis

WHO   World Heatlh Organization

WT   Wild Type

XDR-TB   Extensively Drug Resistant tuberculosis

2D   Two- dimensional

3D   Three-dimensional

TABLE OF AMINO ACIDS

Full name	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S

Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

CHAPTER ONE

1. LITERATURE REVIEW

1.1 TUBERCULOSIS

1.1.1 Introduction

Tuberculosis (TB) was declared a global health emergency in 1993 by the World Health Organization (WHO) (Floyd *et al.,* 2018), a disease claiming approximately 1.7 billion infections and 10 million sicknesses each year (WHO 2019). According to Singh *et al.,* (2018), a third of the world's population has been infected with TB and there is a documented case of TB infection in almost every country. WHO reported TB-related deaths of up to 1.4 million in 2019 out of the 10 million infections.Majority of the infections were reported from Africa, South-East Asia and West Pacific regions (WHO, 2020). South Africa has been reported to be among the eight high burden countries (India, China, Indonesia, Philippines, Pakistan, Nigeria, Bangladesh) that contribute to two thirds of the global infections (Figure 1.1, WHO 2020). It is also the only high burden country having zoonotic TB cases (*Mycobacterium bovis*) (WHO, 2020).

A target has been set by WHO to eradicate TB by 2035 with the acknowledgment that better diagnostic methods, preventative and therapeutic measures have to be employed (CryPTIC, 2018). The ultimate goal is guided by reducing deaths and incidence rates by 90% and 80% respectively between 2015 and 2030 (WHO 2019; Floyd *et al.,* 2018). However, WHO (2020) reports that due to the current COVID-19 pandemic, the extra pressure on health services may slow or reverse the progress done on TB eradication. Although lockdown measures and physical distancing policies

imposed across all nations may reduce the incidences of TB transmission and infection, this may be an offset of worsening TB therapy outcomes, having longer infectiousness periods and poverty (WHO, 2020). Already, the high burden countries have reported massive reduction in the number of new cases within the first few months of COVID-19 lockdown. The high burden countries reported 25–30% lower TB cases in India, Indonesia and the Philippines within the first six months of 2020 compared to 2019 while South Africa had a decrease of up to 50% (WHO, 2020).
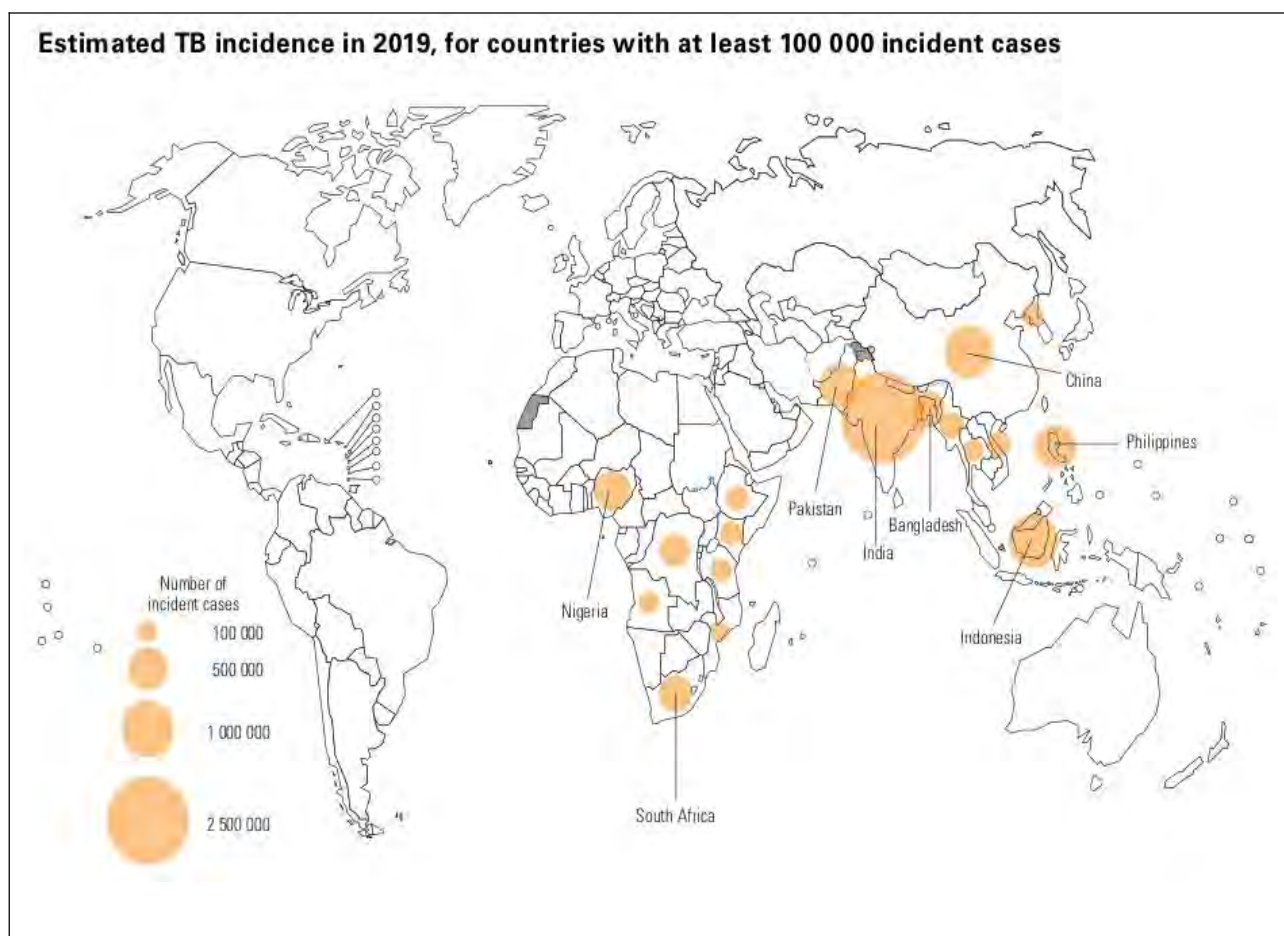
Figure 1. 1: Estimated TB incidence cases for 2019 (image from WHO, 2020)

The main population at risk is adults with underlying conditions such as diabetes, HIV/AIDS or malnutrition (WHO, 2020). Generally, there is a double risk in males than female (WHO, 2019), and according to Jimenez-Corona *et al.,* 2006 this difference is mainly due to men being more

exposed to local transmission in crowded working areas, poor ventilated and imprisonment shelters. Mens' excessive use of tobacco and alcohol also increase the rate at which latent TB progresses to active disease and symptoms such as coughing may be mistaken for tobacco use while one is already symptomatic (Jimenez-Corona *et al.,* 2006). However, McQuaid *et al.,* (2020) states that males and females have the same risk of Multi Drug Resistant (MDR) TB.

TB mainly affects the lungs (pulmonary TB), however, other organs such as the brain, joints, central nervous system, lymph and circulatory system may also be affected. An infected person shows respiratory symptoms such as emaciation, low fever, night sweats among many other symptoms (Zhou *et al.,* 2017). It is mostly spread when an infected person expels the bacteria into the air through coughing or sneezing (Delogu *et al.,* 2013; WHO, 2020). Failure of an individual's innate immune defense mechanism to eliminate the bacteria leads to its replication and spreads to other organs and tissues. The cell-mediated response attempts to control the bacterial replication resulting in latent TB (dormant *bacilli*) with no symptoms or signs of the disease which can last for days or even years. A decrease in cell mediated response mechanism leads to manifestation of the disease and thus active TB (Delogu *et al.,* 2013). This process makes up the stages of TB which are exposure, latent, and active TB. Patients with latent TB are at risk of reactivation of the disease, which is one major problem in controlling TB globally (Smith *et al.,* 2004).

1.1.2 *Tubercle bacilli*

The causative agent of TB was discovered by Robert Koch in 1882 as a bacillus species called *Mycobacterium tubercle bacilli* (*M. tb*) (WHO, 2020). *M. tb* is a small, rod-shaped, aerobic bacillus bacterium with a slow reproductive cycle of 24 to 48 hours under optimal conditions (Delogu *et al,*

2013; Muller, 2016). It has a thick cell wall made up of mycolic acids and waxy components on its inner and outer layers that prevent harmful agents and antibiotics from entering into the cell (Zhang *et al.,* 2003). The synthesis of these membrane components are the target sites of the effective anti-TB drugs such as isoniziad and ethambutol (Muller, 2016). The cell also has a deficiency of pyrazinoic acid (POA) efflux pumps which make it susceptible to POA derivative drugs like pyrazinamide (Zhang *et al.,* 2003). During a disease process, *M. tb* can be found in micro-environments that are acidic, have nutrient deficiency and low or high oxygen content. All these different environmental conditions lead to the development of heterogeneous bacterial populations that are either non-replicating or growing with different levels of susceptibility to anti-TB drugs (Zhang *et al.,* 2012).

1.1.3 Diagnosis of TB

Different methods have been developed to detected TB infection with each technique having different detection approach. TB infection can be identified using a microscope to visualize sputum smears using the Ziehl-Neelsen technique, chest X-rays, molecular tests such as Gene-Xpert, immunology-based technique like TB LAM test and cultured based methods (Lawn *et al,* 2017; Broger *et al.,* 2019; WHO, 2020). Multiple line probe assays tests for drug resistance have also been developed, however, culture-based assays are the standard susceptibility tests frequently used (Floyd *et al.,* 2018).

1.1.4 Drug resistance

Effective treatment and eradication of TB is being hindered by the emergence of Drug Resistant Tuberculosis (DR-TB) such as MDR-TB and Extensively Drug Resistant (XDR-TB). MDR-TB is

defined as resistance of *M. tb* to at least isoniazid and rifampicin, while XDR-TB is the resistance to both first line and second line injectable TB drugs including fluoquinolones (Zhang and Yew, 2015). Resistance against newly discovered drugs targeting both growing and non-growing *M. tb* (bedaquiline, pretomanid, delamanid) has been reported (Zhang and Yew, 2015). Studies revealing the vital factors of TB virulence including the unique components of its cell membrane contributing to virulence and persistence have also been done (Smith *et al*, 2004).

Zhang and Yew (2015) stated that chromosomal gene mutations and protein modifications account for the two types of drug resistance in *M. tb*. These are genetic and phenotypic resistance respectively. Mutations on the genes encoding the proteins that are targeted by the present anti-TB drugs is the major cause of *M. tb* resistant strains which might occur as a result of sub-optimal physician prescription, failure of patient compliance and bacterial efflux pump (Zhang and Yew, 2015; WHO 2020; WHO, 2020). The resistant gene may also be passed from one individual to the other (Zhang and Yew, 2015) and the majority of MDR-TB and XDR-TB cases are due to the transmitted Beijing genotype in China, Europe and Africa (Zhang and Yew 2015). Zhang *et al* (2012) also states that persister bacteria, with the potential and ability to survive antibiotic stress, are one of the main causes of prolonged TB treatments and drug resistance. Multiple clinical experiments have demonstrated that persisters that can be found in lesions, sputum or adipose tissue, are the problem cause in TB relapse and drug resistance (Connolly *et al.*, 2007; Zhang *et al.*, 2012).

1.1.5 TB treatment

Statistics conducted before drugs for TB treatment became available reported that 70% of the people diagnosed smear positive with pulmonary TB died within 10 years and overall 40% of the people that tested positive for all forms of clinical TB died (Floyd *et al.,* 2018). A decrease in incidence and mortality rate was obtained from around the 1940s after the introduction of effective anti-TB drugs, resulting in TB being regarded as a disease of the past. However, it has remained an infectious disease responsible for the highest number of deaths globally (Floyd *et al.,* 2018, WHO, 2020).

Currently, the available TB vaccine, Baccille Calmette-Guerin (BCG), is only effective in preventing severe forms of TB infections in children and is restricted to HIV negative children. No vaccine has been synthesized for adults (Floyd *et al.,* 2018, WHO, 2020). The currently used first line drugs for drug-susceptible TB are isoniziad (INH), rifampicin (RMP), ethambutol (EMB) and pyrazinamide (PZA) (Zhang and Yew, 2015; Zhou *et al.,* 2017; WHO, 2019). These drugs are prescribed over a period of 6 months with a minimum success rate of 85% (WHO, 2020). Generally, RMP interferes with RNA synthesis and binds to *rpoB* forming a hydroxyl radical. On the other hand, INH attacks the enolyl acyl carrier protein reductase enzyme and inhibits mycolic acid synthesis in cell wall while EMB targets arabinosyl transferase, resulting in no synthesis of cell wall arabinogalactan (Zhang and Yew 2015). Table 1.1 gives a summary of some drugs used in TB treatment.

Table 2.1: Summary of the common drugs used in TB treatment (Zumla *et al.,* 2013; Zhang and Yew, 2015; Zhou *et al.,* 2017; WHO, 2019; Barozi, 2020).

Group	Drug name
First line drugs	Isoniziad, Rifampicin, Ethambutol, Pyrazinamide
Injectable drugs	Kanamycin, Aminkacin, Streptomycin, Capreomycin
Fluoroquinolones	Moxifloxacin, Levofloxacin, Gatifloxacin
Second -line drugs	Ethionamide, Terrizidone, Para-amino salicylic acid, Prothionamide
Drugs with unclear efficacy	Amoxicillin, linezolid, clofazimine

Majority of the developed antibiotic drugs are based on their effect on growing bacteria with little or no activity on persister bacteria (Cogan, 2006). Since one of the main causes of different anti-TB drug susceptibility is due to persisters, drugs with mechanisms that target persisters will greatly improve treatment of TB (Zhang *et al.,* 2012). Apart from developing drugs that target persisters, manipulating the host immune system by enhancing and utilizing its defense mechanism with vaccines and immuno-modulating agents, stimulating innate and adaptive immunity may assist in preventing and quickening recovery from the disease (Bishop *et al.,* 2001).

Some anti-TB drugs like rifamycins and fluoroquinolones primarily target growing bacteria but also have limited activity on non-growing persisters. The few identified persister-active compounds are not readily bioavailable and have high toxicity, thus need further studies for optimization (Cogan, 2006). In 1944, Bigger proposed an intermittent drug dosing approach to allow persisters to grow in the absence of antibiotics and become susceptible to drugs (Bigger, 1944; Lewis, 2012), however, his model was discovered to be practical *in vitro* as complex conditions are encountered *in vivo*

(Cogan, 2006). The development of pyrazinamide drug that targets persisters by disrupting vital processes needed for their survival in stressful conditions has greatly improved the treatment regimens in TB (Bishop *et al.,* 2001; Cogan, 2006).

1.2 PYRAZINAMIDE (PZA)

Pyrazinamide is a pro-drug used in the therapy of TB alongside other first line drugs. It is also used to prevent relapse of TB and is incorporated with second line drugs in treating drug susceptible and resistant *M. tb* (Juniad *et al.,* 2018; Sheik Amamuddy *et al.,* 2020). The drug pyrazinamide (PZA) was chemically synthesized in 1936 by Dalmer and Walter, and was later discovered as an anti-TB drug in 1952 based on the effects of its analog, nicotinamide, on mycobacteria in animal models (Zhang and Mitchison, 2003; Zhang *et al.,* 2014). It was initially used as a second line TB drug due to its high hepatic toxicity effects which were caused by the high dosages (3.0 g daily) and prolonged treatment periods. Further studies later discovered lower dosage concentration (1.5 to 2 g daily) with effective sterilizing effects and in synergy with other TB drugs (Zhang *et al.,* 2014). It has a molecular weight of 123.1, melting point of 188-189 ^{0}C and molecular formula $C_5H_5N_3O$ (Zhang and Mitchison, 2003; Zhang *et al.,* 2015). PZA has poor solubility in organic solvents and dissolves in water (15 mg/ml) at room temperature while its active derivative pyrazinoic acid (POA) readily dissolves in organic solvents like dimethyl sulfoxide (DMSO) (Zhang and Mitchison, 2003).

Pyrazinamide has been reported to have excellent sterilizing bactericidal effect *in vivo* against *M. tb* while having no notable activity *in vitro* (Peterson *et al.,* 2015; Gopal *et al.,* 2016). Its activity *in vitro* can be induced in the presence of efflux pump inhibitors, in mild acidic conditions or in media with nutrient deficiency, anaerobic conditions and molecules that alter energy metabolism. It has a bactericidal effect on semi-dormant *tubercle bacilli* in a pH range of (4.8-5.0) (Zhang and Mitchison,

2003). According to Juniad *et al* (2018), the use of PZA in TB treatment cuts the therapy time by 33% from the standard 9 to 12 months to 6 months. Additionally, its removal from TB therapy reduces other drugs' efficiency to destroy bacterial cells (Bishop *et al.,* 2001). However, the 6 months therapy period is still long enough to facilitate drug resistance development and patient noncompliance, thus the goal is to develop drugs for TB treatment within 2 months or less (Gopal *et al.,* 2016).

Pyrazinamide in humans and murine models have been reported to be effective in the first two months of TB treatment in combination with the first line drugs (Zhang and Mitchison, 2003; Ahmad *et al.,* 2013). However, Ahmad *et al* (2013) states that in second line regimen of murine models, PZA contributes sterilizing activity beyond two months when incorporated with streptomycin and isoniazid which also resonates with his obtained study with moxifloxacin and levofloxacin. Failure of PZA activity beyond the first two months may be due to overlapping sterilizing effect with rifamycins or antagonistic effects with isoniazid (Ahmad *et al.,* 2013; Zhang *et al.,* 2014). Apart from the good antibacterial effect exhibited by PZA, it has negative side effects of damaging the liver, therefore alternative compounds with trivial side effects need to be identified (Zhou *et al.,* 2017).

1.2.1 Mechanism of PZA

In spite of the significant role played by PZA in TB treatment and its inclusion in all TB drug combinations, its mechanism is the least understood among all anti-TB drugs (Zhang *et al*., 2013; Gapol *et al*., 2016; Sheik Amamuddy *et al.,* 2020). Its role as a persister drug has attracted a lot of attention in trying to understand its mode of action in TB treatment and also in developing drugs for other persistent infections (Cagon, 2006).

Different mechanisms of PZA on *M. tb* have been suggested (Lamont *et al.,* 2020). Multiple studies revealed that the sterilizing activity of PZA depends on the acidic environment in the lesion caused by inflammation, explaining its use only in the first two months of therapy (Zimhony *et al.,* 2000; Shi *et al.,* 2011; Zhang *et al.,* 2013) while further studies demonstrated the independence of PZA bactericidal effect in nearly neutral and alkaline conditions with critical accumulation concentrations of POA (Dillion *et al.,* 2014).

The accepted model for PZA mechanism involves the acidification of cytoplasmic bacilli by POA mediated proton shuttling (Figure 1.2) (Zhang *et al.,* 2013; Peterson *et al*., 2015). PZA is activated only in acidic conditions (Bishop *et al.,* 2001), and at this low pH, POA becomes toxic and inhibits the growth of *M. tb* (Juniad *et al.,* 2018). PZA enters into *M. tb* by passive diffusion (Zhang and Mitchison, 2003; Junaid *et al.,* 2018) and possibly by active transport (Zhang and Mitchison, 2003). Initially, POA is formed as a charged anionic form with no bactericidal effect in a neutral cytoplasmic environment, however, in acidic conditions, POA is excreted by a weak efflux pump and is converted to uncharged protonated POA (HPOA) which accumulates in the cell and eventually kills the cell (Zhang *et al.,* 2003; Zhang and Mitchison, 2015; Juniad *et al.,* 2017). Protonated POA acidifies the cell cytoplasm and affects bacterial cell activity by inhibiting functioning of vital enzymes such as ribosomal protein S1 *rpsA* and *panD* which are involved in translation, co-enzyme A synthesis and charges on the cell membrane thus affecting membrane transportation (Zhang *et al.,* 2013; Zhang and Yew, 2015).

Several target sites and pathways for TB treatment have been proposed and studied (Shi *et al.,* 2014, Gopal *et al.,* 2020; Smith *et al.,* 2004). Zimhony *et al* (2000) suggested fatty acid synthase1 (FAS1) as a molecular target site of PZA, however these findings were later refuted by Boshoff *et al.,* (2002). Studies by Shi *et al* (2011) identified 30S ribosomal protein S1 (*rpsA*) as a target site for PZA/POA which inhibits trans-translation process. However, Personne (2014) have shown that strains with a defective trans-translation pathway were still susceptible to PZA and Dillion (2017) also concluded that PZA is independent of *rpsA* and trans-translation.
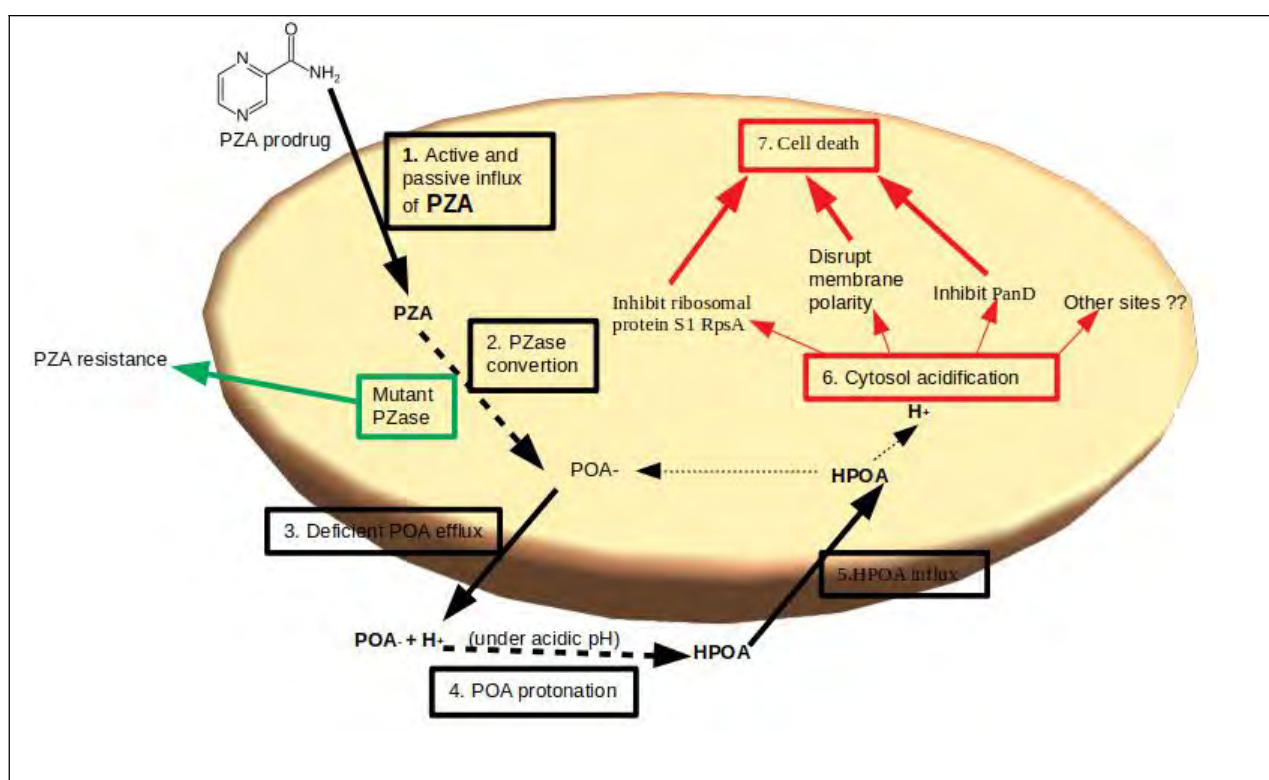
Figure 1.2: Schematic representation of pyrazinamide mechanism in *Mycobacteria tuberculosis* (mechanism generated from Zhang *et al.,* 2013; Peterson *et al*, 2015; Junaid *et al.,* 2018; Zhang and Mitchison, 2003)

rspA and *panD* sites have been targeted as potential drug sites for PZA/POA (Zhang *et al.,* 2003; Zhang and Yew, 2015; Gopal *et al.,* 2020). The study by Shi *et al* (2014) suggested *panD* as the target site for PZA/POA. A recent study by Gopal *et al* (2020) supported Shi *et al* (2014) findings and revealed the mechanism of POA as a weak *panD* inhibitor that triggers degradation of the

enzyme by caseinolytic protease. Degradation of this enzyme blocks the synthesis of co-enzyme A, an enzyme involved in energy metabolism reactions. Gopal studies in 2016 and 2017 have illustrated that the use of PZA/POA in TB treatments reduces the levels of co-enzyme A. This site however, is not the exclusive target site of the PZA drug as conditional susceptibility was observed in strains without the *panD* region (Dillion *et al.,* 2017). The mechanism of PZA remains elusive and further studies are still to be done (Sheik Amamuddy *et al.,* 2020).

1.2.2 Alternatives to PZA

PZA derivatives can be studied for their activity against *M. tb*. Some synthesized PZA derivatives were studied by Zhou *et al* (2017), who discovered N-(3-thiomorpholinopropyl)pyrazine-2-carboxamide as a potential compound, however further analysis is needed. One of the PZA derivatives is morphazinamide (MZA) which has a similar impact on *M. tb* in both acidic and neutral pH conditions. It is converted to PZA, formaldehyde and morphiline in bacterial cells. However, its use in animal TB treatment is inferior to PZA regardless of its high activity in in vitro models thus considered less useful in TB treatment (Zhang *et al.,* 2003). A synthetic analogue of PZA, 5-chloro-pyrazinamide, is also active against *M. tb.* However, its mode of action is independent of the enzyme pyrazinamidase and has no effect on M. tb in mouse models thus makes it different to PZA (Zhang *et al.,* 2003; Zhang *et al.,* 2014). Studies have also shown that esters of POA have anti-TB properties similar to the mechanism of PZA in vitro studies, but have failed to show significant activity against *M. tb* in vivo (Zhang *et al.,* 2003). Therefore, PZA remains the best prodrug for TB treatment regardless of the emerging PZA-resistant *M. tb* strains. However, identification of an alternative drug that mimics the mechanism of PZA with little to no side effects to counteract drug resistance is crucial for progress in TB eradication.

1.3 PYRAZINAMIDASE

Pyrazinamidase (PZase) has been identified as the enzyme responsible for the conversion of prodrug PZA into its active form pyrazinoic acid (POA) (Juniad *et al.,* 2018). This enzyme is present in microorganisms including *M. tb, Saccharomyces cerevisiae, Acinetobacter baumani*i, *Escherichia coli* and *Pyrococcus* (Rasool *et al.,* 2019). The overall structure of *M. tb* PZase is similar to the crystal structures of *A. baumanii* and *P. horikoshi,* however, with crucial variations (Petrella *et al.,* 2011). *P. horikoshii* PZase was shown to contain Zn^{2+} ion in its crystal structure (Du *et al.,* 2001) while *A. baumanii* PZase had Zn^{2+} and Fe^{2+} ions in the ratio 1:1 (Fyfe *et al.,* 2009) and *M. tb* PZase contains Fe^{2+} (Petrella *et al.,* 2011). PZase is located in the cytoplasm and is encoded by the *pncA* gene of *M. tb* (Zhang *et al.,* 2003). According to Zhang *et al.,* (2003), PZase is also responsible for the conversion of nicotinamide to its acidic form nicotinic acid, which is used in making nicotinamide adenine dinucleotide (NAD) in bacterial species (Zhang *et al.,* 2003).

1.3.1 Structure and mechanism

M. *tb* PZase is a metallo-enzyme with an amidase activity, made up of alpha helices surrounding six parallel beta-sheets (Juniad *et al.,* 2018; Sheik Amamuddy *et al.,* 2020) (Figure 1.3) with a substrate binding cavity of approximately 7Å wide and 10 Å deep (Petrella *et al.,* 2011). Its metal binding site (MBS) consists of an iron (Fe^{2+} ion) coordinated by two water molecules (H2O 220, H2O 221), three histidine and one aspartate residues (H51, H57, H71, D49) in a distorted tetragonal bipyramidal arrangement (Figure 1.4a) (Juniad *et al.,* 2018; Petrella *et al.,* 2011). These residues leave a space on the cavity for PZA binding. It has also been reported that *M. tb* PZase might contain low amounts of manganese and zinc or may contain them in the metal binding site (Petrella *et al.,* 2011) and some few metal ions can competitively bind as co-factors altering the structure and function of the enzyme (Rasool *et al.,* 2019).
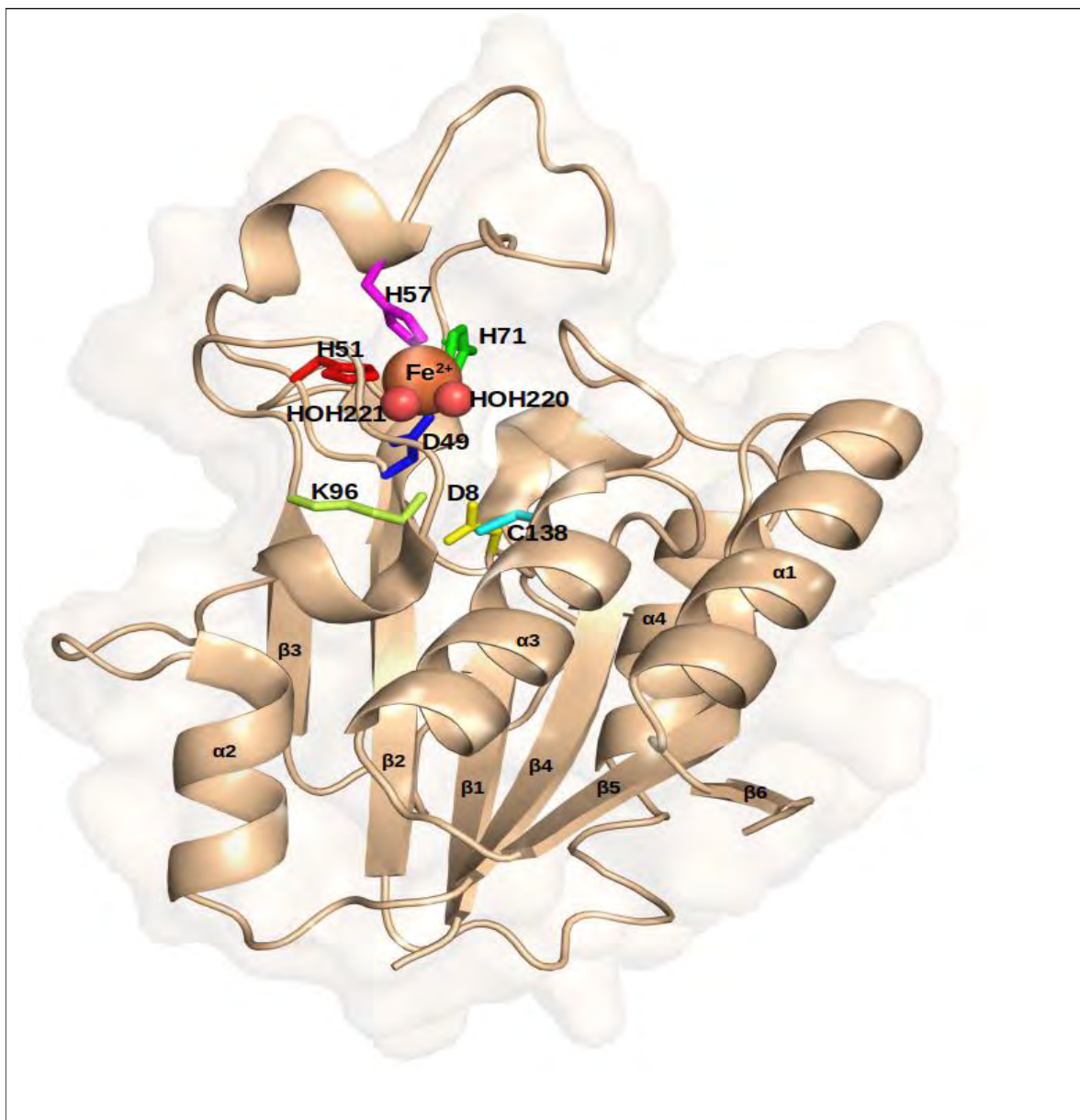
Figure 1.3: Crystal structure of M. tb PZase (PDB ID:3PL1) showing the alpha helices, loops and six beta pleated sheets in wheat. The metal ion (Fe^{2+}) is represented as an orange sphere while water molecules are red spheres. The active site and metal binding site amino acid residues are represented as sticks. The structure was generated in PyMOL.

The active site residues Asp8, Lys96 and Cys138 (Figure 1.4a) located at the end of β strand 1, β strand 3 and at the N-terminal of alpha helix 3 respectively, make up a catalytic triad (Juniad *et al.*, 2018; Petrella *et al.*, 2011). According to Petrella *et al* (2011), the key residue Cys138 is involved in the nucleophillic attack of PZA while Asp8 and Lys96 act as the activating base and stabilizer of the

acylenzyme. The addition-elimination mechanism found in *A. baumanii* also applies in *M. tb* which results in the release of ammonia and the formation of acyl-enzyme intermediate after a nucleophilic attack on PZA carbonyl carbon by the thiolate form of Cys138 supported by Asp8. In the final hydrolytic step, the formed Acyl-enzyme intermediate is hydrolysed by water (H_2O 202) (Petrella *et al.,* 2011). Figure 1.4b shows the general activity of PZase/pncA on PZA and nicotinamide as explained above.
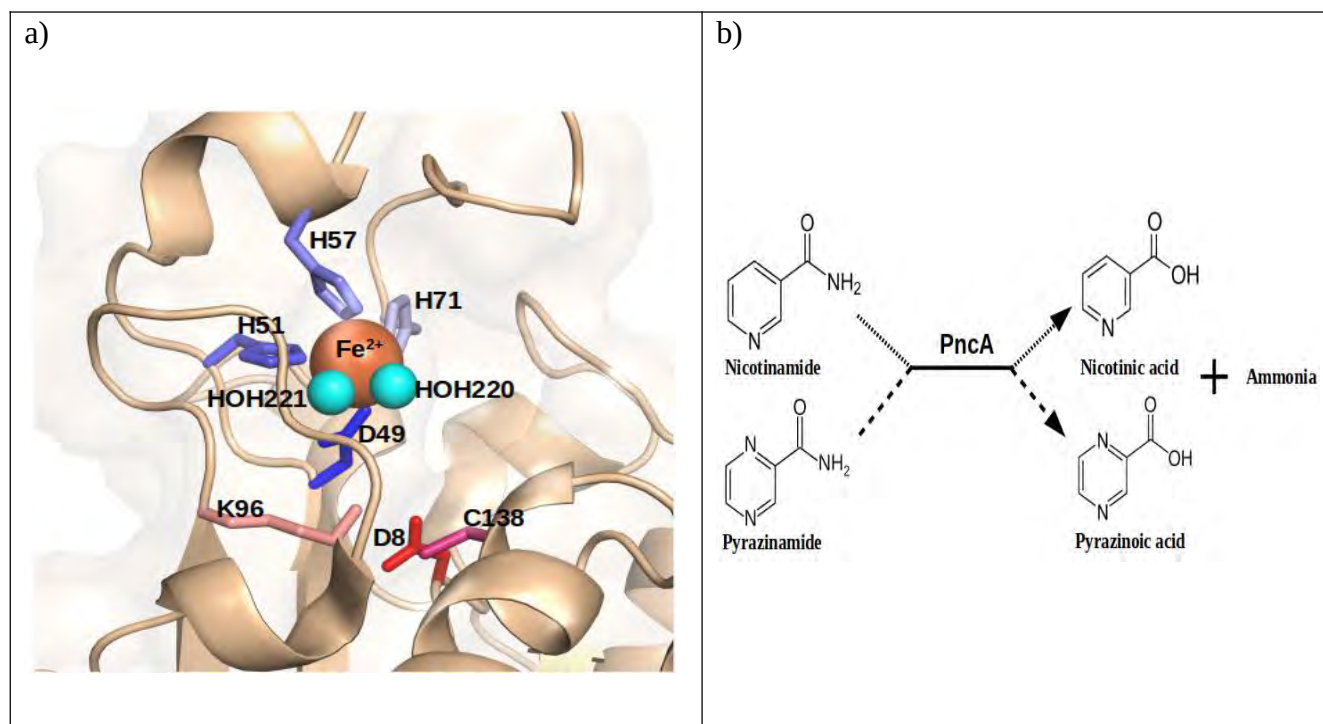
Figure 1.4: **a)** Catalytic triad of *M. tb* pyrazinamidase with metal binding site residues in blues (H51, H57, H71, Asp49) and active site residues in reds (Asp8, Lys96, Cys138). The metal ion (orange) and water molecules (H_2O 220/221) in cyan spheres **b)** Primary function of Pzase/pncA on nicotinamide and pyrazinamide to produce nicotinic acid and pyrazonoic acid with ammonia as a by-product respectively (active site residues visualized in PyMOL, equation generated from Zhang *et al.,* 2013).

The protein has 52-70 residues that form a loop lid controlling access to the active pocket, balanced by H51 and H71 residues (Sheik Amamuddy *et al.,* 2020). On the extended part of this loop, there is a conserved residue found in *M. tb*, *A. baumanii* and *P. horikoshi* species (Trp68 residue), which is located above the catalytic site and delineates binding site sides (Petrella *et al.,* 2011). The active

site also has a cis-peptide bond between Ile33 and Ala134 residues that draws the nitrogen atom of Ala134 towards the center, forming an oxyanion hole with nitrogen from Cys138 that is occupied by a water molecule (H_2O 221) (Petrella *et al.*, 2011).

1.4 PYRAZINAMIDE RESISTANCE

Current studies have reported some resistance in the use of PZA prodrug (Juniad *et al.,* 2018). This can be accounted for by the loss of PZase enzyme activity (Zhang and Mitchson, 2003; Kim *et al.,* 2012; Zhang and Yew, 2015). A recent study by Juma *et al (*2019) in Tanzania revealed that 50% of patients with MDR-TB and 21.3% with drug sensitive TB also had PZA resistance. There have also been reports on increasing rate of PZA resistance cases (Sheen *et al.,* 2020). Determination of resistance of *M. tb* to PZA is difficult to interpret and analyze using acidic media approaches thus sequencing of the gene *pncA* is often done to identify mutations that are related to PZA resistance (Petrella *et al.,* 2011). A study by CRyPTIC (2018) showed that correct sequencing of the *M. tb* isolates on *pncA* gene predicted PZA resistance with 96.8% specificity and 91,3% sensitivity and susceptibility, accommodating other factors as causes for PZA resistance.

Strains of *M. tb* that are resistant to PZA have been found in isolates lacking *pncA* mutations (Simons *et al.,* 2013). These results supported the results obtained by Sreevatsan *et al.,* 1997; Kim *et al.,* 2012; Shi *et al.,* 2014, whose analysis also showed that approximately 0% to 30% (depending on geographical region) of *M. tb* isolates show PZase inactivity in the absence of *pncA* mutation thus suggesting alternative mechanisms for PZA resistance. Multiple studies have been done in attempt to understand the mechanism of PZA resistance (Bishop *et al.,* 2001; Shi *et al.,* 2011; Zhang *et al.,* 2014; Shi *et al.,* 2014; Palmer and Kishony, 2014; Gopal *et al.,* 2016; Dillon *et al.,* 2017).

Analyses of genetic sequences of PZA-resistant strains reveal no PZA-associated prominent clustering mutations apart from those identified in prodrug activating pyrazinamide (Gopal *et al.,* 2016). Although the resistance of PZA can be encoded by three genes (*pncA, rpsA* and *panD*), *pncA* gene contributes up to 72-99% of the mutations (Juniad *et al.,* 2018). Resistance of this drug according to Junaid and colleagues is due to scattered mutations in the coding and promoter region of the *pncA* gene (Kim *et al.,* 2012; Juniad *et al.,* 2018). The most identified type of mutations in the *pncA* gene are missense and nonsense mutations which cause amino acid substitution or nucleotide insertions or deletions (Zhang and Mitchison, 2003).

Mutations on the *pncA* gene are commonly found on three regions 3-17, 61-85 and 132-142 which have an effect on the folding of the active site (Zhang *et al.,* 2003; Sheik Amamuddy *et al.,* 2020). These mutations are responsible for 54% of PZA resistant strains (Sheik Amamuddy et *al.,* 2020). A study on various PZase species by Lemaitre *et al.,* 1999 showed that these three regions are highly conserved therefore alteration on these regions affect the structural or catalytic function of the enzyme. This study also suggested Cys138, Ala134, Thr135, Trp68 and Asp8 as the key residues for PZA hydrolysis and stated that mutations on these specific residues or close to these residues causes active site modifications leading to the loss of PZase activity and PZA resistance (Lemaitre *et al.,* 1999). Apart from acquired PZA resistance through mutations, natural PZA resistance also occurs in some mycobacteria such as *M. bovis* that have a point mutation on the *pncA* gene at nucleotide 169 from Cytosine to Guanine which replaces Histidine (H57) in *M. tb* with Aspartic acid in *M. bovis. The M. kansasii* and *M. smegmatis* species are also naturally resistant to PZA due to the weak activity of PZase enzyme and highly active POA efflux pump respectively (Zhang *et al., 2*003).

1.5 PZase-PZA COMPLEX AND RESISTANT MUTATIONS

The mechanism of resistance to the drug PZA remains unclear and is yet to be determined (Sheik Amamuddy *et al.,* 2020). Due to the increase in *M. tb* PZA resistant strains, there is need for new approaches to identify novel drug target sites and new drugs for TB treatment. Previous studies have been focusing on the wet laboratory experiments such as PZA susceptibility tests (Yoon *et al.,* 2014; Morlock *et al.,* 2017). Computational methods can aid in analyzing the effect of mutagenicity on PZase function (Rasool *et al.,* 2019).

PZase-PZA complex can be analysed by computational methods. Recent studies using computational approaches have been focusing on the effect of mutations on the metal ion and PZA binding (Sheen *et al.,* 2012; Kadem-Maaref *et al.,* 2017; Khan *et al.,* 2018; Rasool *et al.,* 2019; Sheik Amamuddy *et al,* 2020). Mutations that have been identified to have major effects on PZase activity include Ala3 to Gly17, Thr61to Leu85 and Gly132 to Thr142 (Juniad *et al.,* 2017). Juniad *et al* (2017) however stated that mutations on other sites might also have an effect on the solubility, structure and function of the protein. Previous studies have also demonstrated that mutations on the metal binding site lowers the binding affinity of the co-factor ion which is crucial for the activity of PZase (Sheen *et al.,* 2012; Rasool *et al.,* 2019;).

According to Petrella *et al.,* 2011, mutations on *pncA* gene not only affect catalytic functioning of the protein, but also affect the thermal stability and folding properties of the protein. Their study on mutating residues that make up the catalytic triad (Asp8Glu, Cys138Ala, and Lys96Gln), substrate binding site (Phe13Leu and Trp68Leu), metal binding site (His51Ala, Asp49Gly, and His57Asp) and the oxyanion hole formation (Ala134Val) resulted in none or low PZase activity. These results

demonstrated the importance of specific residues for PZase activity and also support that mutations have a great effect on the integrity of the 3D structure and activity. Khan *et al* (2018) study showed that the mutations Leu19Arg, Arg140His and Glu144Lys on PZase caused changes in the protein stability, flexibility, activity and in the binding pocket size when analyzed using molecular dynamics simulations.

A study by Rasool *et al.,* 2019based on the Density Functional Theory (DFT) approach reported that mutagenicity on PZase is detrimental to its activity and results in weak binding of its co-factor metal (iron) and the prodrug PZA. The study revealed that the mutations Asp12Gly, Asp12Ala, Thr135Pro and Asp136Gly weaken the binding of PZA as these mutations occur close to the active site. Junaid *et al* (2018) investigated the effect of mutating Asn11 to Lys, Pro69 to Thr and Asp126 to Asn on the *pncA* gene. Their study showed that these mutations resulted in an increase in fluctuations in the mutant protein compared to the wild type, indicated by a weakened binding of PZA to the active site and an alteration in the active site volume which in turn altered the binding of PZA to PZase.

A study by Kadem-Maaref *et al* (2017), analyzing the effect of different metals on PZase function using DFT model revealed that cobalt and nickel are more active than iron and can effectively replace it as a co-factor while magnesium, zinc and copper decreases its activity. These results resonate with those obtained by (Sheen *et al.,* 2012 and Rasool *et al.,* 2019). Sheik Amamuddy *et al* (2020) study investigated the mechanism of PZA resistance by studying the unbinding events of PZA on the WT and mutant Pzase. Their study revealed that mutations on MBS residues caused iron ion delocalization which led to the opening of the lid and unbinding of PZA.

1.6 DRUG REPURPOSING

Despite the advancement in technology and knowledge on human disease states, the traditional development of new drugs is substantially expensive and time consuming (Rudrapal *et al.,* 2020). Therefore, the use of already discovered drugs (drug repurposing or repositioning) is being adopted for lower costs and shorter timelines to treat common and rare diseases. Drug repurposing is a strategy that identifies new alternative uses on drugs that have been approved, discontinued, abandoned or under experimental investigations to target other medical conditions (Elder and Tindall, 2020; Khan *et al.,* 2020). The approach generally follow three steps of i) identifying a potential molecule for the given state, ii) preclinical assessment of drug effects and iii) efficacy evaluation in clinical trials (Pushpakom *et al.,* 2019). According to Rudrapal *et al.,* (2020) approximately 30% of US Food and Drug Administration approved drugs and vaccines are as a result of repositioned drugs. Some of the common effective drugs from repositioning are minoxidil, aspirin, valproic acid including sildenafil (Viagra) which was initially developed for hypertension and angina pectoris treatment but has been repurposed to treat erectile dysfunction (Pushpakom *et al.,* 2019; Elder and Tindall, 2020; Khan *et al.,* 2020; Rudrapal *et al.,* 2020).

Various computational approaches like signature matching, genome-wide association and molecular docking and/or experimental approaches like phenotypic screening and binding-target assays may be used to identify repositioning opportunities (Pushpakom *et al.,* 2019; Rudrapal *et al.,* 2020). The experimental-based approach is a protein target-based and cell-based screening method of original drugs for new pharmacological effects. The *in silico* approach is based on molecular interactions between protein and drug molecules through virtual screening of drug databases using computational biology tools (Rudrapal *et al.,* 2020). In this study, the computational molecular

docking approach will be used to screen multiple drugs against target protein PZase (conventional docking).

1.6.1 Advantages and limitations of drug repurposing

Advantages of drug repurposing include low chances of failure as the drug would have been approved and successfully tested in preclinical models and humans thus also implies less time and less investment in preclinical tests (Pushpakom *et al.,* 2019; Khan *et al.,* 2020). According to Rudrapal *et al* (2020), the average traditional approach requires 10-16 years to develop a new drug while 3-12 years are required to design a drug through drug repurposing. Repurposed drugs may also reveal off-target or on-target effects, exposing new target sites and pathways that can be further exploited as potential drug-target sites (Pushpakom *et al.,* 2019).

Successful drug repurposing has been achieved through both computational and experimental approaches, with fast screening on large data using computational methods compared to experimental methods. However, there are barriers hindering the wide use and success rate of these techniques. The challenges faced include legal and intellectual property barriers and organizational hurdles which require collaborations from pharmaceutical companies, biotechnology firms and academic communities.

1.7 PROBLEM STATEMENT

Tuberculosis has been declared a global health emergency by the WHO with South Africa being reported to be among the high burden countries that contribute two thirds to the overall TB incidences. The World Health Organization has set a target to eradicate TB by 2035 aiming to reduce deaths and incidence rates by 90% and 80% respectively between 2015 and 2030 (WHO, 2020). The emergence of MDR-TB and XDR-TB *M. tb* strains has become a major public health problem threatening the progress made in TB treatment worldwide. In order to accomplish the goal set, intense research and development of new novel drugs and drug targets has to be done. The prodrug PZA with bactericidal effect on semi-dormant mycobacterium tuberculosis has been identified as a critical drug needed in all TB treatment combinations, reducing therapy time from 9 months to 6 months (Gopal *et al.,* 2016; Juniad *et al.,* 2018). However, various mechanisms contributing to PZA drug resistance have been reported (Zhang and Mitchson, 2003; Kim *et al.,* 2012; Zhang and Yew, 2015) with previous studies focusing on wet laboratory experiments and *in silico* analysis on the effect of mutations on active site and metal binding site of pyrazinamidase (Sheen *et al.,* 2012; Kadem-Maaref *et al.,* 2017; Khan *et al.,* 2018; Rasool *et al.,* 2019; Sheik Amamuddy *et al.,* 2020). Exploration on alternative compounds to PZA has been partially conducted in wet laboratory research (Zhang *et al.,* 2003; Zhang *et al.,* 2014; Zhou *et al.,* 2017) with no *in silico* approaches to identify compounds against PZase active site. This study aims to identify novel selective compounds against the active site of *M. tb* PZase and analyse their behaviour in the presence of mutations. The identification of scaffolds or selective novel compounds against *M. tb* PZase might lead to the design of more effective drugs, for cure and eradication of TB.

1.8 HYPOTHESIS

The study hypothesizes that DrugBank database have compounds that can selectively bind to the active site of *M. tb* PZase, mimicking the behavior of PZA prodrug. Therefore, computational techniques of virtual screening and dynamic simulations may be employed for the identification of novel compounds for drug repurposing in TB.

1.9 AIM

The study aimed to virtually screen DrugBank compounds against the active site of M. *tb* PZase and analyze the effect of point mutations on the identified hit compounds. This is so as to identify scaffolds that may lead to the development of effective TB drugs.

1.10 OBJECTIVES

1. To identify potential DrugBank compounds that selectively bind to PZase active site by performing *in silico* molecular docking studies.

2. To perform Molecular Dynamics (MD) calculations on wildtype PZase to identify stable protein-drug complexes.

3. To introduce mutations on PZase-DrugBank complexes and study the effect of the mutations on the complexes.

CHAPTER TWO

2. STRUCTURE BASED VIRTUAL SCREENING

2.1 CHAPTER OVERVIEW

Compounds that bind on to the active or allosteric site of a target protein can induce conformational changes and either inhibit or promote the protein's functionality. The aim of this chapter (Chapter 2) was to identify potential DrugBank (Wishart *et al*, 2018) compounds that can selectively bind onto the active site of wild type PZase with better binding characteristics compared to the control Pyrazinamide. The entire protein surface was subjected to two thousand and eighty-nine (2089) DrugBank compounds through molecular docking using AutoDOCK 4.2 (Morris *et al.,* 2009). The protein and PZA ligand structures were prepared using AutoDOCK4 Tools (Morris *et al.,* 2009). Pyrazinamide was used as a control and the docking parameters were validated by redocking PZA onto the wildtype PZase and comparing its pose and interactions with those obtained by the previous group study of Sheik Amamuddy *et al.,* (2020). The best hit compounds were selected based on lowest binding energy, active site proximity, low molecular weight and presence of hydrogen bond interactions before being subjected for further analysis. This chapter provides a brief introductory description on high throughput virtual screening, applied docking methodology, discussion of the results obtained and summarized conclusion.

2.2 INTRODUCTION

Proteins interact with other molecules and their functionality in biological processes occur through recognition of other molecules. Identification of the small molecules protein-target site provides insight on the underlying molecular modes of actions, giving information on their pharmaceutical

effects (Wang *et al.,* 2012). The first target-substrate binding was reported by Fischer in 1894, who interpreted using the lock and key analogue, based on the protein-substrate complementary shape. However, his interpretation did not explain allosteric modulation and non-competitive binding which lead to other binding models such as the induced-fit by Koshland in 1958 being proposed (Salmaso and Moro, 2018).

2.2.1 Virtual screening

Computational methods have been applied in drug discovery processes since 1980s, leading to the establishment of (Computer Aided Drug Design) CADD techniques which improved from analysis of a rigid ligand-target binding to flexible ligand-protein complexes (Salmaso and Moro, 2018; Nguyen *et al.,* 2019). These techniques are developed primarily for virtual screening hit/lead optimization as well for the designing of novel compounds (Kitchen *et al.,* 2004; Salmaso and Moro, 2018). Virtual screening, which aims to increase the novel compounds hit rate and reduce experimentally tested compounds, accomplishes its goal by screening a large data set of compounds in search for binding capacity for a targeted molecule.

The CADD techniques are grouped as Ligand-based (LB) or Structure-based (SB) methods. The Ligand-based method only depends on information about the similarity of known ligands (Sliwoski *et al.,* 2014). In this study, the Structure-based method was used, which depend on the crystal structure of the target molecule (obtained by NMR, Xray crystalography or from homology modelling) and on the fact that binding to the target structure may be optimized since ligand-target binding is influenced by structure complementarity (Sliwoski *et al.,* 2014; Salmaso and Moro, 2018). The SB technique has made prominent inhibitors of HIV-1 reverse transcriptase (Vadivelan

et al., 2011), heat shock proteins (Doddareddy *et al.,* 2011) and Plasmodium parasite (Chaudhary and Prasad, 2014) among many other therapy fields.

The availability of databases that consist of chemical compounds from natural and/or synthetic origin have enabled easier identification and screening on potential hit compounds. One of the most comprehensive source of small biological molecules is the CheMBL database that is compiled from publications and other chemical databases like PubChem. Generally, only approximately 11 000 compounds are readily available from pure natural products (Kinghorn *et al,* 2019). Several natural product (NP) databases have been developed, however, majority of these databases are specialized including databases such as Super Natural II (Banerjee *et al.,* 2015), one of the largest online database that mainly consist of purchasable compounds and NPAtlas (Van Santen *et al.,* 2019) database focusing on microbial natural products. Some databases like AnalytiCon provides over 2000 semi-synthetic compounds (Kinghorn *et al,* 2019) while other databases like PubChem (Kim *et al.,* 2016) and ZINC (Irwin *et al.,* 2012) consist of millions of mainly commercially available small and large synthetic molecules. The ZINC database is widely accepted as a meta-database of readily purchasable compounds (Kinghorn *et al.,* 2019).

Recently, a free and open generalistic NP database that inco-operates and curates data from various databases (composed of 401,624 compounds), COlleCtion of Open Natural ProdUcTs (COCONUT) has been designed for diverse and advanced searches for NPs (Sorokina *et al,* 2021). In Africa, some key present natural compound databases are SANCDB, a South African Natural Compound Database that is made up of highly curated natural compounds from plants and marine habitats (Hatherley *et al.,* 2015) and NANPDB, a Northern African Natural Database (Ntie-Kang *et al.,*

2017). In this study, DrugBank database compounds were screened in search of orthosteric compounds to the WT PZase using molecular docking technique. DrugBank contains information on millions of drugs and their target sites, their clinical and drug repurposing trials (Wishart *et al,* 2018).

2.2.2 Molecular docking

In order to identify novel ligand compounds, prediction and interpretation of ligand binding modes is vital. A well developed technique that accomplishes this is molecular docking, which predicts and identifies the best ligand orientation to a protein molecule counterpart. It is an important technique because it reduces time and cost to design novel pharmaceutical drugs. The main goals of molecular docking are prediction of pose, virtual screening and estimation of binding affinity (Guedes *et al.,* 2014). The first algorithm for molecular docking was developed by Kuntz *et al* (1982), which was a fully rigid docking technique. According to (Salmaso and Moro, 2018), docking methods can be grouped based on the molecules degree of flexibility (Pagadala *et al.,* 2017).

2.2.2.1 Molecular docking algorithms

Molecular docking process has two distinct steps which are orientation sampling and scoring function. The sampling process searches conformational space while the scoring function associates the bound conformation to the global minimum energy. One type of sampling method is rigid docking, in which both proteins and ligands are treated as rigid molecules. This is based on the key-lock model and is mainly done in protein-protein docking where there are infinite conformational changes to sample. Semi-flexible docking however considers the ligand as flexible and protein as rigid. This samples the conformational changes of the ligand while maintaining the protein

conformation (Salmaso and Moro, 2018). The third method is flexible docking, which treats both ligand and protein molecules as flexible entities (Salmaso and Moro, 2018). This method is computationally expensive and thus techniques with a balance in accuracy and speed are preferred. Majority of the algorithms consider the protein rigid, that is, allow the bond to rotate yet prohibiting bond angle and lengths (Guedes *et al.,* 2014).

The goal of the scoring function is to determine poses in the sampling engine and distinguish correct binding poses from non-binding modes (Salmaso and Moro, 2018). The scoring functions are divided into three main groups which are Empirical (such as GlideScore and LUDI), Knowledge based (DrugScore and GOLD/ASP) and Force-field based (AutoDOCK) scoring functions (Salmaso and Moro, 2018). The force-field based programs approximate the systems energy by calculating the bonded and non-bonded components as by Lennard-Jones and Coulomb function while the Knowledge-based are based on that the ligand-protein interactions are correlated with favourable interactions (Salmaso and Moro, 2018). Other scoring functions employ algorithms from more than one group to develop a multi-phase approach with better scoring, an example is AutoDOCK4 which is a semi-empirical function (Huey *et al.,* 2007). In this study, AutoDOCK4 was used to dock all the ligand molecules to the wild type PZase protein.

2.2.2.2 AutoDOCK4

AutoDOCK is an automated molecular docking program that is computationally characterized by its use of one CPU (Central Processing Unit) core during execution (Santos-Martins *et al.,* 2019). Its semi-empirical scoring function is made up of a force-field that calculates hydrogen bonding, repulsion, desolvation and electrostatics (Morris *et al.,* 2009). The program has two main

algorithms, which are Autogrid4 and AutoDOCK4 (Morris *et al.,* 2009; Lokesh and Kannbiran, 2016).

In AutoDOCK, intermolecular (ligand and target) interactions are calculated based on the search parameters defined by the user. These intermolecular interactions are calculated using the Autogrid4 program that creates energy maps for all the amino acids in the defined search area (Santos-Martins *et al.,* 2019). AutoDOCK4 then calculates the interactions of the ligand to the amino acids using AMBER force field and linear regression approaches. It provides an option to treat the target as rigid or flexible while automatically determining the ligand's flexibility on rotatable and non-rotatable bonds (Lokesh and Kannbiran, 2016).

Ligand poses are generated using the Lamarckian Genetic Algorithm (LGA) which employs a global and local search in genetic algorithms (Santos-Martins *et al.,* 2019). Execution for ligand pose search is terminated when either the given score evaluations or GA generations are met. The ligand pose results are then clustered based on root mean square deviation (RMSD), where generally, if the first cluster has at least 20% of the ligand poses, the search process is considered a success (Santos-Martins *et al.,* 2019). The binding error found in AutoDOCK4 is approximately 2.5 kcal/mol (Morris *et al.,* 2009; Lokesh and Kannbiran, 2016).

2.3 METHODOLOGY

The flow chart below (Figure 2.1) shows the overall procedure applied to successfully dock and select novel hit compounds. The key steps taken include data retrieval, protein-ligand docking

preparations, setting docking parameters, screening and identifying successfully docked orthosteric compounds.
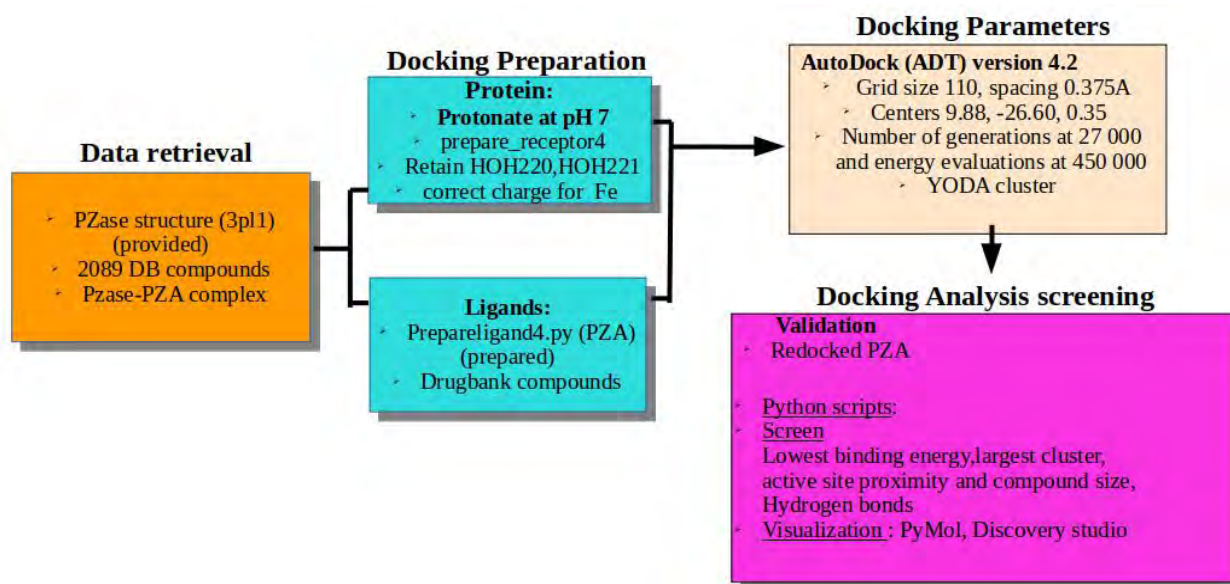
Figure 2.1: Overall applied molecular docking steps, tools and techniques. The docking simulation was perfomed using AutoDOCK 4.2 algorithm. Arrows show the flow of the procedure.

2.3.1 Data retrieval

Due to no available co-crystallized PZase-PZA complex structure in Protein Data Bank (PDB), homology modelling technique was applied to generate a complex structure as described in the study by Sheik Amamuddy et *al.,* (2020). The PZase-PZA complex was provided by the previous group (Sheik Amamuddy *et al.,* 2020). The protein's PDB ID is 3PL1 (Petrella *et al.,* 2011). It is an X-ray diffractioned structure with no co-crystallized molecules obtained at 2.20Å resolution having an R-Value Free of 0.240. A data-set of minimized 2089 DrugBank compounds were also provided by Sheik Amamuddy *et al.,* (2020), who prepared using an in house built script incorporating RDkit tool (Landrum, 2006).

2.3.2 PZase protonation

According to Zhang *et al.,* (2008), PZase is active at a pH of 7.0, therefore in this study, the modelled structure was protonated at pH 7 using the webserver (https://server.poissonboltzmann.org/pdb2pqr) (*Dolinsky et al, 2004).* The protonation states were assigned using PROPKA with the pH set at 7.0. AMBER force field was used as the output naming scheme. Additional selected options were to avoid rebuild of atoms close to existing ones, optimize hydrogen bonding, create an APBS input file and to keep chain IDs in the output file. One key option was not to remove waters from the output file, since there are two crucial water molecules involved in stabilizing the metal binding site. The created .pqr file was edited by adding Iron and renaming the water molecules.

2.3.3 **Protein and compound preparation**

The protein was prepared for docking using the Python command (prepare_receptor4.py -r 'name

of receptor') in AutoDOCKTools/Utilities24 by adding hydrogens and gasteiger charges. The Gasteiger-Huckel method computed the assigning of charges. Non polar hydrogens were mergerd and AutoDOCK atom types were assigned. The output modified file was saved as a .pdbqt file. Since the waters on position 220 and 221 and the Iron metal are required for the functioning of the protein, the water molecules were retained and the metal ion was concatenated to the output file with its charge manually edited to +2.000. The control ligand, PZA was prepared using the Python command (prepare_ligand4.py -l 'name of ligands') which added gasteiger charges, hydrogens and torsions, merged non-polar hydrogens, detected aromatic carbons and rotatable bonds and saved the modified output file in a .pdbqt format. All the 2089 DrugBank compounds were prepared in a

similar manner to PZA and they were provided by the previous group (Sheik Amamuddy *et al.,* 2020).

2.3.4 Docking parameters

Blind docking was done using AutoDOCK (ADT) version 4.2 on all 2089 DrugBank compounds and PZA (as a control) against the PZase enzyme in a Linux based cluster, YODA. The parameters were set in ADT (version 1.5.6). Grid box size was set to cover the whole protein with 110Å in all x,y,z directions (Figure 2.2 a). The grid center box spacing was set to 0.375Å while the x,y,z centers were set to 9.88, -26.60 and 0.35 respectively (Figure 2.2 b). LGA was used to search for 100 conformations with the maximum number of generations at 27 000 and energy evaluations at 450 000. The semi empirical scoring algorithm calculated the interaction energies. All compounds were docked using the above mentioned parameters and tools. The best poses were selected based on largest cluster and lowest energy, hydrogen bonds and center of mass distance (active site proximity).
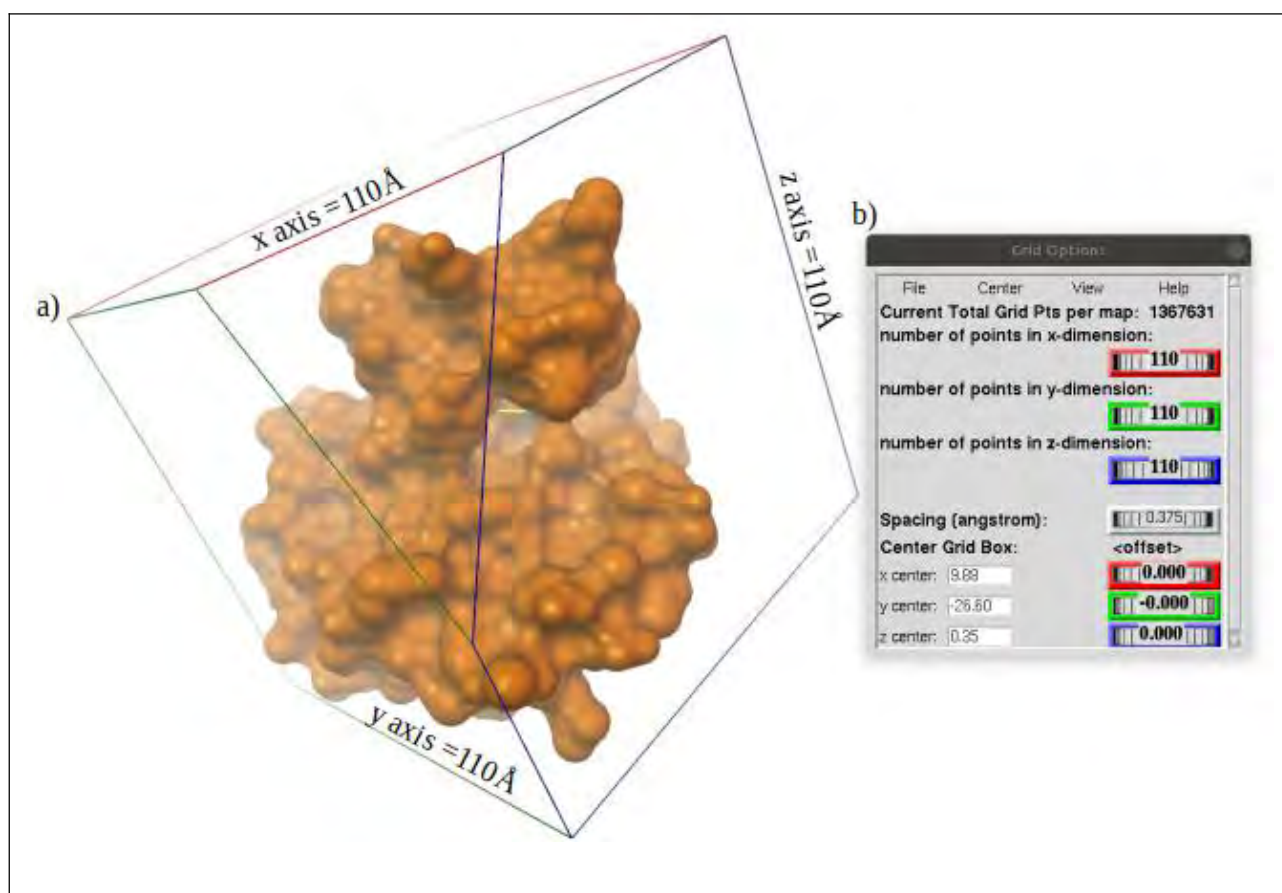
Figure 2.2: Molecular docking box size covering whole protein as used in blind docking. **a)** Visualized grid box dimensions viewed in AutoDOCK Tools. The protein is represented as an orange surface while its center is marked by a yellow line. **b)** The spacing, x,y,z centers and dimension parameters assigned to obtain the required box size. All x,y,z dimensions are colour co-ordinated as red, green and blue respectively.

2.3.5 Initial docking validation

In order to validate the set parameters and the ability of AutoDOCK 4.2 to reproduce the correct poses, the prepared PZA ligand was docked on to PZase using the above mentioned docking parameters. Due to the absence of a co-crystallized PZA-PZase complex, the obtained pose and interactions were compared to those obtained by Sheik Amamuddy *et al.,* (2020). Ligand pose and residues interactions were visualized in BIOVIA Discovery Studio Visualizer and PyMOL.

2.3.6 Criteria for compound selection

Python scripts were used on the resultant files containing the ligand binding energies to separate and extract the best ligand pose based on lowest energy and/or largest cluster. The docking scores predicts the binding potential of the ligand to the target molecule. From the extracted ligands, selection was done based on (i) lowest binding energy, which is the minimum energy required to destabilize system particles, (ii) the presence of conventional hydrogen bonds, (iii) compound size and iv) compound proximity to the active site, which allows for ligand interaction with active site residues.

Proximity of the ligand to the active site was calculated using the center of mass by applying Equation 1 below in a Python script. The Euclidean distance cut off was set at 8Å. The selected ligands were further screened using a cut off binding energy $\leq$ -7.0 kcal/mol, molecular weight $\leq$ 500g/mol and having at least one conventional hydrogen bond with residues in the active pocket. All selected ligands were further subjected for further stability analysis (Chapter 3).

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Equation 1: The Euclidean equation used in the calculation of the center of mass distance. From this equation, x_1, y_1, and z_1 represent the coordinates of the active site and x_2, y_2, and z_2 represent the central point of each ligand.

2.4 RESULTS AND DISCUSSION

The docking procedure was completed in three distinct steps. The steps were (i) docking validation by referencing PZA interactions (ii) blind docking of 2089 DrugBank compounds against PZase and (iii) virtual screening, identification and visualization of the successfully docked potential compounds.

2.4.1 Docking validation

Docking validation was done to determine the reproducibility of the results and to validate the docking parameters used. The redocked PZA best ligand was selected based on the ability to mimic the pose and interactions exhibited by that obtained from Sheik Amamuddy *et al.,* (2020). The PZA pose with binding energy of -4.12 kcal/mol had the best pose. Its interactions with the protein were visualized in PyMOL and BIOVIA Discovery Studio Visualizer. The refRSMD was 0.6 which indicates the difference between the expected outcome and the exhibited pose. A low value shows less divergence from the expected pose.

Figure 2.3 shows the superimposed 3D of the redocked ligand with the expected PZA pose and interactions. The redocked ligand in this study (Figure 2.3 c) exhibited a similar pose with that obtained from the previous study (Figure 2.3 b). All the interactions were similar except for Trp68 that interacted with a Pi-Pi bond instead of van der Waal interaction. The key interactions are the four conventional hydrogen bonds, out of which two were with the active site residues, Asp8 and Cys138. The conventional hydrogen bonds were formed by the PZA carbonyl oxygen with Ala134 and Cys138 while its amide hydrogens interacted with Ile133 and Asp8. The ligand also formed a water hydrogen bond with H_2O220, crucial for the hydrolytic activation of the prodrug PZA. Some

active site residues (Lys96, Asp49) and metal binding site residue (His71), iron (Fe^{2+} 188) and other proximal residues (Leu19, Phe13, His137, Val7) interacted via van der Waal forces of attraction.
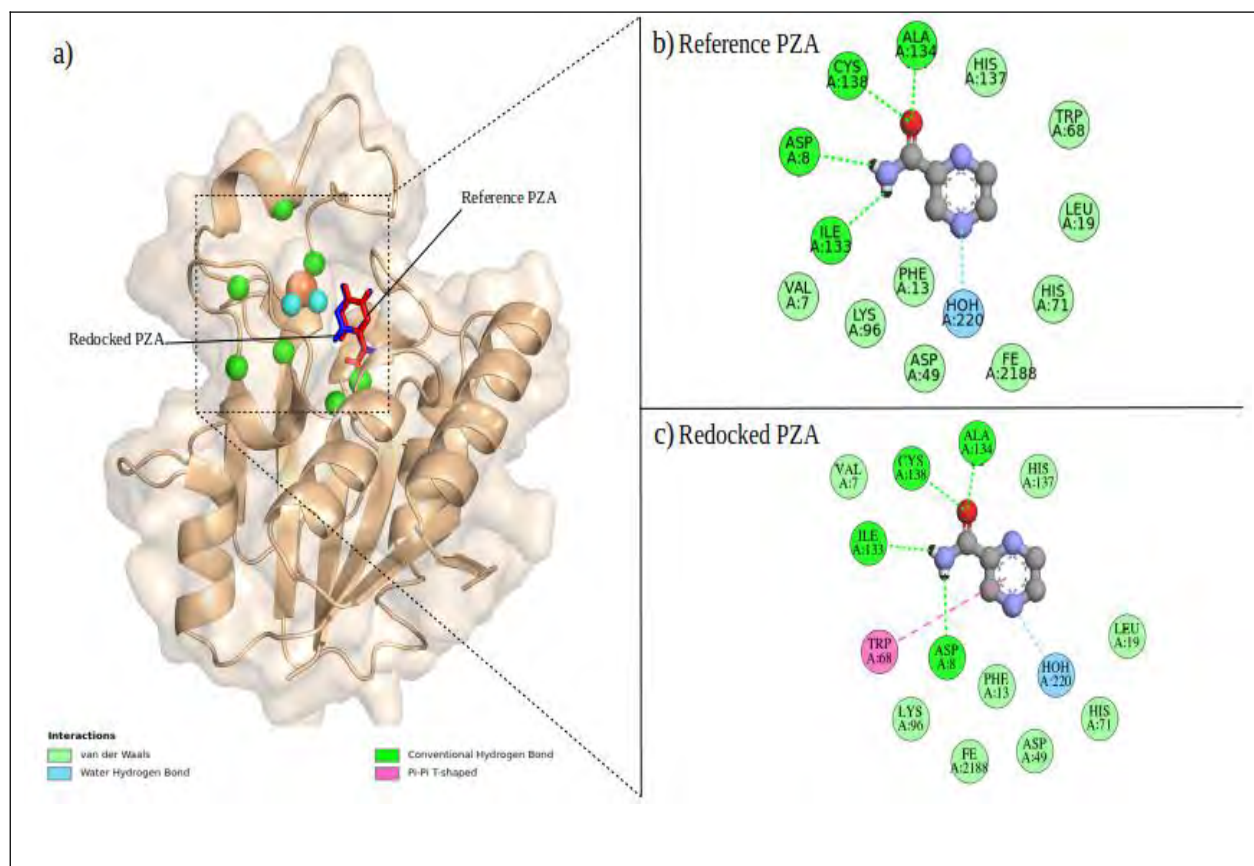
Figure 2.3*: Docking validation of PZA on wild type PZase (wheat cartoon and surface).* **a)** Visualization of superimposed redocked PZA structures from this study (blue) against the redocked pose from Sheik Amamuddy *et al.,* 2020 (red) as visualized in PyMOL. The active site residues are represented in green, Fe^{2+} metal in orange and the water molecules as cyan spheres. The 2D interactions of the redocked PZA against PZAse **b)** from the reference study by Sheik Amamuddy *et al.,* 2020 and **c)** from this study, as visualized in BIOVIA Discovery Studio Visualizer.

2.4.2 Blind docking screening

Since the redocking parameters exhibited the desired pose and key interactions, 2089 DrugBank compounds were docked against the WT PZase protein. Screening was done using various ad hoc Python scripts. After extracting the lowest energy and/or highest cluster poses, further compound screening was done by selecting ligands that passed the 4 selection criteria of binding energy ≤ -7.0

kcal/mol, Euclidean distance ≤ 8 Å, molecular weight ≤ 500 g/mol and having at least one conventional hydrogen bond with proximal amino acid residues (Figure 2.4).
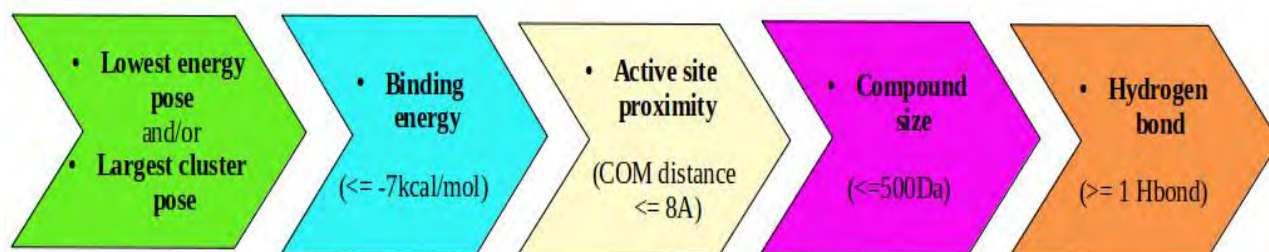
Figure 2.4: Systematic order summarizing the flow of blind docking compound selection. Ligand filtering was done in 5 stages to extract and identify the best docked DrugBank compounds.

The lowest energy and/or largest cluster poses were both selected in order to reduce bias on selecting ligand poses. The lowest energy poses have higher binding affinity which means that the ligand is more stably bound to the receptor compared to those with higher molecular docking energy. However, the lowest energy cluster might have few conformations and exhibit less reproducibility in the cluster. The largest cluster ligand poses thus counteract this bias by representing the cluster with the highest pose reproducibility, regardless of not falling in the lowest energy cluster. Ideally, the best ligand poses would be those having the lowest energy while also falling in the largest cluster. Thus, in this study, in cases where the largest cluster pose was not in the lowest energy cluster, both poses were extracted and manually analysed to identify the best pose exhibiting the desired key interactions. Appendix 1 shows an example of compound DB11793 in which the largest cluster pose was not the lowest energy pose. In that instance (Appendix 1), both poses were extracted and selection of the best pose was made after visualizing and further analysis. For this ligand, the largest cluster conformation was selected over the lowest energy pose. The cut off binding energy was set at most -7 kcal/mol, to increase the chances of extracting ligands with high

receptor-ligand stability. The Euclidean 8 Å COM distance was to selectively identify compounds within the active site pocket, increasing the probability of forming molecular bonds with some key active site residues, hence the additional screening criteria of at least one conventional hydrogen bond.

2.4.2.1 Blind docking outcome
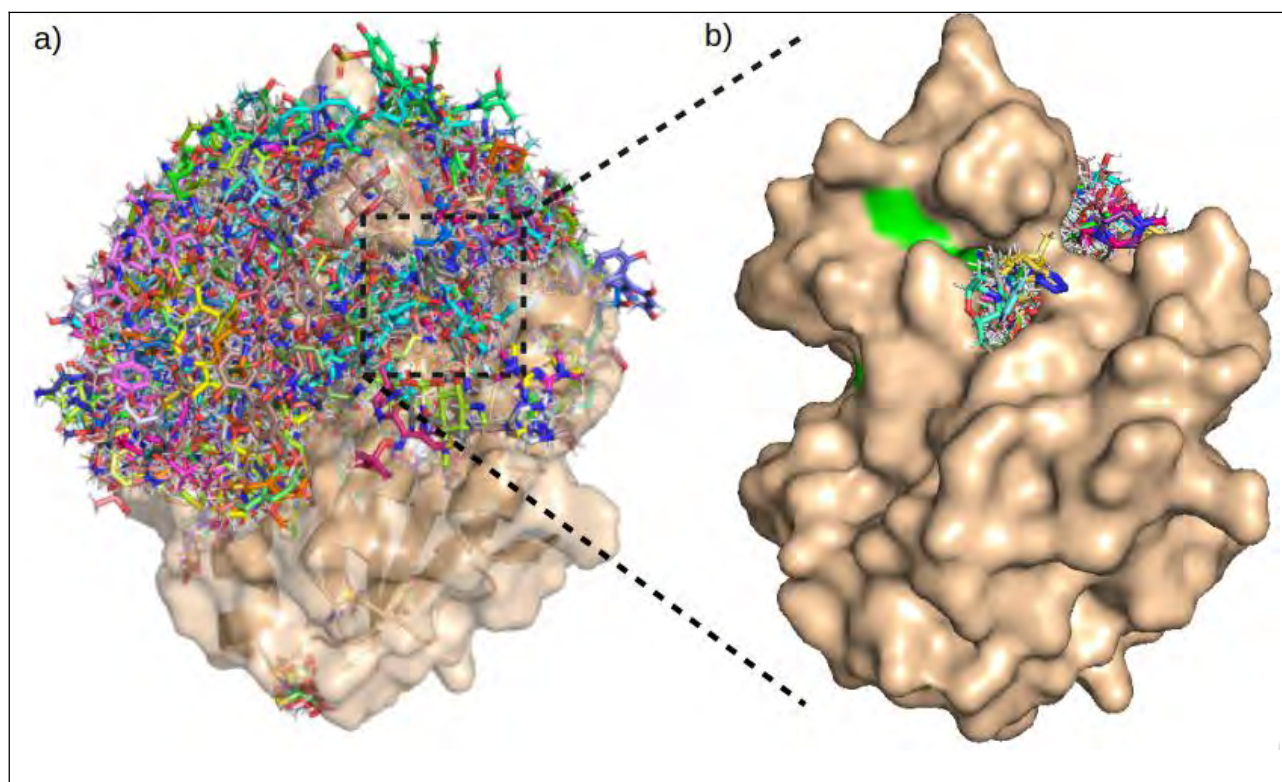
Figure 2.5: Blind docking screening of the 2089 DrugBank compounds against wild type PZase (wheat surface) **a)** Lowest energy and highest cluster ligand poses. **b)** Hit compounds (93 ligands) on the active pocket of PZase with the active site residues highlighted in green.

Extraction of lowest energy and/or largest cluster ligand poses resulted in potential 3516 ligand poses from the 2089 docked compounds. After the screening process, a total of 93 ligands passed the selection criteria. Figure 2.5 (a) shows the 3516 compound poses covering majority of the protein, with majority of the ligands binding around the catalytic pocket (the top parts of the protein). The down surface sites of the protein were hardly bound to a ligand Figure 2.5 (a). This

could be due to the presence of bigger binding pockets around this receptor's surface area (top parts of the protein, Figure 2.5 a) compared to the almost empty regions (down area). The compounds that met all the criteria are shown in Figure 2.5 (b), bound within the catalytic cleft with the active site residues highlighted green.

From the potential 93 ligands (Figure 2.5 (b)), after 2D visualization of the compounds (discussed under Section 2.4.3) and initial molecular dynamics stability simulation analysis (further explained in Chapter 3, Section 3.4.1), 47 ligands were identified and selected as the best docked drug compounds with their binding energies, Euclidean distance, number of conventional hydrogen bonds and molecular weights summarized in Figure 2.6.

In Figure 2.6 (a), binding energies are colour coordinated with the highest energy being illustrated by a dark colour (black) fading to the lowest energy of light colour (white). A lower binding energy represents high binding affinity which signifies better receptor-ligand stability. The control PZA had the highest binding energy of -4.12 kcal/mol as indicated by the colour black. All the compounds had binding energy lower than -7 kcal/mol as set off during the screening process, and thus can be deduced that they had better affinity to the receptor compared to PZA. The lowest binding energy was exhibited by compounds DB13943 (-9.84 kcal/mol) and DB00878 (-9.12 kcal/mol), represented with white and lightest yellow colours respectively. A few other DrugBank compounds (DB13953, DB13952, DB09132, DB06414, DB01623, DB01104, DB01058, DB00952, DB00655, DB00606, DB00146) ranged below -8 kcal/mol but not above -9 kcal/mol. The rest of the compounds had their energies in the red colour range of approximately -7 kcal/mol.
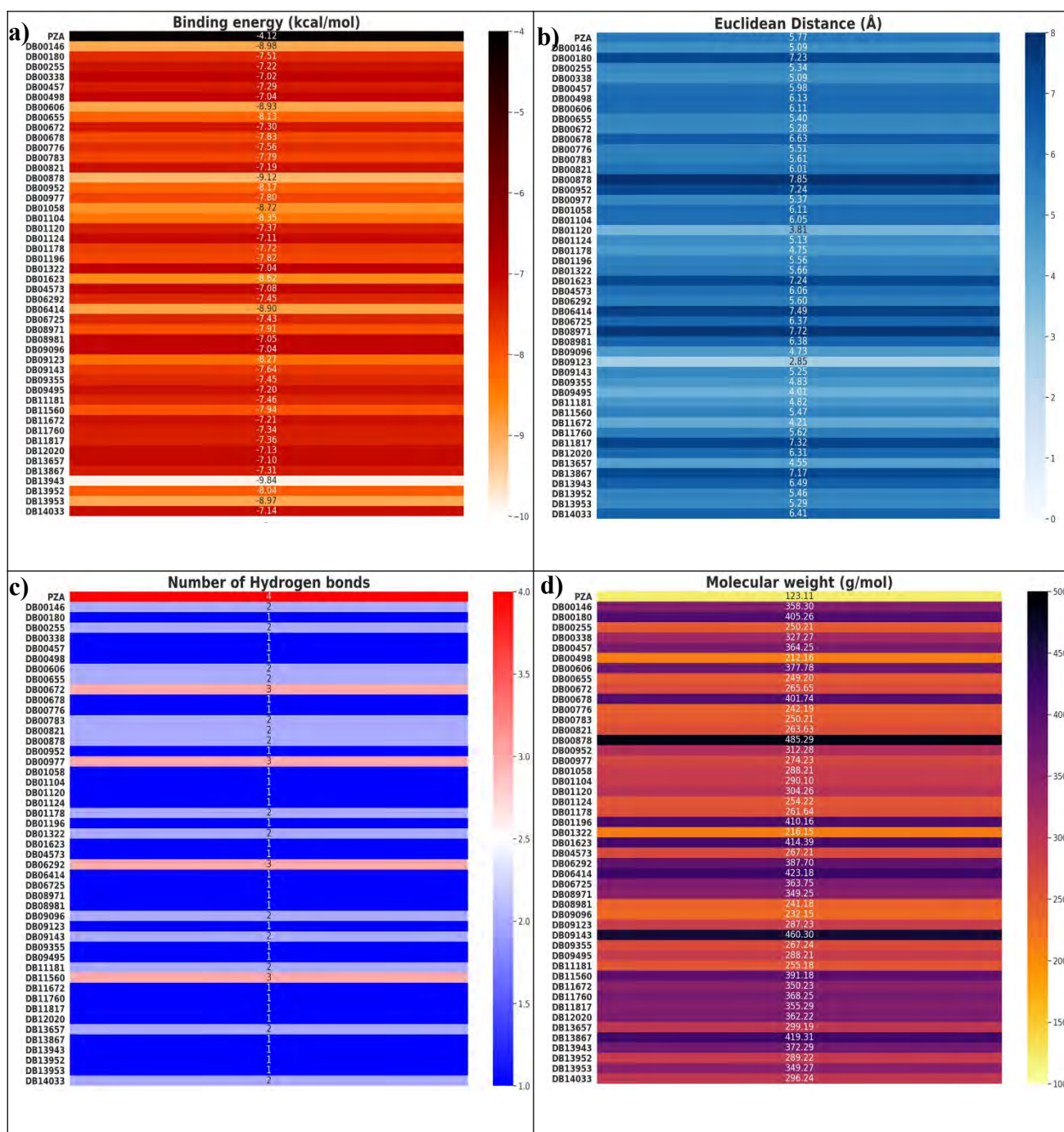
Figure 2.6: Characteristic heatmaps of the selected 47 compounds in reference to the control PZA. **a)** Docking binding energies of the ligands ranging from -4 kcal/mol for the PZA and the rest of the compounds below -7 kcal/mol. **b)** Euclidean distances of the ligands, all ligands within 8 Å. **c)** The number of hydrogen bonds formed between ligand and protein. **d)** The calculated compound molecular weight values.

Proximity of the ligand to the catalytic cleft allows for the formation of strong molecular interactions between the receptor active site residues and the ligand. Compound DB09123 and

DB01120 were the closest with distance 2.85 Å and 3.81 Å respectively. These compounds would therefore be expected to have stronger bond formation with the protein. However, Figure 2.6 (c) shows that the ligands only formed one hydrogen bond (the strongest molecular bond) with the receptor. While the furthest compound DB00878 (7.85 Å; Figure 2.6 b) formed two hydrogen bonds. All the other selected compounds were within the 8 Å distance (Figure 2.6 b) of the set protein center (x=9.88, y=-26.60, z=0.35; Figure 2.2) and formed at least one hydrogen bond (Figure 2.6 c). PZA had the highest number of hydrogen bonds (four), followed by compounds (DB11560, DB06292, DB00977, DB00672) with three hydrogen bonds. Analysis based on these two criteria show that proximity of the ligand does not guarantee formation of many hydrogen bonds but however only provides higher chances for interactions (Chapter 2, Section 2.4.3).
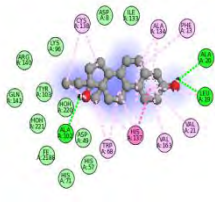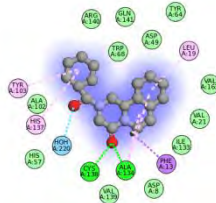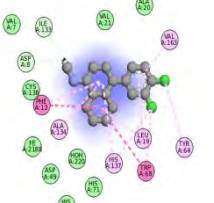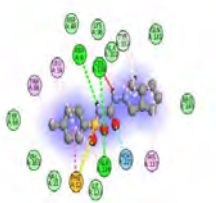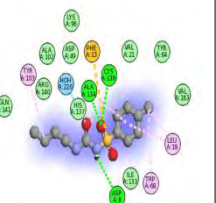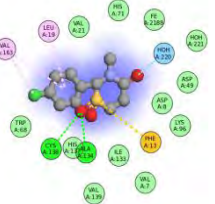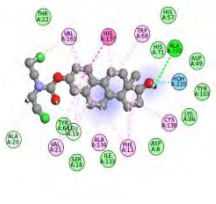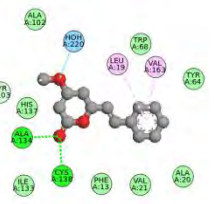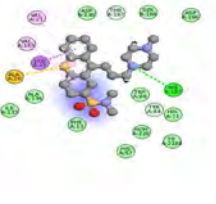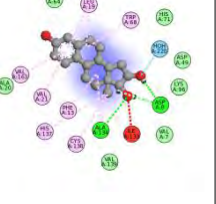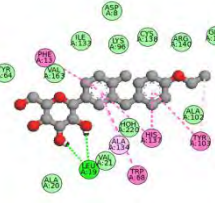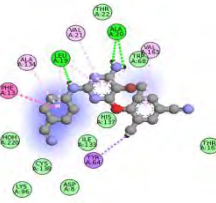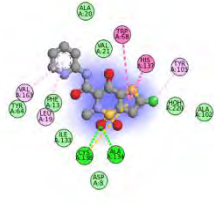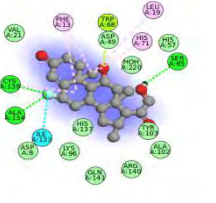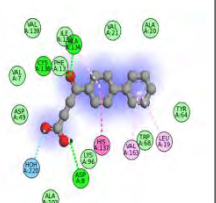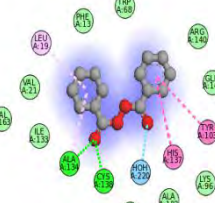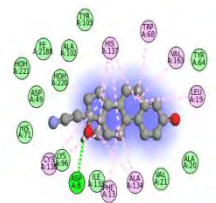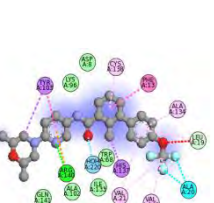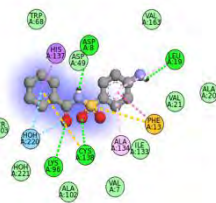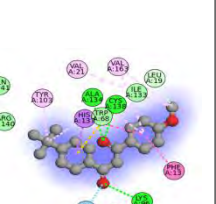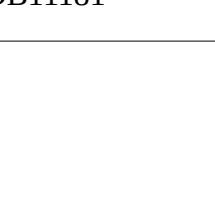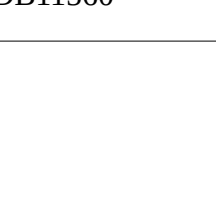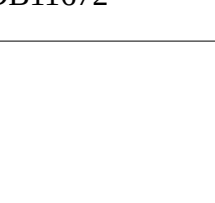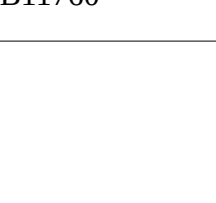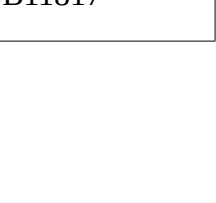
All the selected compounds had molecular weight less than 500g/mol. PZA had the smallest weight of 123.11g/mol represented by the yellow colour, followed by DB00498 (212.16 g/mol) (Figure 2.6 d). The biggest molecules were DB00878 and DB09143 with molecular weight of 485.29 and 460.30 g/mol respectively. Majority of the ligands were in the range 200 to 400g/mol Figure 2.6 (d).

2.4.3 Protein-ligand interactions

Interactions between the hit compounds and the protein residues were analyzed using BIOVIA Discovery Studio Visualiser. Due to the 8 Å cut-off distance used in selecting ligands, majority of compounds formed similar molecular bond interactions while a few had unique interactions (Figure 2.7). The most proximal compounds (DB09123 and DB01120; Figure 2.6 b) were expected to exhibit more molecular interactions. However, there was no distinct difference with other compounds (DB00977, DB00783, DB00672, DB00255) with average Euclidean distance including the furthest compound DB00878. Each ligand formed at least one conventional hydrogen bond with

receptor residues, one of the main stabilizing forces in molecular structures which also play a key role in dynamic processes (Alapour *et al.,* 2017). Key conventional hydrogen bonds between PZA and receptor formed with active site residues Cys138 and Asp8 (Chapter 2, Figure 2.3 c). Similar interactions were exhibited by compounds (DB00180, DB00255, DB00498, DB00606, DB00776, DB00821, DB00952, DB01058, DB01120, DB01124, DB01178, DB01322, DB04573, DB06725, DB08971, DB08981, DB09096, DB09123, DB09355, DB09495, DB11181, DB11817, DB13657, DB13943) while others formed the bond with other residues in proximity. Compounds (DB09335, DB9495, DB11181, DB12020, DB13867) also formed a conventional hydrogen bond with the active site residue Lys96, which stabilizes the enzyme while majority interacted with Lys96 through van der Waal forces.

DB00977	DB01058	DB01104	DB01120	DB01124
DB01178	DB01196	DB01322	DB01623	DB04573
DB06292	DB06414	DB06725	DB08971	DB08981
DB09096	DB09123	DB09143	DB09355	DB09495
DB11181	DB11560	DB11672	DB11760	DB11817
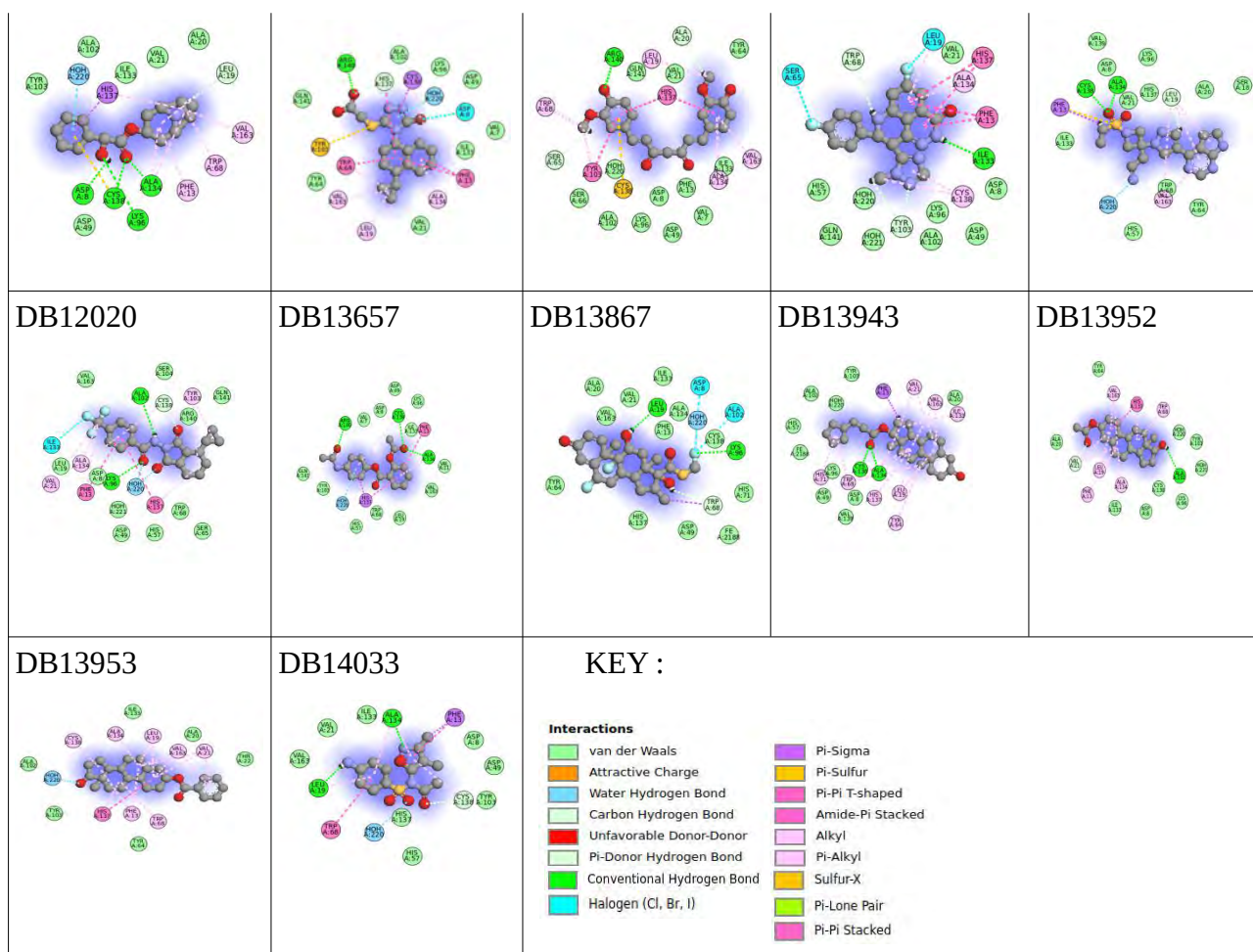
Figure 2.7: Discovery Studio 2D visualization of the 47 best docked DrugBank compounds against PZase. The colour key represents the different bonds formed within each complex. The ligand is represented as balls and sticks while the receptor residues are disc shaped.

Just as PZA formed a water hydrogen bond with H_2O220, some compounds (DB00146, DB00338, DB00498, DB00606, DB00655, DB00672, DB00776, DB01058, DB01120, DB01124, DB01178, DB01196, DB04573, DB09096, DB09143, DB09355, DB09495, DB11181, DB11560, DB11817, DB12020, DB13657, DB13867, DB13953, DB14033) also had a similar bond. This water hydrogen bond is crucial in PZA compounds as it activates the compound by hydrolysis, thus a similar mechanism may be expected to occur in these other compounds, an aspect that can be studied in the future to understand the mechanism of successfully selected hit compounds. Compounds DB00338,

DB00678, DB04573 formed unfavourable donor-donor interactions with Ala134, H_20220 and Ile133 residues respectively.

Other intermolecular interactions that were formed are pi (*π)* interactions, which are noncovalent bonds that can occur with other $π$ systems, metals or anions (represented by the purple colour). These interactions can be used in recognition of protein-ligand complexes and also increases their binding stability (Meyer *et al.,* 2003). Majority of the receptor residues interacted through van der Waal forces of attraction represented by the light green colour. A few other systems (DB13867, DB12020, DB11760, DB11560, DB11181, DB09143, DB08971) had halogen bonds (cyan colour) with Bromine, Iodine, Fluorine or Chlorine ligand atoms. The colour key shows other respective bonds (sulfur-X, alkyl, *π-sigma, π-sulfur, π-alkyl, π-lone pair*, *π-donor hydrogen, amide-π, attractive charge, carbon hydrogen)* formed between the docked compounds and the protein receptor.

2.4.4 Pharmacology of the hit compounds

Since the identified 47 potential anti-TB compounds are approved drugs already in use or under clinical investigation, their current medicinal uses were searched and tabulated in the table below (Table 2.1). From this table, generally, majority of the drugs are used in the treatment of hypertension (DB00457, DB00606, DB00678), manage menopause disorders (DB00255, DB00655, DB00783, DB00977, DB04573, DB09123, DB13952, DB13953), inflammations (DB00180, DB00878, DB06725, DB08971, DB08981, DB13867) and non-insulin-dependent diabetes mellitus (DB00672, DB01120, DB06292) among many other disease conditions. Since all these compounds

successfully bound to PZase active site, it implies that they may be subjected for further tests which

may eventually lead to their repurposing for TB therapy.

Table 2. 1: Summary of general clinical uses of the selected 47 hit compound. The compound
names and uses are from the DrugBank online database (www.DrugBank.ca).

Compound		Use
ID	Name	
DB00146	Calcifediol	Used in the treatment of rickets and osteomalacia
DB00180	Flunisolide	It is an anti-inflammatory for the treatment of allergic rhinitis
DB00255	Diethylstilbestrol	Synthetic nonsteroidal estrogen for treating menopause and postmenopause disorders
DB00338	Omeprazole	Treat gastric acid disorders like peptic ulcer diseaseand gastroesophageal reflux disease
DB00457	Prazosin	Treatment of hypertension and controlling post-traumatic stress disorders
DB00498	Phenindione	It is used as an anticoagulant for the treatment of pulmonary and cerebral embolism and mural thrombosis
DB00606	Cyclothiazide	Used in therapy of hypertension and edema associated with hepatic cirrhosis or congestive heart failure
DB00655	Estrone	Managemnet of peri- and post-menopausal symptoms
DB00672	Chlorpropamide	Treats non-insulin-dependent diabetes mellitus
DB00678	Losartan	Treament of hypertension
DB00776	Oxcarbazepine	Anti-epileptic medication for treatment of partial onset seizures
DB00783	Estradiol	Supplement estrogen amounts such as vulvovaginal atrophy and hot flashes
DB00821	Carprofen	Reduces arthritic symptoms in geriatric dogs
DB00878	Chlorhexidine	Antiseptic, and treats dental inflammatory conditions.

DB00952	Naratriptan	Treatment of migraine headaches.
DB00977	Ethinylestradiol	Used as contraceptive pills
DB01058	Praziquantel	Used in schistosome and many cestode infestations.
DB01104	Sertraline	Functions as an antidepressant
DB01120	Gliclazide	Treats non-insulin-dependent diabetes mellitus
DB01124	Tolbutamide	Used for the treatment of non-insulin-dependent diabetes mellitus
DB01178	Chlormezanone	Used in management of anxiety
DB01196	Estramustine	Used to treat prostatic neoplasms
DB01322	Kava	-
DB01623	Thiothixene	Used as an antipsychotic agent for management of schizophrenia.
DB04573	Estriol	Menopausal and Postmenopausal Disorders and determine the general health of an unborn fetus.
DB06292	Dapagliflozin	Managing diabetes mellitus type 2
DB06414	Etravirine	Treatment of human immunodeficiency virus type 1 (HIV-1) infection
DB06725	Lornoxicam	Treats pain and inflammation of the joints.
DB08971	Fluocortolone	Anti-inflammatory drug for various skin disorders
DB08981	Fenbufen	Non-steroidal anti-inflammatory drug for inflammation treatment and backaches, sprains and fractures.
DB09096	Benzoyl peroxide	Treatment for acne and bleaching hair and teeth whitening
DB09123	Dienogest	It is used as a contraceptive
DB09143	Sonidegib	Anti-cancer agent for treatment of basal cell carcinoma.
DB09355	Sulfabenzamide	Antimicrobial agent for intravaginal infections
DB09495	Avobenzone	-
DB11181	Homatropine	Anticholinergic drug for induction of mydriasis in ophthalmic solutions.
DB11560	Lesinurad	treatment of hyperuricemia associated with gout
DB11672	Curcumin	It has antibacterial, anti-inflammatory,

		hypoglycemic, antioxidant, wound-healing, and antimicrobial activities used in care clinical conditions
DB11760	Talazoparib	Used in germline BRCA mutated, HER2 negative
DB11817	Baricitinib	Treatment of rheumatoid arthritis in adults.
DB12020	Tecovirimat	Treatment of smallpox
DB13657	Benorilate	-
DB13867	Fluticasone	Treatments for various inflammatory indication, nonallergic rhinitis
DB13943	Testosterone cypionate	Supplement deficiency or absence of endogenous testosterone
DB13952	Estradiol acetate	Pro-drug ester of Estradiol used in oestrogen hormone balancing
DB13953	Estradiol benzoate	Treating menopause and postmenopause disorders
DB14033	Acetyl sulfisoxazole	Antibiotic to treat acute otitis media

2.5 CONCLUSION

In order to achieve the set objective for this chapter, 2089 DrugBank compounds were docked using AutoDOCK4. The whole protein (blind docking) was subjected to all DrugBank compounds (Figure 2.2 a) after validation of the docking parameters by referencing the pose and molecular interactions of the control, PZA, obtained by Sheik Amamuddy et *al.,* 2020 (Figure 2.3 b) with those obtained in this chapter (Figure 2.3 c). From the blind docking, several ligands were positioned in the active site as well as other sites. These other docked sites (other than the active site) where the compounds were located might be allosteric sites and can be further studied and analysed. The specific selection parameters for each ligand were summarised on (Figure 2.6). The selected 47 compounds had the highest binding affinity to the PZase active site and also had the most interactions with the active pocket residues. These two characteristics implies that the ligand will be stably bound to the active cleft and thus less likely to exit during protein dynamics. This screening approach is significant as it allows for best potentially bound compounds with strong

intermolecular interactions to be identified.

The current clinical use of the hit compounds is summarized in Table 2.1. Majority of these compounds are currently approved drugs in the treatment of hypertension, inflammation, diabetes and menopause disorders. All the identified 47 compounds may potentially be repurposed as anti-TB drugs and can be further investigated on their mode of action, in molecular dynamic simulations and other stability check methods.

CHAPTER THREE

3. WILD TYPE - MOLECULAR DYNAMICS

3.1. CHAPTER OVERVIEW

Protein-protein or protein-ligand interactions influences the protein's stability and dynamic processes. The use of techniques such as Molecular Dynamic (MD) simulations have been employed to study the dynamics of protein complexes as a function of time, analyzing their interactions and flexibility behaviours. In this chapter, all-atom MD simulations was performed on the WT Pyrazinamidase-ligand complexes and the dynamics of the selected hit DrugBank compounds from Chapter 2 in each system were analyzed. The 93 compounds from blind docking screening were subjected to short MD simulations of 20 ns after which the stable compounds were subjected for longer MD runs of 150 ns. A total of 142 MD runs were performed (93 runs for 20 ns and 48 runs for 150 ns). PZase-PZA system was used as the control. Due to the presence of the iron metal, the AMBER force-field was used to calculate forces that act on the metal binding site. The output files were converted to GROMACS (using GROMACS v2018) compatibles using tLeap and ACPYPE algorithms. All systems were subjected to energy minimization, equilibration and finally the MD run set for 150ns at 2fs on Centre for High Performance Computing (CHPC) cluster using GROMACS v2016. The protein-ligand dynamic trajectories were analysed using the *gmx* tools for Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), gyration and hydrogen bonds. Ligands that remained stable throughout the simulation period were identified and subjected for further analytical tests.

3.2 INTRODUCTION

3.2.1 Protein dynamics

Nucleic acids and proteins are dynamic entities, their behaviour in the presence of environmental forces play a significant role in their functionality (Karplus and Kuriyan, 2005; Hospital *et al,.*2015). Majority of proteins are unfolded in the absence of ligands, however, in the presence of other molecules or environmental changes, their structures undergo conformational changes. Among many other theoretical methods, MD simulations are one of the most powerful developed techniques to understand the relationship between structure and biological function of macromolecules (Rueda *et al.,* 2007). These simulations give details on the vital conformational changes and structural rearrangements made when macromolecules form complexes with other molecules (Hospital *et al.,* 2015, Junaid *et al.,* 2020). MDs treat all entities in the simulation system as flexible entities and give their dynamic behaviour as a function of time (Salmaso and Moro, 2018). These simulations can provide insight on ligand binding and protein folding, mutations, protonation and phosphorylation (Karplus and Kuriyan, 2005; Hollingsworth and Dror, 2018). It is a key aspect in the study of biomolecules as it provides important information on the molecules dynamics at different time scales (Shaw *et al.,* 2009).

Since the establishment of the first MD simulation of simple gasses (Alder and Wainwright, 1957) and protein MD of 9.2 ps by McCammon *et al* (1977), there has been great advancement in MD algorithms that cover long time scales of up to a millisecond where major critical phenomena occur, providing details that are inaccessible in experimental studies (Shaw *et al.,* 2009; Hollingsworth and Dror, 2018). MD algorithms of fine-tuning energy calculations and graphical processing units (GPU) have enabled simulation of systems with better biological relevance through the use of HighPerformance Computing (HPC) computer facilities (Shaw *et al.,* 2009; Hospital *et al,* 2015).

3.2.2 Simulation parameters

Molecular mechanics uses the laws of classical mechanics to counteract computational cost required in quantum mechanical calculations (Vanommeslaeghe *et al.,* 2014). In each MD system, positions and velocities of the atoms are calculated using classical mechanics, where Newtons second law (motion equation) is employed to calculate the systems time evolution. In order to accurately solve the motion equation, a time-step smaller than the fastest time scale in the system has to be used to avoid unstable and atomic collisions which results in rapid rise in energy and forces. A time-step of 1 fs is often used in classical systems with fastest motion being bond vibrations while a 2fs time-step is used in simulations with altered lengths of covalent bonds (Hospital *et al.,* 2015; Weiergräber *et al.,* 2017). A smaller timestep is advised because majority of biochemical structural changes occur at nanoseconds or microseconds timescales (Hollingsworth and Dror, 2018). Larger time-steps can be used to accommodate longer simulation periods, however, these result in less accurate simulation ensembles (Hospital *et al.,* 2015).

The motion equation is also based on that the system has to conserve its total energy based on the NVE ensemble of constant number of atoms (N), volume (V) and energy (E). However, in experimental systems, temperature causes fluctuations in total energies. Therefore during MD simulations, kinetic energy of the system is controlled to desired temperature (T), room temperature (298K) or physiological temperature (310K), using the NVT ensemble while pressure (P) is maintained at 1 bar using the NPT ensemble (Weiergräber *et al.,* 2017). Periodic boundary conditions are used to prevent the system from having unexpected borders in a vacuum and to create an infinite system by replicating the polyhedron shape (Weiergräber *et al.,* 2017).

3.2.3 Force fields

Calculation of the protein model is carried out in the presence of one empirical potential, known as molecular mechanics force fields (FF). These force fields represent protein conformations by fitting their parameters to the calculations of quantum mechanics and small molecules experimental properties (Weiergräber *et al.,* 2017; Hollingsworth and Dror, 2018). In each FF, electrostatic interactions, bond length and other interatomic interactions are incorporated. Some of the mostly used force-fields include AMBER (Assisted Model Building with Energy Refinement) (Case *et al.,* 2005), CHARMM (Brooks *et al.,* 1983), NAMD (Phillips *et al.,* 2005), GROMACS (Groningen Machine for Chemical Simulations) (Pronk *et al.,* 2013) and OPLS (Jorgensen *et al.,* 1996). These force fields differ in the way they are parameterized, thus not all FFs represent all types of molecules (Hospital *et al*, 2015). In this study, the AMBER and GROMACS (GROMACS v2016.4) force fields were used to study the simulation of the protein complexes.

3.2.3.1 GROMACS

GROMACS is an open source and free software widely used in the study of biomolecule dynamic simulations. This force field uses a set of equations describing time evolution of torsions, bond angles and lengths, electrostatic and van der Waal interactions between atoms. Completion of the MD run generates a trajectory file (composed of information on the movement of atoms over the simulation period) that can be analysed using the *gmx* commands and tools for the systems. The goal of the FF is to have the highest absolute efficiency on any hardware and to provide high realworld throughput (Karplus and Kuriyan, 2005; Abraham *et al*, 2015). The system can be used in parallel using Message Passing Interface (MPI) or threads as a results of its flexibility accounted for by support from different force fields.

3.2.3.2 AMBER

AMBER is also a biomolecular dynamic simulation force-field which requires the parameter values such as force constants, charges, bond lengths and angles. Each parameter is specific for the different types of molecules like peptides, proteins, nucleic acids, carbohydrates, lipids and small organic molecules (Dickson *et al.,* 2014). For the modeling of proteins, the primary FF used is ff14SB (Maier *et al.,* 2015), this FF was therefore used in this study for modeling PZase. A set of programs like LEaP, MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area), NAB (for complex nucleic acids) and Antechamber among many others are incorporated in the AMBER software. In this study, the LEaP (tleap) and Antechamber programs were used so as to prepare input files for production calculations and parameterize small molecules using the General AMBER Force Field (GAFF2) respectively.

3.2.4 Advantages and limitations of molecular dynamics

Difficulties in understanding protein dynamics in wet labs such as inconsistent results using different experimental techniques and results that are difficult to interpret are encountered (Shaw *et al.,* 2009; Abraham *et al.,* 2015). MD simulations can be used in linking with experimental data by calculating experimental observables from MD data (Shaw *et al.,* 2009). The current MD biological modelling provides a better realistic environment by including ions, solvents and lipid membranes (Weiergräber *et al.,* 2017) and also account for flexibility and thermodynamics properties in the systems (Hospital *et al.,* 2015). It provides spatial and temporal resolution of data not observed in laboratory experiments (Abraham *et al.,* 2015). However, there is still room to improve the representation standards, optimize analysis tools and create storage for the created huge trajectory

files. The lack of parameters for cofactors also challenge the use of force-fields in many simulations. However, webservers such as ATB (Malde *et al.,* 2011), ProDRG (van Aalten *et al.,* 1996) and ACPYPE (Sousa da Silva and Vranken, 2012) have been developed to bridge this gap.

3.3 METHODOLOGY

Summary of all the steps executed to successfully calculate acting forces, simulate biological environmental conditions and analyze the obtained trajectories from the Molecular Dynamics simulation (Figure 3.1).
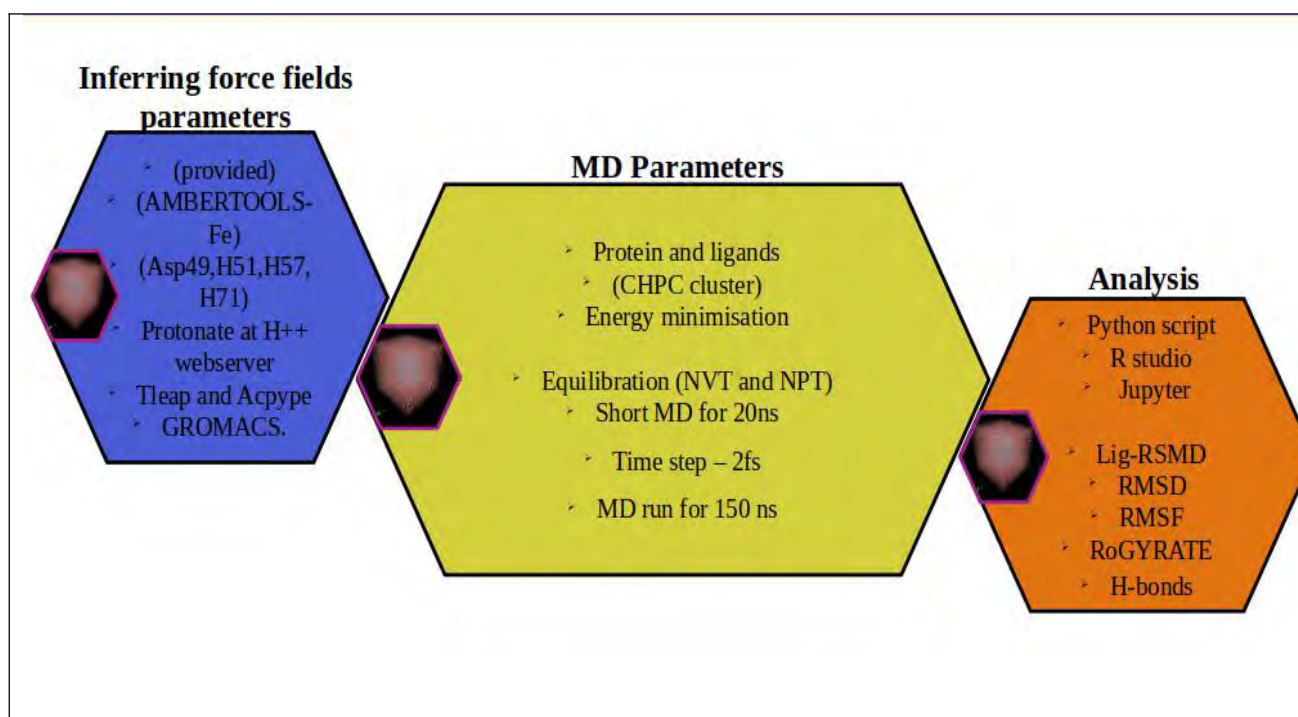
Figure 3.1: Overall summary of the performed Molecular Dynamics simulation methods. The three key steps are Force-field inferring, Molecular dynamics run and analysis obtained results

3.3.1 Inferring force fields parameters

Due to the presence of the cofactor Iron metal (Fe^{2+}) that is crucial for the protein functionality, there was need for a force-field that accounted for the ion. The coordinating MBS residues (H51,

H57, H71 and D49 including the Iron) forcefield parameters were provided by the previous group (Sheik Amamuddy *et al.,* 2020). The AMBER FF was used to perform all atoms MD simulation (AMBER SQM V19) (Walker *et al.,* 2008). The importance of all atom MD simulations is to study and explore conformational flexibility and stability of protein and/or ligand systems (Musyoka *et al*, 2016).

3.3.2 MD simulation runs

3.3.2.1 Protein and ligand preparation

The WT PZase protein was protonated on H++ webserver (http://biophysics.cs.vt.edu/) at pH 7. The other parameters were left at default of 0.15 salinity, 10 internal dielectric, 80 external dielectric and no preparation of explicit solvent box topology/coordinate files (AMBER). The obtained topology and coordinate files were concatenated to a pdb file using the *ambpdb* command.

An ad hoc Python script was used to reduce the ligand, delete extra hydrogens, rename the HIS and ASP atoms to HID/HIE and AP1 respectively, correctly number all atoms, merge the protein, iron and ligand coordinates files and execute tLeap and AMBER ACPYPE (antechamber Python parser interface). tLeap and ACPYPE were used to infer parameters and import AMBER files to generate GROMACS compatibles. The tLeap step solvated the system using TIP3P mode (Mahoney and Jorgensen, 2000).

3.3.2.2 Energy minimization

Energy minimization was done using the GROMACS (v2018) *gmx mdrun* to avoid steric clashes that might have formed due to inappropriate geometry and/or addition of ions and water molecules. This step was done so as to relax the system using the steepest descents method with a force tolerance of 1000 kJ/mol/nm capped to an upper limit of 50 000 steps.

3.3.2.3 Equilibration

After minimization, the systems were equilibrated to mimic the biological environmental conditions necessary for the functioning of the protein. Equilibration of temperature and pressure were done on CHPC cluster. The temperature of the system was equilibrated using a constant number of particles, volume and temperature (NVT) ensemble over a period of 100 ps at 300 K while pressure was equilibrated at 50 000 steps for 100 ps at 1 atm using the NPT ensemble.

3.3.2.4 MD simulation

CHPC cluster was used for the dynamic simulations of the protein complexes using GROMACS (gromacs/v2016). The initial 93 simulations were run for 20 ns at 2 fs timestep. From the 93 simulation systems, compounds that portrayed unimodal conformations (47 compounds) in the last 10 ns of the 20 ns simulation period were extended to 150 ns using the same timestep.

3.3.3 Post MD trajectory analysis

After completion of 150 ns MD simulations, the whole system was removed from the periodic boundary conditions and centered using the MD trajectory analysis (*trjconv*) tool. Analysis of the

trajectories were computed using Radius of gyration (Rg), hydrogen bond,RMSD and RMSF. The dynamics of the systems throughout the simulation period were visualized in Visual Molecular Dynamics (VMD) (Humphrey *et al.,* 1996). The ad hoc analysis Python and R scripts used were provided by the previous group (Sheik Amamuddy *et al.,* 2020), the codes were edited to suit the current data.

3.3.3.1 RMSD

RMSD measures the average distance between atoms of superimposed proteins. It is often used in globular protein studies measuring similarities of alpha carbon atomic co-ordinates and molecular configuration of ligands when bound to macromolecules (Zhao *et al.,* 2015). To determine how each ligand was behaving throughout the simulation, initial analysis of the ligand RMSD was done. The ligands that portrayed stable and unimodal conformations were identified and reviewed in further analysis. Alpha carbon RMSD was also computed. A Python script was used to create compatible R data files which were analysed in R studio to generate a data frame for use in Jupyter Lab. The RMSD data was then plotted as violin plots using Jupyter Lab.

3.3.3.2 RMSF

RMSF measures local chain flexibility by calculating the deviation of protein residues from the averaged position of the particle over time (Zhao *et al.,* 2015). To monitor the protein residues motions in each system, RMSF was computed based on alpha carbon atoms using the GROMACS *rms* command. A Python script was used to create compatible R data files which were analysed in R studio to generate a data frame for use in Jupyter Lab. The RMSF data was then plotted as a heat map in Jupyter Lab.

3.3.3.3 Radius of gyration

Radius of Gyration measures the protein compactness by calculating the distance between protein centre of mass and its terminals. A stably folded structure roughly maintains a steady Rg value. The whole protein compactness and active site compactness were computed. Active site Rg was computed by selecting residues interacting within 8 Å of the PZA ligand. The results were represented as violin plots plotted in Jupyter Lab after the use of a Python script and R compatible files in R studio.

3.3.3.4 Hydrogen bonding profiling

Since a hydrogen bond is one of the main stabilizing forces in molecular structures, the number of present hydrogen bonds were computed throughout the simulation period using the GROMACS *gmx hbond* command. Python scripts and R studio were used to extract and analyse the data while Jupyter Lab was used to construct the *hbond* plot. To determine the precise residues that formed the hbonds, the *cpptraj4* command was used.

3.3.3.5 VMD visualization

Protein dynamics results were visualized on VMD, a tool used to animate and analyze trajectories of MD simulations.

3.4 RESULTS AND DISCUSSION

Molecular Dynamic simulation on the WT PZase was successfully executed. The obtained data

from the procedure was analyzed based on RMSD, RMSF, RoG, hydrogen bonds and the results were represented as violin plots and heat maps.

3.4.1 Initial MD simulation

After successful docking and initial screening of active site hit compounds (Chapter 2, Section 2.4.2.1, Figure 2.5 b), the 93 identified compounds were subjected for further short stability check through Molecular Dynamics. The initial step was to screen by analysing the behaviour of the ligand within the last 10 ns of the 20 ns simulation period.

Figure 3.2 gives the ligand structural conformations based on RMSD values. RMSD values predicts the deviation or average distance between any two points. PZA was set as the control (blue). It had an RMSD value of approximately 0.2 nm and portrayed a thin distribution of a unimodal shape ranging from around 0.1 to 0.3 nm. Since the ligand poses were selected based on largest cluster and/or lowest energy (Chapter 2, Section 2.4.2), the compound naming had the suffix *le* for lowest energy, *lc* for largest cluster and *vs* to indicate that the lowest energy cluster was also the largest cluster. Ligands were selected based on compounds that had RMSD values less than or slightly higher than that exhibited by PZA and also exhibiting unimodal conformations. The unimodal conformation (represented by one peak) shows that the ligand conformed to a single conformation, an indication of less fluctuation and more stability (Sheik Amamuddy *et al.,* 2020). The selected ligands are highlighted in yellow while the rejected compounds are highlighted in red (Figure 3.2). From the selected group, majority of the compounds had RMSD values lower than 0.2 nm with some of the best compounds (DB00498_le, DB00672_vs, DB00977_le, DB11760_le) being distinctly below 0.1 nm. A few selected compounds (DB00457_le, DB00606_le, DB06292_le, DB11560_le, DB12020_le, DB13867_le, DB14033_le) had RMSD values slightly higher than 0.2

nm but however, exhibited the unimodal conformation. The ligands highlighted in red portrayed higher RMSD values and/or multimodal conformations. Compounds DB00322_le, DB0125_le, DB11750_le had RMSD values close to 0.8 nm with bimodal distributions. Visualization of these systems in VMD showed a lot of fluctuations throughout the simulation. Thus, these ligands were excluded as they were not stable and left the system at some point during the simulation.
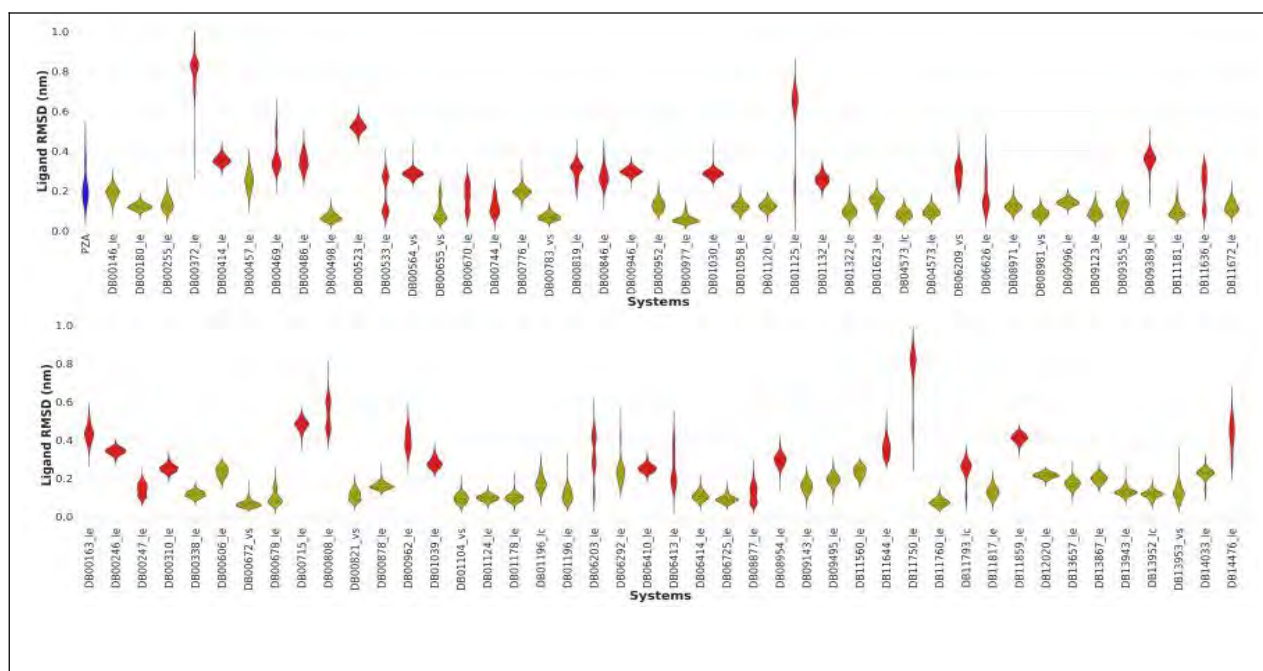
Figure 3.2: Violin plots for Ligand RMSD of the last 10ns of the 20ns simulation period. The control PZA in blue, selected stable compounds in yellow and rejected compounds in red.

3.4.2 Root Mean Square Deviation

Screening from the initial MD (20 ns simulation) analysis identified 47 ligand compounds set for an extended simulation of 150 ns. RMSD calculations were done on both the ligand (Figure 3.3 a) and backbone protein residues (Figure 3.3 b). Ligand RMSD calculation was crucial to understand the behaviour of the ligand throughout the simulation, this was used as the basis to highlight potentially stable systems.
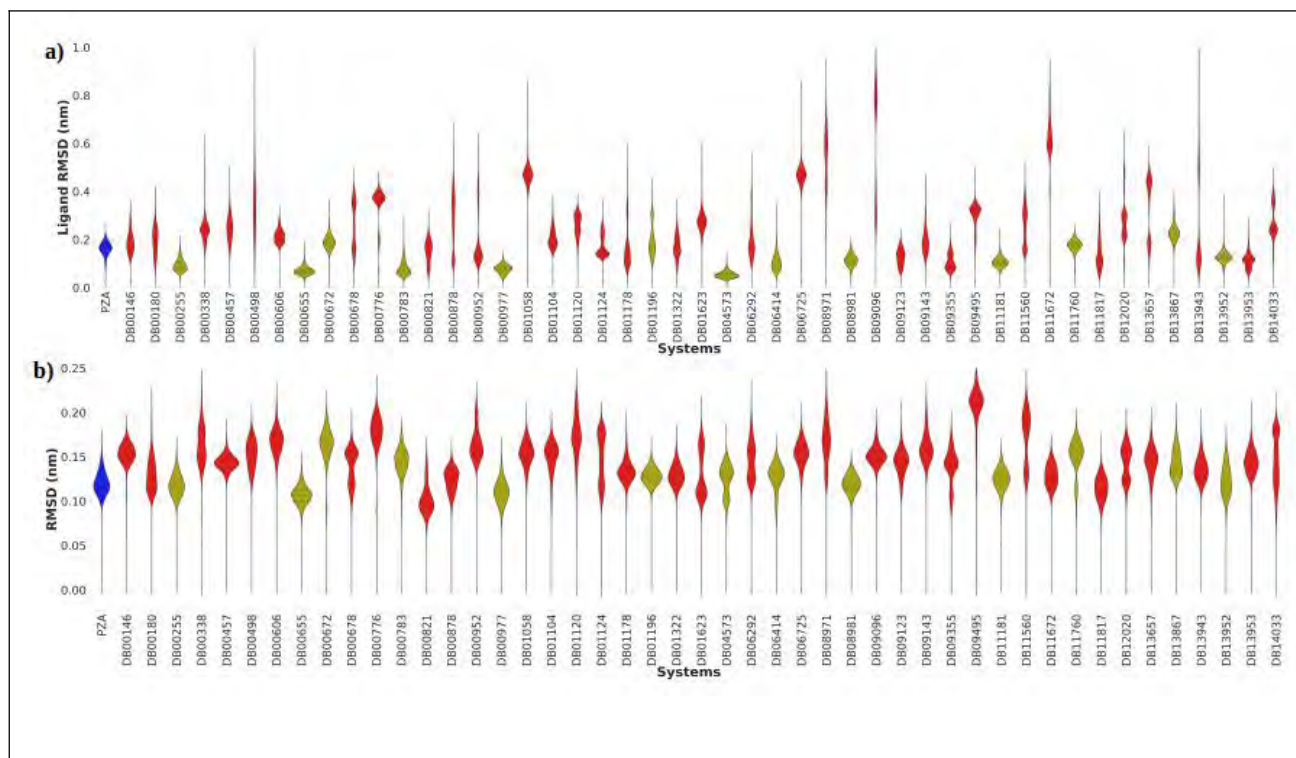
Figure 3.3: Violin plots of **a)** Ligand RMSD and **b)** Protein backbone RMSD of the selected 47 compounds. PZA is in blue, potential hit compounds in yellow and excluded compounds in red.

After the 150 ns simulation period, ligand RMSD for PZA (Figure 3.3 a, blue) was around 0.2 nm and the ligand exhibited a distinct unimodal conformation, predicting stability of the ligand throughout the simulation. Not all the 47 ligands that passed the first 20 ns simulation (Figure 3.2) exhibited unimodality and low RMSD value at the end of the 150 ns simulation. Compounds with unimodal conformations and RMSD values close to that of PZA (0.2 nm) were selected as the stable systems. Out of 47 ligands, only 13 compounds (DB00255, DB00655, DB00672, DB00782, DB00977, DB01196, DB04573, DB06414, DB08981, DB11181, DB11760, DB13867, DB13952) were identified (Figure 3.3, yellow) while the majority mainly had multimodal conformations (DB00678, DB00776, DB00821, DB00878, DB01120, DB01124, DB09355, DB11560, DB12020, DB13675, DB14033) or high RMSD values (DB01058, DB08971, DB09096, DB11762) as highlighted in red.

The systems backbone RMSD values had small differences as majority ranged between 0.1 nm and 0.2 nm. This provided less correlation between the ligand RMSD data with the backbone RMSD data. PZA had a backbone RMSD value around 0.12 nm and the selected systems also exhibited their RMSD values around that value. In some systems (DB00776, DB00878, DB01178, DB01322, DB09096, DB11817, DB13657, DB13953), ligand RMSD plots portrayed bimodal conformation but however exhibited unimodal conformation in backbone RMSD calculations while the opposite occurred in systems like DB01623. This may be due to ligand fluctuations that did not have a great impact on the protein backbone residues resulting in the system still obtaining a unimodal state regardless of the ligand fluctuations.

3.4.3 Root Mean Square Fluctuation

The degree of protein-ligand complex fluctuations were analysed by calculating the deviation of the protein residues from an average set position (RMSF). The calculation predicts the most and least flexible residues throughout the simulation process. From Figure 3.4(a), a higher value (ranging from 0.05 to 0.45 nm) shows more flexibility and it is represented by a dark colour on the heat map. The system with PZA showed less fluctuations with the darkest colour ranging around 0.25 nm. In PZA bound system, residues 34-41 had the most flexibility. The selected 13 compounds are marked by an asterisk. Majority of these systems had similar flexibility trend to PZA except for DB00338, DB13687 and DB13952 that displayed high flexibility on residues 98-102 that were not identified in PZA and other selected systems. Across all the systems, the most fluctuating residues were amino acids 14-17, 32-41, 52-54, 59-66, 98-102, 182-184. These residues were mapped on the protein structure (Figure 3.4b), represented in red. From Figure 3.4(b), as expected, the highlighted

areas are loop regions around the catalytic cleft and on the protein N-terminal. These loop area regions form part of the side flaps that opens during ligand release (14-17, 32-41, 52-54, 59-66, 98-102) and N-terminal (182-184) regions. The beta pleated sheets and alpha helices regions had less flexibility represented by an orange colour. This shows that the protein's secondary structure is more stable and was less affected by ligand binding. The active site residues (8, 49, 51, 57, 71, 96, 138) were not affected by the fluctuations, signifying their stability for protein functionality.
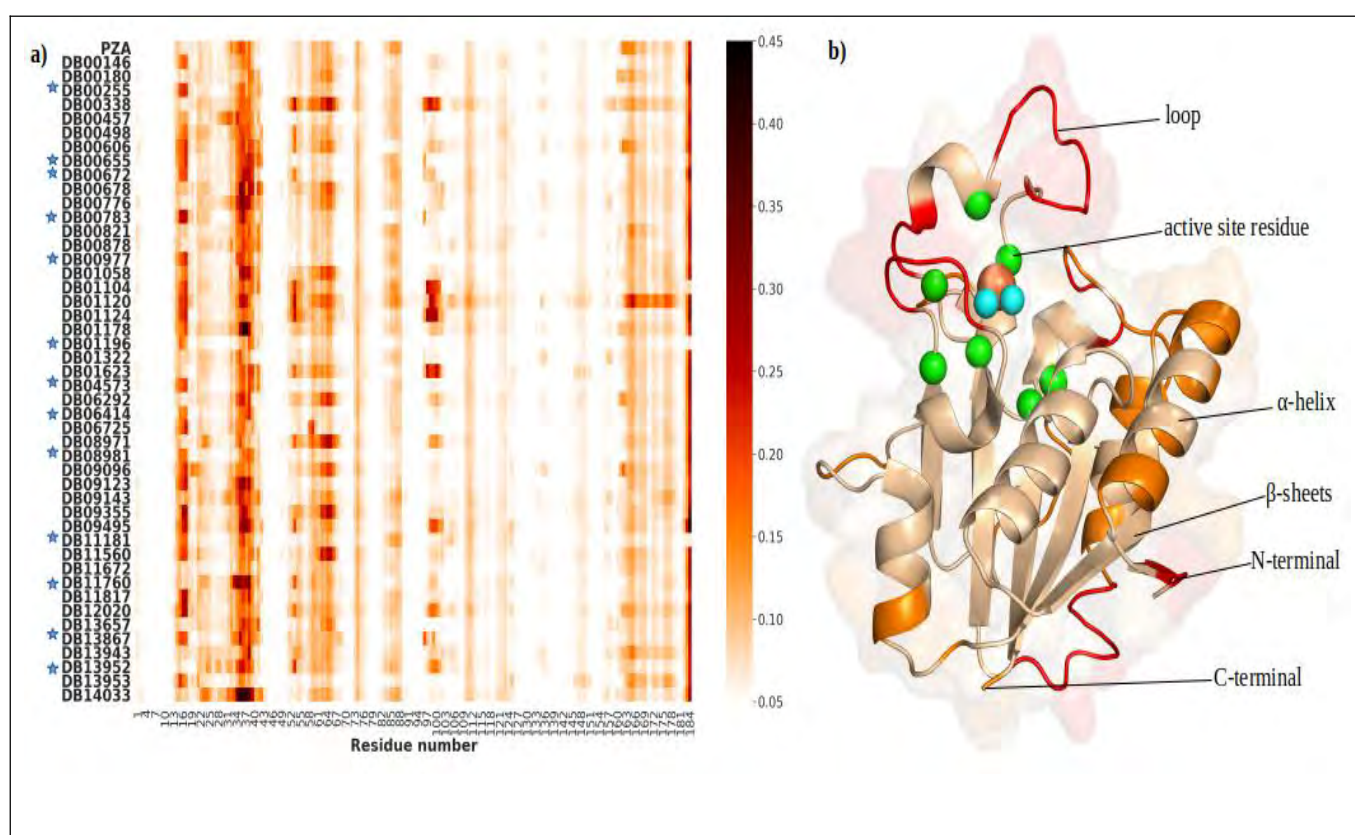
Figure 3.4: RMSF analysis. **a)** A heat map showing the local residue fluctuations during the 150ns simulation period across all hit compounds. A high RMSF value represents more flexibility. The 13 potential hit compounds are identified by an asterisk. **b)** Mapped fluctuating regions on the protein, green spheres - active site residues, red - most fluctuating, orange- moderate fluctuating.

3.4.4 Radius of gyration

After deducing the most flexible regions, the protein's compactness was analysed by computing

radius of gyration on the whole protein complex system and active site area. This measure gives insight on how closely the protein is packed. A lower Rg value shows more structural compactness while a larger value indicates less compact. The lowest exhibited backbone Rg value on the selected 13 compounds was in DB00655 system similar to that of PZA (1.53 nm) (Figure 3.5a). The active site compactness was also analysed (Figure 3.5b) where PZA system had the lowest value of 0.95 nm. Across all systems, in both active site and whole protein gyration analysis, unimodal conformations were displayed, which generally shows close packing, except for DB01623 system that displayed a bimodal shape. Generally, the systems Rg values are within a small range of 0.925-1.100 nm in active site and 1.52-1.60 nm in whole protein gyration. Thus, there are extremely small differences across the systems resulting in little conclusion and deduction based on gyration analysis.
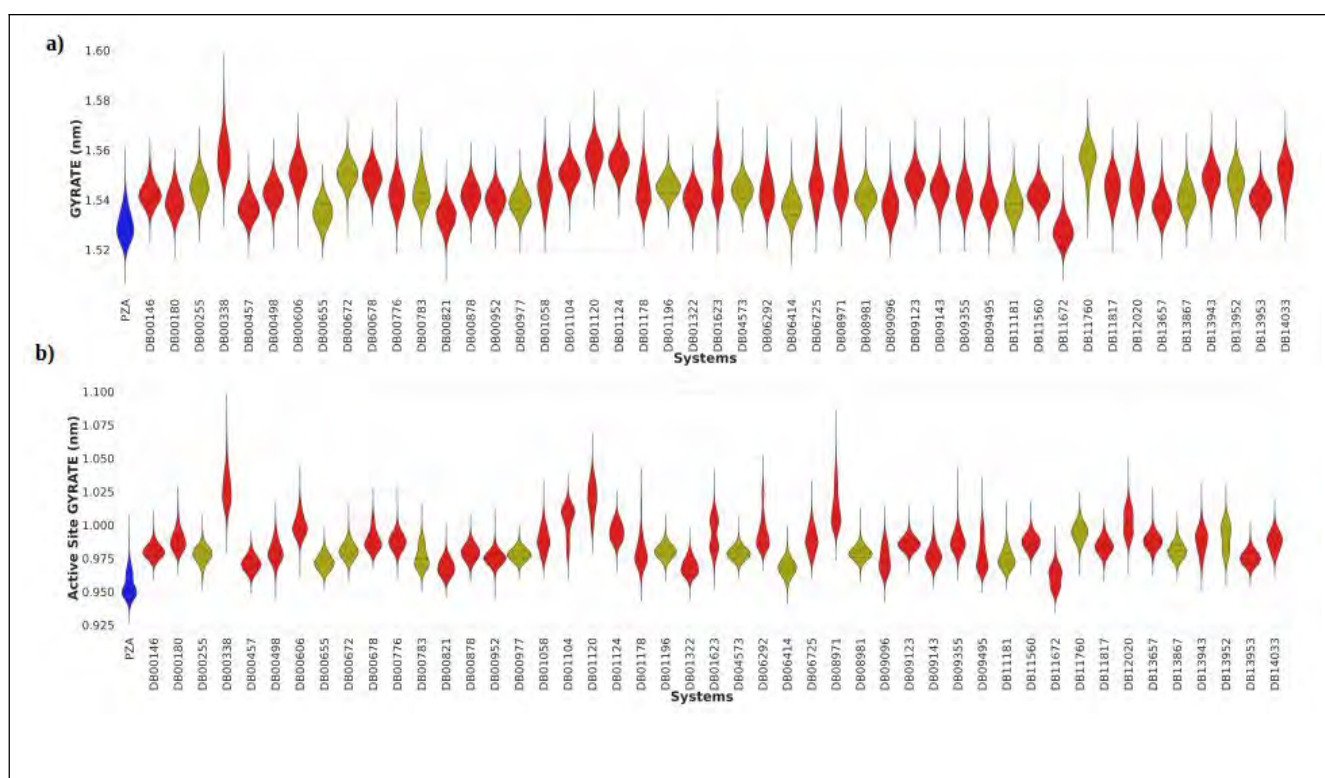
Figure 3.5: Radius of Gyration **a)** for the protein and **b)** active site residues within 8Å of PZA in the catalytic cleft. PZA is in blue, selected potential hit compounds in yellow, compounds in red.

3.4.5 Hydrogen bonds

Throughout the MD simulation, multiple intermolecular interactions such as van der Waals, ionic, pi and hydrogen bonds are formed among many other interactions. Since hydrogen bonds are one of the main stabilizing forces in molecular structures, the number of hydrogen bonds present over the period of 150 ns was counted (Figure 3.6). This analysis predicts that more hydrogen bonds tend to make the receptor-ligand complex more stable since more energy is required to break the strong bond.

The control PZA had a constant number of bonds throughout the simulation of about 2 to 3 bonds and at some points formed 4 bonds (around 65 ns, 115 ns ,135 ns). Majority of the selected hit compounds (identified by an asterisk) had at least one hydrogen bond. The most number of bonds were formed in systems DB00783, DB04573 and DB08981 and great consistence was observed in DB00255, DB06414, DB13867. Of the selected compounds, the least number of hydrogen bonds were observed in DB00672 and DB11181 systems, however, these systems were selected as the ligand RMSD analysis portrayed distinct unimodal and low RMSD values.

Other unselected systems (DB00457, DB00821, DB00878, DB01120, DB09143, DB09355, DB13953) however had more hydrogen bonds compared to the selected hit compounds regardless of showing bimodal or multimodal conformations and high ligand RMSD values (Figure 3.3). Visualization of these systems showed that fluctuations of these ligands occurred within the catalytic cleft and at no point left the system thus still maintained molecular bonding throughout the simulation. However, in DB00498, DB01104, DB09096 and DB13943 systems, the ligand would remain in the active site for the first 20 to 40 ns after which it was completely released from the

system or takes a conformation position that does not allow the formation of hydrogen bonds. Hence the total absence or sparsely presence of hydrogen bond until the end of the simulation, respectively.
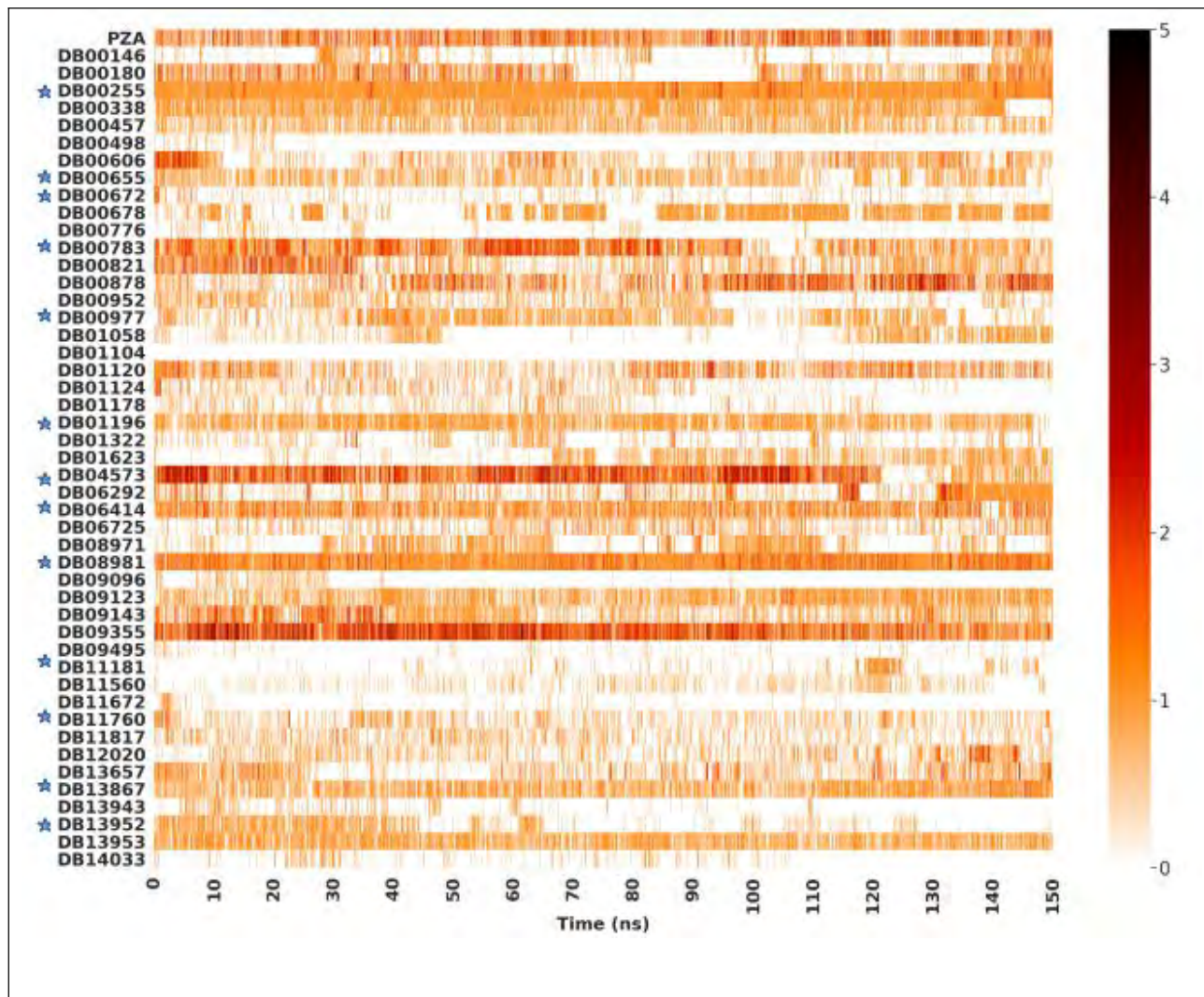
Figure 3.6: Number of hydrogen bonds throughout the 150 ns simulation period. The light colour represent fewer bonds while a dark colour shows more hydrogen bonds at a specific time.

Figure 3.7 summarizes the hydrogen bonds between the ligand and amino acid residues within 8 Å of the catalytic cleft center. The occupancy percentage values of these interactions are shown corresponding to each residue. Occupancy values estimates the fraction of time at which the bond occurred throughout the simulation (Helen *et al.*, 2000). Thus, a higher value means the bond occurred for a longer period. PZA had a total of 4 hydrogen bonds out of which 2 were with the

active site residues (CYS138-3% and Asp8-70%). From the H bonds that were formed between the ligands and the protein, majority of the ligands had high occupancy with Leu19 residue. The only systems with hydrogen bonding with the active site residues were DB00672, DB11181 and DB13867. The behaviour of the selected 13 compounds were further analyzed for stability in the mutants structures.
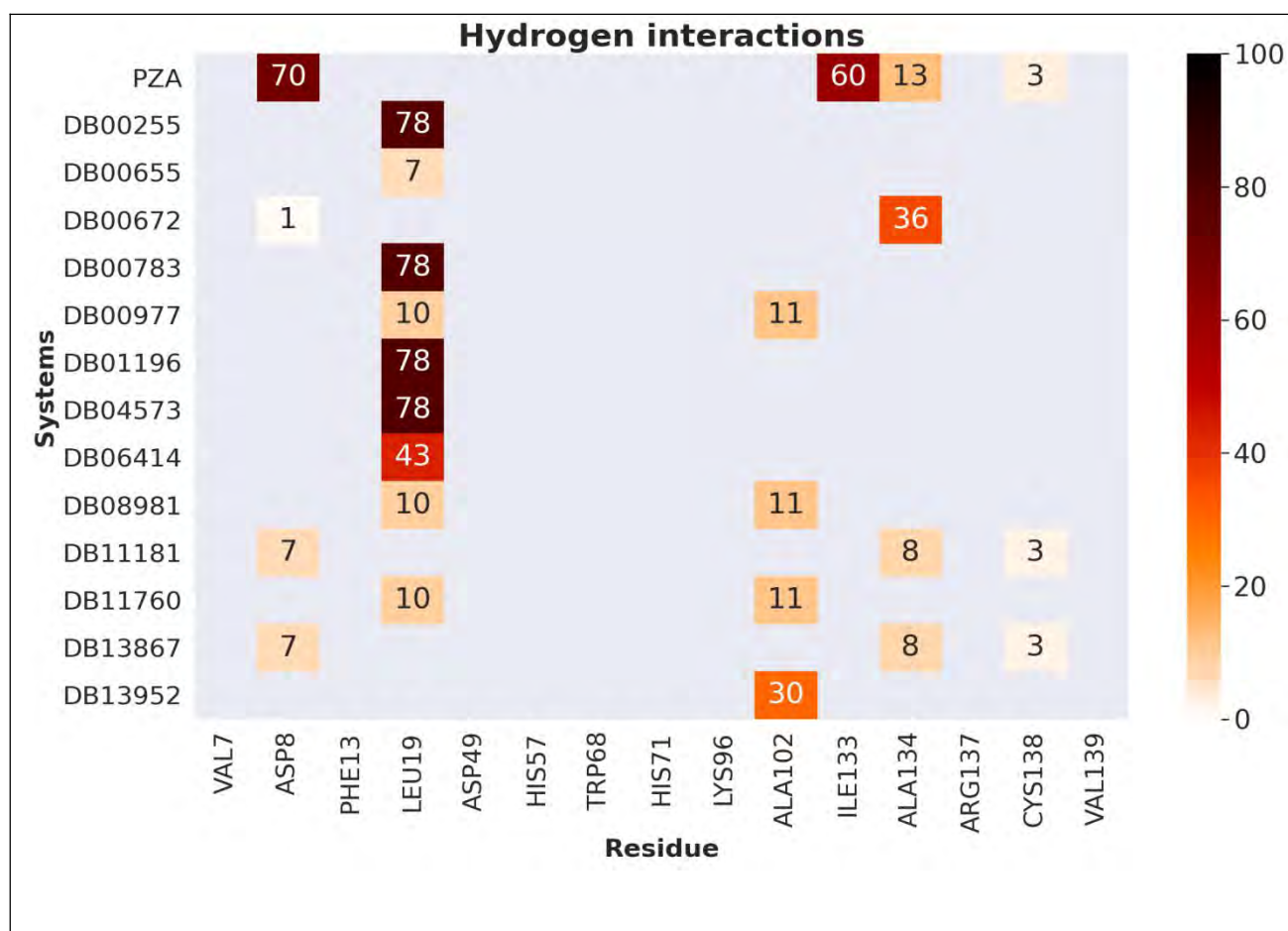
Figure 3.7: Summary of hydrogen bonds formed throughout the 150 ns simulation. The interactions were between the ligand and active site residues within 8 Å. The percentage occupancy values are stated for each residue interaction (0-100%). The occupancy values are also colour ranked from white-low to black-high.

3.4.6 Drug use and target site

Table 3.1 below gives the general overview on the uses and target sites of the selected compounds. Out of the 13 compounds, 6 are estrogen steroid supplements (DB00255, DB00655, DB00783,

DB00977, DB04573 and DB13952) that target estrogen receptors in the treatment of menopause disorders. The other compounds generally treat inflammations, diabetes and HIV-1 infections. Mild side effects like nausea, vomiting and rash are related with some of these drugs with extreme effects being carcinogenic.

Table 3. 1: Summary of the uses and target sites of 13 identifiied stable DrugBank compounds in WT PZase dynamic simulations.

Uses	Compound ID and name	Target area	Side effetcts
Syntehtic or natural estrogen steroid for perimenopause, menopause and postmenopause treatment	DB00255 Diethylstilbestrol	The drug targets the female reproductive tract, hypothalamus, pituitary and mammary gland.	It is a carcinogen
	DB00655 Estrone		Nausea, breast tenderness, edema, poor contact lenses
	DB00783 Estradiol	It binds to the Estrogen Receptor (ER) including ERα and ERβ subtypes, located in tissues like breasts, uterus, ovaries, skin, prostate, bone, faf and brain	Breast cancer and cardiovascular related diseases
	DB00977 Ethinylestradiol		No adverse effects except risks of thrombotis
	DB04573 Estriol		--
	DB13952 Estradiol acetate		Nausea and vomiting, and withdrawal bleeding in women
Treatment of non-insulin-dependent diabetes mellitus	DB00672 Chlorpropamide	Stimulating β cells of the pancreas to release insulin	May cause hypoglycemia, weight gain
Treatment of patients with metastatic and/or progressive carcinoma of the prostate	DB01196 Estramustine	It *is* a combination of estradiol with nitrogen that alkylates DNA leading to apoptosis.	--
Treatment of human immunodeficiency virus type 1 (HIV-1)	DB06414	Inhibits reverse transcriptase enzyme of HIV-1 and blocks DNA-	Rash, Nausea, Diarrhea and peripheral

infection	Etravirine	and RNA-dependent polymerase activity.	neuropathy
Treats inflammation	DB08981 Fenbufen	Prevents cyclooxygenase from producing prostaglandins which cause inflammation.	--
	DB13867 Fluticasone	Activates glucocorticoid receptors and inhibits lung eosinophilia in rats	Adrenal suppression
Anticholinergic drug for induction of mydriasis in ophthalmic solutions.	DB11181 Homatropine	Acts as an antagonist at muscarinic acetylcholine receptors	--
Used in germline BRCA mutated, HER2 negative, advanced or metastatic breast cancer	DB11760 Talazoparib	Binds to and inhibits PARP1 and PARP2 at the NAD+ binding site	--

3.6 CONCLUSION

Molecular dynamic simulations on the WT PZase-DrugBank compound systems were performed successfully and the behaviour of the systems were determined. From the initial MD runs of 20 ns on the 93 ligands (Section 3.4.1), 47 compounds were identified as stable by the end of the 20 ns short MD simulation (Figure 3.2). Extension of the simulation on the selected compounds revealed only 13 compounds with low ligand RMSD value and unimodal distribution. These ligands were represented in green throughout the whole analysis steps of Backbone RMSD, RMSF and radius of gyration, where they indicated more stable dynamic motions compared to the control PZA system. Further studies were done based on the number of hydrogen bonds and interaction between PZase and the ligand throughout the simulation (Figure 3.6 and Figure 3.7). Compounds that displayed the best stable dynamic motions across all analytical steps were DB00255, DB00655, DB00783,

DB00977, DB04573 and DB13952 (Figure 3.3 - Figure 3.7). These 6 compounds are all eostrogen derivatives currently used to supplement oestrogen levels in female and in the treatment of menopause disorders. The other compounds also behaved better than the PZA control and these drugs are generally approved clinical drugs for inflammation, diabetes and hypertension (Table 3.1). Since these compounds were found to be stable in the WT PZase systems, further MD analysis can be done in the mutant PZase systems.

CHAPTER FOUR

4. MUTANTS - MOLECULAR DYNAMICS

4.1. CHAPTER OVERVIEW

Mutations are alterations in DNA sequence that may result in changes in amino acid sequences, eventually affecting the functional and structural purpose of the organism. Techniques such as molecular dynamics have been implemented to understand the dynamic effects of mutations in biomolecules. The aim of this chapter was to mutate WT PZase and investigate the effect of those mutations on PZase-ligand systems using MD simulations. This was done by mutating WT PZase based on the identified mutations from TB resistant database and previous group study by Sheik Amamuddy *et al* (2020). The identified hit compounds from Chapter 3 were further studied in subjection to mutations. The simulations and calculations were performed just as in Chapter 3, for 150 ns across all systems. The WT system for each specific ligand was used as the control. Calculations were done using RMSD, RMSF, radius of gyration and hydrogen bond count. The structure of this chapter begins with background introduction on mutations, detailed methodology, followed by results, discussion and chapter conclusion.

4.2 INTRODUCTION

4.2.1 Mutations

Mutations are changes in the genetic make-up of a cell that may be hereditary and these changes are the main cause of diversity among organisms (Loewe *et al.*, 2008; Griffiths, 2020). Mutations may occur spontaneously due to errors in DNA replication or may be induced by exposure to

environmental conditions with detrimental effects such as ultraviolet radiation or chemical carcinogens (Lodish and Zipursky, 2001).

The different types of mutations include nonsynonymous and synonymous. The nonsynonymous mutations are categorized as either nonsense mutations in which an amino acid is replaced with a stop codon leading to early termination of protein synthesis or frameshift mutations which changes the reading frame and introduce new unrelated amino acids resulting in the production of a different protein (Loewe *et al,* 2008; Khan and Malik, 2020). The synonymous point mutations are due to alteration of a single base pair, producing a missense mutation whereby one amino acid is substituted with another amino acid. These mutations alter electrostatic nature of protein surfaces which may eventually affect the proteins' folding, functionality, stability and ligand binding affinity (Khan and Malik, 2020).

4.2.2 Pyrazinamidase mutations

Drug resistance has been reported in all first line TB drugs (Khan *et al.,* 2019) and it is a major hinderance to TB eradication (WHO 2019; Khan and Malik, 2020). Resistance of the first line drugs in M. *tb* treatment has been reported in various studies and this resistance has been attributed to mutations in the target protein coding or promoter regions (Whitfield *et al.,* 2015)

A database that consists of all TB related mutations (in DNA and proteins) has been developed with the goal to easily provide mutations responsible for *M. tb* antibiotics resistance and gives biological and therapeutic interpretation of mutations. MUBII-TB-DB is a TB database that consist of seven major loci mutations, having a great effect on the treatment and management of tuberculosis. The

major mutations gene loci are *rrs* for amikacin drug, *rpoB* for rifampcin, *KatG* and *mabA* operon for isoniazid, *gyrA* and *gyrB* for fluoroquinolones and *pncA* for pyrazinamide (Flandrois *et al.,* 2014). Out of all the reported cases, 85% of resistance to the essential prodrug (PZA) has been discovered to be caused by mutations in the coding and promoter region of the *pncA* gene (Khan *et al.,* 2019; Khan and Malik, 2020). These mutations mainly affect three regions, which are residues 3–17, 61–85 and 132–142, resulting in the proteins' loss of activity, although mutations on other regions may also alter the protein's structural and functional behaviour (Petrella *et al.,* 2011; Khan *et al.,* 2019; Khan and Malik, 2020).

4.2.3 Mutation study

Analysing and studying various protein mutations may aid in understanding the molecular mechanisms and dynamics that result in drug resistance (Khan and Malik, 2020). Biomolecular dynamics is a technique used to study atom interactions in proteins or other systems over a stipulated time period, providing insight on target-ligand interactions and structural changes. It is also useful in determining and studying the effects of perturbates such as protonation, ligand binding or unbinding, phosphorylation and mutations (Hollingsworth and Dror, 2018). Thus, this *in silico* approach may be used to elucidate the dynamics within mutant systems and predict the effects of the mutations on the macromolecules functionality and stability.

A number of studies have reviewed and analysed PZA prodrug susceptibility in mutants (Chang *et al.,* 2011; Miotto *et al.,* 2014; Whitfield *et al.,* 2015; Khan *et al.,* 2019; Wu *et al.,* 2019) and studied the PZase molecular mechanism of resistance using computational techniques such as MD simulations (Junaid *et al.,* 2019; Juniad *et al,* 2020; Sheik Amamuddy *et al.,* 2020). Information on

the mechanism of resistance and mode of action will assist in the design and identification of novel effective drugs (Juniad *et al.,* 2020). Recent computational methods by Sheik Amamuddy studied the binding and unbinding of PZA on a selected number of mutations based of mutant location (Sheik Amamuddy *et al.,* 2020). Their study determined the unbinding time of PZA from the mutants and thus predicted the mutations in which PZA was more or less stable. The study by Sheik Amamuddy *et al* (2020) was used as a reference point for selecting the protein mutants. MD simulations were done on some selected mutants to investigate the dynamic behaviour of the various mutant-ligand systems, using the 13 potential compounds identified from Chapter 2 and Chapter 3 analysis.

4.3 METHODOLOGY

Generally, the same MD steps taken in Chapter 3 were repeated with minor alterations after completing the initial step of mutating PZase. The overall workflow of MD simulation and dynamic calculations on the PZase mutants is displayed below (Figure 4.1).
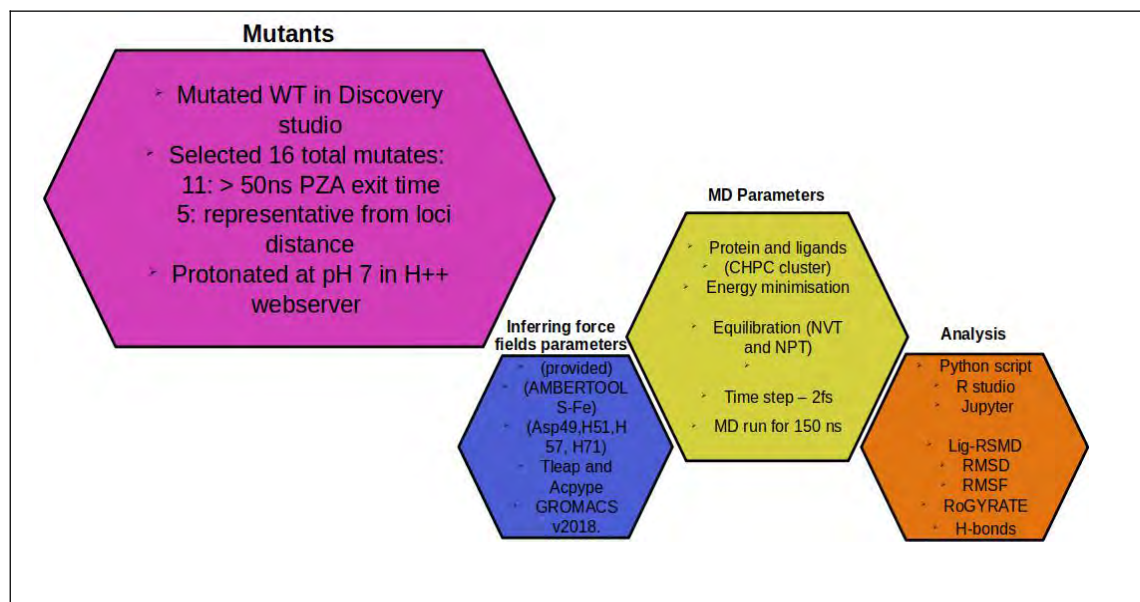
Figure 4.1: Summary of the performed molecular dynamics simulation on the mutant systems. The first step was introduction of mutations followed by the general MD steps of inferring force-field, running molecular dynamic simulations and analysis of obtained results.

4.3.1 Mutants preparation

A total of sixteen PZase mutations were selected from those studied by the previous group (Sheik Amamuddy *et al.,* 2020), which they identified from the MUBII-TB-DB database. Twelve of these mutations were selected based on systems that had PZA unbinding time of less than 50 ns (Table 4.1). Additional mutants were selected by selecting representatives from the given groups, which were catergorised based on location of the alpha carbon of the mutation from PZA center of mass distance (Table 4.2). Group 1 represented mutations within a radius of 6.7 Å while group 2 comprised of mutations between 6.7 Å and 11.0 Å. Group 3 had mutations in a distance above 11 Å while group 4 consisted of mutations in the metal binding site. The position of these mutations relative to the PZA ligand are illustrated in Figure 4.2.

Table 4. 1: Selected mutations from the previous group study by Sheik Amamuddy *et al* (2020) where PZA-bound protein complexes had an estimated exit time point of less than 50 ns.

Mutant	Estimated exit time (ns)	Group
A134V	48.71	1
D8G	18.26	2
Y103S	27.70	2
V139M	25.94	2
A3P	33.80	3
L116R	26.59	3
T61P	48.66	3
A146V	7.34	3
D49A	12.31	4
D49G	12.16	4
H51P	29.80	4

Table 4.2: Selected mutants from Sheik Amamuddy *et al* (2020) from group 1-3. Mutants were selected based on systems where ligand appeared stable prior to release.

Mutant	Group
H137R	1
Q10P	2
R140S	2
D63G	3
L85R	3
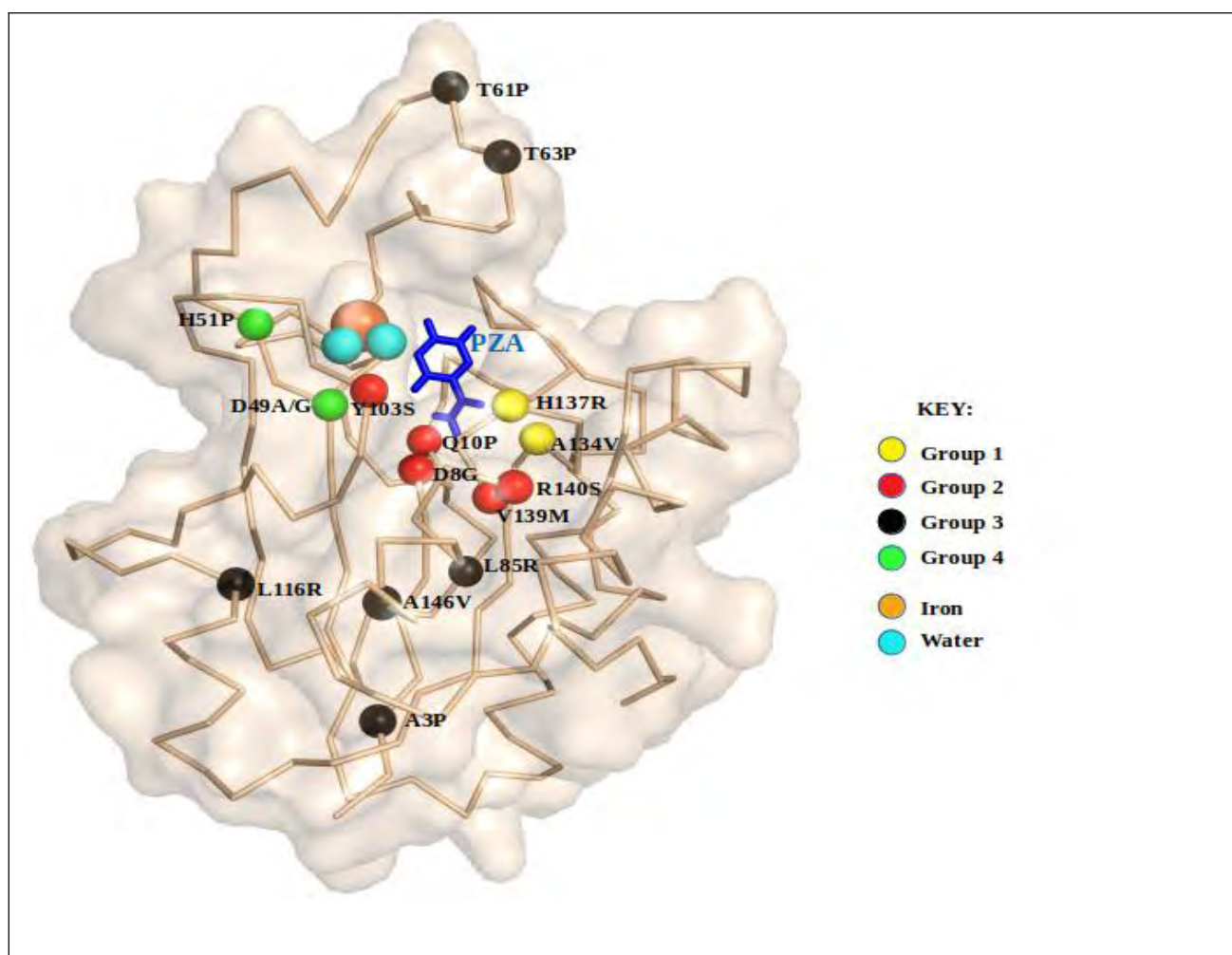
Figure 4.2: Representation of positions on the mutated amino acid residues in PZase, based on the center of mass of the docked PZA ligand. The amino acid residues are coloured based on their

designated groups where yellow, red, black and green are for groups 1, 2, 3 and 4 respectively. Labelling of residues shows transition from WT to mutant amino acid residue.

4.3.2 MD preparation and analysis

Point mutations were introduced on to the WT PZase in BIOVIA Discovery Studio following Table 4.1 and 4.2 above. The mutated structures were saved and protonated at pH seven on H++ webserver. Inferring of all metal binding site residue parameters was done across all mutants as described in Chapter 3 (Section 3.3.1) except for mutants D49A, D49G and H51P. In mutants D49A and D49G, Asp49 parameters were not inferred while in H51P, His51 parameters were not included. This was because D49A/G and H51P mutations affected the metal binding site residues hence D49 and H51 residue parameters were not needed for the respective mutants.

Minimization, equilibration, MD simulation and analysis (RMSD, RMSF, radius of gyration, hydrogen bond) steps were performed using the same steps as in WT simulation as described in Chapter 3 (Section 3.3.2 and 3.3.3). An R script was used to calculate the exact medians for ligand RMSD on the WT and mutant ligand systems to determine range of median RMSD values. An additional analysis tool *cpptraj* was used to investigate the fraction of time the ligands and the protein residues formed hydrogen bonds.

4.4 RESULTS AND DISCUSSION

The mutated sixteen PZase in complex with the potential thirteen hit compounds systems were successfully subjected to MD simulations of 150 ns. Analysis using RMSD, RMSF, radius of gyration and hydrogen bonds were done in comparison to the wild type systems.

A number of studies on effects of mutations on PZase revealed alteration in enzyme activity and stability have been done (Sheen *et al.,* 2009; Rajendran and Sethumadhavan, 2014; Junaid *et al.,* 2018; Juniad *et al.,* 2020; Sheik Amamuddy *et al.,* 2020). Previous related studies on some of the mutations in this study has been done. An *in silico* study by Juniad *et al* (2020) on mutation Q10P showed that the mutation greatly affected Gln10-His43, Phe50-Gly75 regions and also disturbed the catalytic triad (Asp8, Lys98 and Cys138) and metal bind site residues (Asp49, His51, His57 and His71). A study on D8G by Rajendran and Sethumadhavan (2014) indicated that the alteration caused PZA rigid binding, resulting in failure of its conversion to pyrazonoic acid, thus the cause for its resistance. As mutations may alter the folding, packaging and overall structure of a molecule, the selected hit compounds may thus behave differently as in the WT systems. Therefore, from this Chapter's analysis, ligands that display similar behaviour to WT systems across majority of the mutants would be potential drugs for repurposing to TB treatment.

4.4.1 Root Mean Square Deviation

This parameter measures the average difference between any two values. A smaller value indicates less deviation and thus signifies higher similarity between the compared states. A higher RMSD value imply great difference between systems, thus less similarities between them. The RMSD values for the ligand and backbone protein residues were calculated.

4.4.1.1 Ligand RMSD

The behaviour of the compounds throughout the simulation were analysed by calculating the ligand RMSD values in comparison to the WT systems (Figure 4.3). The control system (WT) is highlighted in blue while the mutants systems are in yellow. The PZA ligand systems had mutant-

ligand systems with higher RMSD values and thin distribution plots compared to the rest of the ligand systems. In PZA systems, the control had unimodal conformation with RMSD value close to 0.16 nm (Appendix 2). Visualization in VMD of the ligand in mutant systems showed that PZA took a number of conformations and left the system in (A3P, A134V, D49A, D49G, H51P, L116R, Q10P, R140S, T61P, V139M, Y103S) mutants. In mutant systems A146V and D8G, PZA had high RMSD value and bimodal conformation. However, PZA exhibited unimodal conformation in systems D63G, H137R and L85R. The behaviour in these systems shows that the ligand was stable throughout the simulation and similar results were reported by Sheik Amamuddy *et al* (2020) for ligand release time analysis. The aim of this analysis was to identify compounds that remained stable across all mutants. Generally, majority of the ligand systems exhibited unimodal and low RMSD values, an indication of ligand stability. Systems of DB00672, DB06414 and DB11181 showed higher and more multimodal conformations compared to other systems. Ligand systems that portrayed best stability were DB13952, DB13867, DB08981 and DB00255, where the ligands displayed well distributed unimodal plots closely ranging to the WT RMSD value (Figure 4.3 and Appendix 2).
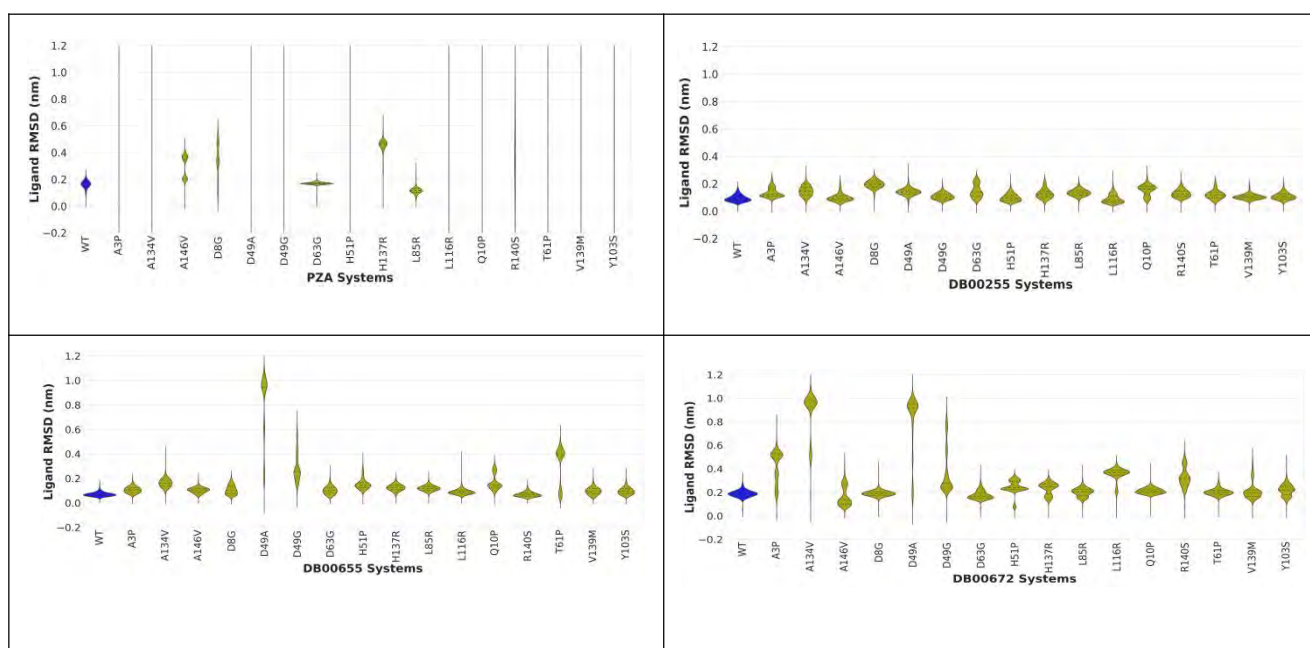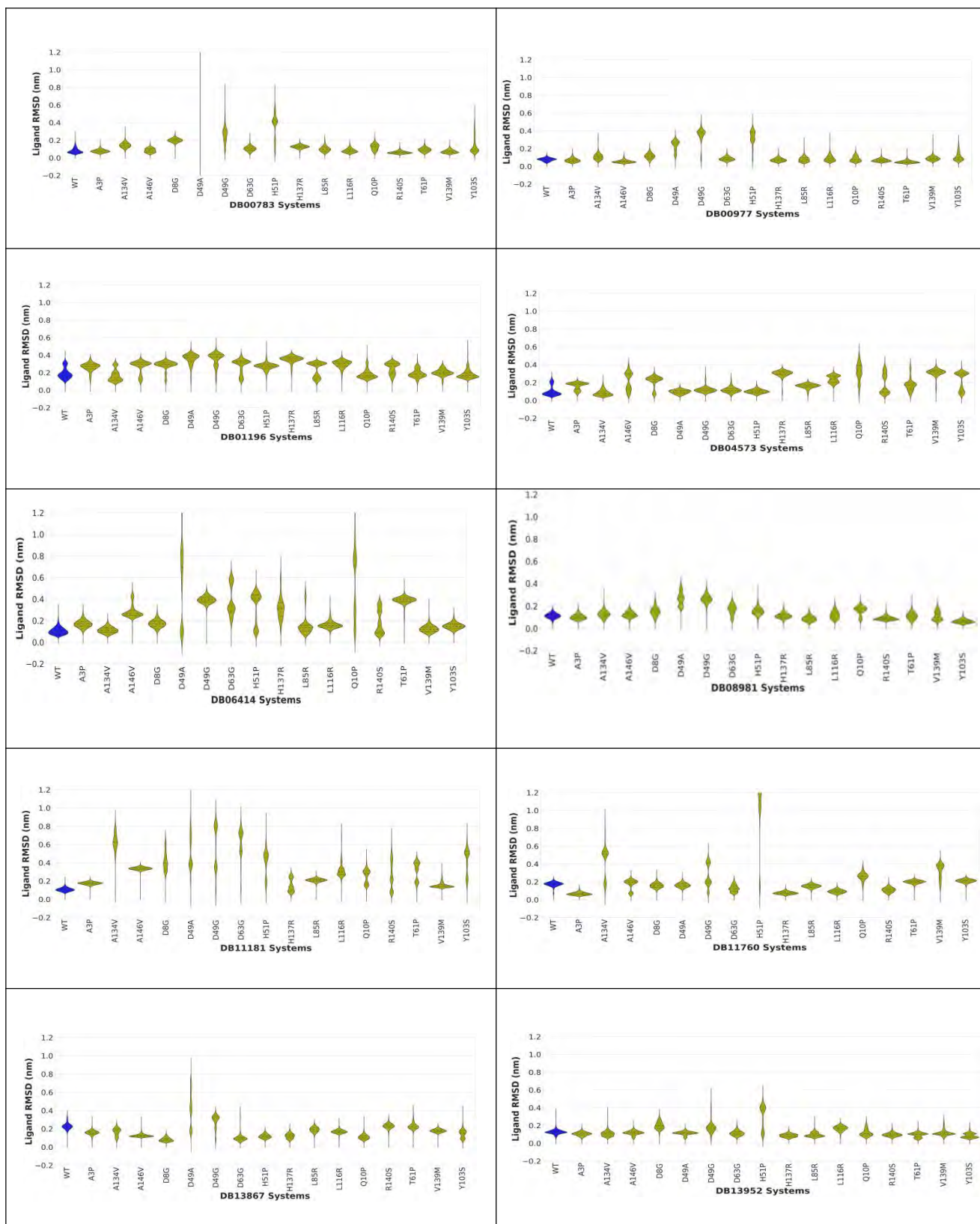
Figure: 4.3: Ligand RMSD violin plots of the hit compounds across mutants. The WT systems are highlighted in blue while mutant ligand systems are in yellow.

A table summarizing the median values is shown on Appendix 2. The table indicates the WT and mutant ligand RMSD values. As already shown in the kernel plots (Figure 4.3), majority of the ligand RMSD values clustered around the same values. Calculation of the exact median values clearly showed that there was little difference between the RMSD values as most of the mutant systems were within 0.5 nm range with the control WT system. Extreme ligand systems with values that were not within 0.5 nm of the WT are highlighted in red (Appendix 2). PZA system had the most number of systems having values out of 0.5 nm with the WT system while systems like DB00255, DDB00977, DB01196, DB04573, DB08981, DB13867, DB13952 had all systems within the range. This overall implies that the ligands were stable in the mutant systems as they behaved and exhibited similar RMSD values as those in the WT systems.

4.4.1.2 Backbone RMSD.

In addition to the Ligand RMSD calculations,protein backbone RMSD was calculated to understand the deviation of the backbone residues. From Figure 4.4 the WT systems (blue) generally had lower RMSD values, however the overall range from 0 to 0.25 nm implies the systems were all in a close range. Unlike in the Ligand RMSD plots, the backbone RMSD data displayed more unimodal conformations with RMSD values all clustered around the WT system of 0.15 nm. Mutant D49G RMSD value was consistently higher and displayed bimodal or multimodal conformations across majority of the systems. In systems such as PZA (Figure 4.3) where the ligand left majority of the systems, the calculated backbone RMSD values does not give true insight on system behaviour as the results only provide data of the protein in the absence of the the ligand. Thus therefore explains the similar unimodal conformations and RMSD values on Figure 4.4 across the mutant systems.
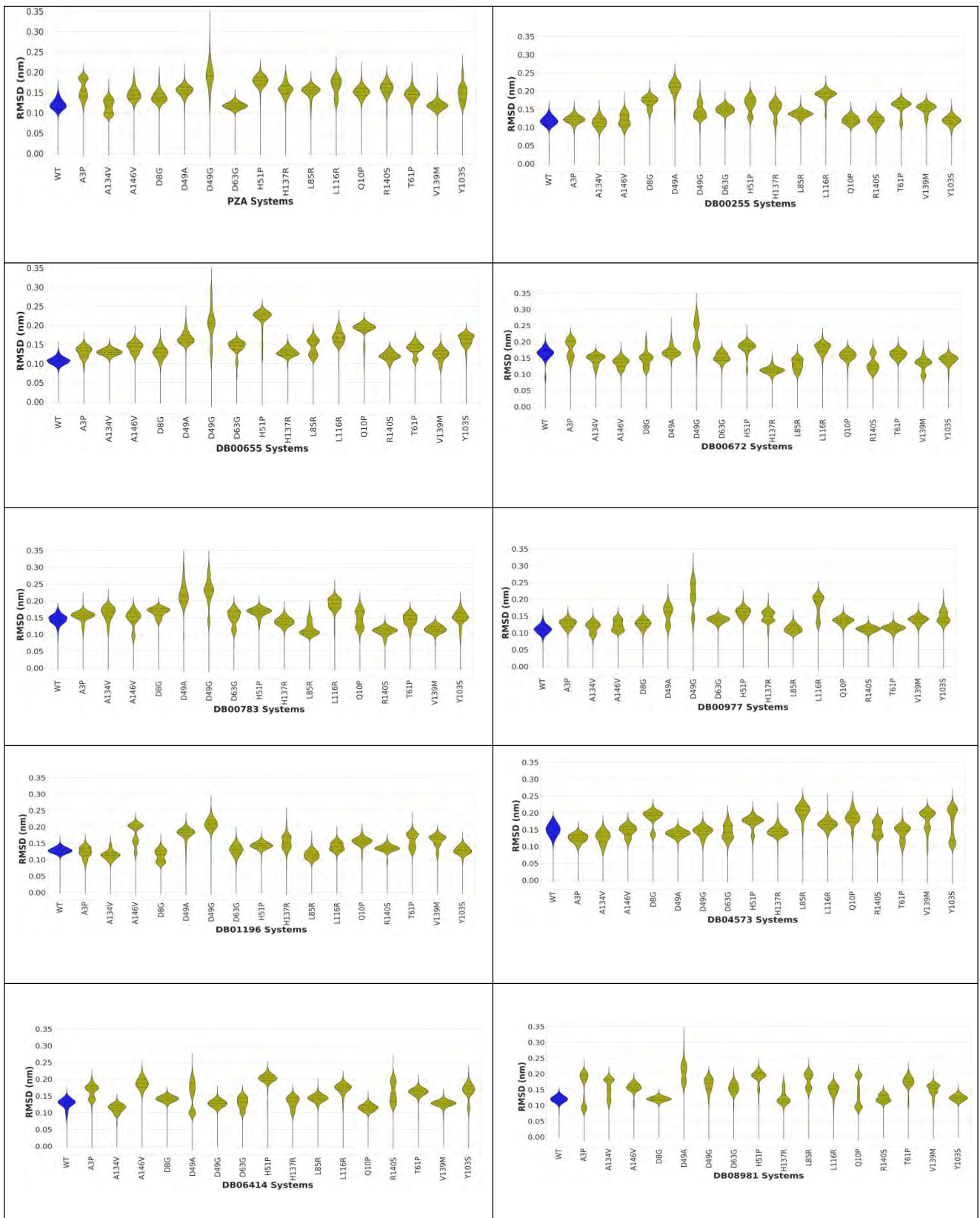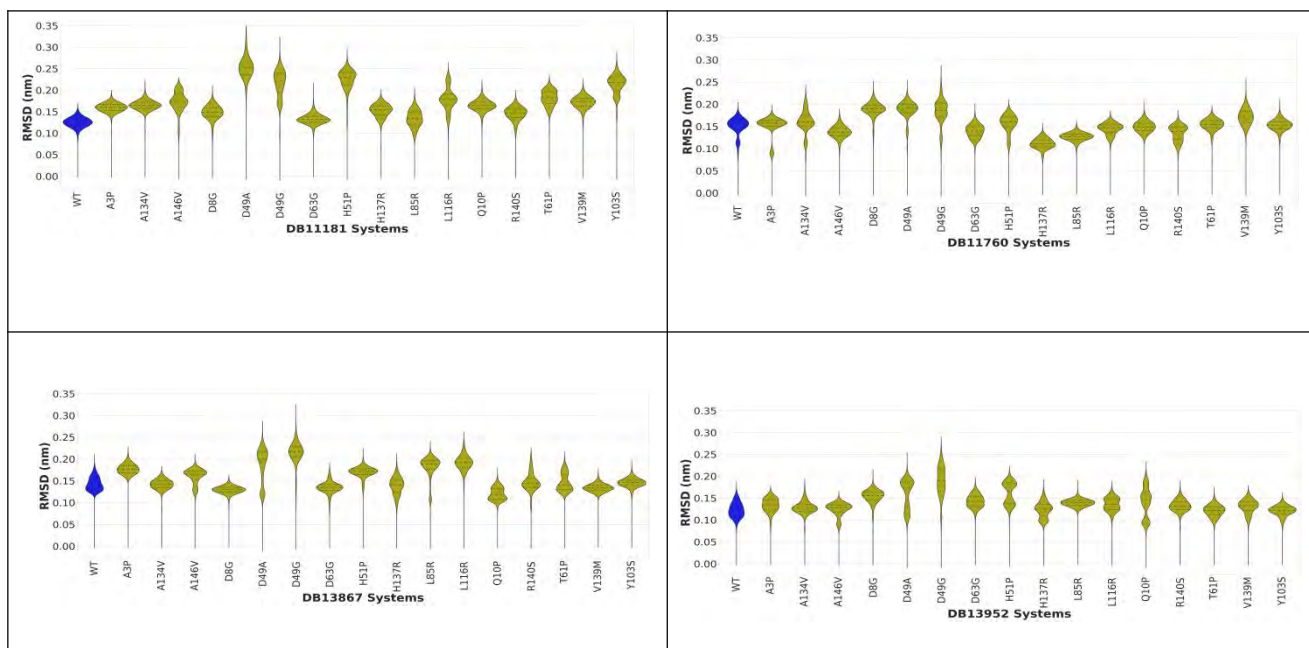
Figure: 4.4: Backbone RMSD violin plots for 150 ns of the hit compounds across mutants. The WT systems are highlighted in blue, mutant systems ligands in yellow.

4.4.2 Radius of gyration

4.4.2.1 Active site radius of gyration

The stability of protein folding is measured in MD simulations by calculating the systems radius of gyration. The compactness of the active site pocket and protein were analysed, where a well packed structure would be expected to have a low Rg value and exhibit a unimodal structure while a loosely packed structure will have a bigger Rg value. A more compact system would represent a more stable receptor-ligand system. From Figure 4.5, across all ligand systems, majority exhibited tight unimodal conformations with Rg values below 1.00 nm. Mutant systems D49A and D49G prominently displayed different behaviour across all systems, exhibiting a higher Rg value and/or bimodal conformations. This is because the mutations affect the metal binding site coordinating residues (D49) in which replacement of Aspartic acid to Alanine or Glycine disrupts the closely packing of the protein residues because of the difference in amino acid structural properties. For the

other systems, the displayed data implies that the change in amino acid residues did not greatly affect the packaging and arrangement on the active site pocket.
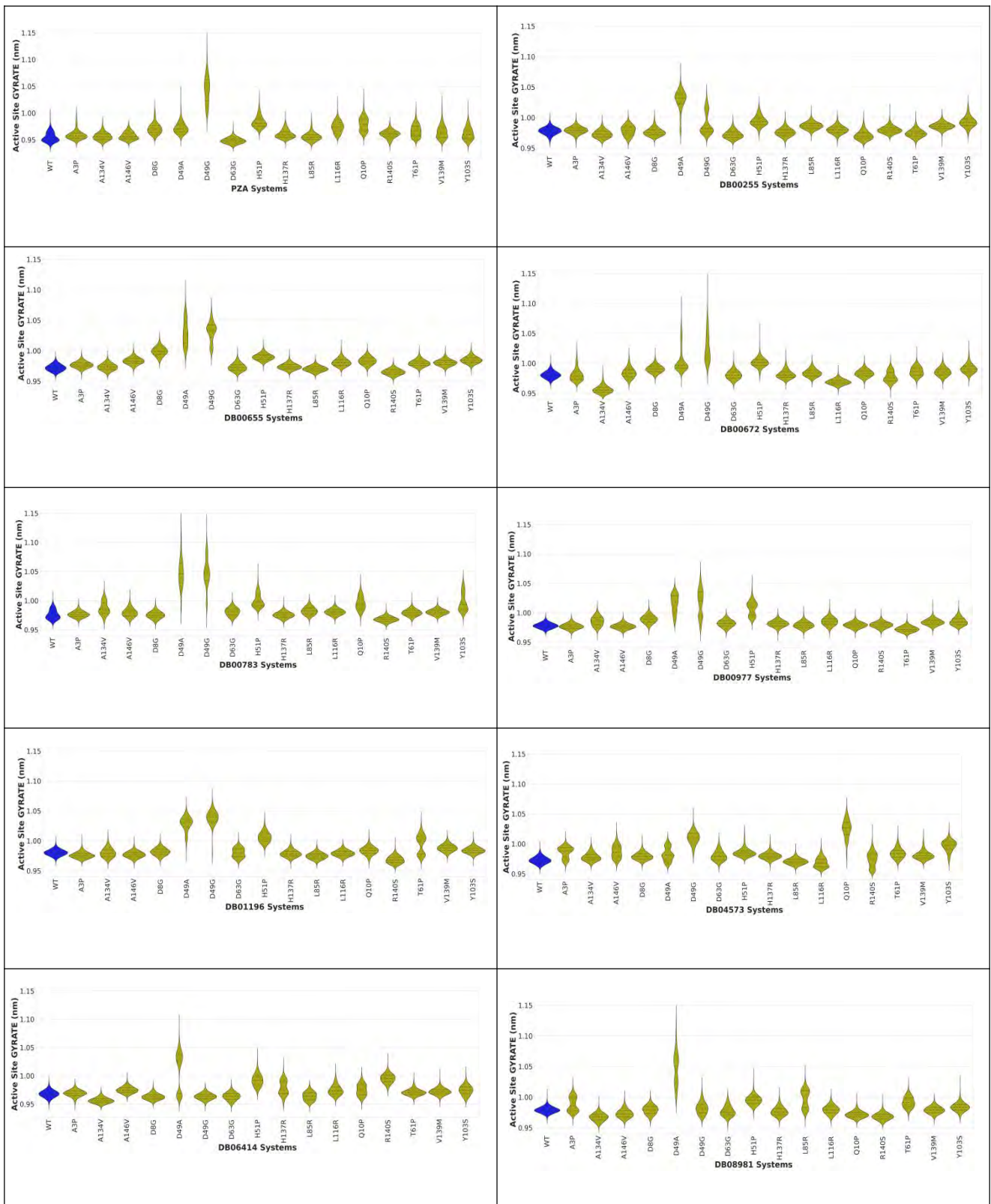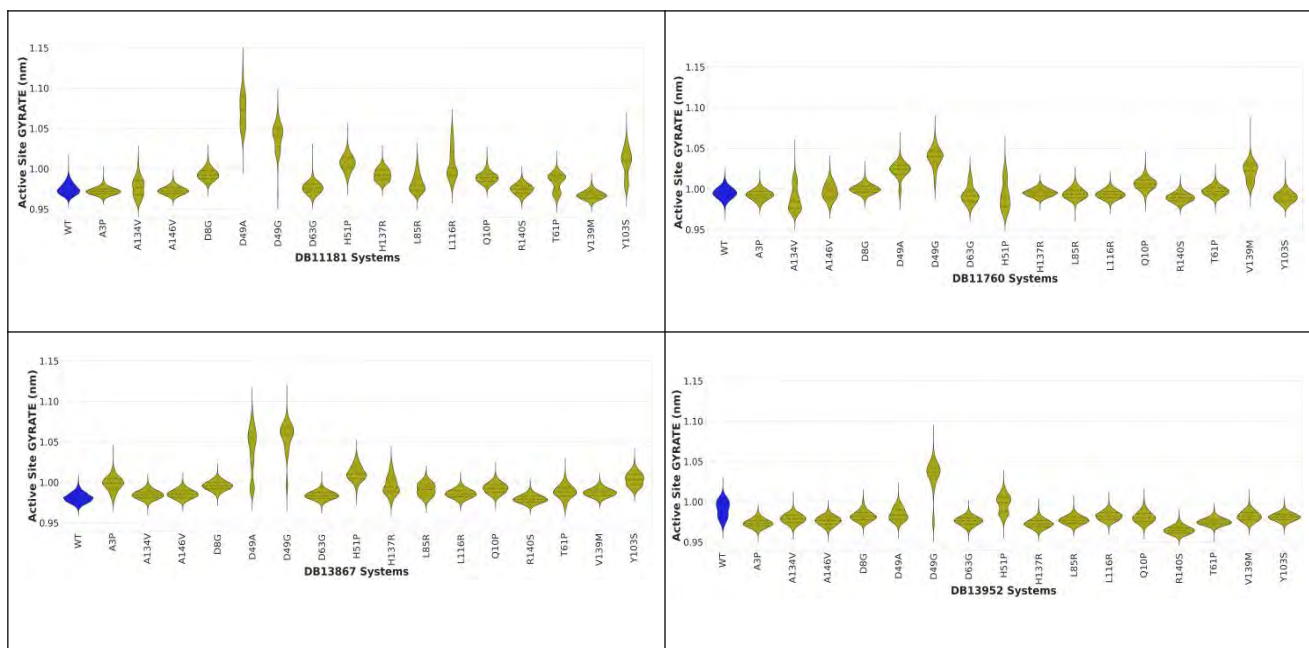
Figure 4.5: Active site radius of gyration violin plots for the MD simulation of 150 ns on the hit compounds across PZase mutants. The WT systems are highlighted in blue while mutant systems are in yellow.

4.4.2.2 Whole system gyration

The compactness of the whole system was also analysed to determine how the protein-ligand complex affected the residues packing throughout the simulation (Figure 4.6). This analysis displayed a similar trend to that observed in active site gyration analysis (Figure 4.5). Generally, the gyration value ranged from 1.52 to 1.60 nm, a small range for distinct comparison between systems while the WT ligand systems were between 1.54 to 1.56 nm. However, systems D49A and D49G still exhibited distinct Rg values in most of the ligand systems (DB00255, DB00655, DB00672, DB00783, DB00977, DB01196, DB11181, DB11760, DB13867). Since the introduction of mutations are expected to alter the active site pocket and whole protein complex compactness of ligand binding, results from this analysis shows that the presence of these ligands maintained similar protein-ligand compactness as that observed in the WT systems.
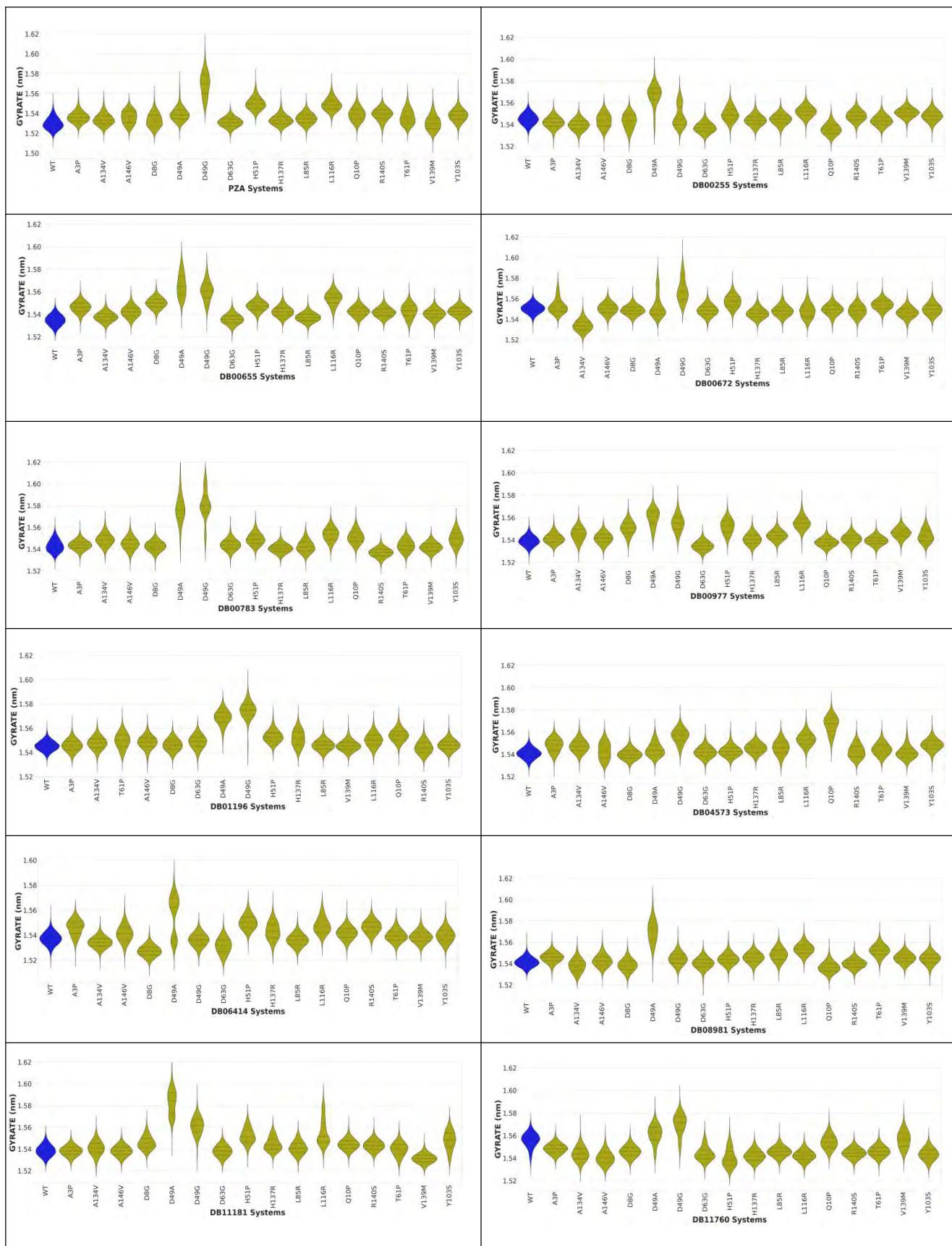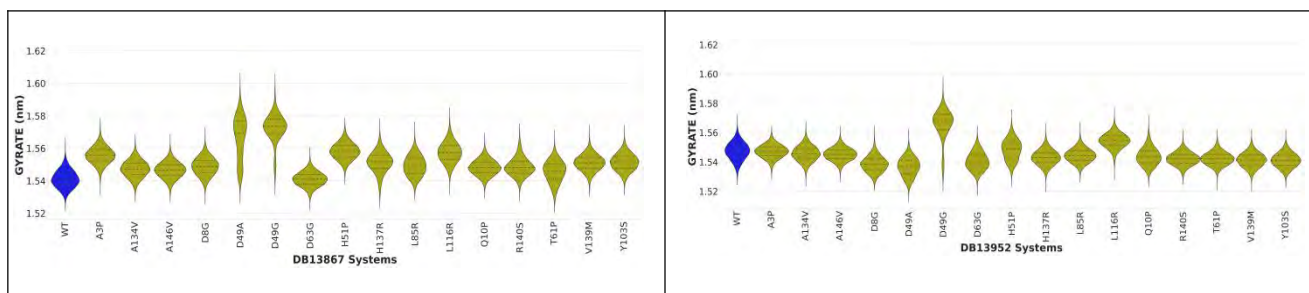
Figure 4.6: Whole protein system radius of gyration for PZase WT and mutant systems. The WT systems are highlighted in blue while mutant systems are in yellow.

4.4.3 Hydrogen bonds

Throughout the simulation period, the number of hydrogen bonds present at each time were counted. Figure 4.7 gives a summary on the number of hydrogen bonds exhibited within each mutant-ligand system. As shown in the PZA system, the ligand had more hydrogen bonds in the WT system and in three of the mutant systems (D63G, H137R, L85R). This data also corresponds to the data provided by the PZA ligand RMSD analysis where only these three mutants showed unimodal conformations with low RMSD values (Figure 4.3). In other mutant systems, PZA hardly reached 50 ns with at least one consistent hydrogen bond, shown by the presence of white space after 50 ns. These results support those obtained by Sheik Amamuddy *et al.,* 2020, who explained PZA unbinding within 50 ns of the simulation and thus explains the cause for PZA resistance in these mutants.

Among the selected compounds, ligands DB00255, DB00655, DB00783, DB00977, DB04573, DB06414, DB08981, DB11760, DB13867, DB13952 displayed more bonds compared to PZA in majority of the mutants. This suggests these ligands as potentially better drugs than PZA. The other ligand systems DB00672, DB01196 and DB11181 had fewer bonds but however had almost at least one bond throughout the simulation an indication that the ligand was bound to the active pocket

through out the simulation regardless of exhibiting high flexibility when visualized VMD.an . The active site mutations (D49A and D49G) distinctly had few bonds in systems DB01196, DB08981, DB13867, results that relate to those observed in the previous analysis techniques of RMSD and gyration.

The hydrogen interactions formed between the ligand and the proximal protein amino acid residues were determined and summarized in Figure 4.8. The number corresponding to each residue represents the bond occupancy percentage values. A bigger value, represented by a darker colour implies that the ligand bond to the precise residue was present for a longer time during the simulation. The best three ligands based on hydrogen bond count (DB00255, DB04573 and DB08981, Figure 4.6) displayed high occupancy values with the active site residue Asp8.
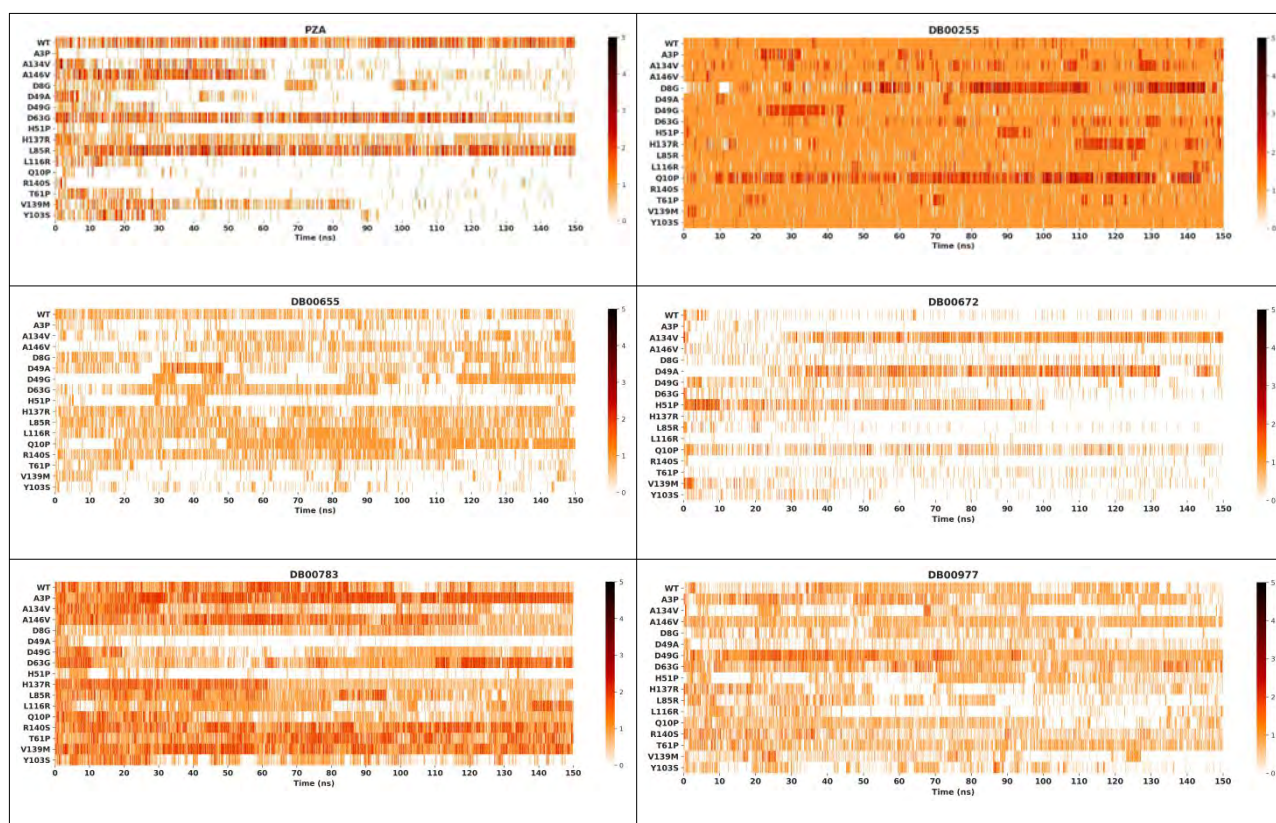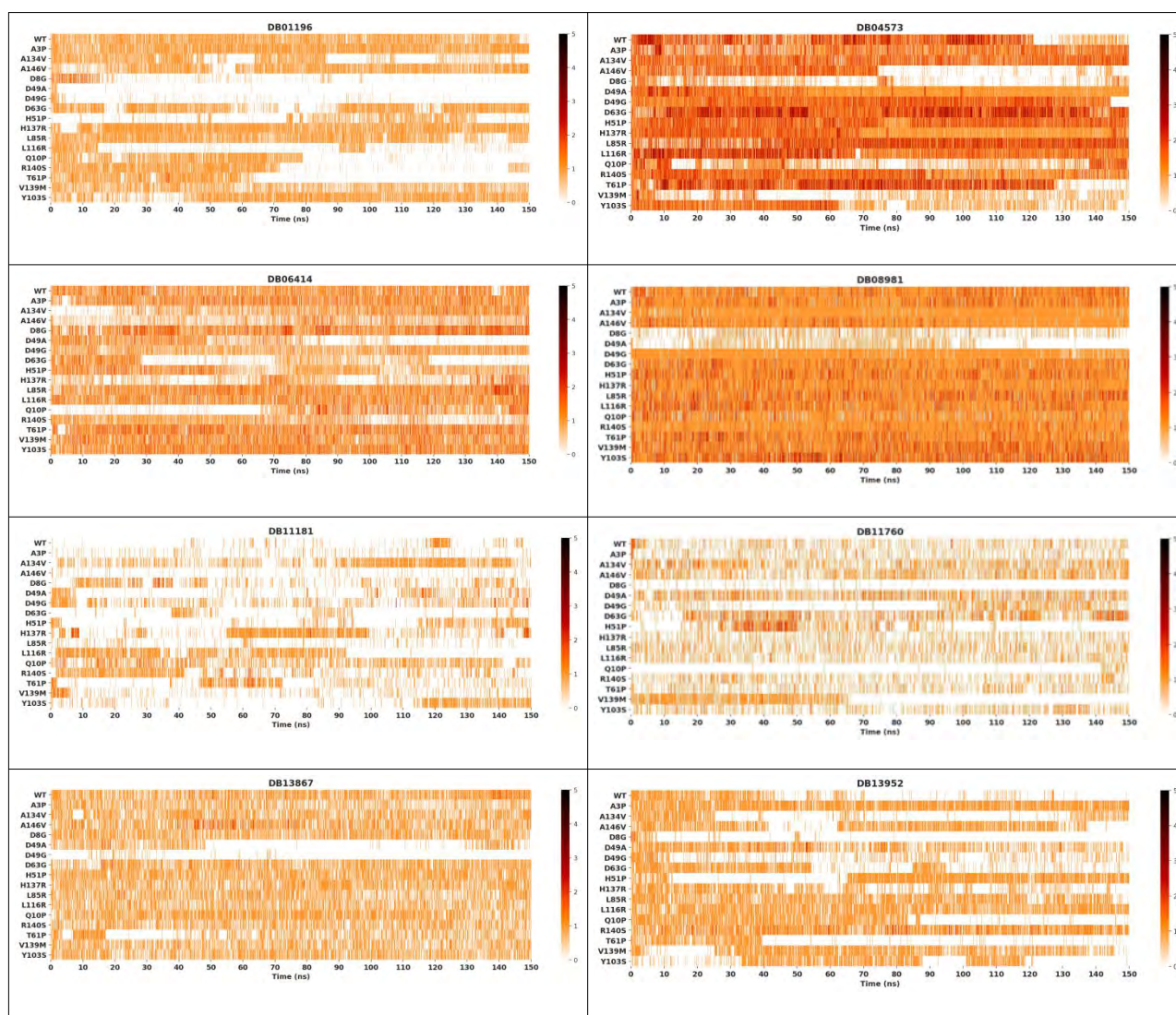
Figure 4.7: Distribution of hydrogen bonds through out 150 ns MD simulation. The light colour represent fewer bonds while a dark colour shows more hydrogen bonds at a specific time.

Compound DB08981 averagely had 100% occupancy with the active site residue Asp8 across all mutants except for mutants D8G, D49A and D49G that displayed 6%, 7% and 6% on Asp49, Lys96 and Asp8 respectively. This is a crucial bond present throughout the simulation occurring with one of the active site residues thus results in the ligand being bound to the catalytic pocket for the whole simulation period. These ligands also interacted with Ala134 and active site Cys138 residues. Compound DB00255 had similar interactions to DB08981 with Asp8 residue. In other ligand

systems, there was a wider distribution of bonds with majority of them occuring with Leu19, Lys96, Ala102, Ile133 and His137 amino acid residues.
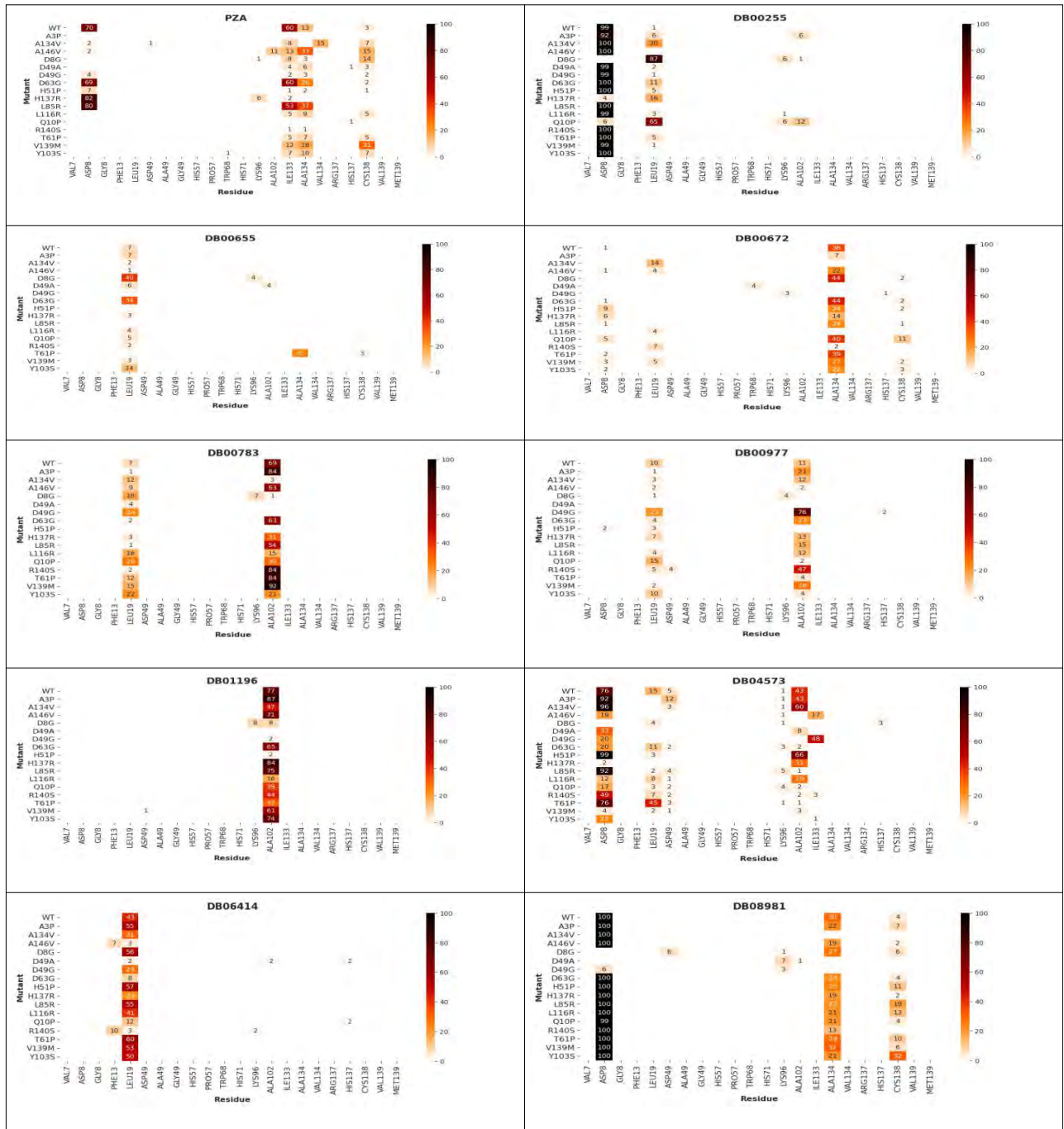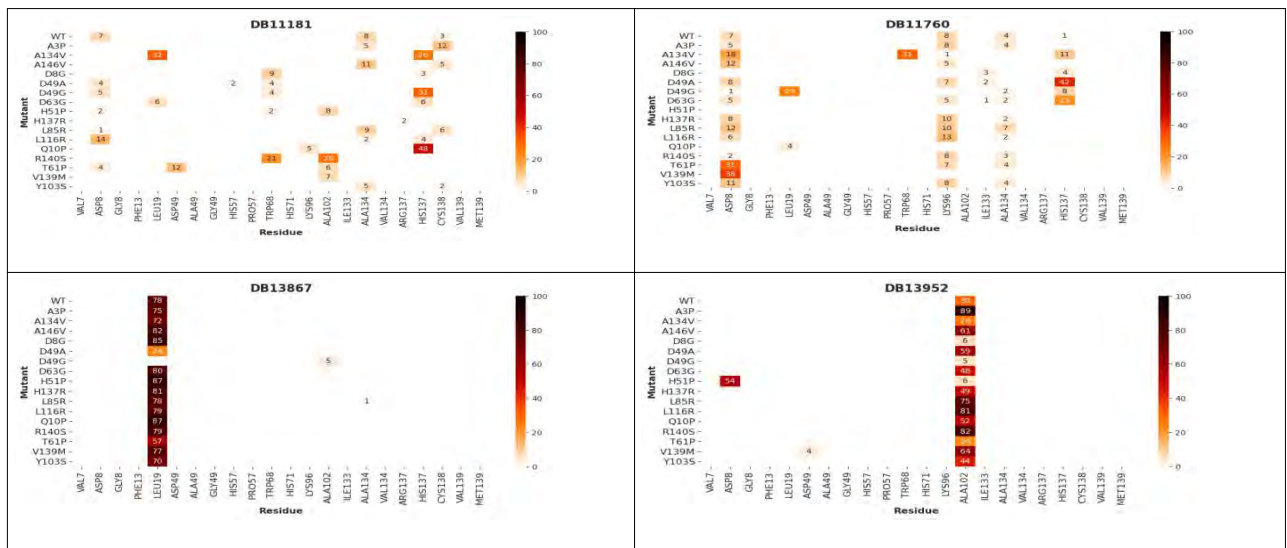
Figure 4.8: Occupancy of hydrogen bonds throughout the 150ns simulation period. The percentage value is highlighted in each box. The light colour represent fewer bonds than dark color.

4.4.4 RMSF

In all systems, the degree of fluctuating residues were calculated using RMSF analysis. This measure predicts the protein regions that are most or least flexible. From Figure 4.9, the most fluctuating areas are highlighted by a dark colour ranging from 0.05 to 0.45. Generally, across all systems, comparison of the WT systems in each ligand system to its respective mutant systems shows little difference. The most highlighted fluctuating regions in majority of the mutants were residues 14-17, 34-40 and 184-186, mapped on Figure 4.10. This analysis also revealed active site mutants D49G and D49A having the most fluctuating regions on residues 52 - 64 in systems of PZA, DB00255, DB00655, DB00672, DB00783, DB00977, DB08981, DB11181, DB11760, DB13876, DB13952. Mapping these fluctuating regions on the proteins structure indicated that these regions were mainly loop regions (Figure 4.10). However, additional fluctuating residues 52-64 in mutants D49G and D49A occurred on some secondary structure regions (highlighted in magenta in Figure 4.10). The overall analysis implies that majority of the mutant-ligand complexes were quite stable as signified by the low RMSF values and light colors on the heat maps.
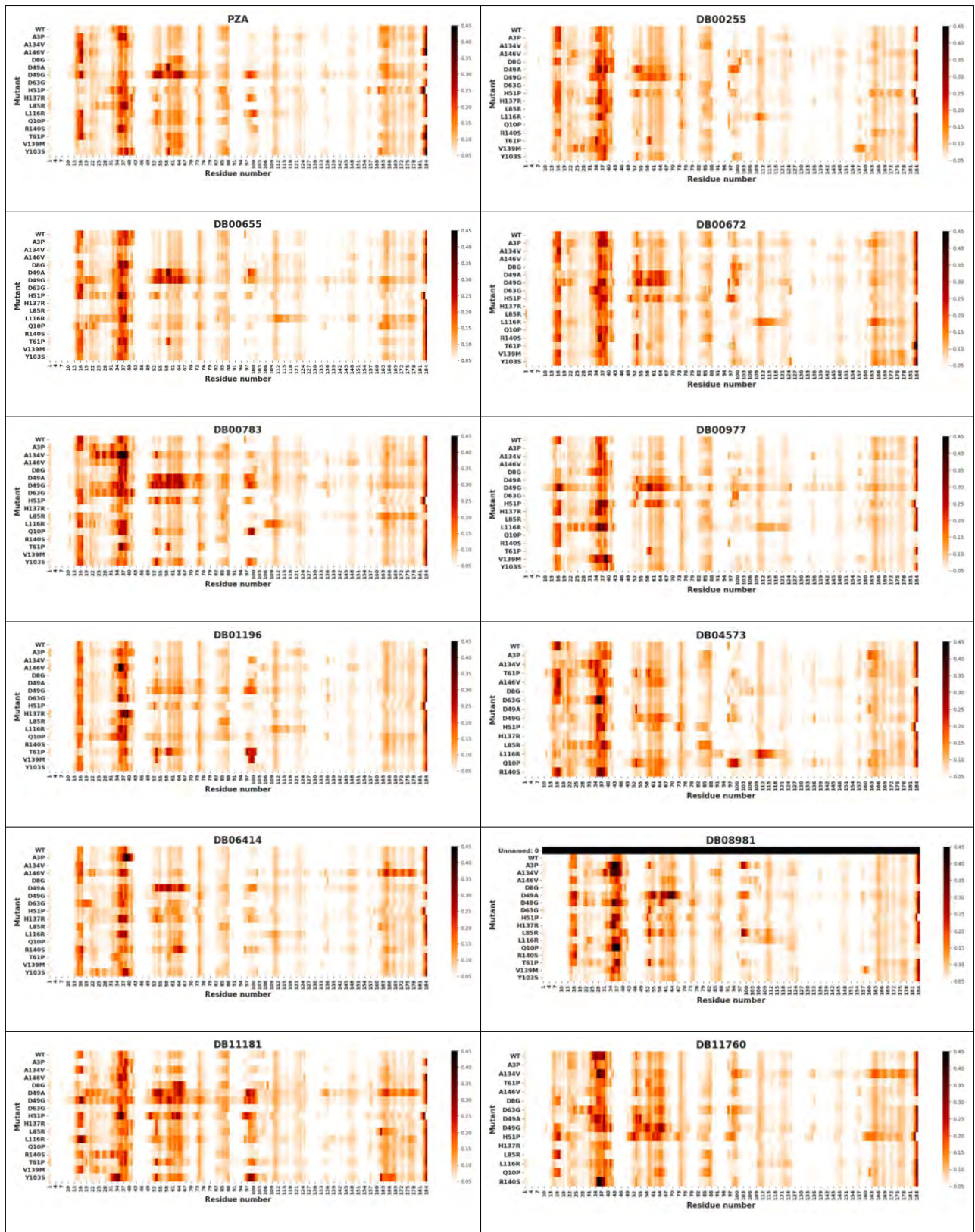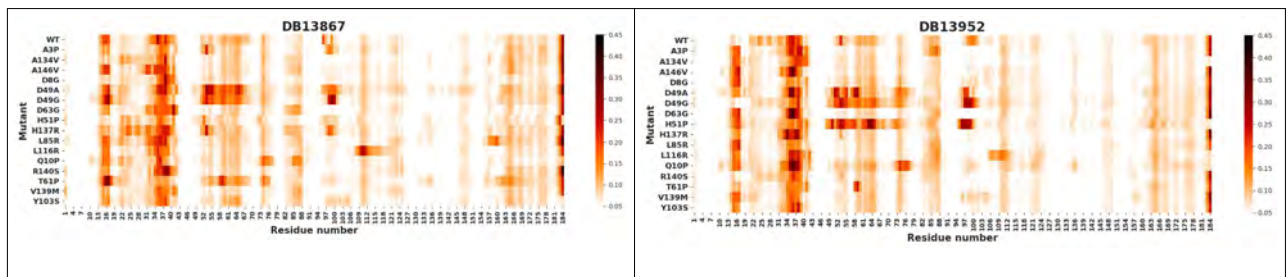
Figure 4.9: RMSF heat maps of local residue fluctuations during the 150 ns MD simulation period in 13 hit compounds across 16 PZase mutations. A higher value (represented by a darker colour) signifies more fluctuations while a lower value (lighter colour) represents less flexibility.
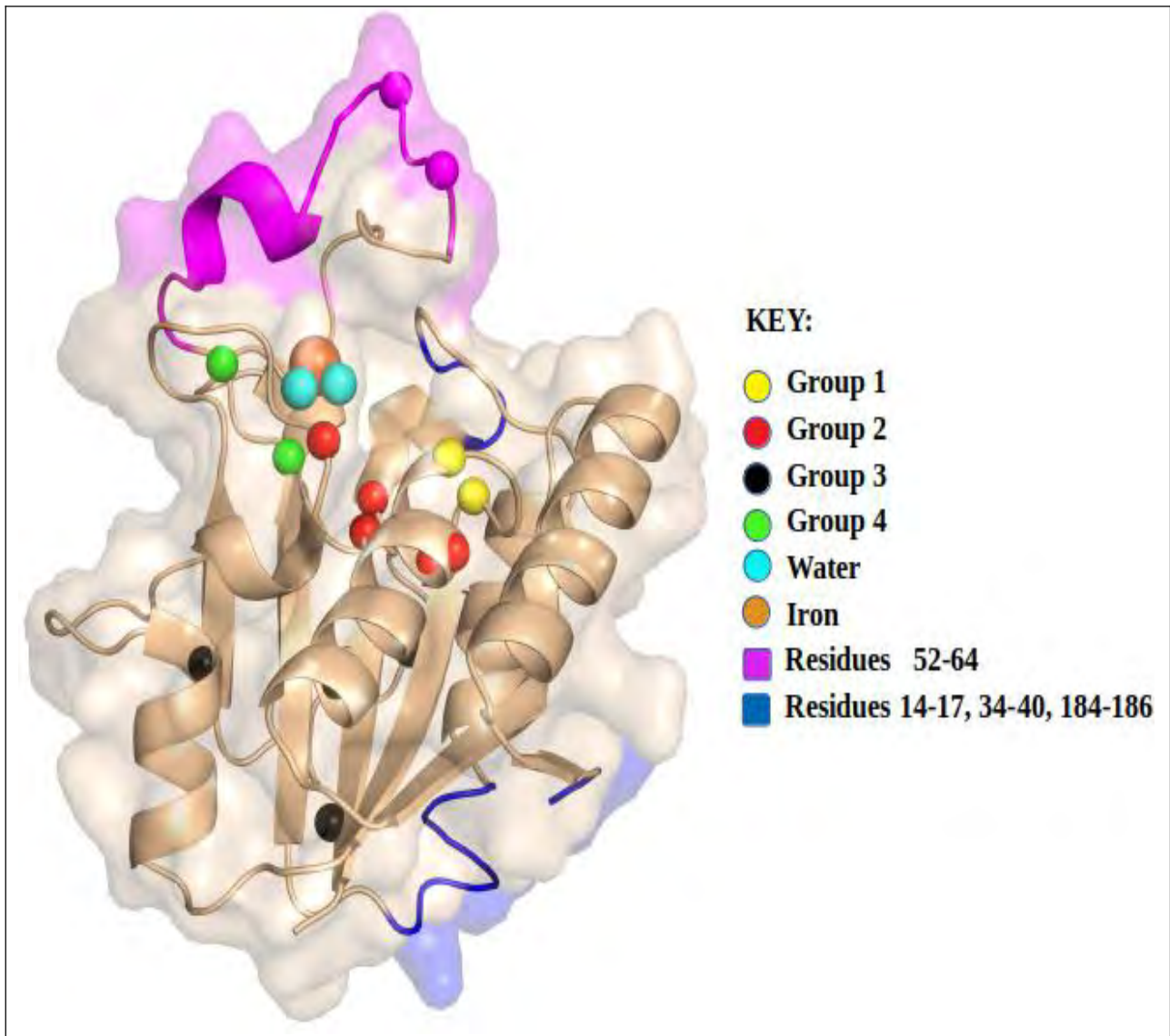
Figure 4.10: Mapped fluctuating regions on mutant PZase protein structure. The coloured spheres represent different mutant groups, iron and water while magenta shows the most fluctuating regions in mutants D49G/A while blue show the general fluctuating residues. The protein cartoon and surface structures is shown in wheat.

4.5 CONCLUSION

Molecular dynamic simulations on the mutants complexes were successfully performed. Analysis using RMSD, RMSF, radius of gyration and hydrogen bonds identified majority of the ligand systems as stable (Figure 4.3 and 4.4), closely packed (Figure 4.5 and Figure 4.6) and less fluctuating (Figure 4.9). All the selected ligands performed better than the PZA ligand across all the analysis measures. The aim of this chapter was to screen the compounds that were stable in WT PZase by identification of those that would also be stable in mutant PZase. Comparison of the ligand WT systems to each respective mutant showed little differences between the WT and mutant systems. The little differences between the dynamic behaviour of the WT and mutant-ligand systems is a positive outcome as it implies that the ligand was stably bound in both WT and mutant PZase. This ultimately suggests that the drug is likely to be functional in both WT and mutant PZase, reducing the rate of drug resistance accounted by mutated *pncA* gene. The best ligands that consistently displayed high stability, low RMSD and RMSF values, high structure compactness and more hydrogen bond interactions were compound systems DB00255, DB04573 and DB13952. Among all the mutants, the active site mutation on D49 to A or G displayed the most fluctuations and unstable results. Mapping of these flexible residues showed that the regions were mainly loop areas with exception of mutant D49 that had fluctuations on alpha helix structure on residues 52 to 64. This was mainly because the mutations were in the metal binding site. Overall, fromthese stable mutant-compound systems, the compounds are identified as potential drugs targeting the WT and mutant M. *tb* PZase strains requiring further analytical and activity studies for TB therapy.

CHAPTER FIVE

5 CONCLUSION AND FUTURE WORK

5.1 CONCLUDING REMARKS

Tuberculosis is an infectious disease caused by *M. tb* and has the leading cause of death especially in low income countries. Effective treatment of this disease has been achieved by the use of first line drugs rifampicin, isoniziad, ethambutol and PZA. The prodrug, PZA, has been reported the most effective drug, reducing therapy time from 9 months to 6 months and thus is incorporated in all TB treatments. However, studies have reported resistance to PZA due to alteration in the *pncA* gene that codes for its target protein, PZase. Various laboratory studies and *in silico* approaches have been done to understand the causes of PZA resistance and mutations and the effects of these mutations on the protein stability and functionality. The emergency of TB drug resistance has lagged successful eradication and treatment of TB, therefore there is need for the development of novel drugs functional in both WT and mutant M. *tb* strains

The aim of this study was to identify compounds that bind to the active site of PZA target protein, PZase, and analyse the behaviour and dynamics of the complexes in the presence of mutations. This was achieved by screening 2089 DrugBank compounds against WT PZase in molecular docking experiments and study the dynamics of the WT and mutant complexes in MD simulations as summarized in Figure 5.1. These are useful techniques in drug repurposing as they identify potential hit compounds within a short period of time and at a low cost.
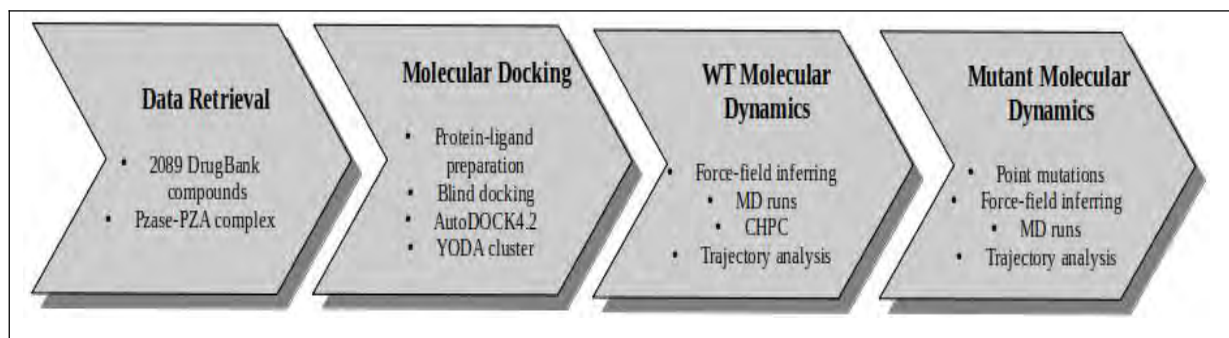
Figure 5.1: Overall flow of compound screening from data retrieval to molecular docking and molecular dynamic simulations of the WT and mutant PZase.

Screening of the docked compounds gave 93 ligands that selectively bind to the active site, where further screening in 20 ns MD simulations gave 47 stable protein-ligand complexes (Figure 2.5). All 47 compounds had Euclidean distance ≤ 8 Å to the active site, binding energy ≤ -7.0 kcal/mol, molecular weight ≤ 500 g/mol and formed hydrogen bond ≥ 1 (Figure 2.6).

Stability of these ligands in WT PZase were analysed using MD simulations where a total of 13 ligands were selected as the most stable compounds (DB00255, DB00655, DB00672, DB00782, DB00977, DB01196, DB04573, DB06414, DB08981, DB11181, DB11760, DB13867, DB13952), based on RMSD (Figure 3.3), radius of gyration (Section 3.4.4), RMSF (Figure 3.4) and number of hydrogen bonds throughout the simulation (Section 3.4.5). The effects of mutations on the potential hit compounds were analysed by introducing mutants (A3P, A134V, A146V, D8G, D49A, D49G, D63G, H51P, H137R, L85R, L116R, Q10P, R140S, T61P, V139M, Y103S) and calculating the deviation of the mutant systems from the WT system. The selected 13 compounds generally had stable conformations across majority of the mutants with exception to mutant D49A/G that consistently displayed high deviation values and more flexibility compared to other systems. The primary goal of the study, to identify drugs that can be repurposed for TB therapy or aid in the development of novel drugs was successfully achieved.

5.2 FUTURE WORK

Further analysis may be to perform dynamic residue network (DRN) calculations to calculate the betweenness centrality and average shortest path of residues. This will provide more information on the overall interactions of residues throughout the simulation period. From the whole protein molecular docking in Chapter 2, potential future work will be to identify allosteric sites and study the effect of compounds that might bind to these sites in relation to the stability and functionality of PZase. The mechanism of the identified potential drugs may be studied to identify those that may mimick PZA as prodrugs or function just as inhibitors of the PZase enzyme.

REFERENCES

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX, 1*, pp.19-25.

Ahmad, Z., Tyagi, S., Minkowski, A., Peloquin, C.A., Grosset, J.H. and Nuermberger, E.L., 2013.Contribution of moxifloxacin or levofloxacin in second-line regimens with or without continuationof pyrazinamide in murine tuberculosis. American journal of respiratory and critical care medicine,188(1), pp.97-102.

Amamuddy, O.S., Musyoka, T.M., Boateng, R.A., Zabo, S. and Bishop, Ö.T., 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.

AutoDock: Theory and Practice. Expert Opin. Drug Discov. 2010, 5, 597–607

Baig, M.H., Sudhakar, D.R., Kalaiarasan, P., Subbarao, N., Wadhawa, G., Lohani, M., Khan, M.K.A. and Khan, A.U., 2014. Insight into the effect of inhibitor resistant S130G mutant on physico-chemical properties of SHV type beta-lactamase: a molecular dynamics study. Plos one, 9(12).

Banerjee, P., Erehman, J., Gohlke, B.O., Wilhelm, T., Preissner, R. and Dunkel, M., 2015. Super Natural II—a database of natural products. *Nucleic acids research, 43*(D1), pp.D935-D939.

Barozi, V., 2020. UNDERSTANDING OF THE UNDERLYING RESISTANCE MECHANISM OF THE KAT-G PROTEIN AGAINST ISONIAZID IN Mycobacterium tuberculosis USING BIOINFORMATICS APPROACHES, Research Unit in Bioinformatics (RUBi), RHODES UNIVERSITY, SOUTH AFRICA

Benkert, P.; Künzli, M.; Schwede, T. QMEAN server for protein model quality estimation. Nucleic Acids Res.2009, 37, W510–W514.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., 2000. The protein data bank. *Nucleic acids research, 28*(1), pp.235-242.

Bigger, J., 1944. Treatment of Staphyloeoeeal Infections with Penicillin by Intermittent Sterilisation. Lancet, pp.497-500.

Bishop K, S., Blumberg, L., Trollip A, P., Smith A, N., Roux, L., York D, F. and Kiepiela, P., 2001. Characterisation of the pncA gene in Mycobacterium tuberculosis isolates from Gauteng, South Africa. The International Journal of Tuberculosis and Lung Disease, 5(10), pp.952-957.

Bitencourt-Ferreira, G., Pintro, V.O. and de Azevedo, W.F., 2019. Docking with AutoDock4. In *Docking Screens for Drug Discovery* (pp. 125-148). Humana, New York, NY.

Boshoff, H.I., Mizrahi, V. and Barry, C.E., 2002. Effects of pyrazinamide on fatty acid synthesis by whole mycobacterial cells and purified fatty acid synthase I. Journal of bacteriology, 184(8), pp.2167-2172.

Bozzani, F.M., Mudzengi, D., Sumner, T., Gomez, G.B., Hippner, P., Cardenas, V., Charalambous, S., White, R. and Vassall, A., 2018. Empirical estimation of resource constraints for use in model based economic evaluation: an example of TB services in South Africa. Cost effectiveness and resource allocation, 16(1), p.27.

Broger, T., Sossen, B., du Toit, E., Kerkhoff, A.D., Schutz, C., Reipold, E.I., Ward, A., Barr, D.A., Macé, A., Trollip, A. and Burton, R., 2019. Novel lipoarabinomannan point-of-care tuberculosis test for people with HIV: a diagnostic accuracy study. The Lancet Infectious Diseases, 19(8), pp.852-861

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S.A. and Karplus, M., 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, *4*(2), pp.187-217.

Chang, K.C., Yew, W.W. and Zhang, Y., 2011. Pyrazinamide susceptibility testing in Mycobacterium tuberculosis: a systematic review with meta-analyses. *Antimicrobial agents and chemotherapy*, *55*(10), pp.4499-4505.

Chaudhary, K.K. and Prasad, C.S., 2014. Virtual Screening of compounds to 1-deoxy-Dxylulose 5-phosphate reductoisomerase (DXR) from Plasmodium falciparum. *Bioinformation*, *10*(6), p.358.Cogan, N.G., 2006. Effects of persister formation on bacterial response to dosing. Journal of theoretical biology, 238(3), pp.694-703.

Connolly, L.E., Edelstein, P.H. and Ramakrishnan, L., 2007. Why is long-term therapy required to cure tuberculosis?. PLoS Med, 4(3), p.e120.

Cosconati, S.; Forli, S.; Perryman, A.L.; Harris, R.; Goodsell, D.S.; Olson, A.J., 2007. Virtual Screening with dimensional structures of proteins. Nucleic Acids Res. 35 (W407–W410)..

CRyPTIC Consortium and the 100,000 Genomes Project, 2018. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. New England Journal of Medicine, 379(15), pp.1403-1415.

DeLano, W.L., 2002. Pymol: An open-source molecular graphics tool. CCP4 Newsletter on protein crystallography, 40(1), pp.82-92.

Delogu, G., Sali, M. and Fadda, G., 2013. The biology of mycobacterium tuberculosis infection. Mediterranean journal of hematology and infectious diseases, 5(1).

Dickson, C.J., Madej, B.D., Skjevik, Å.A., Betz, R.M., Teigen, K., Gould, I.R. and Walker, R.C., 2014. Lipid14: the amber lipid force field. *Journal of chemical theory and computation*, *10*(2), pp.865-879.

Dillon, N.A., Peterson, N.D., Feaga, H.A., Keiler, K.C. and Baughn, A.D., 2017. Anti-tubercular Activity of Pyrazinamide is Independent of trans-Translation and RpsA. Scientific reports, 7(1), pp.1-8.

Dillon, N.A., Peterson, N.D., Rosen, B.C. and Baughn, A.D., 2014. Pantothenate and pantetheine antagonize the antitubercular activity of pyrazinamide. Antimicrobial agents and chemotherapy, 58(12), pp.7258-7263.

Doddareddy, M.R., Thorat, D.A., Seo, S.H., Hong, T.J., Cho, Y.S., Hahn, J.S. and Pae, A.N., 2011. Structure based design of heat shock protein 90 inhibitors acting as anticancer agents. *Bioorganic & medicinal chemistry*, *19*(5), pp.1714-1720.

Du, X., Wang, W., Kim, R., Yakota, H., Nguyen, H. and Kim, S.H., 2001. Crystal structure and mechanism of catalysis of a pyrazinamidase from Pyrococcus horikoshii. Biochemistry, 40(47), pp.14166-14172.

Eisenberg, D.; Lüthy, R.; Bowie, J.U., 1997. VERIFY3D: Assessment of protein models with three dimensional profiles. Methods Enzymol, 277, pp.396–406

El-Hachem, N., Haibe-Kains, B., Khalil, A., Kobeissy, F.H. and Nemer, G., 2017. AutoDock and AutoDockTools for protein-ligand docking: Beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) as a case study. In Neuroproteomics (pp. 391-403). Humana Press, New York, NY.

Elder, D. and Tindall, S., 2020. The many advantages of repurposing existing drugs. *European Pharmaceutical Review*, *25*(3), pp.34-37.

Flandrois, J.P., Lina, G. and Dumitrescu, O., 2014. MUBII-TB-DB: a database of mutations associated with antibiotic resistance in Mycobacterium tuberculosis. *BMC bioinformatics, 15*(1), p.107.

Forli, W., Halliday, S., Belew, R. and Olson, A.J., 2012. AutoDock Version 4.2. *Journal of Medicinal Chemistry, 55*(2), pp.623-638.

Fox, W., Ellard, G.A. and Mitchison, D.A., 1999. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946–1986, with relevant subsequent publications. The International Journal of Tuberculosis and Lung Disease, 3(10), pp.S231-S279.

Fyfe, P.K., Rao, V.A., Zemla, A., Cameron, S. and Hunter, W.N., 2009. Specificity and mechanism of Acinetobacter baumanii nicotinamidase: implications for activation of the front-line tuberculosis drug pyrazinamide. Angewandte Chemie International Edition, 48(48), pp.9176-9179.

Gagneux, S., 2018. Ecology and evolution of Mycobacterium tuberculosis. Nature Reviews Microbiology, 16(4), p.202.2730.

Gopal, P., Sarathy, J.P., Yee, M., Ragunathan, P., Shin, J., Bhushan, S., Zhu, J., Akopian, T., Kandror, O., Lim, T.K. and Gengenbacher, M., 2020. Pyrazinamide triggers degradation of its target aspartate decarboxylase. Nature communications, 11(1), pp.1-10.

Gopal, P., Yee, M., Sarathy, J., Low, J.L., Sarathy, J.P., Kaya, F., Dartois, V., Gengenbacher, M. and Dick, T., 2016. Pyrazinamide resistance is caused by two distinct mechanisms: prevention of coenzyme A depletion and loss of virulence factor synthesis. ACS infectious diseases, 2(9), pp.616-626.

Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S. and Onufriev, A., 2005. H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic acids research, 33*(suppl_2), pp.W368-W371.Guedes, I.A., de Magalhães, C.S. and Dardenne, L.E., 2014. Receptor–ligand molecular docking. *Biophysical reviews, 6*(1), pp.75-87.

.

Hatherley, R., Brown, D.K., Musyoka, T.M., Penkler, D.L., Faya, N., Lobb, K.A. and Bishop, Ö.T., 2015. SANCDB: a South African natural compound database. *Journal of cheminformatics, 7*(1), pp.1-9.

Hollingsworth, S.A. and Dror, R.O., 2018. Molecular dynamics simulation for all. *Neuron, 99*(6), pp.1129-1143.

Hospital, A., Goñi, J.R., Orozco, M. and Gelpí, J.L., 2015. Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry: AABC, 8,* p.37.

Huey, R., Morris, G.M., Olson, A.J. and Goodsell, D.S., 2007. A semiempirical free energy force field with charge-based desolvation. *Journal of computational chemistry, 28*(6), pp.1145-1152.

Humphrey, W., Dalke, A. and Schulten, K., 1996. J. lVlolec. *Graphics, 14,* pp.33-38.Irwin, J.J., Sterling, T., Mysinger, M.M., Bolstad, E.S. and Coleman, R.G., 2012. ZINC: a free tool to discover chemistry for biology. *Journal of chemical information and modeling, 52*(7), pp.1757-1768.Jimenez-Corona, M.E., Garcia-Garcia, L., DeRiemer, K., Ferreyra-Reyes, L., Bobadilla-del-Valle, M., Cano-Arellano, B., Canizales-Quintero, S., Martinez-Gamboa, A., Small, P.M., Sifuentes Osornio, J. and Ponce-de-León, A., 2006. Gender differentials of pulmonary tuberculosis transmission and reactivation in an endemic area. *Thorax, 61*(4), pp.348-353.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L., 2008. NCBI BLAST: a better web interface. *Nucleic acids research, 36*(suppl_2), pp.W5-W9..

Jorgensen, W.L., Maxwell, D.S. and Tirado-Rives, J., 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society, 118*(45), pp.11225-11236.

Juma, S.P., Maro, A., Pholwat, S., Mpagama, S.G., Gratz, J., Liyoyo, A., Houpt, E.R., Kibiki, G.S., Mmbaga, B.T. and Heysell, S.K., 2019. Underestimated pyrazinamide resistance may compromise outcomes of pyrazinamide containing regimens for treatment of drug susceptible and multi-drug-resistant tuberculosis in Tanzania. BMC infectious diseases, 19(1), pp.1-6.

Junaid, M., Khan, M.T., Malik, S.I. and Wei, D.Q., 2018. Insights into the mechanisms of the pyrazinamide resistance of three pyrazinamidase mutants N11K, P69T, and D126N. Journal of chemical information and modeling, 59(1), pp.498-508.

Junaid, M., Li, C.D., Li, J., Khan, A., Ali, S.S., Jamal, S.B., Saud, S., Ali, A. and Wei, D.Q., 2020. Structural insights of catalytic mechanism in mutant pyrazinamidase of Mycobacterium tuberculosis. *Journal of Biomolecular Structure and Dynamics*, pp.1-14.

Karplus, M. and Kuriyan, J., 2005. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences, 102*(19), pp.6679-6685.

Khadem-Maaref, M., Mehrnejad, F. and Phirouznia, A., 2017. Effects of metal-ion replacement on pyrazinamidase activity: A quantum mechanical study. Journal of Molecular Graphics and Modelling, 73, pp.24-29.

Khan, M.T. and Malik, S.I., 2020. Structural dynamics behind variants in pyrazinamidase and pyrazinamide resistance. *Journal of Biomolecular Structure and Dynamics, 38*(10), pp.3003-3017.

Khan, M.T., Chinnasamy, S., Cui, Z., Irfan, M. and Wei, D.Q., 2020. Mechanistic analysis of A46V, H57Y, and D129N in pyrazinamidase associated with pyrazinamide resistance. *Saudi Journal of Biological Sciences*, *27*(11), pp.3150-3156.

Khan, M.T., Junaid, M., Mao, X., Wang, Y., Hussain, A., Malik, S.I. and Wei, D.Q., 2019. Pyrazinamide resistance and mutations L19R, R140H, and E144K in Pyrazinamidase of Mycobacterium tuberculosis. *Journal of cellular biochemistry, 120*(5), pp.7154-7166.

Khan, H.J., Rohondia, S.O., Ahmed, Z.S.O., Zalavadiya, N. and Dou, Q.P., 2020. Increasing opportunities of drug repurposing for treating breast cancer by the integration of molecular, histological, and systemic approaches. In *Drug Repurposing in Cancer Therapy* (pp. 121-172). Academic Press.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A. and Wang, J., 2016. BS The PubChem Project. *Nucleic Acids Research*, *44*(D1), pp.D1202-13.

Kim, H.J., Kwak, H.K., Lee, J., Yun, Y.J., Lee, J.S., Lee, M.S., Min, S.Y., Park, S.K., Kang, H.S., Maeng, Y.H. and Kim, S.Y., 2012. Patterns of pncA mutations in drug-resistant Mycobacterium tuberculosis isolated from patients in South Korea. *The International journal of tuberculosis and lung disease, 16*(1), pp.98-103.

Kinghorn A.D, gibbons H F S, Asakawa J K Y, Liu J .2019. Progress in the chemistry of organic natural products. Cheminformatics in Natural Product research. Spinger Nature Switzerland AG.

Kitchen, D.B., Decornez, H., Furr, J.R. and Bajorath, J., 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery*, *3*(11), pp.935-949.

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., 1982. A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*, *161*(2), pp.269-288

Lamont, E.A., Dillon, N.A. and Baughn, A.D., 2020. The Bewildering Antitubercular Action of Pyrazinamide. Microbiology and Molecular Biology Reviews, 84(2).

Landrum, G., 2006. RDKit: Open-source cheminformatics.

Laskowski, R.A. and Swindells, M.B., 2011. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery.
Lawn, S.D., Kerkhoff, A.D., Burton, R., Schutz, C., Boulle, A., Vogt, M., Gupta-Wright, A., Nicol, M.P. and Meintjes, G., 2017. Diagnostic accuracy, incremental yield and prognostic value of Determine TB-LAM for routine diagnostic testing for tuberculosis in HIV-infected patients requiring acute hospital admission in South Africa: a prospective cohort. BMC medicine, 15(1),p.67.

Lemaitre, N., Sougakoff, W., Truffot-Pernot, C. and Jarlier, V., 1999. Characterization of new mutations in pyrazinamide-resistant strains of mycobacterium tuberculosisand identification of 28conserved regions important for the catalytic activity of the pyrazinamidase PncA. Antimicrobial agents and chemotherapy, 43(7), pp.1761-1763.

Lewis, K., 2012. Persister cells: molecular mechanisms related to antibiotic tolerance. In *Antibiotic resistance* (pp. 121-133). Springer, Berlin, Heidelberg.

Lodish, H. and Zipursky, S.L., 2001. Molecular cell biology. *Biochem Mol Biol Educ*, *29*, pp.126 133.

Lokesh, R. and Kannabiran, K., 2016. A HANDBOOK ON PROTEINğLIGAND DOCKING TOOL: AUTODOCK4.

Machaba, K.E., Mhlongo, N.N. and Soliman, M.E., 2018. Induced mutation proves a potential target for TB therapy: a molecular dynamics study on LprG. *Cell biochemistry and biophysics*, *76*(3), pp.345-356

Mahoney, M.W. and Jorgensen, W.L., 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics*, *112*(20), pp.8910-8922.

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C., 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.

McCammon, J.A., Gelin, B.R. and Karplus, M., 1977. Dynamics of folded proteins. *Nature*, *267*(5612), pp.585-590.

McQuaid, C.F., Horton, K.C., Dean, A.S., Knight, G.M. and White, R.G., 2020. The risk of multidrug-or rifampicin-resistance in males versus females with tuberculosis. *European Respiratory Journal*, *56*(3).

Meyer, E.A., Castellano, R.K. and Diederich, F., 2003. Interactions with aromatic rings in chemical and biological recognition. *Angewandte Chemie International Edition*, *42*(11), pp.1210-1250.

Miotto, P., Cabibbe, A.M., Feuerriegel, S., Casali, N., Drobniewski, F., Rodionova, Y., Bakonyte, D., Stakenas, P., Pimkina, E., Augustynowicz-Kopeć, E. and Degano, M., 2014. Mycobacterium tuberculosis pyrazinamide resistance determinants: a multicenter study. *MBio*, *5*(5).
Morlock, G.P., Tyrrell, F.C., Baynham, D., Escuyer, V.E., Green, N., Kim, Y., Longley-Olson, P.A., Parrish, N., Pennington, C., Tan, D. and Austin, B., 2017. Using reduced inoculum densities of Mycobacterium tuberculosis in MGIT pyrazinamide susceptibility testing to prevent false-resistant results and improve accuracy: a multicenter evaluation. Tuberculosis Research and Treatment, 2017.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, *30*(16), pp.2785-2791.

Muller A.,2016, A description of the TB germ: Mycobacterium tuberculosis, Retrived from https://www.tbonline.info/posts/2016/3/31/description-tb-germ-mycobacterium-tuberculosis-1/#:~:text=It%20is%20a%20small%20bacillus,is%20extremely%20slow%20for%20bacteria.

Nguyen, N.T., Nguyen, T.H., Pham, T.N.H., Huy, N.T., Bay, M.V., Pham, M.Q., Nam, P.C., Vu, V.V. and Ngo, S.T., 2019. Autodock vina adopts more accurate binding poses but autodock4 forms better binding affinity. *Journal of Chemical Information and Modeling, 60*(1), pp.204-211.

Ntie-Kang, F., Telukunta, K.K., Döring, K., Simoben, C.V., A. Moumbock, A.F., Malange, Y.I., Njume, L.E., Yong, J.N., Sippl, W. and Günther, S., 2017. NANPDB: a resource for natural products from Northern African sources. *Journal of natural products, 80*(7), pp.2067-2076.Pagadala, N.S., Syed, K. and Tuszynski, J., 2017. Software for molecular docking: a review. *Biophysical reviews, 9*(2), pp.91-102.

Palmer, A.C. and Kishony, R., 2014. Opposing effects of target over expression reveal drug mechanisms. Nature communications, 5(1), pp.1-8.

Parrinello, M. and Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics, 52*(12), pp.7182-7190.Patterns of pncA mutations in drug-resistant Mycobacterium tuberculosis isolated from patients in South Korea. The International journal of tuberculosis and lung disease, 16(1), pp.98-103.

Personne, Y. and Parish, T., 2014. Mycobacterium tuberculosis possesses an unusual tmRNA rescue system. Tuberculosis, 94(1), pp.34-42.

Peterson, N.D., Rosen, B.C., Dillon, N.A. and Baughn, A.D., 2015. Uncoupling environmental pH and intrabacterial acidification from pyrazinamide susceptibility in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy, 59(12), pp.7320-7326.

Petrella, S., Gelus-Ziental, N., Maudry, A., Laurans, C., Boudjelloul, R. and Sougakoff, W., 2011. Crystal structure of the pyrazinamidase of Mycobacterium tuberculosis: insights into natural and acquired resistance to pyrazinamide. PloS one, 6(1), p.e15785.Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. and Schulten, K., 2005. Scalable molecular dynamics with NAMD. *Journal of computational chemistry, 26*(16), pp.1781-1802.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D. and Hess, B., 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics, 29*(7), pp.845-854.

Pushpakom, S., Iorio, F., Eyers, P.A., Escott, K.J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C. and Norris, A., 2019. Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, *18*(1), pp.41-58.

Rasool, N., Husssain, W. and Khan, Y.D., 2019. Revelation of enzyme activity of mutant pyrazinamidases from Mycobacterium tuberculosis upon binding with various metals using quantum mechanical approach. Computational biology and chemistry, 83, p.107108.

Rudrapal, M., Khairnar, S.J. and Jadhav, A.G., 2020. Drug Repurposing (DR): An Emerging Approach in Drug Discovery. In *Drug Repurposing-Hypothesis, Molecular Aspects and Therapeutic Applications*. IntechOpen.

Rueda, M., Ferrer-Costa, C., Meyer, T., Pérez, A., Camps, J., Gelpí, J.L. and Orozco, M., 2007. A consensus view of protein dynamics. *Proceedings of the National Academy of Sciences*, *104*(3), pp.796-801.

Salmaso, V. and Moro, S., 2018. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Frontiers in pharmacology*, *9*, p.923.

Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B.K., Church, G.M. and Murray, M.B., 2009. Tuberculosis drug resistance mutation database. *PLoS Med*, *6*(2), p.e1000002.Santos-Martins, D., Solis-Vasquez, L., Koch, A. and Forli, S., 2019. Accelerating autodock4 with gpus and gradient-based local search.

Shaw, D.E., Dror, R.O., Salmon, J.K., Grossman, J.P., Mackenzie, K.M., Bank, J.A., Young, C., Deneroff, M.M., Batson, B., Bowers, K.J. and Chow, E., 2009, November. Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the conference on high performance computing networking, storage and analysis* (pp. 1-11).

Sheen, P., Ferrer, P., Gilman, R.H., Christiansen, G., Moreno-Román, P., Gutiérrez, A.H., Sotelo, J.,Evangelista, W., Fuentes, P., Rueda, D. and Flores, M., 2012. Role of metal ions on the activity of Mycobacterium tuberculosis pyrazinamidase. The American journal of tropical medicine and hygiene, 87(1), pp.153-161.

Sheen, P., Monsalve, A., Campos, J., Huerta, R., Antiparra, R., Arteaga, H., Duran, P., Bueno, C., Kirwan, D.E., Gilman, R.H. and Zimic, M., 2020. Metallochaperones are needed for Mycobacterium tuberculosis and Escherichia coli nicotinamidase-pyrazinamidase activity. Journal of bacteriology,202(2).

Shi, W., Chen, J., Feng, J., Cui, P., Zhang, S., Weng, X., Zhang, W. and Zhang, Y., 2014. Aspartate decarboxylase (PanD) as a new target of pyrazinamide in Mycobacterium tuberculosis. Emerging microbes & infections, 3(1), pp.1-8.

Shi, W., Zhang, X., Jiang, X., Yuan, H., Lee, J.S., Barry, C.E., Wang, H., Zhang, W. and Zhang, Y., 2011. Pyrazinamide inhibits trans-translation in Mycobacterium tuberculosis. Science, 333(6049), pp.1630-1632.2958.

Simons, S.O., Mulder, A., van Ingen, J., Boeree, M.J. and van Soolingen, D., 2013. Role of rpsA gene sequencing in diagnosis of pyrazinamide resistance. Journal of clinical microbiology, 51(1), pp.382-382.

Singh, P., Kant, S., Gaur, P., Tripathi, A. and Pandey, S., 2018. Extra pulmonary Tuberculosis: An overview and review of literature. Int. J. Life. Sci. Scienti. Res, 4(1), pp.1539-1541.

Sliwoski, G., Kothiwale, S., Meiler, J. and Lowe, E.W., 2014. Computational methods in drug discovery. *Pharmacological reviews*, *66*(1), pp.334-395.

Smith, C.V., Sharma, V. and Sacchettini, J.C., 2004. TB drug discovery: addressing issues of persistence and resistance. Tuberculosis, 84(1-2), pp.45-55.

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M.A. and Steinbeck, C., 2021. COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics*, *13*(1), pp.1-13.

Sreevatsan, S., Pan, X., Zhang, Y., Kreiswirth, B.N. and Musser, J.M., 1997. Mutations associated with pyrazinamide resistance in pncA of Mycobacterium tuberculosis complex organisms. *Antimicrobial agents and chemotherapy*, *41*(3), pp.636-640.

The UniProt Consortium UniProt: A hub for protein information. Nucleic Acids Res. 2015, 43, D204–D212.

Vadivelan, S., Deeksha, T.N., Arun, S., Machiraju, P.K., Gundla, R., Sinha, B.N. and Jagarlapudi, S.A., 2011. Virtual screening studies on HIV-1 reverse transcriptase inhibitors to design potent leads. *European journal of medicinal chemistry*, *46*(3), pp.851-859..

Van Santen, J.A., Jacob, G., Singh, A.L., Aniebok, V., Balunas, M.J., Bunsko, D., Neto, F.C., Castaño-Espriu, L., Chang, C., Clark, T.N. and Cleary Little, J.L., 2019. The natural products atlas: an open access knowledge base for microbial natural products discovery. *ACS central science*, *5*(11), pp.1824-1833.

Walker, R.C., Crowley, M.F. and Case, D.A., 2008. The implementation of a fast and accurate QM/MM potential method in Amber. *Journal of computational chemistry*, *29*(7), pp.1019-1031.

Wang, J.C., Chu, P.Y., Chen, C.M. and Lin, J.H., 2012. idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a   divide-and-conquer docking approach. *Nucleic acids research*, *40*(W1), pp.W393-W399.

Weiergräber, O.H., Schwarten, M., Strodel, B. and Willbold, D., 2017. Investigating structure and dynamics of Atg8 family proteins. In *Methods in enzymology* (Vol. 587, pp. 115-142). Academic Press.

Whitfield, M.G., Soeters, H.M., Warren, R.M., York, T., Sampson, S.L., Streicher, E.M., Van Helden, P.D. and Van Rie, A., 2015. A global perspective on pyrazinamide resistance: systematic review and meta-analysis. *PloS one*, *10*(7), p.e0133869.

Wiederstein, M. and Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research, 35(suppl_2), pp.W407-W410.

Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z. and Assempour, N., 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research*, *46*(D1), pp.D1074-D1082.World Health Organization, 2020. Global tuberculosis report 2020.

World Health Organization, 2019. GLOBAL TUBERCULOSIS REPORT 2019.

Wu, X., Lu, W., Shao, Y., Song, H., Li, G., Li, Y., Zhu, L. and Chen, C., 2019. pncA gene mutations in reporting pyrazinamide resistance among the MDR-TB suspects. *Infection, Genetics and Evolution*, *72*, pp.147-150.

Xing L, McDonald JJ, Kolodziej SA, Kurumbail RG, Williams JM, Warren CJ, O'Neal JM, Skepner JE & Roberds SL (2011) Discovery of potent inhibitors of soluble epoxide hydrolase by combinatorial library design and structure-based virtual screening. J. Med. Chem. 54, 1211–1222.

Yoon, J.H., Nam, J.S., Kim, K.J. and Ro, Y.T., 2014. Characterization of pncA mutations in pyrazinamide resistant Mycobacterium tuberculosis isolates from Korea and analysis of the correlation between the mutations and pyrazinamidase activity. World Journal of Microbiology and Biotechnology, 30(11), pp.2821-2828

Zhang, Y. and Mitchison, D., 2003. The curious characteristics of pyrazinamide: a review. The international journal of tuberculosis and lung disease, 7(1), pp.6-21.

Zhang, Y. and Yew, W.W., 2015. Mechanisms of drug resistance in Mycobacterium tuberculosis: update 2015. The International Journal of Tuberculosis and Lung Disease, 19(11), pp.1276-1289.

Zhang, Y., Shi, W., Zhang, W. and Mitchison, D., 2014. Mechanisms of pyrazinamide action and resistance. Molecular Genetics of Mycobacteria, pp.479-491.

Zhang, Y., Yew, W.W. and Barer, M.R., 2012. Targeting persisters for tuberculosis control. Antimicrobial agents and chemotherapy, 56(5), pp.2223-2230.

Zhao, Y., Zeng, C. and Massiah, M.A., 2015. Molecular dynamics simulation reveals insights into the mechanism of unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain. PloS one, 10(4)

Zhou, S., Yang, S. and Huang, G., 2017. Design, synthesis and biological activity of pyrazinamide derivatives for anti-Mycobacterium tuberculosis. Journal of enzyme inhibition and medicinal chemistry, 32(1), pp.1183-1186.

Zimhony, O., Cox, J.S., Welch, J.T., Vilchèze, C. and Jacobs, W.R., 2000. Pyrazinamide inhibits the eukaryotic-like fatty acid synthetase I (FASI) of Mycobacterium tuberculosis. Nature medicine, 6(9), pp.1043-1047.

Zumla, A., Raviglione, M., Hafner, R. and von Reyn, C.F., 2013. Current concepts. N Engl J Med, 368, pp.745-55.

APPENDICES

Appendix 1: AutoDOCK4 clustering on ligand DB11793_lc. The lowest energy cluster was not the highest energy, thus both poses were selected for assessing.

```
8958 _____
8959      |          |     |          |     |
8960 Clus | Lowest   | Run | Mean     | Num | Histogram
8961 -ter | Binding  |     | Binding  | in  |
8962 Rank | Energy   |     | Energy   | Clus|   5   10   15   20   25   30   35
8963 _____|_____|_____|_____|_____|___:____|____:____|____:____|____:___
8964   1  |   -8.58  | 75  |   -7.71  | 10  |##########
8965   2  |   -8.50  | 14  |   -8.16  |  7  |#######
8966   3  |   -8.08  | 97  |   -7.48  |  4  |####
8967   4  |   -7.79  | 91  |   -7.44  |  8  |########
8968   5  |   -7.79  | 89  |   -7.32  | 11  |###########
8969   6  |   -7.49  | 65  |   -7.36  |  5  |#####
8970   7  |   -7.47  | 73  |   -7.24  | 14  |##############
8971   8  |   -7.23  | 11  |   -6.87  |  3  |###
8972   9  |   -7.20  | 69  |   -7.07  |  2  |##
8973  10  |   -7.09  | 12  |   -7.05  |  2  |##
8974  11  |   -7.01  | 51  |   -6.97  |  2  |##
8975  12  |   -6.91  | 37  |   -6.91  |  1  |#
8976  13  |   -6.86  | 79  |   -6.70  |  3  |###
8977  14  |   -6.75  | 55  |   -6.75  |  1  |#
8978  15  |   -6.72  | 61  |   -6.72  |  1  |#
8979  16  |   -6.71  |  2  |   -6.71  |  2  |##
8980  17  |   -6.68  | 43  |   -6.14  |  2  |##
8981  18  |   -6.60  | 63  |   -6.60  |  1  |#
8982  19  |   -6.43  | 19  |   -6.22  |  3  |###
8983  20  |   -6.43  | 93  |   -6.43  |  1  |#
8984  21  |   -6.42  | 81  |   -6.42  |  1  |#
8985  22  |   -6.40  |100  |   -6.40  |  1  |#
8986  23  |   -6.38  | 48  |   -6.38  |  1  |#
8987  24  |   -6.32  | 44  |   -6.32  |  1  |#
8988  25  |   -6.26  | 49  |   -6.26  |  1  |#
8989  26  |   -6.26  | 39  |   -6.26  |  1  |#
8990  27  |   -6.19  | 17  |   -5.89  |  3  |###
8991  28  |   -6.11  | 30  |   -6.11  |  1  |#
8992  29  |   -6.02  | 45  |   -6.02  |  1  |#
8993  30  |   -5.97  |  3  |   -5.97  |  1  |#
8994  31  |   -5.93  | 26  |   -5.93  |  1  |#
8995  32  |   -5.92  | 64  |   -5.92  |  1  |#
8996  33  |   -5.82  | 53  |   -5.82  |  1  |#
8997  34  |   -5.62  | 29  |   -5.62  |  1  |#
8998  35  |   -5.44  | 85  |   -5.44  |  1  |#
8999 _____|_____|_____|_____|_____|_____
```

Appendix 2: Tabulated median values for all the mutants across all the ligand systems in comparison to the WT median. The WT medians are in blue. Median values that fell above 0.5 range to the WT syteme are highlighted in red.

Ligand system	WT-Median	Mutant	Median (nm)	Mutant	Median (nm)	Mutant	Median (nm)	Mutant	Median (nm)
PZA	0.164445	A3P	21.20423	D49A	0.6434385	H137R	0.4629939	R140S	0.3357925
		A134	1.311614	D49G	9.502109	L85R	0.1141979	T61P	18.33244

		V							
		A146V	0.3438485	D63G	0.1684 77	L116R	12.14709	V139M	0.3197724
		D8G	0.3524665	H51P	14.09235	Q10P	9.55299	Y103S	0.406138
DB00255	0.0884032	A3P	0.1231214	D49A	0.1391764	H137R	0.121591	R140S	0.1253088
		A134V	0.1447465	D49G	0.1057436	L85R	0.1330266	T61P	0.1162062
		A146V	0.0925909	D63G	0.1375335	L116R	0.0807351	V139M	0.1015875
		D8G	0.1951947	H51P	0.0933415	Q10P	0.1623088	Y103S	0.1046375
DB00655	0.0674602	A3P	0.1045995	D49A	0.942853	H137R	0.1247657	R140S	0.0652588
		A134V	0.1617667	D49G	0.2543127	L85R	0.1186377	T61P	0.389969
		A146V	0.1046888	D63G	0.0995842	L116R	0.0867077	V139M	0.095414
		D8G	0.1018965	H51P	0.1444355	Q10P	0.1472665	Y103S	0.0928249
DB00672	0.1873462	A3P	0.4895436	D49A	0.9189393	H137R	0.2462417	R140S	0.3176509
		A134V	0.9574616	D49G	0.2758844	L85R	0.2034416	T61P	0.198714
		A146V	0.1355111	D63G	0.1679863	L116R	0.3678121	V139M	0.1935503
		D8G	0.1895045	H51P	0.2396258	Q10P	0.2088178	Y103S	0.2198232
DB00783	0.071924	A3P	0.0752198	D49A	9.07433	H137R	0.1269564	R140S	0.0591639
		A134V	0.1430783	D49G	0.273259	L85R	0.0967755	T61P	0.0914366
		A146V	0.0825138	D63G	0.1040287	L116R	0.0764279	V139M	0.0680901
		D8G	0.1989668	H51P	0.4105033	Q10P	0.1247638	Y103S	0.0988449
DB00977	0.0780896	A3P	0.0662447	D49A	0.2542534	H137R	0.0706548	R140S	0.0647683
		A134V	0.1056988	D49G	0.3737917	L85R	0.0756264	T61P	0.047412
		A146	0.0509272	D63G	0.0823686	L116	0.0810443	V139	0.0868392

		V				R		M	
		D8G	0.1144794	H51P	0.3393904	Q10P	0.0761071	Y103S	0.0934659
DB011196	0.1771652	A3P	0.2722431	D49A	0.3717174	H137R	0.3581638	R140S	0.2758004
		A134V	0.1458451	D49G	0.3796869	L85R	0.2870666	T61P	0.1803126
		A146V	0.2966298	D63G	0.3134519	L116R	0.3054149	V139M	0.1985128
		D8G	0.2971003	H51P	0.2745737	Q10P	0.1649797	Y103S	0.1667653
DB045753	0.0829458	A3P	0.178061	D49A	0.0982053	H137R	0.3064429	R140S	0.1271223
		A134V	0.0753513	D49G	0.1150339	L85R	0.1645294	T61P	0.168364
		A146V	0.2490833	D63G	0.1138185	L116R	0.2391505	V139M	0.3174464
		D8G	0.2365497	H51P	0.0993769	Q10P	0.3244765	Y103S	0.284879
DB064144	0.1000941	A3P	0.1678196	D49A	0.6907824	H137R	0.3091336	R140S	0.1519863
		A134V	0.1100709	D49G	0.3924699	L85R	0.132307	T61P	0.3917673
		A146V	0.2645432	D63G	0.3291633	L116R	0.1550753	V139M	0.1212497
		D8G	0.1715504	H51P	0.4043808	Q10P	0.7045259	Y103S	0.1487681
DB089881	0.1131341	A3P	0.0999697	D49A	0.2666149	H137R	0.1108027	R140S	0.0870794
		A134V	0.1272178	D49G	0.2632805	L85R	0.0862399	T61P	0.111763
		A146V	0.1215842	D63G	0.1745371	L116R	0.1188489	V139M	0.1020893
		D8G	0.1506397	H51P	0.151953	Q10P	0.1712145	Y103S	0.0616769
DB111181	0.104633	A3P	0.1753278	D49A	0.4332178	H137R	0.1344773	R140S	0.2365587
		A134V	0.6288372	D49G	0.7045431	L85R	0.2098789	T61P	0.3460547
		A146V	0.3368713	D63G	0.6804121	L116R	0.2986856	V139M	0.1434226
		D8G	0.3952104	H51P	0.4592027	Q10P	0.2793811	Y103S	0.4879319

DB11760	0.1767728	A3P	0.0644074	D49A	0.1617042	H137R	0.0740717	R140S	0.1105821
		A134V	0.5080385	D49G	0.2160882	L85R	0.1533011	T61P	0.2012245
		A146V	0.1947073	D63G	0.1220111	L116R	0.0924843	V139M	0.3541162
		D8G	0.1586422	H51P	1.108168	Q10P	0.2645459	Y103S	0.2118541
DB13867	0.2260094	A3P	0.1588301	D49A	0.4197661	H137R	0.1216939	R140S	0.2311636
		A134V	0.1743439	D49G	0.3108039	L85R	0.1924449	T61P	0.2234677
		A146V	0.1217397	D63G	0.0926795	L116R	0.167852	V139M	0.1779705
		D8G	0.0779648	H51P	0.1161776	Q10P	0.1078896	Y103S	0.1521243
DB13952	0.1257215	A3P	0.1064138	D49A	0.1158549	H137R	0.0872469	R140S	0.0953429
		A134V	0.1053755	D49G	0.1713789	L85R	0.0895303	T61P	0.1000443
		A146V	0.1184868	D63G	0.1113061	L116R	0.1698281	V139M	0.1087708
		D8G	0.1946213	H51P	0.368363	Q10P	0.1076136	Y103S	0.0813479