Understanding the underlying resistance mechanism of *Mycobacterium tuberculosis* against Rifampicin by analyzing mutant DNA-directed RNA polymerase proteins via bioinformatics approaches

A mini-thesis submitted in partial fulfillment of the requirements of a degree

of

Master of Science

in

Bioinformatics and Computational Molecular Biology

(Coursework and Thesis)

by

Mokgerwa Zacharia Monama

Department of Biochemistry and Microbiology, Rhodes University,

Grahamstown, 6140, South Africa.

Abstract

Tuberculosis or TB is an airborne disease caused by the non-motile bacilli, *Mycobacterium tuberculosis* (*MTB*). There are two main forms of TB, namely, latent TB or LTB, asymptomatic and non-contagious version which according to the World Health Organization (WHO) is estimated to afflict over a third of the world's population; and active TB or ATB, a symptomatic and contagious version which continues to spread, affecting millions worldwide. With the already high reported prevalence of TB, the emergence of drug-resistant strains has prompted the development of novel approaches to enhance the efficacy of known drugs and a desperate search for novel compounds to combat *MTB* infections. It was for this very purpose that this study was conducted. A look into the resistance mechanism of Rifampicin (Rifampin or RIF), one of the more potent first-line drugs, might prove beneficial in predicting the consequence of an introduced mutation (which usually occur as single nucleotide polymorphisms or SNPs) and perhaps even overcome it using appropriate therapeutic interventions that improve RIF's efficacy. To accomplish this task, models of acceptable quality were generated for the WT and clinically relevant, RIF resistance conferring, SNPs occurring at codon positions D516, H526 and S531 (*E .coli* numbering system) using MODELLER. The models were accordingly ranked using GA341 and z-DOPE score, and subsequently validated with QMEAN, PROCHECK and VERIFY3D. MD simulations spanning 100 ns were run for RIF-bound (complex) and RIF-free (holo) DNA-directed RNA polymerase (DDRP) protein systems for the WT and SNP mutants using GROMACS. The MD frames were analyzed using RMSD, Rg and RMSF. For further analysis, MD-TASK was used to analyze the calculated dynamic residue networks (DRNs) from the generated MD frames, determining

both change in average shortest path (ΔL) and *betweenness centrality* (ΔBC). The RMSD analysis revealed that all of the SNP complex models displayed a level instability higher than that of the WT complex. A majority of the SNP complex models were also observed to have similar compactness to the WT holo when looking at the calculated Rg. The RMSF results also hinted towards possible physiological consequences of the mutations (generally referred to as a fitness cost) highlighted by the increased fluctuations of the zinc-binding domain and the *MTB* SI α helical coiled coil. For the first time, to the knowledge of the authors, DRN analysis was employed for the DDRP protein for both holo and complex systems, revealing insightful information about the residues that play a key role in the change in distance between residue pairs along with residues that play an essential role in protein communication within the calculated RIN. Overall, the data supported the conclusions drawn by a recent study that only concentrated on RIF-resistance in *rpoB* models which suggested that the binding pocket for the SNP models may result in the changed coordination of RIF which may be the main contributor to its impaired efficacy.

Keywords: AMBERSB-ILDN, average shortest path, *betweenness centrality*, DOPE, DRN, DDRP, fitness cost, GA341, homology modeling, *Mycobacterium tuberculosis*, MD, *MTB* SI α helical coiled coil, PROCHECK, QMEAN, Rg, RIF, RMSF, RMSD, SNP, TB, VERIFY3D, zinc-binding domain

Declaration

I, Mokgerwa Zacharia Monama, hereby declare that the thesis submitted to Rhodes University for the purpose of fulfillment of the one year MSc in Bioinformatics and Computational Molecular Biology requirements is my own work and has not been submitted to this or any other university for the same purpose.

_____ _____

Signature Date

Dedication

I dedicate this work to my loving parents,

my dear sister and supportive friends…

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Özlem Tastan Bishop, whose constant inquiries and guidance has led me to produce exceptional work.

Secondly, I would like to thank my co-supervisors, Dr. Thommas Mutemi Musyoka and Dr. Amamuddy Olivier Sheik, whose tireless efforts throughout the course of the study have been nothing less than excellent.

I would also like to thank my friends (both new and old), my family, and my colleagues at RUBi for their continuous support throughout this time.

Lastly, I would like give my deepest appreciation and thanks to the University of Rhodes for selecting me for this great opportunity. Salud!

Table of contents

List of figures

List of tables

List of abbreviations

ATB	Active TB

BC	*Betweenness centrality*
BCG	Bacille Calmette-Guérin
CA	Alpha carbon
CDC	Center for Disease Control
CPU	Central processing unit
DCC	Dynamic cross correlation
DDRP	DNA-directed RNA polymerase
DNA	Deoxyribonucleic acid
DOPE	Discrete optimized protein energy
DOTS	Directly Observed Treatment, Short Course
DRN	Dynamic residue network
EMB	Ethambutol
GB	Gigabyte
GPU	Graphical processing unit
HCM	High confidence mutation
IGRA	Interferon Gamma Release Assay
INH	Isoniazid
L	Average shortest path

LTB	Latent TB
MA	Mycolic acid
MDR-TB	Multiple-drug-resistant TB
Mg	Magnesium
MTB	*Mycobacterium tuberculosis*
*MTB*C	*MTB* complex
nt	Nucleotide
PDB	Protein Data Bank
PE	Proline-glutamine
PPE	Proline-proline-glutamine
PRS	Perturbation response scanning
PZA	Pyrazinamide
QMEAN	Qualitative Model Energy ANalysis
RIF	Rifampicin
RNAP	RNA polymerase
RPTIC	RNA polymerase transcription initiation complexes
RRDR	RIF-resistance-determining region
SNP	Single nucleotide polymorphism

TB	Tuberculosis
TBDReamDB	Tuberculosis drug resistance mutation database
TPT	TB preventative therapies
WT	Wild type
XDR-TB	Extremely-drug-resistant TB
Zn	Zinc

Chapter 1

1. Literature review

1.1. Tuberculosis

1.1.1. What is tuberculosis?

Tuberculosis or TB (formerly known as consumption) is an airborne disease generally caused by infection of *Mycobacterium tuberculosis* (*M. tuberculosis* or *MTB*), a discovery that was made public in the year 1882 by German microbiologist, Robert Koch, effectively paving the path for over a century's worth of TB research (Keshavjee and Farmer, 2012).

1.1.2. Pathogenesis of *MTB*

Infection mainly occurs through inhalation of proximal *MTB* bacterial cells in the form of an aerosol. The bacterial cells will then make their way through either the mouth and/or nasal cavities, to the upper respiratory tract, followed by the bronchi, and then arrive at the alveolar of the lungs (CDC[a], 2016). If an exposed individual's immune system, or more specifically the macrophages (part of the innate immune response) within the alveolar are only able to mount a less than effective attack on the pathogen (only succeeding in suppressing the growth of *MTB*) symptoms of TB itself might be suppressed. That is, through attempted phagocytosis and destruction by merging of phagosomes with synthesized organelles called lysosomes (present a low pH environment containing destructive compounds e.g. toxic peptides, reactive oxygen intermediates etc.) resulting in a structure referred to as a phagolysosome, may cause a cascade of events that result in a dormant form of the disease. This dormant or non-progressive form of TB is called latent TB (LTB), which happens to be asymptomatic and cannot be transmitted to others

(Nathan and Hibbs, 1991; Thillai et al., 2014). LTB is characterized by the formation of structures called granulomas (formed with the intention of isolating and destroying the pathogen) which are formed when macrophages release cytokines that ultimately recruit more cells of the immune system such as dendritic cells (present T-cells with antigens thus initiating a T-cell response) and additional macrophages to the site of infection. This action results in the encapsulation of the pathogenic bacteria (preventing them from spreading to other parts of the body) where they are only mostly destroyed, and where surviving bacteria may remain dormant for several years. LTB may switch to active TB or ATB at any point during the course of infection. If, however, the initial immune attack on the bacteria is unable to effectively destroy or deter the growth of *MTB*, then the disease is said to be in its ATB state which presents with a number of symptoms and is transmittable (Silva Miranda et al., 2012).

1.1.3. Parts of the body affected by TB

Although, the most commonly affected area of the body is the lungs (pulmonary TB), the disease can affect other areas of the body (extra-pulmonary TB) (WHO, 2014). These include the kidneys, bones, uterus, fallopian tubes i.e. almost any part of the body may be infected by *MTB* (Crowley, 2013). When an individual has active pulmonary TB, symptoms may include weight loss, breathing difficulties, fever and fatigue, severe coughing, night sweats, and coughing up blood. For those individuals with extra-pulmonary TB, symptoms depend on where the infection is localized, although general symptoms include fever, weight loss and night-sweats (WHO, 2014).

1.1.4. Prevalence of TB

According to the World Health Organization (WHO, 2018), TB remains one of the leading causes of mortality in the modern world, with millions of people from all walks of life in both low-and-high income regions around the world still falling sick with the disease. Although, the majority of reported mortalities due to *MTB* infection, tend to occur in middle-to-low income areas, where resources tend to be limited. In 2017 alone, over 1,3 million deaths were reported for individuals with a HIV-negative status, while approximately 300,000 deaths were associated with those individuals with a HIV-positive status. It is however estimated that over 10 million people developed TB in the year 2017, with the majority being men accounting for about 5,3 million of the infected, women accounting for about 3,2 million, and the minority of the infected being children who accounted for about 1 million of all infected individuals. The epidemic has only worsened with the emergence of drug-resistant, multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains of the pulmonary-TB-causing agent, *MTB*. It is estimated that about 558,000 cases of TB reported in 2017 fell into the rifampicin (one of the more potent of the first line drugs)-resistant category, with over 80% and less than 10% of the aforementioned cases confirming infection by MDR-TB and XDR-TB respectively (WHO, 2018).

1.1.5. MDR-TB and XDR-TB

The first-line drugs RIF and isoniazid (INH) are the most potent drugs that are usually used to treat new patients infected with *MTB*. If the tested *MTB* isolate is found to be resistant to both RIF and INH, the *MTB* isolate is referred to as MDR-TB. Given that the majority of cases confirm that over 90% of all *MTB* infections with resistance to RIF are

also found to be resistant to INH, RIF-resistance can be used as a surrogate marker for MDR-TB (Amita and Geeta, 2015). If however, the *MTB* isolate is found to be resistant to both RIF and INH with the addition of at least one fluoroquinolone and one of the injectable drugs, either amikacin, capreomycin or kanamycin, it is referred to as an XDR and results in XDR-TB when infecting a mammalian host (Wright and Zignol, 2008).

1.1.6. TB diagnosis

There are a total of five key aspects that are investigated in order to fully conclude whether an individual is infected with *MTB* or not. That is, an individual's medical history is checked, they are physically examined, screened for *MTB*, receive a chest x-ray, and isolated bacterial specimens are tested (CDC[b], 2016). Regarding medical history, the patient's symptoms may be investigated to try to determine if they are presenting with symptoms that match either pulmonary TB or extra-pulmonary TB. Although symptoms might be more consistent for patients experiencing pulmonary TB, and even though symptoms such as fever, night sweats and a loss of appetite are generally experienced by individuals with ATB infection, symptoms associated with extra-pulmonary TB will be determined by the part of the body affected (Lee, 2015). The physical examination is meant to respectively inform and guide the medical practitioner on the overall condition of the patient and a way forward in terms of diagnostic methods the patient will be able to undertake. Testing for *MTB* infections can be done through skin and blood tests. The skin test is done through the injection of a certain amount of tuberculin fluid into the skin of the patient's lower arm and is known as the Mantoux tuberculin test (CDC[a], 2016). Analysis of the test is usually done within two to three days of administration. Results depend on the diameter of the skin induration (hardened area of the skin) at the site injected. That

is, if the diameter of the area is 0-4 mm, it will be interpreted as a negative test. A diameter of 5-9 mm is interpreted as an uncertain test. The diameter of 5 mm or more might still be considered a positive result if the patient that was being tested is at that particular time, immunologically compromised. However, under normal circumstances, if the diameter is anything equal to or above 10 mm, it is considered a positive test (Nayak and Acharjya, 2012.). The blood test that is used to measure the strength of the immune response upon encountering *MTB* is called the Interferon Gamma Release Assay (IGRA) (CDC[c], 2016). Further, tests such as a chest x-ray and bacteriological exams should be taken to distinguish which form of TB the patient has. In addition, drug susceptibility testing (DST) can be used to gain further insight whether the patient has a resistant form of TB (Agyeman and Ofori-Asenso, 2017).

1.1.7. Modern TB diagnostics tools

Even though the use of DST is useful in determining the appropriate regimen for individuals infected with *MTB*, the cost and turnaround time linked with the required facilities and availability of lab results respectively prompted the development of rapid genome testing tools such as GeneXpert. The GeneXpert tool essentially scans for genetic markers that infer resistance, thereby providing a useful insight into the medical prognosis of an infected patient, although not giving much information about the mechanisms involved in the cause of such mutations to begin with (Dheda et al., 2017).

1.1.8. TB treatment

The treatment that a patient receives will depend on a number of factors. For instance, does the patient have LTB or ATB? Is it pulmonary, or extrapulmonary, or both? Does the infection exhibit resistance to treatment, and if so, to what extant? According to guidelines

provided by WHO, the treatment of drug-susceptible TB comprises of a 2-month intense course of a combination of first-line drugs, preferably including RIF and INH, followed by 4-6 months of the continuation phase. The treatment for MDR-TB is normally informed by the needs of the patient but usually consists of at least four core drugs with one being of the fluoroquinolones and another being a second-line injectable (WHO, 2018).

1.1.9. TB management

With over a third of the world's population suspected of having LTB (Cruz-Knight and Blake-Gumbs, 2013), and with the risk of LTB having a 10% likelihood of changing to ATB (Friedman et al., 2013), it cannot be stressed enough how important a system for managing an *MTB* infection is. This is complicated by the phenotypic resistance (resistance displayed without genetic alterations) displayed by *MTB* against a number of antibiotics. For instance, a drug's effectiveness may be minimal when the *MTB* cells are not in a metabolically active state even when dosed at high concentrations. To further complicate matters, the physical and chemical properties of the *MTB* cell wall do not easily allow drugs to penetrate the cell wall, rendering a number of antibiotics useless against *MTB* infections. The drugs found to be effective against metabolically active *MTB* includes the first line drugs INH, RIF, pyrazinamide (PZA) and ethambutol (EMB). First line drugs are used on patients who have newly been exposed to TB and are less likely to harbor a resistant strain. Second-line drugs are used on patients who have a drug-resistant form of an *MTB* infection (WHO, 2010).

1.1.10. Directly Observed Treatment, Short Course or DOTS

Adherence to the TB regimen is of vital importance in the fight against the development of drug-resistant strains of TB. Lack of adherence to medication is a trend associated with

poor treatment management leading to lengthened treatments with associated complications of treatment and the addition of adverse side effects linked with the use of second-line drugs (CDC, 2006). According to the work compiled by (Munro et al., 2007), 8 factors were identified as being the main contributors to a patient's adherence to TB treatment. These are the organization of treatment and care for those with TB, how the patient interprets illness and wellness, the finances involved in the treatment of TB, what the patient knows about the treatment along with their attitude and beliefs towards the treatment, the patient's legal standing and immigration status, the patient's gender and whether they have a drug-abuse problem, the side effects associated with treatment, and the people that influence the patient. To overcome these barriers, DOTS (Directly Observed Treatment, Short Course) was developed. DOTS is an adherence treatment regimen that is highly supported by the World Health Organization and entails the assistance of health workers, volunteers or family members in making sure that each dose of the TB chemotherapeutic treatment is observed and subsequently recorded when taken by the patient. The program insures aims to accomplish five main objectives in order to stop the spread of TB. In essence, the DOTS program is aimed at creating a system where TB monitoring along with training and recording is both centralized and prioritized; the detection of new cases using sputum smear microscopy analysis; anti-TB drugs to be swallowed under the direct observation of a healthcare worker or a volunteering community DOT provider; a constant supply of free anti-TB drugs for the patient; a system in which the progress of the patient pertaining to treatment is reported and recorded throughout the anti-TB regimen (Shah et al. 2007). The DOTS-plus program

is simply an addition to the DOTS program that aims to halt the spread of MDR-TB (Parish and Brown, 2009).

1.1.11. TB prevention

There are a number of ways in which modern medicine can be used to prevent new *MTB* infections or their progression from LTB to ATB. One method that has proven rather effective has been the use of TB preventative therapies (TPTs). The method entails the use of a single or a combination of drugs over the course of several months (post confirmation of *MTB* infection) (Badjie et al, 2017; Churchyard and Swindells, 2019). Although the list of TPTs is continuously growing, most of the people meant to be receiving care are not availing themselves to medical centers where treatment is offered. These include persons who are immunocompromised such as those living with HIV along with children under 5 years of age who are living with someone with tested and confirmed pulmonary TB (WHO, 2018).

Another means of protection against *MTB* infection is through the use of Bacille Calmette-Guérin (BCG) which is primarily used to vaccinate children (Kandasamy et al., 2016). The efficacy of the vaccine is less in adults as compared to children (Colditz et al.,1994; Mangtani et al., 2013; Mangtani et al., 2017; Roy et al., 2014; Nguipdop-Djomo et al. 2016; Nieuwenhuizen et al., 2018). It is also well documented that the efficacy of BCG is much greater in preventing TB meningitis or disseminated TB in infants and less so in preventing pulmonary TB (Rodrigues et al.,1993). Given that BCG is a live attenuated form of *M. bovis*, it harbors with it the risk of resulting in a local or systemic infection in those individuals in an immunocompromised state (Talbot et al., 1997). Furthermore, in conducted meta-analytical studies, it is indicated that on average, vaccinated individuals

will receive about 50% percent of protection over the course of 10 to 20 years with factors such as geographical location and pre-exposure to *MTB* playing major roles on how well the vaccine will work. However, given that BCG has been in use since the 1920's when it was first used to vaccinate a newborn prior to mass vaccinations (Calmette et al., 1927), it only serves as a testament to how much BCG fails to effectively protect against pulmonary TB given the state of the current global TB burden (WHO, 2017). More research into novel, more effective vaccines is therefore required.

1.2. *MTB*

1.2.1. *MTB* taxonomy and characteristics

MTB belongs to the class Actinobacteria, and is of the order Actinomycetales, a member of the Mycobacteriaceae family. *MTB* is characterized as rod-shaped, non-motile, slow-growing, non-spore forming, chemoorganotrophic and as highly aerobic. When observed in a laboratory under normal physiological conditions, the bacterium can reproduce every 24 hours or so. The two-layered cell wall of *MTB* is composed of mycolic acids or MAs (Daffé and Draper, 1997). It is for this very reason that *MTB* is generally visualized with the Ziehl-Neelson acid fast stain given that its thick and waxy cell wall is able to retain carbol-fuschin (which has affinity for MAs) when washed with acid-alcohol (Gordon and Parish, 2018).

1.2.2. *MTB* genome

Since sequencing the *MTB* genome, it was revealed that *MTB* has a genome spanning 4,4 million bases in length, with a GC content of 65,6% and codes for 4018 genes (Cole et al., 1998; Gordon and Parish, 2018). Of the discoveries made, two large protein families were identified with tallied members exceeding160 in total. Their C-termini were found to

be variable, however, the N-terminal regions were conserved, displaying either proline-glutamine (PE) or proline-proline-glutamine (PPE) (Cole et al., 1998; Gordon and Parish, 2018). These PE and PPE proteins were found to have diverse functionalities with specific members playing roles in cell wall architecture, lipid metabolism and immune modulation. Another discovery was that at least 9% of the coding regions of the genome are dedicated to lipid metabolism, relaying the importance of lipid metabolism in the synthesis of the complex didermal *MTB* cell wall along with the ability to use of host lipids as a food source (Gordon and Parish, 2018). A number of Esx type VII protein secretion systems were also discovered in *MTB*. These specialized secretion systems were initially discovered in Gram-negative bacteria and have been found to overcome the problem of transporting proteins across a two layered cellular envelope (Ates et al., 2016).

1.2.3. *MTB* phylogeny

MTB is closely related to a group of seven mycobacterial species also known to cause TB in mammals, namely, *M. africanum*, *M. bovis*, *M. caprae*, *M. canetti*, *M. orygis, M. microti*, *M. mungi* and *M. pinnipedii*. Together this group of mycobacterial species form the *MTB* complex (MTBC) (Niemann et al., 2000; Brosch et al., 2002; Forrellad et al., 2013; Bansal et al., 2018). Based on studies conducted using comparative genome analysis, it was determined that the MTBC displayed low clonal evolution and sequence diversity, and a sequential loss of regions of difference (RG), first described by (Mahairas et al., 1996) as regions within the genomes of *M. bovis* and *MTB* of all tested virulent laboratory and clinical strains. This suggested that *MTB* is more closely related to the common ancestor of MTBC (Gordon and Parish, 2018). Additional studies using whole

genome sequence analysis revealed seven key lineages isolated in specific regions across the world which happen to match human migration patterns (Comas et al. 2013).

1.3. Rifampicin

1.3.1. What is Rifampicin?

Rifampicin or rifampin (RIF) is a broad spectrum semi-synthetic antimicrobial agent and is derived from rifamycins, a group of naturally produced antibiotics obtained from the fermentation process involving the producer organism, *Norcadia mediterrenei* (Sensei et al., 1959). RIF is a cyclic compound that consists of three main regions, namely, the napthol ring, an ansa bridge and a methylated tail. RIF is highly effective (at low concentrations) in treating tuberculosis (including infections caused by Gram-positive bacteria) and has been widely used for that very purpose (Binda et al., 1971). It is especially useful in the fight against TB because it is effective against slow -growing and non-replicating bacteria (Xie et al., 2005, Sala et al., 2010; Piccaro et al., 2013; Koch et al., 2014), a property that is crucial for the virulence of *MTB* (Connolly et al., 2007; Barry et al., 2009; Rittershaus et al., 2013; Koch et al., 2014). The emergence of MDR-TB, which is resistant to RIF, has thus been an upsetting complication with regards to TB diagnosis and treatment.

1.3.2. RIF binding target

A series of experiments spanning years were used to ultimately arrive at the conclusion that the β subunit of RNA polymerase is the binding site of RIF. That is, based on studies conducted by (Wehrli et al., 1976) it was confirmed that the four subunit RNA polymerase core, made up of α 2, β, and β' subunits (encoded by the genes *rpoA*, *rpoB* and *rpoC* respectively) tightly bind RIF hence through a process of elimination the conclusion that

the σ (encoded by the gene *rpoD*) that completes the holoenzyme is not directly involved in the binding of RIF. Again, with the discovery of a modified β subunit in RIF-resistant *Escherichia coli* (*E. coli*) gave an additional clue to the binding site of RIF (Rabussay and Zillig, 1969). However, the more conclusive experiments were conducted by (Rabussay and Zillig, 1970). They showed through reconstitution experiments that RNA polymerase that consisted of α and β' subunits from a RIF-resistant strain and β subunit from a RIF-sensitive strain was RIF sensitive when tested, and how RNA polymerase that consisted of α and β' subunits from a RIF-sensitive strain and β subunit from a RIF-resistant strain was RIF-resistant when tested.

1.3.3. RIF mechanism of action

RIF binds to the β subunit of DNA-directed RNA polymerase (DDRP) with a binding constant of about 10^{-9} M under normal physiological conditions. This is in contrast with other antibacterial agents like actinomycin which affect both bacterial and mammalian RNA polymerase activity. That is, RIF does not affect mammalian RNA polymerases when tested for activity at similar concentrations, an advantage that is linked with its mode of action (Hartmann et al., 1967) which according to (Johnston and MacClure, 1976) was thought to be the induction of abortive initiation of the transcription process by forming a pppApU dinucleotide and preventing the formation of a second phosphodiester bond (Mcclure and Cech, 1978).

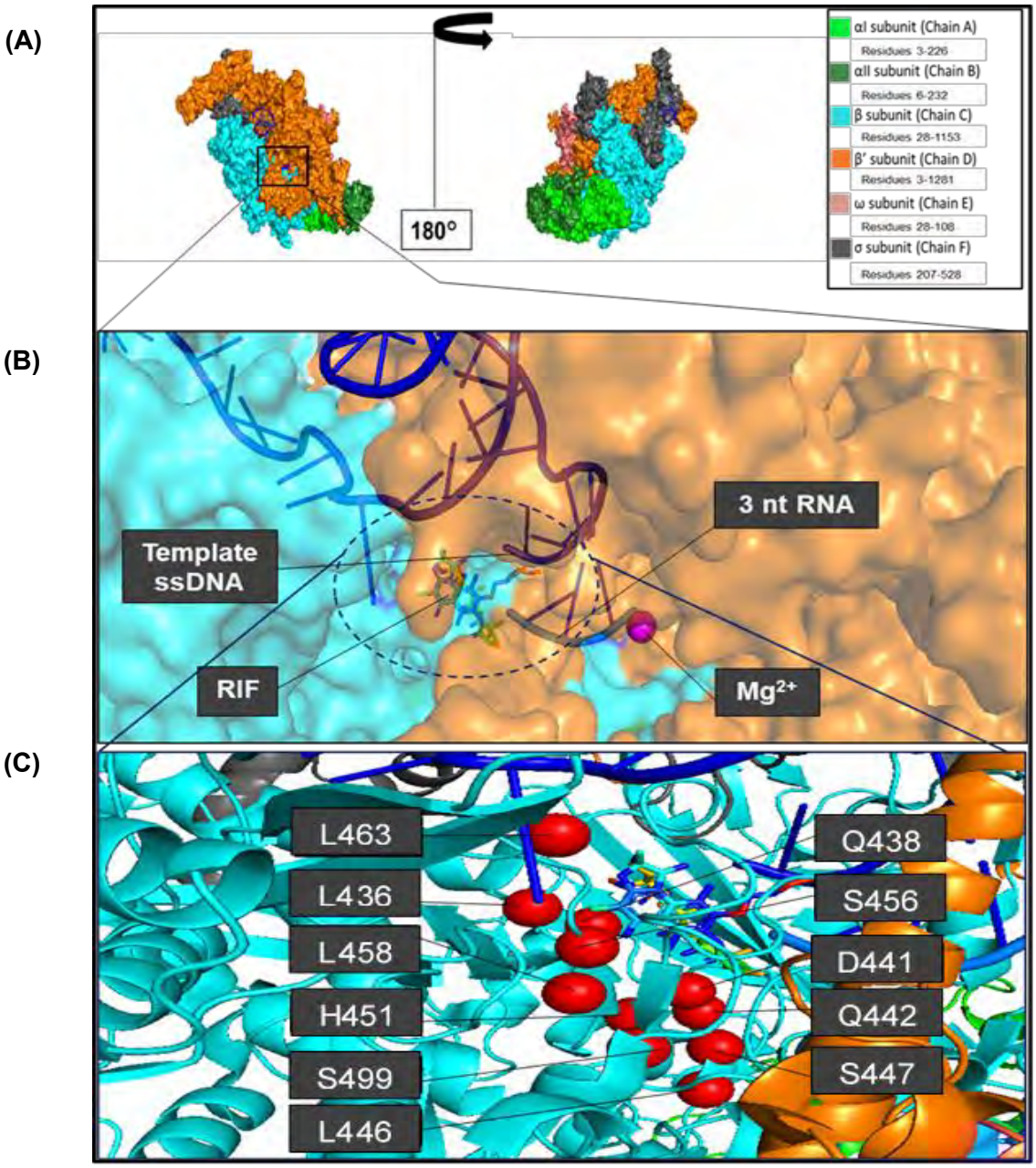


Figure 1. **A)** 3-dimensional image of multisubunit DDRP (surface). Chain A (αI): green, chain B (αII): forest green, chain C(β): cyan, chain D (β'): orange, chain E (ω): deep salmon, and chain F (σ): grey40. The co-crystallized molecules were also uniquely distinguishable. DNA: blue (cartoon), RNA: teal (cartoon), catalytic Mg^{2+}: magenta (sphere), Zn^{2+}: yellow (sphere), and RIF: rainbow (stick). The corresponding ranges of residue positions for the chains were also displayed. **B)** Captured prevention of 3nt RNA elongation by DDRP through RIF steric occlusion (circled) as viewed by PyMol (semi-transparent surface). **C)** Binding pocket of RIF with annotated residues (red spheres) showing identified SNP positions.

With more research being carried out into the mechanism of action, it was further revealed that in the presence of RIF, a minute amount of larger oligonucleotides are formed in addition to the aforementioned dinucleotides, thus leading to the idea that perhaps RIF might in fact compete either directly or allosterically with the binding site of the growing RNA product, resulting in the destabilized binding of intermediate oligonucleotides on the RNA-polymerase-DNA complex (Schulz et al., 1981; Kessler et al., 1982). It was however, only recently that the proposed mechanism of action by (Campbell et al., 2001), which is the steric occlusion of 2 and 3 nucleotide (nt) RNA transcripts (Figure 1B), as demonstrated by (Lin et al., 2017) through the co-crystallization of RIF with RNA polymerase transcription initiation complexes (RPTIC) with bound DNA scaffold and different length RNA directly or allosterically with the binding site of the growing RNA product, resulting in the transcripts.

1.3.4. RIF resistance

Back in the 1980s when the treatment of *MTB* infections was decreased from 24 months to only 6 months, a strict adherence to the new short-course treatment regimen was not taken to heart by all patients, and thus the first in a long line of drug-resistant *MTB* would begin to spring up across the globe. In as little as a decade or so, the first cases of MDR-TB were reported (Sandgren et al., 2009). Given the wide use of RIF since its introduction for clinical purposes in the year 1965, owed to by its highly effective broad spectrum, bactericidal properties, the number of reports of resistant strains has continued to increase (Long, 2000). According to a number of studies conducted the likes of (Lai et al., 2002) to understand the basis of RIF-resistance, the drawn conclusion was that it is usually (about 95% of the time) caused by mutations in the *rpoB* gene which usually

occurs within a 81bp mutation hotspot region termed the RIF-resistance determining region (RRDR). These mutations tend to vary in both pattern and frequency and tend to depend on the geographical location in which the *MTB* isolate was obtained (Adikaram et al., 2012). Of the observed mutations in the region, the commonly occurrences are those of missense and nonsense mutations and less so are insertions and deletions observed (Rienthong, 2017). These mutations decrease the overall level of affinity for the drug to its target site, the β subunit (*rpoB* protein) of bacterial RNA polymerase (Campbell et al., 2001). This results in a decreased efficacy due to the drug being less tightly bound to the target site.

1.3.5. Mutations of *rpoB* gene in South Africa

South Africa is a country that has experienced persistently high incidences of TB in the past few years. In 2008 alone, the Eastern Cape, one of the nine provinces in South Africa, experienced one of the worst epidemics in recorded history when 80% of all reported TB cases occurring in the country, were found to be from the Eastern Cape (Qazi et al., 2014).

The rise in cases is said to have been mostly due to the province being the poorest in the country which makes the location conducive to forming the necessary conditions for easy transmission. The recorded SNPs that encoded RIF-resistance were found to be especially unique, with a majority of them occurring outside the RIF-resistance determining region (RRDR) region of the *rpoB* gene. The recorded SNPs are as follows: Tyr42Asp, Gly52Ala, His87Gly, Leu92Ser, Val441Gly, Leu450Ser, and Leu457Pro (Bhembe et al., 2014).

Another province that has seen its fair share of TB cases is KwaZulu-Natal (Moodley et al., 2011). A number of genetic mutation studies on *rpoB* gene (among other genes) pertaining to resistance observed in MDR-TB and XDR-TB and were conducted using *MTB* isolates from patients that were taking treatment in Church of Scotland Hospital between the years 2005 and 2009. The SNPs of the rpoB gene that were identified to encode RIF-resistance in MDR-TB are as follows: Asp516Tyr, His526Leu, Ser531Leu, Leu533Pro, Pro535Thr, ILe572Met, Tyr645His (Dookie et al., 2016). Of the 28 SNPs of the *rpoB* gene that were identified to encode RIF-resistance in XDR-TB, Asp516Gly and Leu533Pro in combination appeared to be the most frequently observed in each of the tested XDR-TB isolates while only three displayed Asp516Gly and another displayed a Ser531Leu SNP (Dookie et al., 2016).

1.4. Research problem statement

Wet laboratory experiments have yet to fully elucidate the DDRP-RIF interaction dynamics in regard to RIF-resistance-conferring SNPs at the RIF-binding site (Molodtsov et al., 2017; Srivastava et al., 2018). Therefore, by tapping into the bioinformatics toolkit and making use of molecular dynamics, structural models of drug-target interactions can be generated (taking SNPs as listed in the TBDReamDB into account), to gain insight into the structural mechanism of inhibition of RIF and to gain a better appreciation of the effect that SNPs have on RIF binding energy pertaining to the mutant forms of the DDRP protein, and how these can be improved using newer, or well-established or slightly modified well-established drugs to combat RIF-resistance. This insight might help in

predicting the effect a detected *rpoB* SNP might have on a patient's treatment regimen and thus possibly improve their prognosis through the timely prescription of an appropriate drug-therapy by the patient's physician.

1.5. Aim and objectives

1.5.1. Overall aim

To gain insights on the binding behavior of RIF to DDRP in the presence and absence of clinically relevant resistant mutations in terms of binding affinity, conformation sampling and intraprotein communication network, for the purpose of finding a solution to RIF-resistance-conferring SNPs.

1.5.2. Specific objectives

a) To retrieve the 3D structure file of DDRP holoenzyme with the bound RIF ligand (PDB code: 5uhc) and SNP mutation data conferring RIF-resistance from PDB and TBDReamDB, respectively.
b) To apply comparative modelling to construct valid DDRP models of the WT and its respective SNP mutants using MODELLER and to accordingly rank and validate the models using MODELLER's z-DOPE and GA341 scores, and QMEAN, VERIFY3D and PROCHECK tools, respectively.
c) Generation of molecular dynamics simulations using GROMACS with the AMBER force field (to account for bound nucleic acids) and to make use of ACPYPE to

generate the topology files for the RIF ligand since GROMACS does not have topology data for RIF.

d) To analyze the RMSD, RMSF and Rg data of WT and mutant DDRPs to determine changes in the respective stability, residue-level fluctuations, and compactness of the SNP models relative to the WT.

e) To make use of MD-TASK tools to calculate change in average shortest path (ΔL) and change in *betweenness centrality* (ΔBC) for DDRP WT and mutants to reveal changes in residue-pair distances and importance of residues in protein communication within the calculated DRN, respectively.

1.6. Expected outcomes

To establish structural and functional changes in binding characteristics of RIF, pertaining to clinically relevant SNPs in DDRP β subunit. We looked forward to identifying mutation-specific influence on RIF binding-affinity, DDRP structural mobility, and intraprotein communication patterns.

Chapter 2

Homology modelling and structure validation

2.1. Introduction

This chapter was aimed at retrieving the verified TB-drug-resistance mutation data of the RV0667 gene. This includes filtering the data and producing validated protein models that can be used to ascertain the effect that specific mutations have on RIF efficacy. All the tools, steps and approaches used to accomplish this task are described below.

2.1.1. Data retrieval

2.1.1.a) Protein Data Bank

X-ray crystallography is one of the most commonly used techniques used to visualize three-dimensional (3D) structures of biomolecules such as proteins and nucleic acids. It entails the visualization of 3D structures of biomolecules by passing a beam of X-rays through a crystallized biomolecule. The X-rays interact with the electron cloud of the biomolecule's atomic constituents within the crystal and result in an array of diffraction beams that form a complex pattern detected as spots. These spots, coupled with further experimentation and a fair amount of mathematical computations, generate an accurate electron density map indicating the positions of each atom within the crystal (Brito and Archer, 2013).

Since the development of the X-ray crystallography technique in the early 1920's and the structural determination of the protein, pepsin, back in the 1930's (Bernal and Crawfoot, 1934), a platform to share an ever-increasing collection of biomolecular coordinate

information among the structural biology community became a necessity. Hence, it was in the early 1970's that after attending a Cold Harbor Symposium entitled "Structure and Function of Proteins at the Three Dimensional Level" that a crystallographer by the name of Walter Hamilton offered to host the Protein Data Bank (PDB) archive (Protein Data Bank, 1971). It was only a few years later that a guideline for the uniform descriptor for biomolecule structural information depositions was officially made public (International Union of Crystallography,1989; Berman et al., 2014). This allowed for the easier sharing of biomolecule coordinate information, irrespective of whether their structure was elucidated by other techniques such as nano magnetic resonance or NMR, and thus the number of structures that would be determined would increase exponentially, from the initial launch of 7 structures to reaching an astounding figure of 100,000 entries in the year 2014 (Berman et al., 2014).

2.1.1.b) TB Drug Resistance Mutation Database

The TB Drug Resistance Mutation Database (TBDReamDB) website (http://www.tbdreamdb.com/), which was first launched in late 2008, serves as a platform where drug resistance-conferring mutations reported in high quality scientific papers are organized in a comprehensive format, allowing for easy access of the most relevant data to one's research focus, which is intended to effectively speed up the rate at which discoveries are made in terms of novel approaches in treating and preventing TB (Sandgren et al., 2009).

2.1.1.c) Mycobrowser

Mycobrowser is an online resource (https://mycobrowser.epfl.ch/) that provides comprehensive information generated from *in silico* and peer reviewed sources from

databases that concentrate their efforts on the study of *M. leprae*, *M. marinum*, *M. smegmatis*, and *MTB* genomes (Kapopoulou et al., 2011)

2.1.2.) Homology modelling

One of the recurring problems still faced in biology has to do with the functional characterization of proteins, an issue that is usually aided by the use of accurately elucidated 3D structures of proteins homologous to the protein of interest. However, due to the gap that still remains due to the ever-growing number of known primary sequences and less so for 3D structures of the same protein sequences, comparative modelling or homology modelling becomes the next ideal route if a 3D model of the protein is required (Fiser, 2004; Misura and Baker, 2005; Petrey and Honig, 2005; Misura et al., 2006; Webb and Sali, 2014). That is, homology modelling is built on the notion that the structural features of proteins are generally more conserved than the evolving sequences that form them (Illergård et al, 2009).

2.1.2.a) Template identification

In order to quickly identify potential templates for closely related sequences with a sequence identity from 30 to 40%, programs such as the Basic Local Alignment Search Tool or BLAST (Altschul et al., 1997) should be employed. BLAST is a sequence similarity search program which happens to be one of the most widely used bioinformatics research tools, is available as a standalone program, and is available via web interface. The program follows a heuristic method, identifying short local matches between the query sequence and a database sequence and subsequently generates an alignment with the addition of supportive statistical information that provides confidence in the detected hit (as indicated by the expect value or E-value) (Johnson et al., 2008). Another useful tool

comes in the form of HHpred, which is a structure prediction server that employs a profile comparison of hidden Markov models (HMMs) and specializes in distant homology detection (Söding et al., 2005). A useful online homology modeling tool that provides the option to use either one of the aforementioned search programs, and does so on a single platform, is none other than the recent webserver, PRIMO (Hatherley et al., 2016).

2.1.2.b) Template selection

Selection of template structures can be prioritized based on the purpose of homology modelling. That is, a set of criteria may be used for the selection of an appropriate template, for instance, selection based on sequence identity between the template and target, observing PDB validation data (e.g. resolution, R-work and R-free) for accuracy of the potential template, the conservation of active site residues, ligands bound to the protein, and any other reported relevant biological information that might inform the decision. It is also important to note that multiple templates may be selected for a single target sequence. In fact, it is said to produce a more accurate model (Srinivasan and Blundell, 1993; Sánchez and Šali, 1997; Webb and Sali, 2014).

2.1.2.c) Model building

Structure modelling by satisfaction of spatial restraints is a means of deriving a 3D protein model of a particular conformation from a sequence alignment using a homologous protein sequence(s) with known structure(s). The idea behind the method is based on the EMDEB distance geometry algorithm (Crippen and Havel, 1998; Havel et al., 1983) which has been used in multiple instances for the determination of protein structures employing the use of derived restraints from NMR data (Wüthrich, 1986; Webb and Sali,

2014). The restraints are further supplemented with stereochemical restraints derived from molecular mechanics force field calculations. The protein model is then produced by minimizing the violations caused by the restraints (Havel and Snow, 1991; Webb and Sali, 2014).

There are a number of ways one might accomplish the comparative modelling of a protein structure. There are online tools such as SWISS-MODEL (Arnold et al., 2006) or PRIMO (Hatherley et al., 2016) and standalone programs such as MODELLER (Šali, 2011; Webb and Sali, 2014). MODELLER is a script-based program that takes input in the form of aligned sequences of target and template(s), the 3D coordinate file(s) of the template(s), and then together with a script (which may be edited to fit the needs of the user) is able to generate models for all non-hydrogen atoms. However, if a coordinate structure file of a homolog does not exist in the PDB, an *ab initio* approach such as RaptorX (Källberg et al., 2012) may be warranted.

2.1.4. Model evaluation

2.1.3.a) MODELLER DOPE score and z-DOPE score

The MODELLER Discrete Optimized Protein Energy or DOPE (Shen and Šali, 2006) is a statistical potential dependent on interatomic distances and is used to determine the accuracy of fold assignment. The DOPE scoring system is based on a physical reference state that takes into account a protein's finite size along with its spherical shape and also assumes that a protein chain consists of non-interacting atoms within a uniform sphere of radius equal to the protein in question. A normalized DOPE score or z-DOPE score is a general statistical score derived from raw DOPE scores (Eramian et al, 2008).

2.1.3.b) MODELLER GA341

The MODELLER GA341 score is a composite scoring function used to assess the accuracy of protein folding. It is determined by combing a Z-score calculated using a statistical potential function, the sequence identity between template and target, and the measure of a biomolecule's compactness. The GA341 score ranges between 0 and 1, with 0 inferring an incorrect fold while 1 generally relays high similarity between the structure in question and elucidated low-resolution structures. It has also been observed that models with a GA341 score greater than 0.7 tend to a display correct fold when compared to their experimental structures (Eswar et al. 2008).

2.1.3.c) QMEAN

The Qualitative Model Energy ANalysis score or QMEAN score is a composite scoring function used to perform global and local structural model quality assessments based on a weighted linear combination of six structural terms, namely, solvation potential, which is a measure of the burial status of a residue; torsion angle potential, which is indicative of the local geometry of the protein structure by observing three consecutive residues; a pair of distance-dependent interaction potentials for Cβ atoms and all atoms, respectively, for the assessment of long range interactions; a pair of terms that highlight the agreement in predicted and experimental secondary structures and solvent accessibility. QMEAN is a size-independent method that makes use of a set of high quality non-redundant structures resolved by X-ray crystallography retrieved from PDB to produce a normalized score which is then converted to a QMEAN-Z score. The QMEAN-Z score is used to determine how close the model is to being native-like, which can then be used to ascertain whether it is of similar quality when compared to experimentally determined structures or

not. QMEAN6-Z scores around 0 represent structures considered to be native-like while those that are less than -4 are considered low quality structures (Benkert et al., 2008, 2009, 2010).

2.1.3.d) VERIFY3D

VERIFY3D is a method for determining the accuracy of a 3D protein model, by comparing the model with its own amino acid sequence through the use of a 3D profile. The 3D profile consists of atomic coordinates observed in the 3D model, with each residue position represented by 20 numbers also referred to as 3D-1D scores. The scores reflect each residue's environment and represent the statistical preferences for the 20 amino acids for each residue in this environment. 3D-1D scores are defined by a total of 3 parameters, namely, the buried region of a residue, the portion of a side-chain covered by oxygen and nitrogen atoms i.e. polar atoms, along with the local secondary structure of the model. High scoring models are those have high compatibility with their amino acid sequence whilst those that score poorly show less compatibility with their amino acid sequence (Lüthy et al., 1992).

2.1.3.e) PROCHECK

PROCHECK is a suite that can used to determine the stereochemical properties of a protein model by comparing values observed in the model to those of high-resolution structures from the Protein Data Bank (Bernstein et al., 1977). The output comes in the form of a number of plots in a comprehensive format, highlighting regions that might require further inspection (Laskowski et al.,1993). Of the plots generated by PROCHECK, the Ramachandran plot, which displays residue torsion angles, might reveal more data regarding the stereochemical properties of the model of interest (Laskowski et al., 2006).

2.2. Research approach/methodology

2.2.1. Data retrieval and filtering

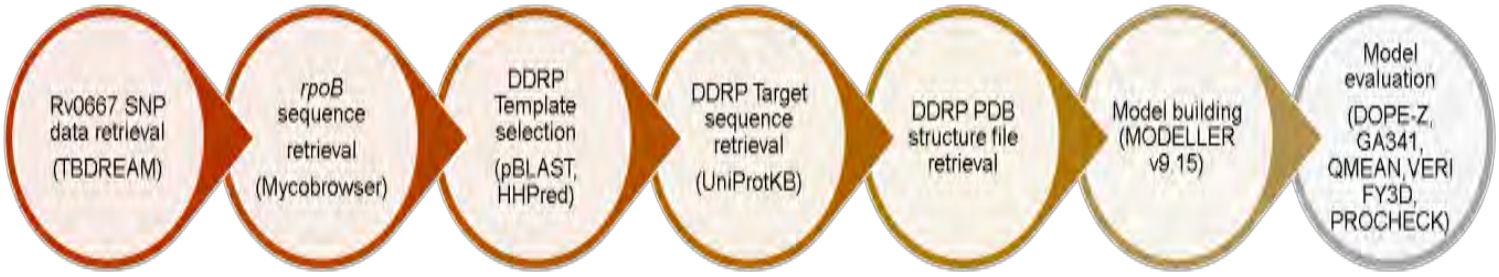


Figure 2: Summary of homology modeling workflow.

2.2.1.a) Rv0667 (*rpoB*) mutation data

The mutation data conferring RIF-resistance was retrieved from the TBDReamDB on the 7th of August 2019. Given that the study would only concentrate on SNPs conferring RIF-resistance, a script was written to sort the data based on the types of mutations reported. The data was further sorted on the basis of what the TBDReamDB refer to as “high confidence” mutations (stated as such through the selection of 10 high quality representative papers that gave further support to the recurrently observed mutations in studies conducted all over the world to perhaps imply their therapeutic significance) and those not considered high confidence (i.e. lacking enough data to meet the criterion of ”high confidence”).

2.2.1.b) DDRP coordinate structure file

A PDB structure search was firstly conducted using a sequence of the H37Rv *rpoB* protein (as displayed in Mycobrowser) on protein-BLAST (BLASTp) and then on HHpred in an attempt to identify the *MTB* DDRP complex co-crystallized with RIF or homologous complex structures also co-crystallized with RIF. The *MTB* DDRP complex with the co-crystallized ligand was found (PDB ID:5uhc) and subsequently assessed based on its respective resolution, R-value and R-free.

2.2.2. Homology modelling and validation

Due to the non-terminal missing residues in the PDB ID:5uhc coordinate structure file, namely, the regions 156-157 (chain B) and 1012-1025 (chain D), MODELLER v9.15 (Šali and Blundell, 1993) was used to produce models of both the WT and mutants (in accordance with the selected group of reported SNPs on the TBDReamDB) of the DDRP using the 5uhc coordinate file as a template. Missing terminal residues were found mainly on the periphery of the DDRP complex and were thus not modelled. The missing residues in chain B were modelled using the residues observed in the UniProtKB sequences while the region of missing residues in chain D was corrected by removing the gap and introducing two glycine residues to complete the loop formed in that region. Firstly, the alignment data (generated using T-Coffee (Notredame et al., 2000)) between the PDB residue sequences and the respective UniProtKB sequences were organized in accordance with instructions stipulated in the MODELLER manual for creating a .pir alignment file (using "." to indicate BLK residues to represent DNA, RNA, RIF, and Mg2+ and Zn2+ metal ions (Index A-1)), and was subsequently saved in the aforementioned

format. The MODELLER script (Index-A2) (Šali, 1993) was used to generate 100 models with very slow refinement. The best model was selected based on the lowest calculated normalized DOPE score (z-DOPE score) and highest GA341 score among the 100 models. The following validation servers were used to further assess the models: PROCHECK (Laskowski et al., 1993), QMEAN (Benkert et al., 2010) and VERIFY3D (Eisenberg et al., 1997).

2.3. Results and discussion

2.3.1. Mutation data filtering

Exactly 133 mutations of the *rpoB* gene encoding the β subunit of *MTB* DNA-directed RNAP (DDRP) were reported in the TBDReamDB. Of the 133, 114 were identified as SNPs, which are mutations that present as variations that occur at a single position within a DNA sequence and are found in at least 1% of a population (Scitable, 2014). Of the 114 SNPs, a total of 17 mutations were categorized as “high confidence mutations” (HCMs). HCMs included the Asp441, His451 and Ser456 SNPs, well documented in literature to confer RIF-resistance and noted as the most frequently occurring mutations among all the observed SNPs conferring RIF-resistance in studies conducted worldwide (Ramaswamy et al. 1998; Chikaonda et al., 2017). For the purpose of this study, given the time restraints, the three aforementioned codon positions were selected for further research. The identified knowledge gap regarding how a change in these residues within a large DDRP complex affects the binding of RIF by MD analysis remains to be fully explored. Furthermore, to the knowledge of the authors, this is the first paper to explore the mechanism of changes in binding efficacy of RIF in *MTB* DDRP mutants.

2.3.2. Template selection

Following a comparison of the listed hits from protein-BLAST and HHpred, the best *MTB* DDRP complex co-crystallized with RIF was found to be PDB ID: 5uhc (Lin et al., 2017). It is stipulated that the protein was prepared through x-ray crystallization and had a resolution of 3.764 Å which was relatively better than any of the observed hits with the protein complex co-crystalized with the RIF ligand available on the PDB server. The pdb validation data were 0.204 and 0.267 for the R-value and R-free, respectively, indicating a good protein model (Kleywegt and Jones, 1997). It also displayed the highest sequence identity and coverage when comparing each chain to the complete WT sequences of H37Rv ATCC 25618 displayed on the UniProtKB database (Bairoch et al., 2005). UniProtKB codes: P9WGZ1(×2), P9WGY9, P9WGYZ, P9WGYS, and P9WGI1, for chains A and B, C, D, E, and F respectively.

2.3.3. Model selection

The best DDRP models were selected from the respective set of100 models generated. The selection was mainly based on the normalized DOPE score or z-DOPE score (Table 1). Structures with a z-DOPE score of -0.5 or less tend to be near-native structures with correct folding and those with a score of -1 and less are considered native-like, consistent with high resolution experimental structures. Based on the z-DOPE values obtained it is clear to see that the models produced can be regarded as being of acceptable quality, however, further assessment using external tools was required to confirm these results.

2.3.4. Model evaluation

The quality of the generated DDRP models using MODELLER (as informed by Tables 1 and 2 (Annex A-3), and Figures 3, 4 and 5) was further assessed using QMEAN,

PROCHECK and VERIFY3D. Based on the results gathered pertaining to the QMEAN local quality assessment (Figure 3), the generated WT model was found to be of similar quality to the WT template, managing to have a normalized QMEAN6 score of 0.72 (ranging between 1 and 0 for good and bad models, respectively) and a QMEAN6 Z-score of -1.226 compared to a normalized QMEAN6 score of 0.71 and a QMEAN6 Z-score of -1.327 for the template, indicating near-native structures. VERIFY3D was used to assess the accuracy of the protein folds by comparing the 3D structure of the protein complex with its primary sequence, also called the 3D-1D assessment scoring system. Looking at the problematic local regions outlined on the graphs (Figure 4), it was clear that the chain D corrected region was the most problematic and was most likely the largest contributor to the 3D-1D score observed (56.77% for the generated WT model compared to 70.71% for the WT template), given that the 3D-1D scores for the respective residues were so low, they did not even appear on the graph using the general scale, which is in contrast to the template. However, given that this region was a loop region corrected by the elimination of a 14 residue long gap and the addition of two glycine residues, introducing any more uncertainty to an already flexible structure to improve the score would not have yielded a more favorable structure, hence it was decided that no further modifications would be made to the DDRP model given that a second template to account for the missing region could not be found in the PDB archive. PROCHECK was used to generate Ramachandran plots (Figure 5) revealing the stereochemical correctness of the generated model and template structures. It was revealed that the template had 89.5% residues within the favorable region with and no residues in the unfavored regions while

the model had 93.9% residues in the favored region and only 0.1% of its residues in the unfavored region, thus also indicating an acceptable model.

Table 105: Best scoring DDRP models based on z-DOPE score and GA341 scores.

Model	z-DOPE score	GA341
WT (template)	-0.870	-
WT	-0.531	1.00
D441V	-0.518	1.00
D441Y	-0.529	1.00
H451D	-0.540	1.00
H451L	-0.530	1.00
H451N	-0.528	1.00
H451R	-0.524	1.00
S456L	-0.523	1.00
S456W	-0.533	1.00

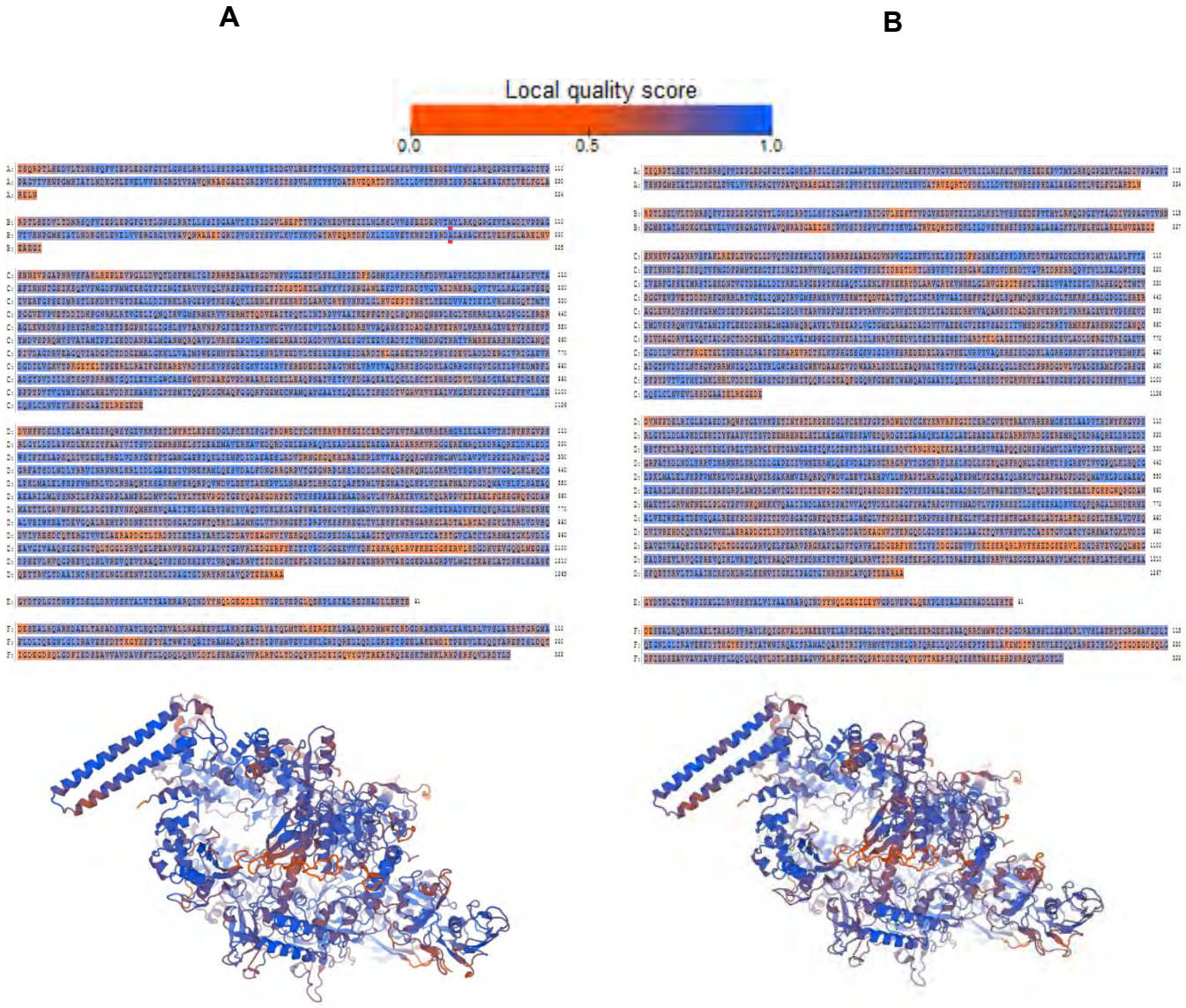
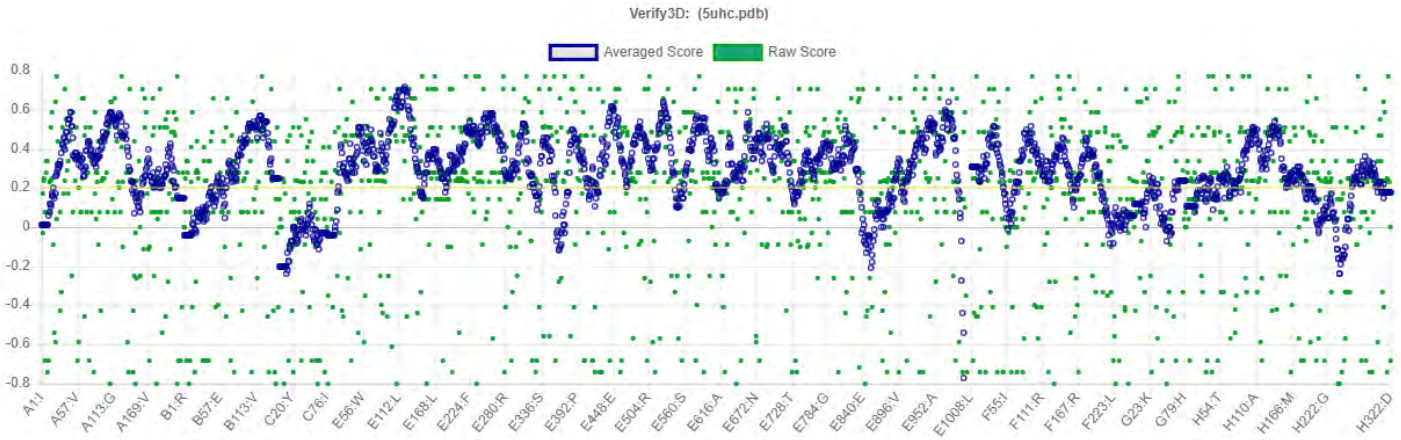


Figure 3: QMEAN model evaluation of the A) WT template and B) WT model. The WT template and the WT model scored a normalized QMEAN6 score of 0.71 and 0.72, respectively. The orange indicates local regions of the protein that score poorly and are likely to stray from the expected native structure. The blue indicates regions that score well and are more likely to agree with the native structure and are thus considered accurate or non-problematic regions.

A



B

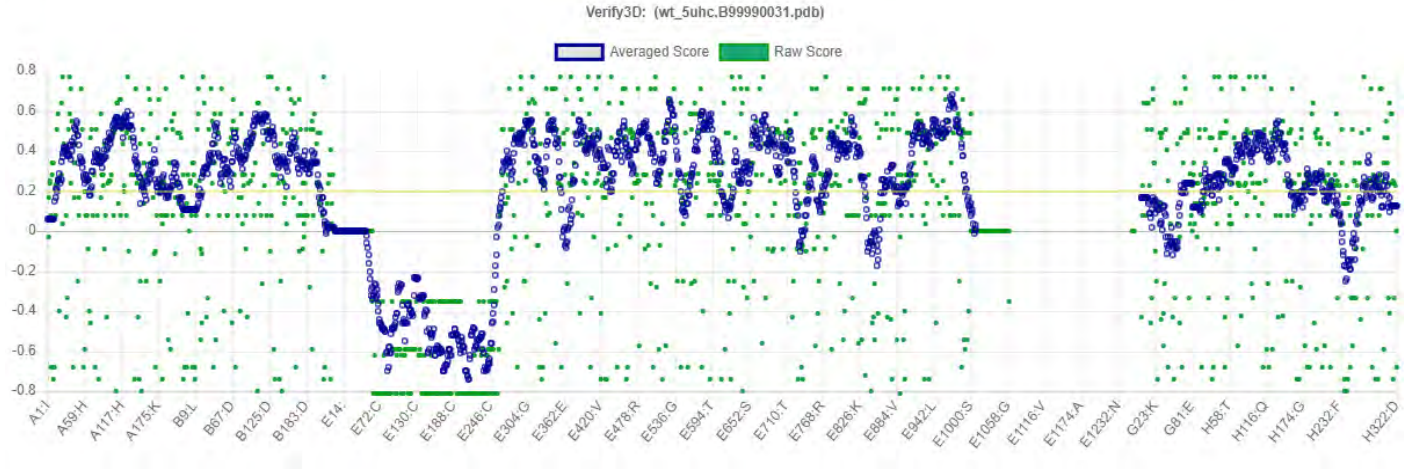


Figure 4: VERIFY3D model evaluation plot of the A) WT template and B) WT model. The horizontal axes describe the residue positions and the vertical axes represent average 3D-1D scores. The 3D-1D score compares the 3D profile of the protein in question with its amino acid sequence and relays a score for each residue. The blue and green dots represent the averaged and raw 3D-1D scores, respectively, for the protein residues. A protein model with a score of 80% and above, is considered a good model while a model that scores less than 80%, is considered to be faulty. The WT template managed to fail with a 3D-1D score of 70.71%, while the WT model failed with a score of 56.77%.

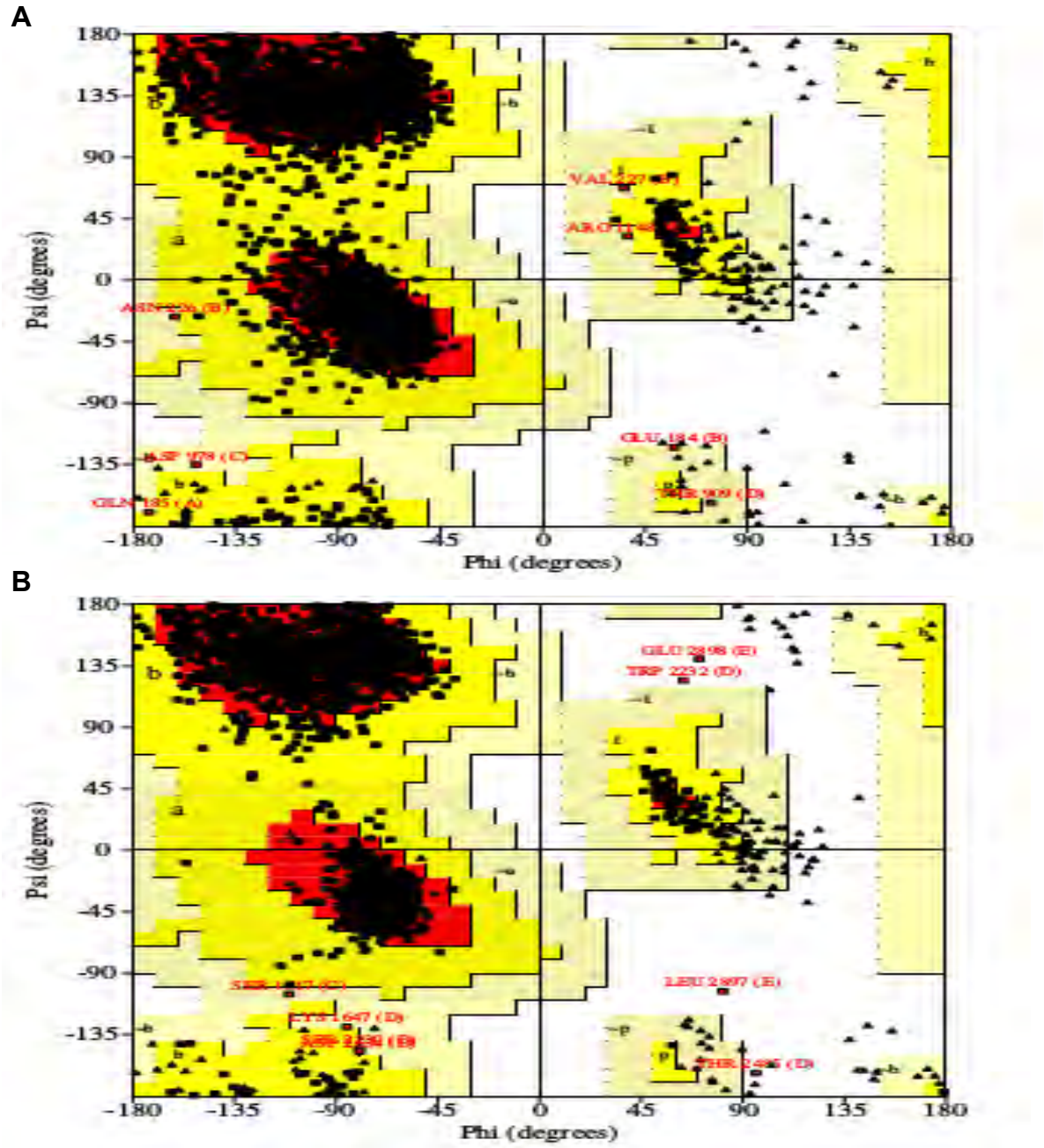


Figure 5: PROCHECK model evaluation of the A) WT template and B) WT model via Ramachandran plot. The horizontal values represent phi (φ) angle values whilst the vertical values represent psi (ψ) angle values. The glycine residues are uniquely represented as triangles whereas the other residues are represented as squares. The different colored regions represent the different parts of the plot as outlined by (Morris et al., 1992; Laskowski et al., 1993). The red color represents regions with favorable φ-ψ combinations with less intense colors signifying the less favored combinations. The expected outcome for a good structure is where over 90% of the residues are concentrated in the designated A, B and L regions regarded as the favored regions (Laskowski et al., 1993). The WT template and WT model had 89.5% and 93.9% of the residues in the favored region, respectively.

Chapter 3

MD simulations and trajectory analysis

3.1. Introduction

The aim of this section is to relay the molecular mechanics calculations that have been conducted on the validated WT and various SNP models, and to attempt to explain the data through a set of analytical tools that might give more useful information on what happens to DDRP at both a global and residue level when clinically relevant mutations are introduced to specific β subunit codon positions.

3.1.1. GROMACS

GROMACS or the GROningen MAchine for Chemical Simulation software package was developed during a collaborative effort conducted in the 1990's by members of the Chemistry department and the Computer Science Department of the University of Groningen (Netherlands) to create a dedicated parallel computer system for the purpose of running molecular simulations. The development of the GROMACS program was aimed at providing an adaptable biomolecule simulation program on either single processors or parallel systems. The program also provides microcanonical Hamiltonian mechanics, stochastic dynamics, Langevin and Brownian dynamics, energy minimization and a number of coupling methods and trajectory analysis tools. The force fields used to define interatomic interactions include GROMOS96 (van Gunsteren et al., 1996), OPLS (Jorgensen et al, 1996), Amber (Case et al., 2005) and CHARMM (Brooks et al., 1983).

GROMACS has grown to become one of the most widely used open-source codes for molecular dynamics simulations (Van Der Spoel et al., 2005).

3.1.2. AmberSB-ILDN force field

Molecular dynamics simulations have faced two main problems that have somewhat impeded their application. The first issue has to do with the insufficient processing power of computers, essentially limiting the amount of time that can be used for conformational sampling. The second problem is the accuracy of potential energy functions, which tends to bias the system towards certain conformations that may be incorrect if the potential energy function fails to accurately describe the system (Klepeis et al., 2009). As the former issue continues to be improved through the development of faster processors, it is of vital importance that the force field calculations required to generate the wealth of conformations that may become available for sampling are not biased by incorrect calculations, therefore force fields need to be continuously improved. It was for this very reason that the Amber ff99SB-ILDN was developed (Lindorff-Larsen et al., 2010). The Amber ff99SB-ILDN which has parameters for both nucleic acids and proteins, consistent with previous versions of Amber force fields, improves on the description of torsion potentials of side chains of amino acids from its predecessor, Amber-ff99SB (Hornak et al., 2006), which is a version of the Amber force fields that improved the description of backbone torsion potentials allowing for a better match to experimental data (Lindorff-Larsen et al., 2010).

3.1.3. Discovery studio

Discovery studio is a comprehensive predictive science suite that provides Life Science researchers with a range of research tools described as a unique concoction of

collaborative, open and scalable set of tools (Dessault Systemes, 2019). One such tool is the Discovery Studio Visualizer (DS Visualizer) which offers the user a variety of options that span from the advanced visualization of molecular systems, ligand-based design structure-based design, macromolecule design and a web-based platform to enable sharing and collaborative work (Dessault Systemes, 2017).

3.1.4. Avogadro

Avogadro is a program developed for the purpose of building and visualizing biomolecules and was designed to accommodate use over a number of platforms including bioinformatics and computational chemistry (Hanwell et al., 2012).

3.1.5. ACPYPE

AnteChamber PYthon Parser interfacE or ACPYPE is a tool based on ANTECHAMBER (Wang et al., 2006) which is used to generate topologies and parameters in formats compatible with multiple molecular mechanics programs including GROMACS. The tool is also able to calculate partial charges (da Silva and Vranken, 2012).

3.1.6. Open Babel

Open Babel is an open-source chemical data file format interconversion software that attempts to solve the problem of the ever-growing chemical file formats. The Open Babel 2.3 is designed to accommodate over 111 interconversion formatting, along with a number of other utilities that include, but are not limited to, 2D depiction, batch conversion and conformer searching (O'Boyle et al., 2011).

3.1.7. MD-TASK

MD-TASK is a user-friendly software suite that makes use of graph theory techniques, perturbation response scanning (PRS), and dynamic cross correlation (DCC) to provide a more detailed picture than would be provided by the traditional means of MD trajectory analysis (Brown et al., 2017).

3.2. Research approach/methodology

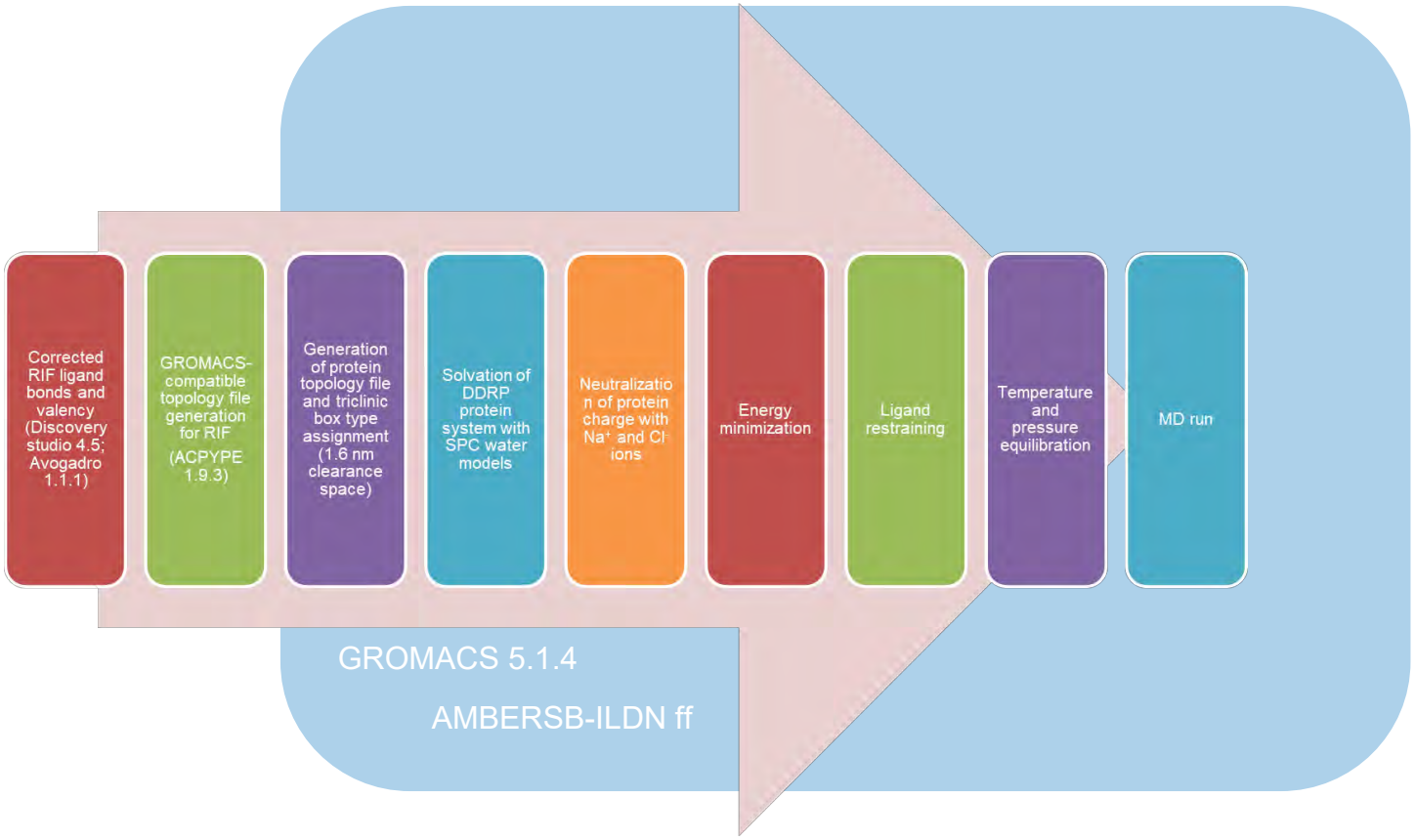


Figure 6: MD simulation workflow summary.

3.2.1. MD simulations

MD simulations were carried out using GROMACS 5.1.4 with a triclinic box type and a clearance space of 1.6 nm. The Amber ff99SB-ILDN (Lindorff-Larsen et al., 2010) force field was used because it has parameters for both proteins and nucleic acids. The models with the ligand were opened in Discovery Studio 4.5 where the bonds and valency of the RIF molecule were corrected based on the PubChem PDB:1I6v entry (National Center for Biotechnology Information, 2019) and subsequently saved as a mol file. The mol file was opened and inspected in Avogadro 1.1.1 and saved as a mol2 file. Given that GROMACS 5.1.4 does not have the required parameters for RIF, the ACPYPE 1.9.3 tool was used with Open Babel 2.3.2 to generate the GROMACS-compatible topology file for the RIF ligand. The long-range cutoffs for the van der Waals interactions and electrostatic forces were set to 1.3 nm. The system was solvated using the rigid extended simple point charge (SPC/E) water model, and the protein's overall charge was neutralized by replacing water molecules with Na^+ and Cl^- counterions. To ensure that the protein model was in a near-native state by preventing high-energy interactions and unwelcomed steric clashes, the energy of the system was minimized up to 50 000 steps of steepest-descent and was terminated when the maximum force was <1000 kJ mol^{-1} nm^{-1}. Following minimization, each system was equilibrated. Firstly, each system was temperature equilibrated in the NVT ensemble with V-rescale coupling set to 300 K for 100ps, and then pressure equilibrated in an NPT ensemble with Parrinello-Rahman barostat and the pressure and temperature set to 1.0 and 300 K for 100ps, respectively. The electrostatic interactions were calculated using the Paritcle Mesh Ewald (PME) algorithm by applying a grid spacing of 0.16 nm and a cutoff distance of 1 nm while using the LINCS method to

restrain the bond length. The 100 ns MD simulations were finally run for both holo and complex systems with trajectories saved every 1000 ps.

3.2.2. Dynamic residue network analysis

MD-TASK tool was employed to generate the DRN data. Two essential residue interaction network metrics were analyzed. That is, change in average shortest path (ΔL) which highlights the accessibility index of residues, and change in *betweenness centrality* (ΔBC), which describes the residues with the highest connectivity within the interaction network when comparing the WT to the mutant. Following calculations, the residues exceeding the mean by 2 and 3 standard deviations were mapped onto a DDRP model. The blue spheres were used to indicate values that were greater in the mutant as compared to the WT, red spheres were used to indicate values that were greater in the WT as compared to the mutant, while a black sphere was used to indicate the position of the SNP.

3.3 Results and discussion

3.3.1. Molecular dynamics simulations

In order to gain a better understanding of the effect of SNPs at codon positions D441, H451 and S456, 100 ns MD simulations for both holo (RIF-free system) and complex system (RIF-system) were successfully conducted for each of the 8 respective SNPs (D441Y, D441V, H451R, H451N, H451D, H451L, S456W, S456L) on the CHPC cluster using the large queue with system provisions of a maximum of 2400 nodes.

3.3.1.a) RMSD analysis

Root mean squared deviation or RMSD which is generally described as the measure of the average distance between the backbone atoms of superimposed proteins over the course of an MD simulation, was determined for both holo and complex systems. RMSD provides valuable insight into the stability of a protein by comparing variable conformations (as a result of a change in the protein of interest's backbone) sampled during the simulation. That is, the more minute the deviations, the more stable the protein structure is said to be. Pertaining to the WT, based on the graphs (Figure 7), it appears that the holo is less stable as compared to the complex. However, the H451D SNP complex appears to attempt to mimic the more stable state displayed by WT complex from around 50 ns into the simulation, a pattern also seen in H451R and H451N SNP complex models, and less so in D441Y, D441V, H451L, S456L and S456W SNP complex models. The D441Y SNP complex registers a higher instability from around 50 ns into the simulation compared to the WT holo and complex, and the D441Y holo. The D441V and S456L SNP complexes appear to exhibit a similar destabilizing effect, unlike the S456W SNP complex which appears to be relatively unstable as compared to the WT complex for a good portion of the simulation but becomes more stable from around 70 ns into the simulation. The H451L SNP complex RMSD does not stray too much from the WT holo, WT complex and H451L holo, although it does become less stable towards the end of the 100 ns simulation. Overall, the SNP complexes appear to be similar in terms of stability to that of the WT holo, indicating shifts in conformation that suggest that the effect of RIF binding becomes less of a factor with regards to the induced conformational shifts that may be required for RIF's efficacy. The SNP holo for a majority of the plots, with the

exception of the S456W and S456L, tends to be as unstable as or more unstable than the WT without RIF.

The distribution of protein conformations throughout the course of the 100 ns trajectory is clearly represented in Figure 8. The WT holo and complex seem to lean towards a unimodal distribution, relaying a single main conformation observed throughout the course of the simulation. The SNP holo and complex models also observe a single main conformation, with the exception of H451L and H451R SNP holo models, and D441V and S456L complex models that lean towards a bimodal distribution, indicating two main conformations. In addition, the figures effectively capture and indicate a higher RMSD observed across the SNP models for both holo and complex systems, a pattern similar to that observed for the WT holo, again (as previously stated) suggesting a shift towards conformations that disregard the presence of RIF, possibly accounting for the driving force behind RIF-resistance.

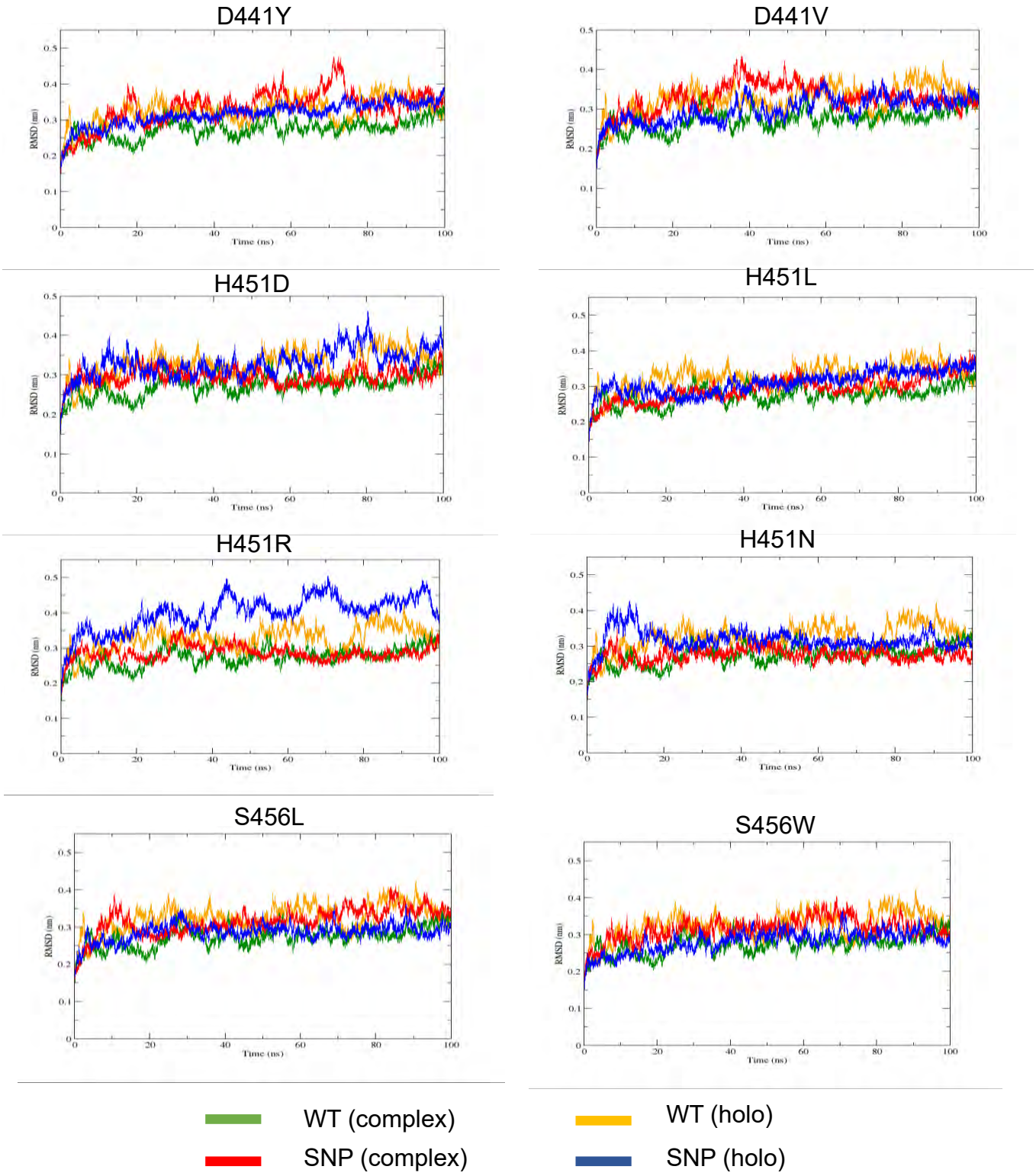


Figure 7: Graphical representation of RMSD over a 100ns trajectory for the WT and respective SNP models viewed with Xmgrace.

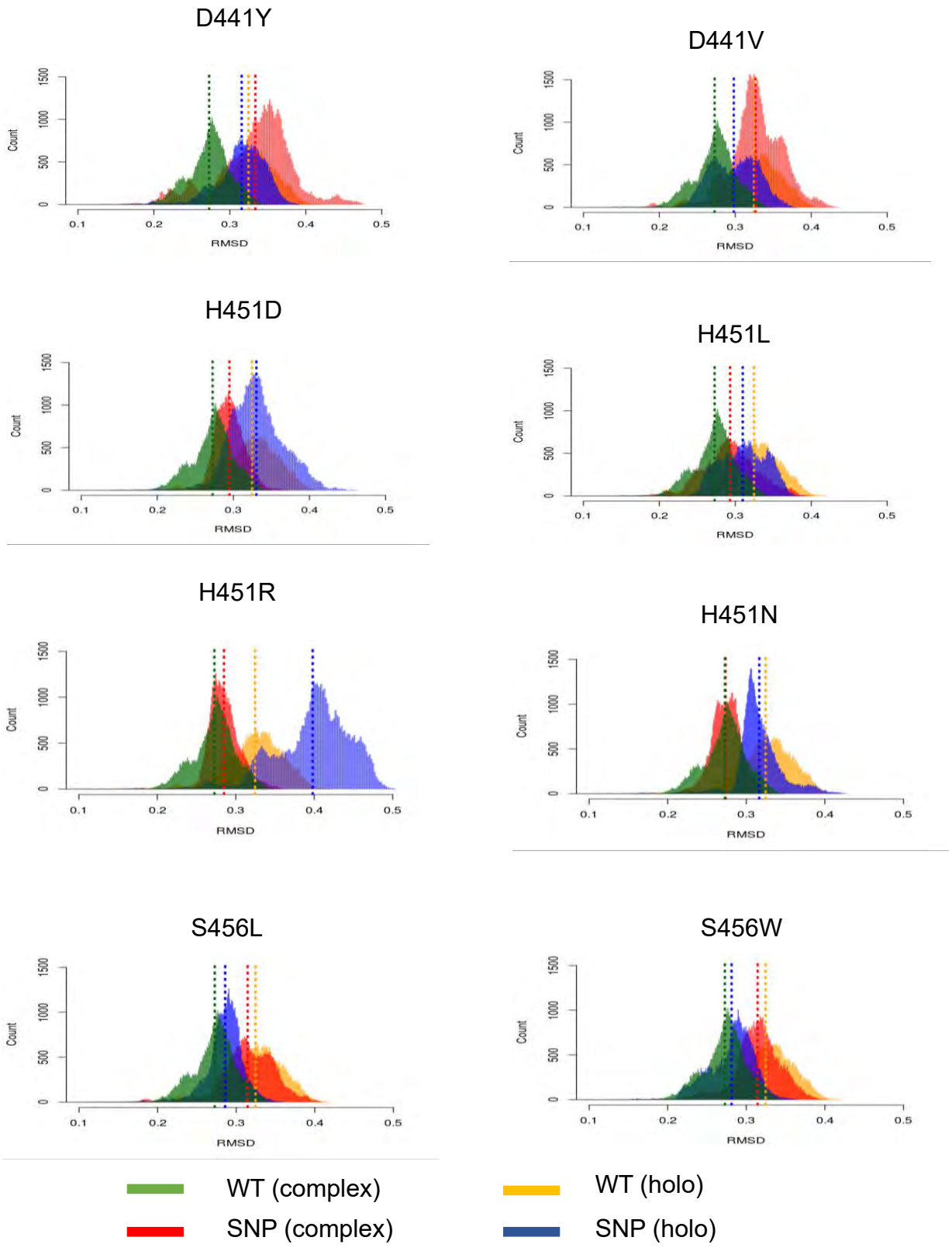


Figure 8: Histogram representation of conformational sampling of DDRP structures over a 100ns trajectory for the WT and respective SNP models as plotted with RStudio. The dotted lines indicate the average RMSD.

3.3.1.b) Rg analysis

Radius of gyration or Rg which describes a molecule's atomic positions with respect to the molecule's center of mass, i.e. relaying a measure of protein compactness, was calculated for the DDRP WT and DDRP mutants over a 100 ns simulation. Figure 9 displays a steady increase in protein compactness throughout the course of the simulation for both holo and complex systems, with the WT observing a relatively higher level of compactness over time compared to the WT holo, the SNP holo and SNP complex models. Exceptions to this trend include the D441Y and D441V holo, and H451R holo which seem to mimic the steady increase in compactness of the WT holo throughout the simulation and from around 50-60 ns into the simulation, respectively. It is also evident that the SNP complexes, with the exception of H451N SNP complex model, are similar to the trend observed in the WT holo, hence suggesting a behavior that favors a state not conducive or promoting of RIF binding.

The plots in Figure 10 show the distribution of populations of holo and complex structures with varying levels of compactness over a 100 ns simulation. By observing the mean values for the models, the unambiguous interpretation (also clearly seen in Figure 9) that the WT complex model has a higher level of compactness compared to the SNP holo and complex models is clear. The exceptions of course, are that of the SNP complex of H451N, displaying a level of compactness even greater than that of the WT complex, a model with the highest registered compactness across the rest of the dataset, and the H451R holo model. This representation of the dataset again provides a better visual of the effect that the presence SNPs have on the expected folding of the DDRP structure.

That is, showing that a majority of the SNP complex models all stray from the WT complex model by either showing a more relaxed state, or showing a more compact state.

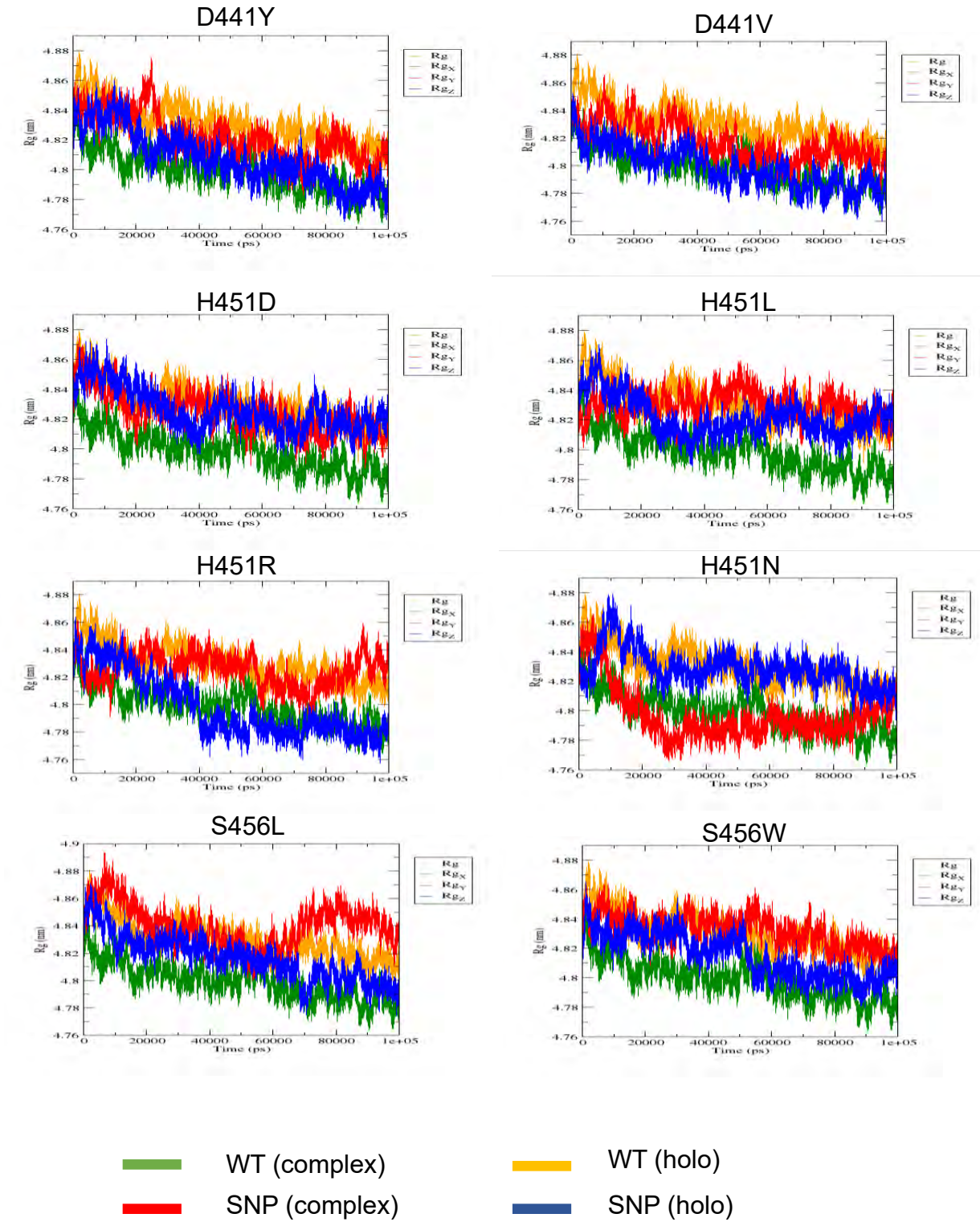


Figure 9: Graphical representation of Rg over the course of 100ns (1000 ps) for the WT and respective SNP models viewed with Xmgrace.

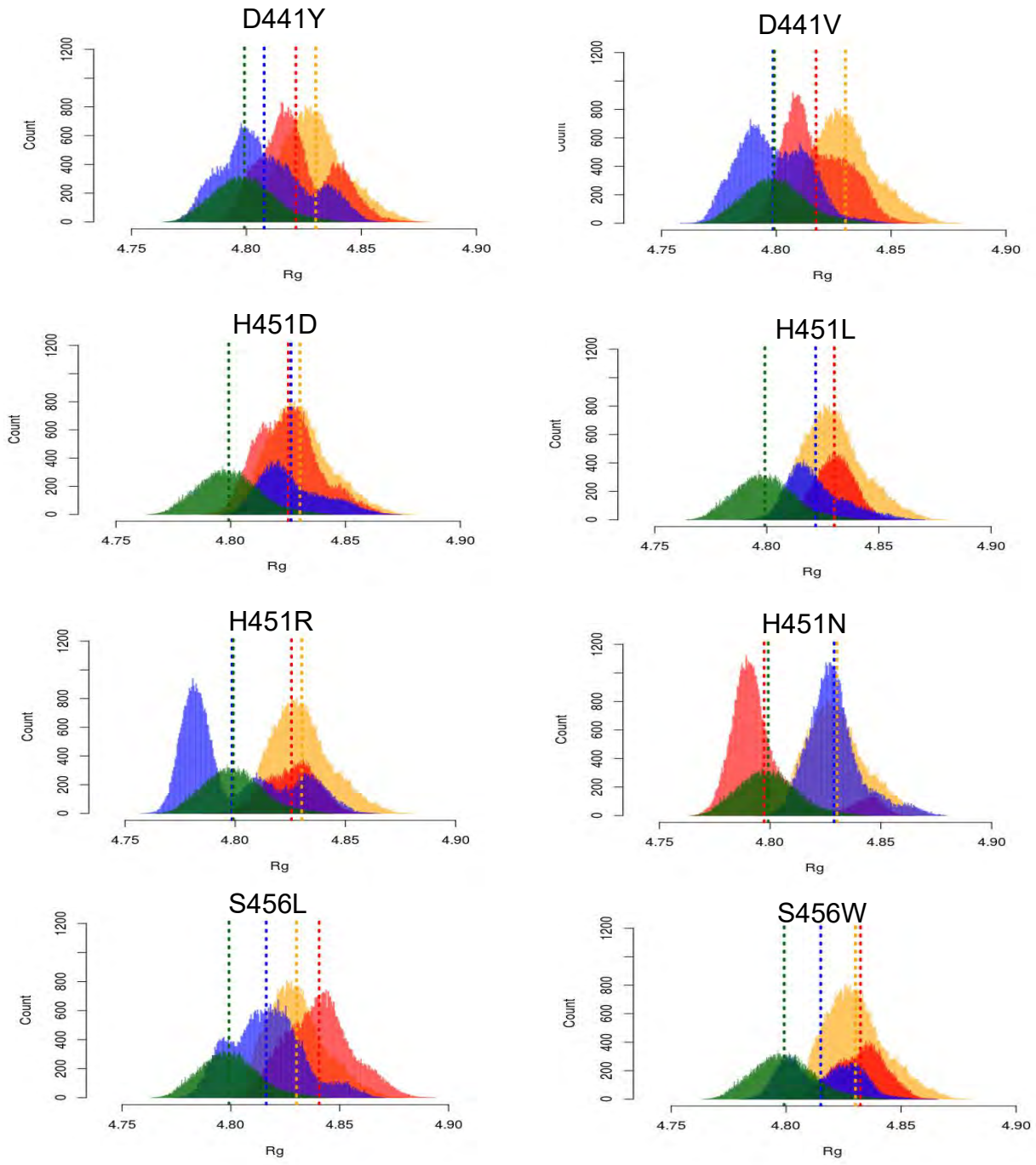


Figure 10: Histogram representation of DDRP structures with varying compactness sampled over a 100ns trajectory for the WT and respective SNP models as plotted with RStudio. The dotted lines indicate the respective average Rg values.

3.3.1.c) RMSF analysis

To observe residue level fluctuations of the WT and SNP models over the 100 ns MD simulation, the RMSF was calculated. The residues were largely showing the same patterns of fluctuation across the whole dataset with only a few regions showing increased levels of fluctuation. These regions, as seen in Figure 11, include inherently flexible loop regions, the N-terminal zinc-binding domain (chain D region 1600-1674 (*MTB* residues 24-98)) said to play a role in termination and antitermination of transcription (King et al., 2004), and the taxon-specific *MTB* β' SI (sequence insertion) α-helical coiled coil (chain D region 1716-1805 (*MTB* residues 140-229)) (Lane and Darst, 2010; Lin et al., 2017), which serves as a "gate", which along with σR1.1 is said to hold dsDNA in the active center cleft for the initial stages of transcription, and is also said to be required for the stability of the open complex (Lin et al., 2017). This increase in flexibility may explain the increased conformational sampling observed in the RMSD plots for the SNP models (Figure 8). The resulting conformations may be leading to a transiently poorer physical obstruction of the 2-3nt RNA transcript by RIF during the initial stages of transcription. These fluctuations may also be the result of the fitness cost of the introduced SNPs being investigated and may thus result in a number of physiological consequences informed by the specific functional roles played by the aforementioned regions.

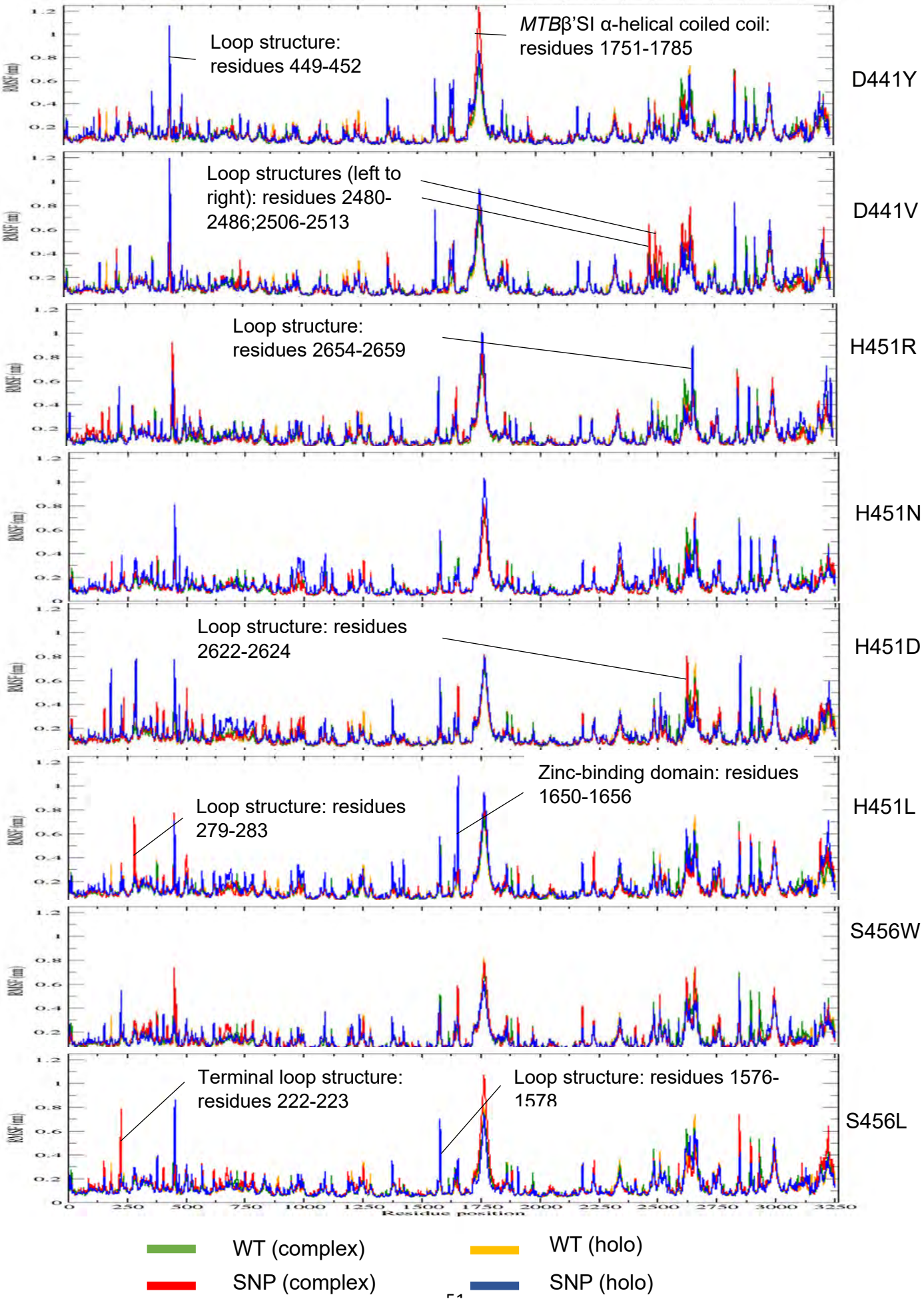


Figure 11: Graphical representation of RMSF over the course of 100ns for the WT and respective SNP models viewed with xmgrace.

3.3.1.d) DRN analysis

i. Average shortest path

Average shortest path or *L* of the WT and SNP models for both holo and complex systems was calculated using a cutoff distance of 6.7 Å to determine residue accessibility using MD frames representative of 5 to 100 ns of the 100 ns MD simulation. Plots (Figure 12 and 13) were generated using R 3.2.3 and RStudio 1.2.1335 to illustrate the change in *L* (Δ*L*) calculated by subtracting respective mutant residue *L* values from WT residue *L* values as shown in a Microsoft Excel sheet (supplementary data). Given the noise level of the indiscriminately visualized residue Δ*L* data that was greater than or less than 2 standard deviations (SD) from the mean for the respective holo and complex datasets (Figure 12), henceforth referred to as 2SD data, and the residues that showed a Δ*L* value greater than or less than 3SD (Figure 13 and 14) from the mean for the respective datasets, henceforth referred to as 3SD data, the 3SD data was therefore used for further analysis.

Pertaining to the holo and complex systems, Tables 3 and 4 (Annex B-1) discriminately lists positive Δ*L* (indicative of a decrease in space between the SNP model residues relative to corresponding WT residues) and negative Δ*L* residues (indicative of an increase in space between SNP model residues relative to corresponding WT residues). Tables 3 and 4 also indicate the most prevalent residues showing significant changes in accessibility (3SD) mapped indiscriminately in Figure 13, along with the differences and commonality observed in the holo and complex systems indiscriminately mapped in Figure 14. The tables further show Δ*L* residues

that would be in contact with RIF within a 7 Å cutoff distance as determined by PyMol 2.4.0a0.

Upon initial observation of the Δ*L* residues for the holo models, as shown in Table 3, it is clear that the negative Δ*L* residues are found to be more than the positive Δ*L* residues, however the positive Δ*L* residues display more commonality as compared to negative Δ*L* residues. Given the wealth of data gathered, and the limited time factor, only Δ*L* residues common in upwards of at least 4 models were investigated further for both holo and complex systems. In addition, the corresponding *MTB* residue positions for the MODELLER residue positions described below were illustrated in Table 9 (Annex C-3).

The following structures show commonality in positive Δ*L* residues for the holo system: chain A loop structure (residue E181) consistent throughout the dataset, chain B loop structure (residue D398) in 4 models (D441Y,D441V,H451L,S456L), chain D loop structure (residues D2519) in close proximity to the N-terminal zinc-binding domain is common among 5 models (D441Y, D441V, H451R, H451D,S456W), and chain D loop structure (residues I2644, S2645, K2646) also common among 5 models (D441V, H451R, H451N,H451D). Regarding the negative Δ*L* residues, only 2 were seen to show consistency among the models. That, is the chain B terminal loop structure (residue I451) marked across 5 models (D441Y, D441V, H451R, H451N, H451L) and the chain D loop structure (residue E1652) which forms part of the N-terminal zinc-binding domain, is marked across 4 models (D441Y, H451N, H451L, S456W).

Regarding the Δ*L* residues displayed in Table 4 (Annex B-1) for the complex models, the number of negative Δ*L* residues is more than the positive Δ*L* residues, suggesting

a bigger change in the structure of the SNP models compared to the WT. The same trend was observed in the holo system. However, in contrast to the holo system, the level of consistent residues observed for negative Δ*L* residues are higher than the positive Δ*L* residues in the complex system. The consistent positive Δ*L* residues are as follows: chain D loop structure (residue E2498) in proximal distance to N-terminal zinc domain and is common in 4 models (D441Y, H451R, H451D, S456W), and chain D loop structure (residue E2645) which is common in 4 models (D441V, H451R, H451N, S456L). The negative Δ*L* residues are as follows: chain B terminal loop structure (residue R255) common in 4 models (H451R, H451N, H451L, S456L), loop structure (residues R401, Q404, R405 and T406, and E403 and D407) common in 4 (H451R,H451D,H451L, S456L) and 5 (H451R, H451N, H451D, H451L, S456L) models respectively, terminal loop structure (residues G450, I451) which is common in 6 models (D441Y, D441Y, H451R, H451N, H451L, S456W), chain D residues forming part of the N-terminal zinc-binding domain (residues C1651, G1655, and C1654) common in 4(H451R, H451D, H451L, S456W; D441V, H451R, H451D,H451L) and 5 models (D441V, H451R, H451D, H451L, S456W) respectively, chain D loop structure (residues E2656, and H2655 and D2657) common in 4(D441V, H451R, H451N, S456W) and 5 models (D441Y, D441V, H451R, H451N, S456W; D441V, H451R, H451N, H451D, S456W), and the chain F loop structure (residues G3151, D3152, G3154) that seems to form part of the interphase residues forming the channel formed by β and β' subunits was found in 7 models (D441Y, H451N, H451D, H451L, S456W, S456L).

The significant positive Δ*L* residues that were observed in both holo and complex systems as indiscriminately mapped in Figure 14 and discriminately outlined in Tables 3 and 4 (Annex B-1) are as follows: D398, I2644, S2645 and K2646; while the negative Δ*L* residues are as follows: R225, R401, E403, Q404, R405, T406, D407, G450, C1651, C1654, G1665, H2655, E2656, D2657, G3151, D3152 and G3154.

The commonality observed for the negative Δ*L* residues suggested similar compensatory motions that may have resulted in a regionally less or more compact binding pocket in the SNP models relative to the WT model. This pattern, albeit globally, is supported by the Rg analysis section. The same conclusion that there may be an altered binding pocket in the SNP models can be drawn when observing Δ*L* residues (both significant and not) for the holo and complex systems (Figure 15). Hence the data appears to support the findings of (Srivastava et al., 2018). As suggested in the (Srivastava et al., 2018) study, the introduction of a SNP may cause a shift in RIF from its native binding site (the binding to which is said to form interactions involving the O9 and O10 of RIF which has been noted as being instrumental in RIF's efficacy in other studies) allowing enough space such that the elongation process of RNA transcription is able to proceed largely uninterrupted even in the presence of RIF, given that apart from the binding stability of RIF, the binding pattern of RIF observed in the WT complex is also essential for its efficacy.

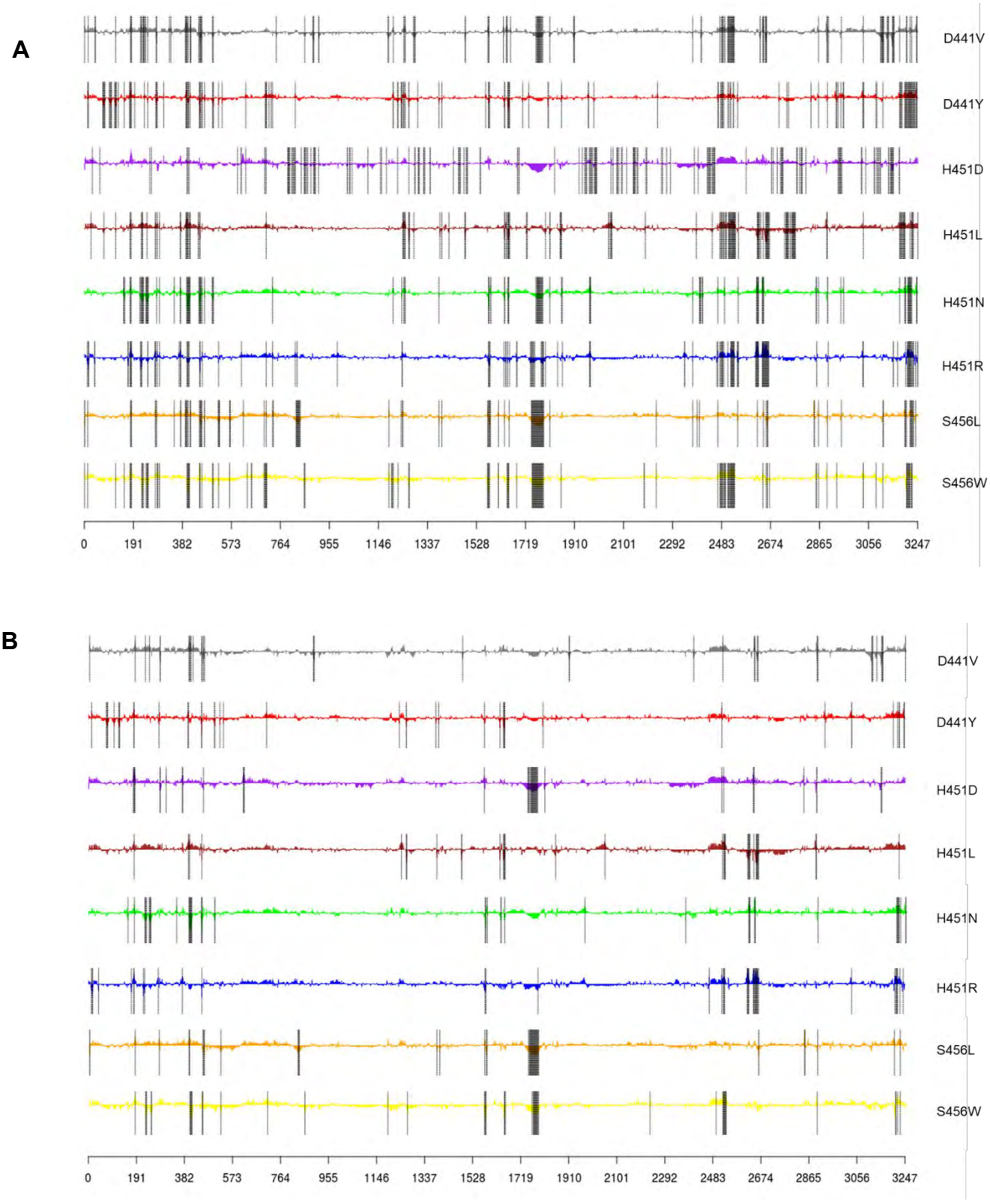


Figure 12: Graphical representation of average ΔL for the holo (RIF-free system). The black dotted lines represent residues with greater than or equal to and less than or equal to **A)**2 and **B)**3 standard deviations from the calculated mean.

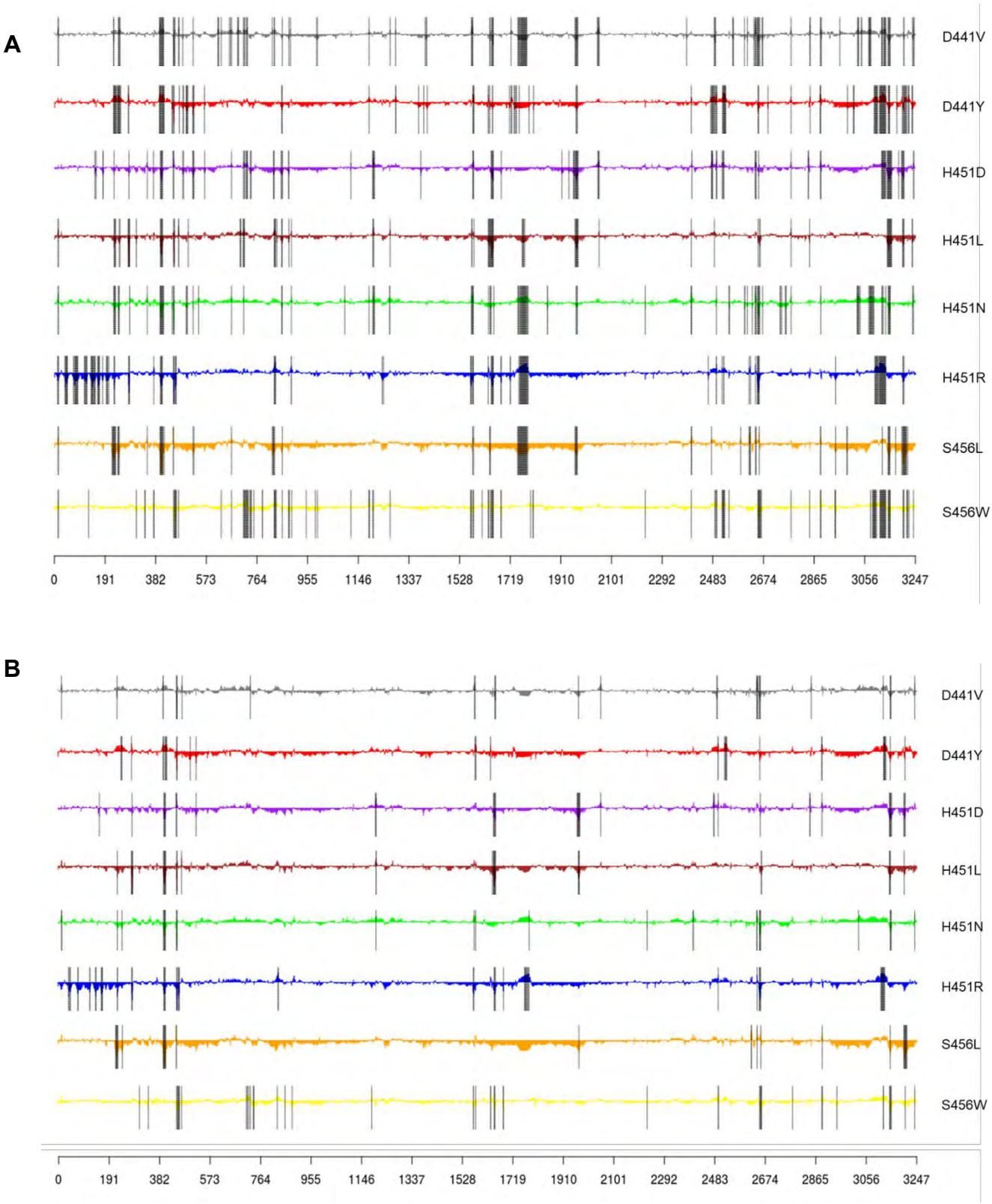


Figure 13: Graphical representation of average ΔL for the complex (RIF system). The black dotted lines represent residues with greater than or equal to and less than or equal to **A)**2 and **B)**3 standard deviations from the calculated mean.

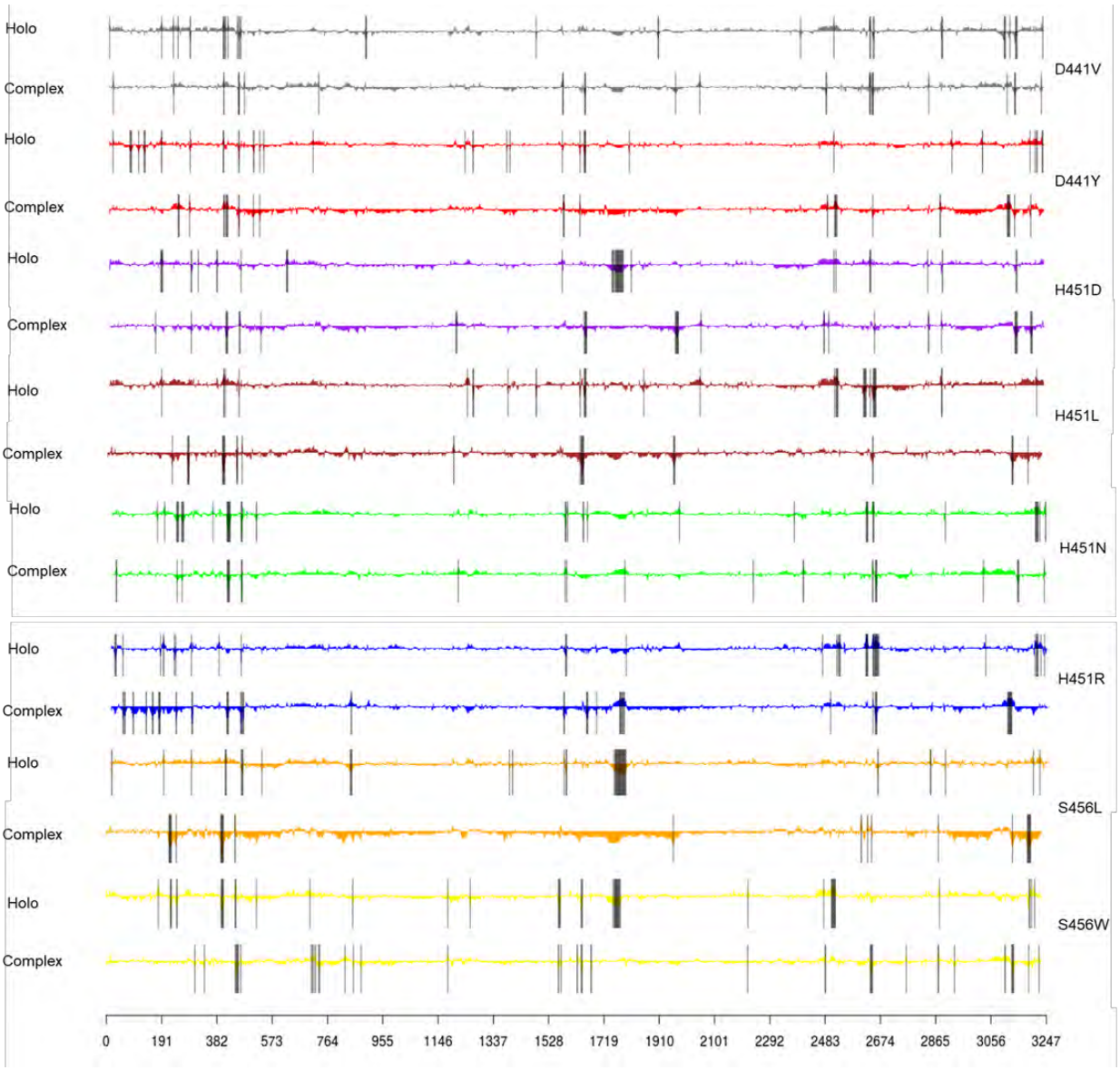


Figure 14: Graphical representation of average ΔL for the holo and complex systems. The black dotted lines represent residues with greater than or equal to and less than or equal to 3 standard deviations from the calculated mean.

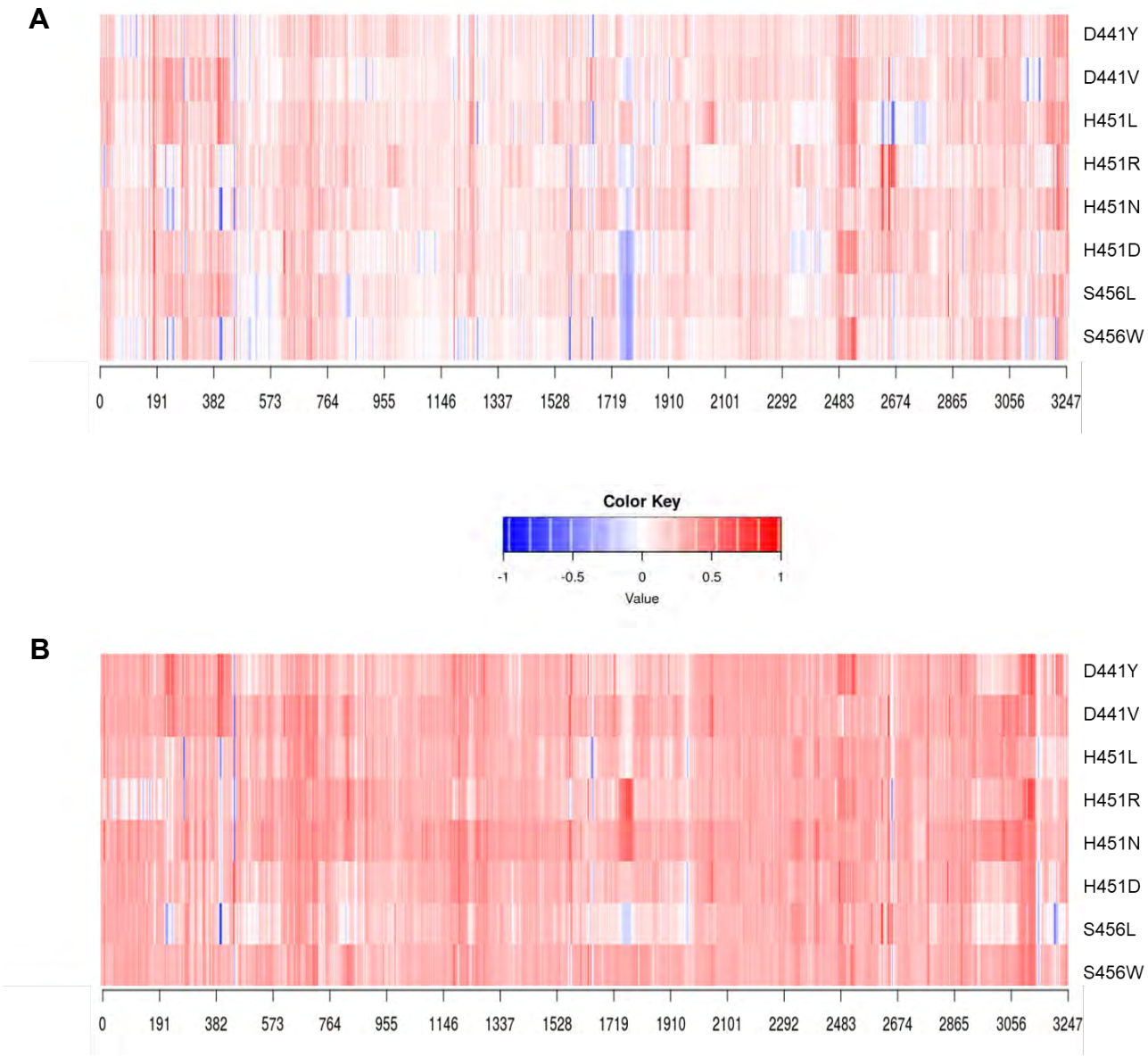


Figure 15: Heat map representation of normalized ΔL for the A) holo and B) complex systems. The red color indicates a greater ΔL for the WT which translates to a shorter distance between the SNP model residues relative to the corresponding WT residues. The blue color indicates a lower ΔL for the WT model which translates to a greater distance between SNP model residues relative to the corresponding WT residues. White indicates no change in distance between SNP model residues and the corresponding WT residues.

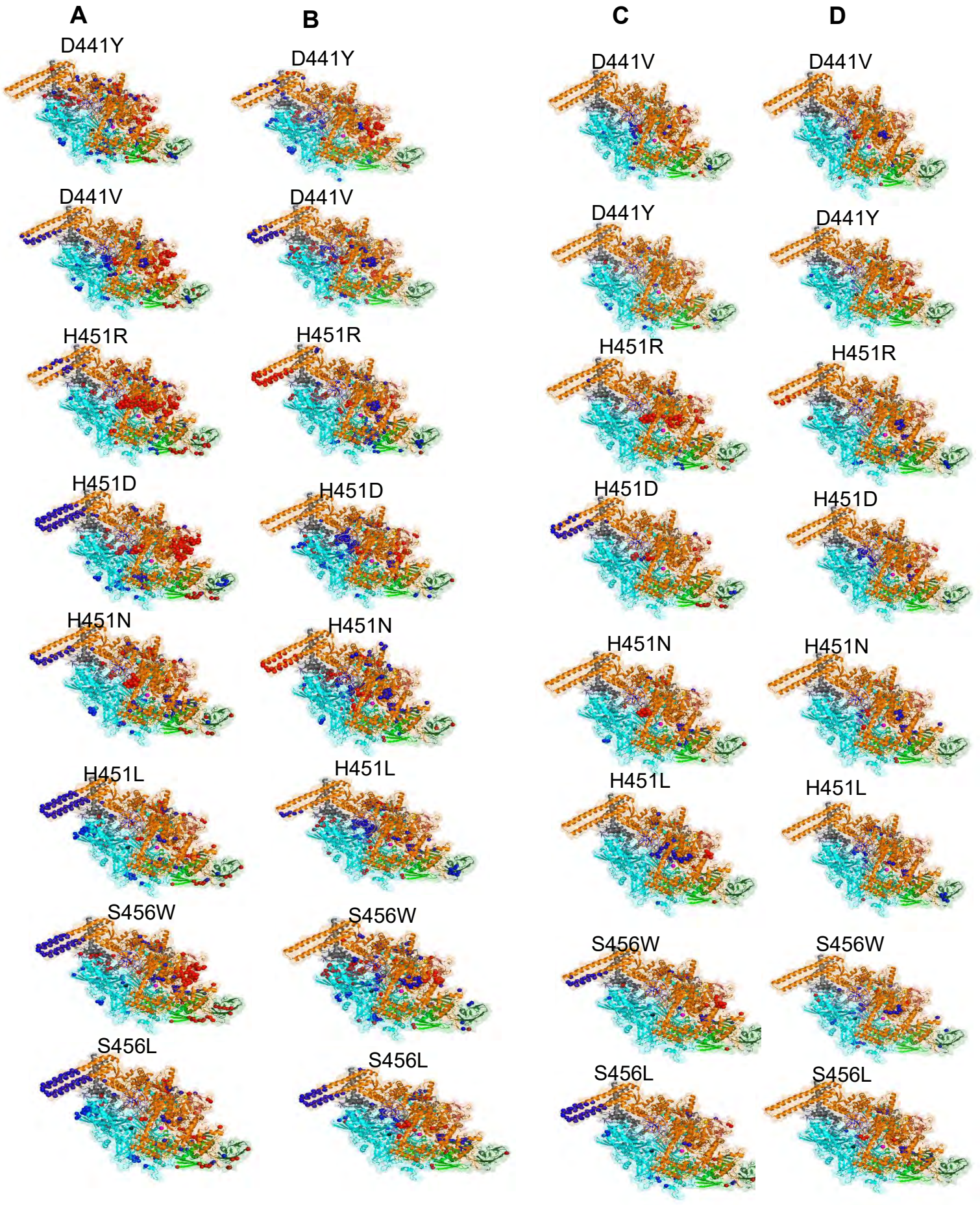


Figure 16: Mapped DDRP models for average ΔL for the A&C) holo and B&D) complex systems.

ii. ∆*BC* analysis

Betweenness centrality or *BC* is a measure of residue importance for protein communication and is calculated by observing the difference in the number of shortest paths passing through a particular node within a protein network. To accomplish this, MD frames representative of 5 to 100 ns of the 100 ns MD simulation were used to calculate the DRN. Change in *BC* or ∆*BC* (calculated by WT-mutant) 2SD and 3SD residues for holo and complex systems are indiscriminately mapped in Figures 17 and 18, respectively, and again in a discriminative manner onto the DDRP protein model (Figure 21). This illustrates the need to opt for the more significant 3SD residues for further study given the noise level, again a consequence of the size of the protein system. The 3SD *BC* residues for the holo were thus compared to the complex (Figure 19) to highlight commonality and differences across the respective holo and complex systems. The 3SD *BC* residues were also noted in Tables 5 and 6 (Annex B-1) for the holo and complex systems, respectively, while heat maps (Figure 20) were generated using the normalized (between -1 and 1) residue *BC* values for the whole protein system.

Upon initial inspection of the 3SD ∆*BC* residues for the holo system (Annex B-1,Table 5), it is clear that there are more positive ∆*BC* residues than there are negative ∆*BC* residues. In other words, there are a greater number of residues that play a significant role in protein communication in the WT holo relative to the SNP holo, indicating a loss of significant ∆*BC* residues involved in protein communication with the introduction of a SNP in the WT holo. It is important to note that most of the identified *BC* residues for both holo and complex systems are interface residues (Tables 7 and

8), as confirmed by observation of the mapped DDRP structures (Figure 21) and using a PyMol script (Annex C-1) highlighting the residues within a 6.7 Å cutoff distance between each of the respective protein chains. The pattern of residue communication displayed a flow of information occurring about the active center channel of DDRP.

Regarding ΔBC residues for the complex system (Table 6) it is clear that the same pattern observed in the holo system (pertaining to the loss of significant ΔBC residues that are important in protein communication) is also observed in the complex structures, further suggesting that the addition of a SNP leads to a loss of interactions in some of the essential coordinating residues that form the RIF binding site, thereby resulting in a decrease or loss of RIF activity.

Looking at the generated heat maps (Figure 20) relaying the ΔBC residues (both significant and not) for the whole protein system, a slightly different overall conclusion can be drawn regarding protein communication. That is, in the holo system, the SNP models appear to have more ΔBC residues than the WT. However, this pattern is not maintained in the complex system. That is, there is again a loss of protein communication observed in the complex system.

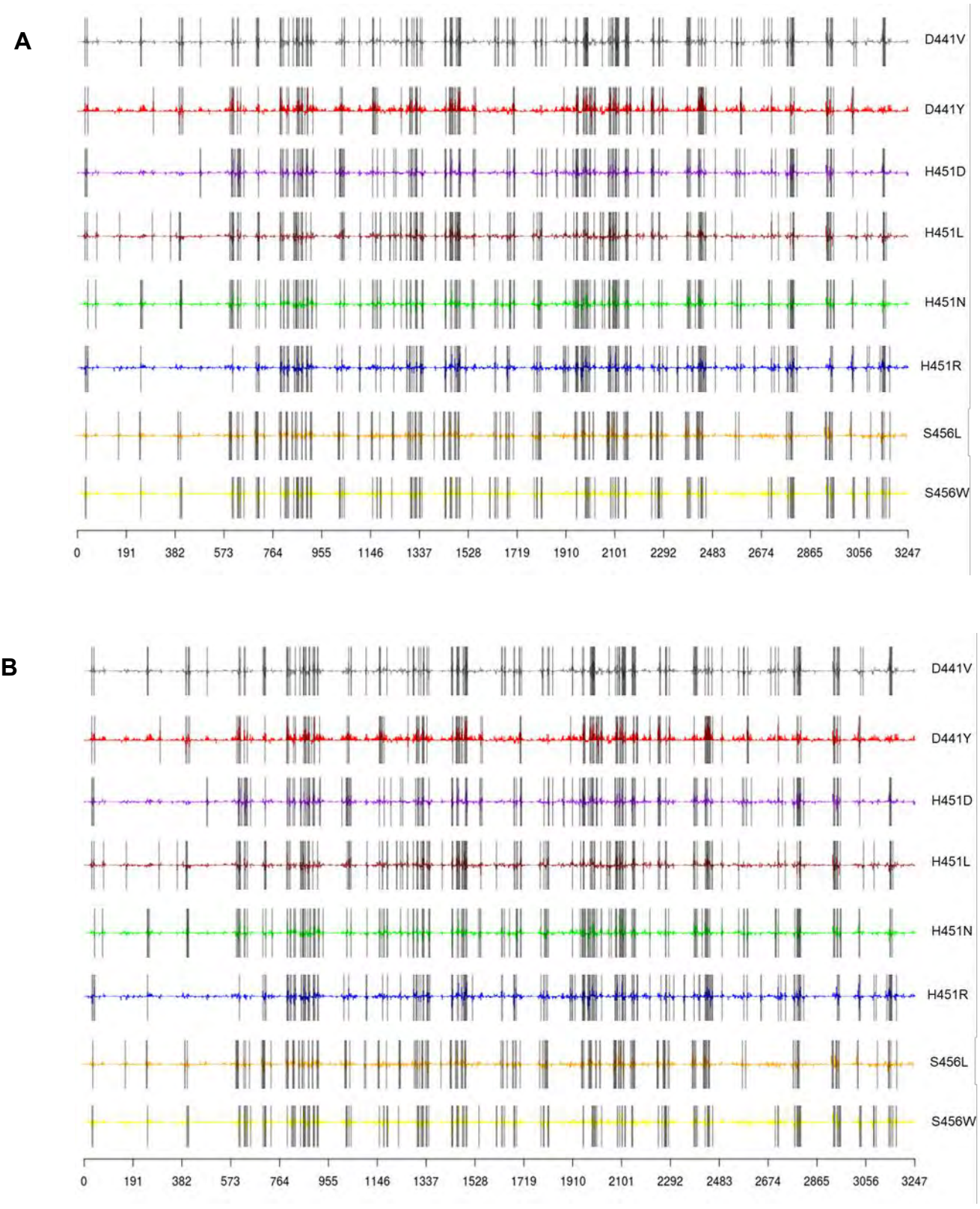


Figure 17: Graphical representation of average ΔBC for the holo system. The black dotted lines represent residues with greater than or equal to and less than or equal to A)2 and B)3 standard deviations from the calculated mean.

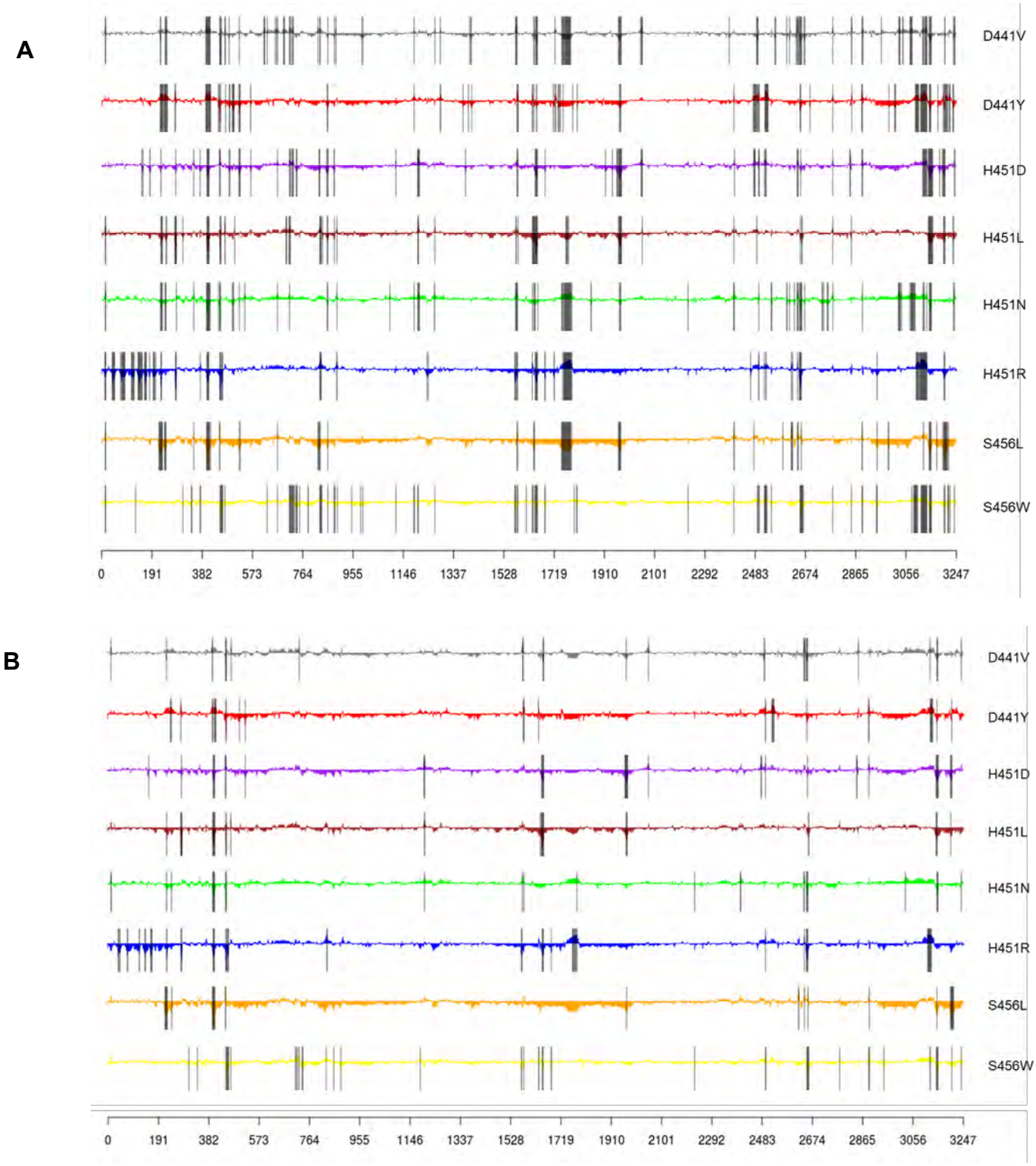


Figure 18: Graphical representation of average ΔBC for the complex system. The black dotted lines represent residues with greater than or equal to and less than or equal to A)2 and B)3 standard deviations from the calculated mean.

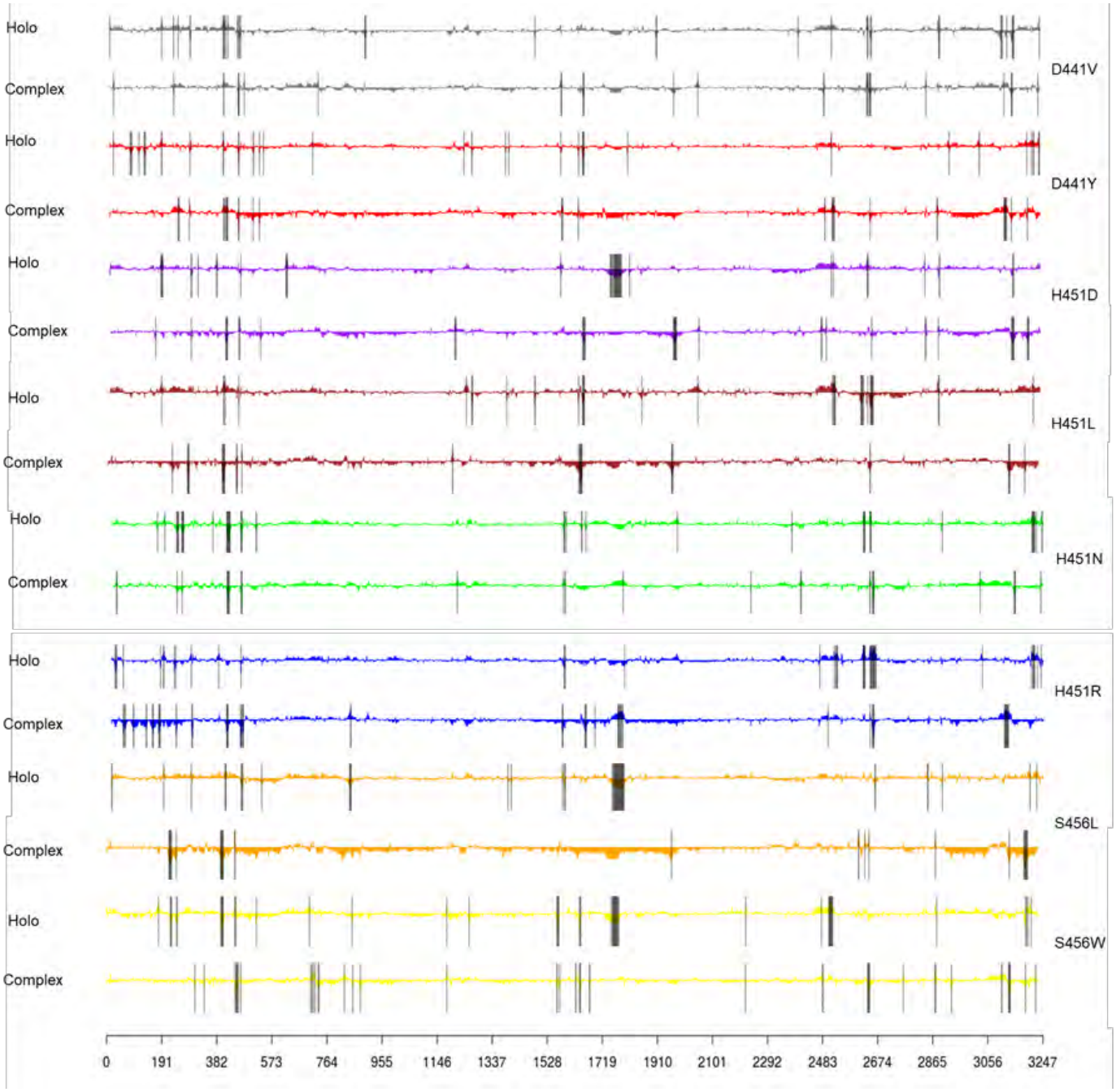


Figure 19: Graphical representation of average ΔBC for the holo and complex systems. The black dotted lines represent residues with greater than or equal to and less than or equal to 3 standard deviations from the calculated mean.

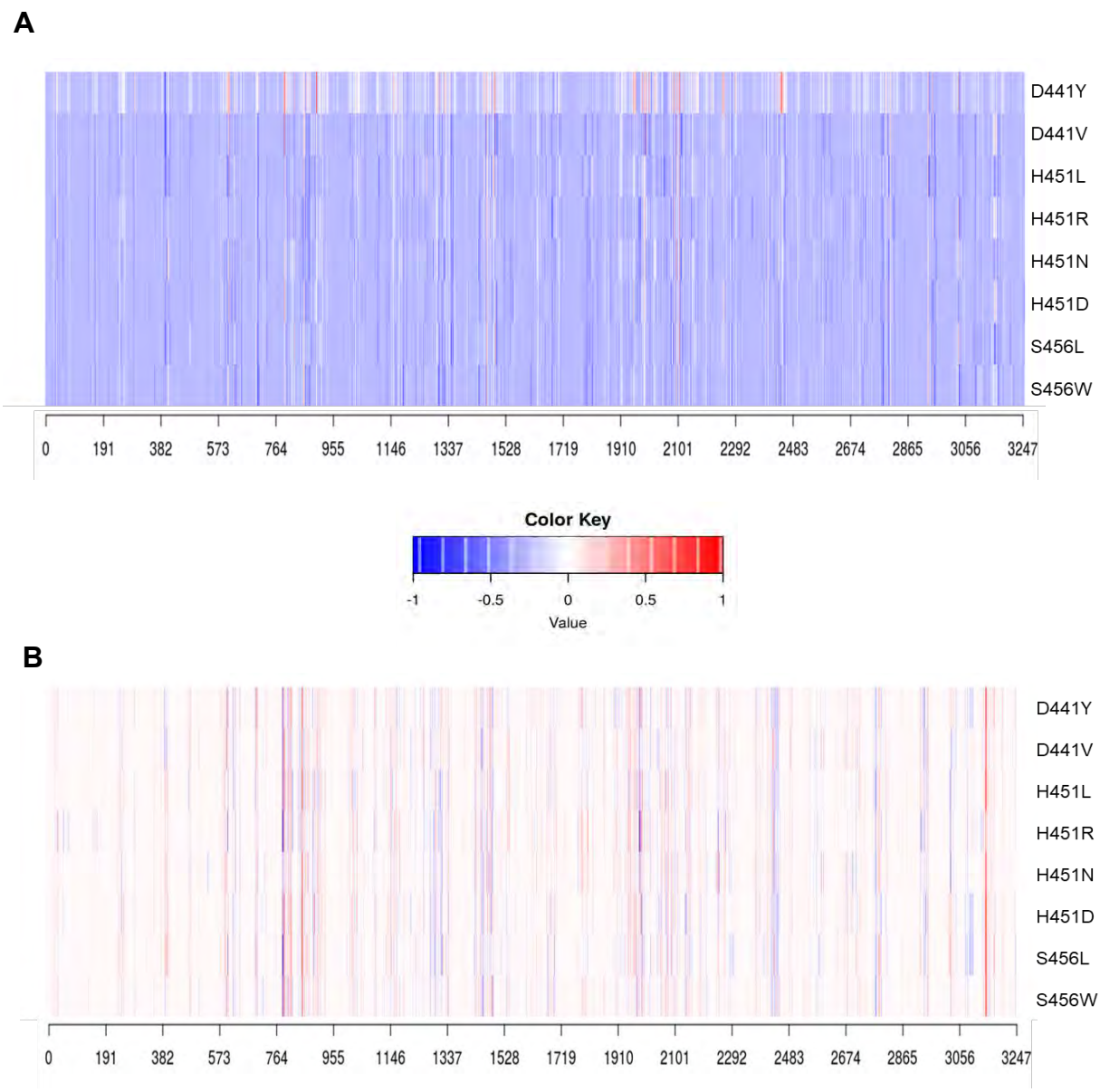


Figure 20: Heat map representation of normalized ΔBC (calculated by WT - mutant) for the A) holo and B) complex systems. The red color represents a higher ΔBC for the WT residue which translates to a greater number of average shortest paths passing through a single node, entailing a corresponding decrease in importance for communication for that particular residue relative to the WT. The blue color represents a lower ΔBC which translates to a lower number of average shortest paths passing through a single node, entailing a loss in communication for the WT residue and an increase in importance for communication in the corresponding SNP residue. White represents no change in importance of a particular residue in communication within the network.

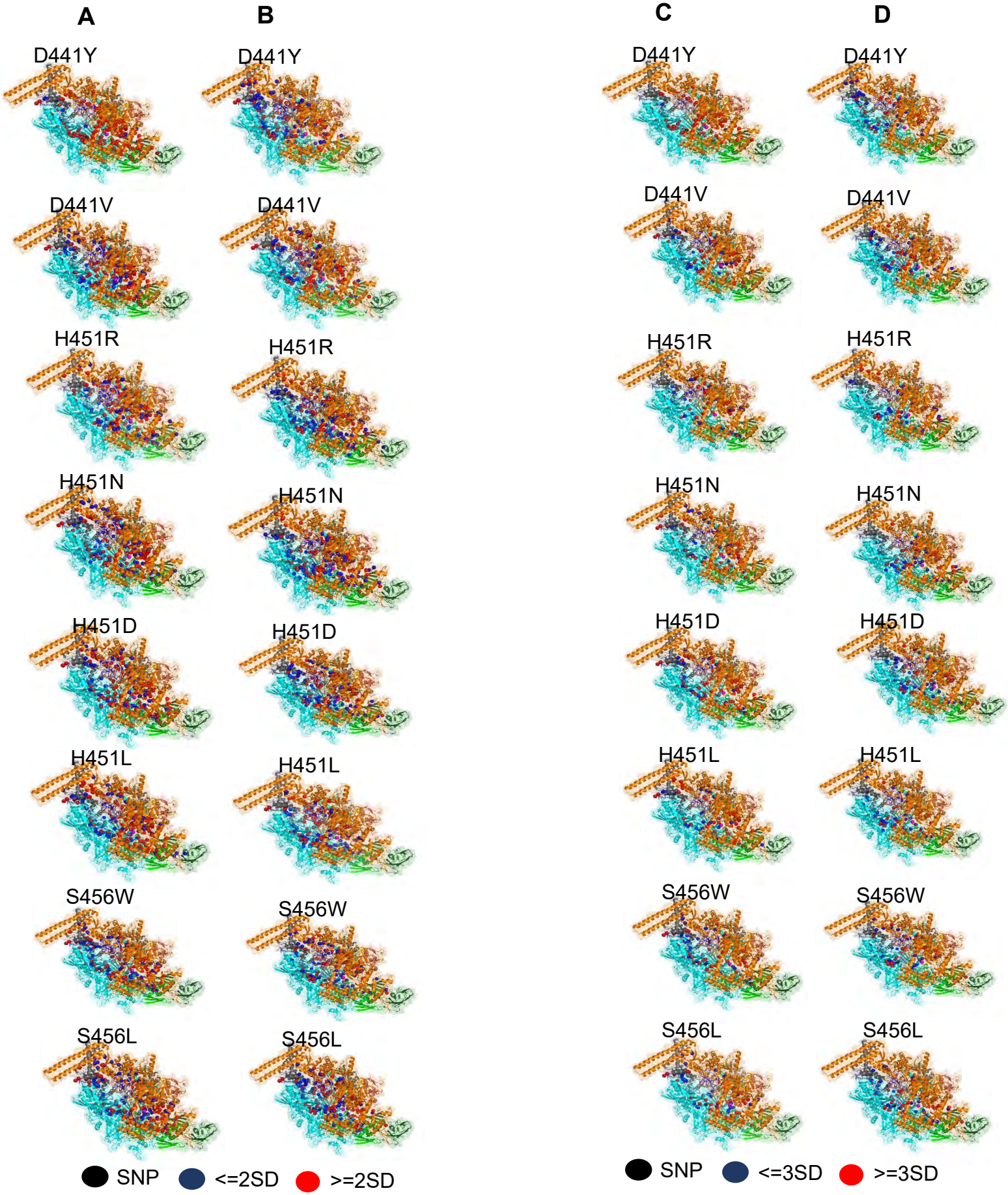


Figure 21: Mapped DDRP models for average Δ*BC* for the A&C) holo and B&D) complex systems.

3.4. Main challenges faced

The complexity of the protein system came with it certain challenges. The following describes some of those challenges and the solutions found to overcome them.

3.4.1. MD runtime

Another issue that arose was the time that the MD run would take to complete on the CHPC cluster. Initially, the CHPC normal queue and its provisioned 240 central processing units (CPUs), 20 queued jobs, and 10 running jobs capped at 48 hours per job was used to run the MDs for both the holo and complex systems. However, given the provision of processors, the MD simulation could hardly reach a 15 ns of the 100 ns run in two days. This after considering the changes from a cubic box type to a triclinic box type and a change in clearance space from 2 to 1.6 nm, respectively, meant to reduce the number of atoms in the system. Hence, it was decided that more nodes would be required to complete the run in a timely fashion. From then on, the MDs would be run on the large queue, using the 2400 processors available. The runs would then take roughly 2 days to complete. The only delay that occurred from this point was that as the year drew closer to the end, and traffic on the cluster was starting to pick up, a representative of CHPC required the halt of all MDs to be run and recommended the use of the Nvidia V100 graphics processing units (GPUs) over the 5^{th} generation Intel CPUs that were being used prior, suggesting it might improve the performance of the runs. The results were however dismal at best, and so the runs were continued on the Intel CPUs. Given the inefficiency of some MD runs, extensions were also required for a number of incomplete runs.

3.4.2. Visualization of MDs

Following the complete generation of the MDs, it was observed that the respective MD files created were around 90-100 gigabytes (GBs) in size. Hence, it was decided that the waters would be removed from each system in an attempt to reduce the number of atoms. This resulted in files ranging between 9 and 10 GBs. The Visual Molecular Dynamics (VMD) program (Humphrey et al., 1996) v1.9.3 was used to visualize the generated MDs for both the holo and complex systems, however, even with the reduced sizes the computers at our disposal were soon overwhelmed by the sheer memory requirements. The decision to split the MDs into two then into three parts seemed to work out, but the task was also somewhat tedious. Hence, the alpha and beta carbons of the MD files, which effectively reduced the file sizes to less than 1 GB, were used for visualization purposes.

3.4.3. DRN calculations

As mentioned above, the DRN analysis was conducted using MD-TASK. Given the inability of the computers at our disposal to handle the memory requirements of the calculations, again a consequence of the size of the protein system, a local cluster was used. However, although the calculations could be conducted concurrently, they took several days to complete for both the holo and complex systems. Once completed, they had produced close to two terabytes of data.

Chapter 4

Final remarks and conclusion

TB remains one of the leading causes of mortality in the modern world and the emergence and continued increase of drug-resistant strains has more than warranted research into combatting said resistance. As already stated before, RIF is among the most potent of the first-line drugs and has been well noted for its broad-spectrum bactericidal properties. It is however especially important in the fight against *MTB* because of its capacity to kill both slow-replicating and non-replicating *MTB* bacilli, making RIF an important drug in the arsenal against *MTB* infections, given that *MTB* has characteristically low metabolic rates, an essential aid in its virulence. Revealing the mechanism of resistance for this drug for clinically relevant mutations found in both the binding region and outside the binding region would be of great importance in finding a solution to the effects of SNPs introduced in the β subunit of the DDRP protein. For the purpose of this study however, only the most frequently occurring SNPs reported by literature to be at codon positions D441, H451 and S456, were investigated for their effects on RIF activity. The listed high confidence mutations (D441Y, D441V, H451D, H451L, H451R, H451N, S456L, and S456W) as reported on TBDReamDB were thus used as the dataset for the study.

The RMSD results revealed that the introduction of RIF leads to a more stable WT protein. It also shows that the introduction of a SNP resulted in varying degrees of conformational sampling, and increased instability across all the SNP complex models relative to the WT complex (even if by a minute amount in the H451 complex models). In addition, the H451

complex models appeared to be more stable than the H451 holo models, a pattern that is in contrast with the rest of the SNP models and suggests a different mode of action with regards to the resistance they impart.

The Rg analysis also revealed a majority of the SNP complex models (with the exception of H451N) registered a level of compactness slightly less than that of the WT holo, suggesting that the core of the SNP complex models was becoming increasingly less spacious relative to the WT complex. It was also confirmed that some of the residues contacting RIF (as confirmed by PyMol using a 6.7Å) were also registering a greater distance between residue pairs within the calculated DRN. Thus, further suggesting a larger than required surface area for the binding and action of RIF. The H451N SNP complex model however, displays a more compact core relative to the WT complex, suggesting a more constricted space for the binding and action of RIF. All these changes may also be leading to different actions in the SNP models compared to WT models as mentioned in a recent study (Srivastava et al., 2018).

With the RMSF analysis, it was observed that mainly loop regions fluctuated more in the SNP models relative to the WT, however, in addition to the loop regions, the highly conserved N-terminal zinc-binding domain, said to be responsible for termination and antitermination of transcription, as well as the *MTB* SI α helical coiled coil, said to play a role in the stability of the DDRP open complex and holding dsDNA in place during the initial stages of transcription, also fluctuate more in the SNP models. It is worth noting that the same regions highlighted in the RMSF results were also highlighted in the *L* results. This was in no way surprising, given that a direct correlation between RMSF and *L* has been confirmed in previous studies (Penkler et al., 2018; Sanyanga et al., 2019).

As stated already, these regions of increased fluctuation, more specifically the zinc-binding domain and the *MTB* SI α helical coiled coil may be the result of the incurred fitness cost of the introduced mutations.

BC results served to indicate that there was loss of residue importance in protein communication across the SNP models relative to the WT models, most of which happened to be interface residues that form the active center channel. This further suggested loss of important residue communication observed in the WT models.

All these results further support the findings of the study conducted by (Srivastava et al., 2018) on *rpoB* mutations and their protein destabilizing effects and hampered efficacy caused by a spatially modified RIF binding pocket leading to unfavorable interactions. The invaluable DRN results in the current study however also effectively highlighted regions that play an essential role in the effects observed with the introduction of the aforementioned SNPs in DDRP.

Further analysis with other tools such as alanine scanning to predict any potentially detrimental mutations, weighted contact maps to observe specific changes in short range residue interactions (among other residue level analytical techniques), and motif analysis coupled with domain level analysis to derive possible functionality in DDRP, and analysis of more accurate models of DDRP through parameterization of the bound metals for those mutations yet to be studied and those that were the focus of the current study, might prove highly beneficial in the fight against RIF-resistance, and will therefore be conducted in future research.

References

Adikaram, C.P., Perera, J. and Wijesundera, S.S., 2012. Geographical profile of rpoB gene mutations in rifampicin resistant Mycobacterium tuberculosis isolates in Sri Lanka. Microbial Drug Resistance, 18(5), pp.525-530.

Agyeman, A.A. and Ofori-Asenso, R., 2017. Tuberculosis—an overview. Journal of Public Health and Emergency, 1(1).

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research, 25(17), pp.3389-3402.

Amita, R. and Geeta, D., 2015. Evaluate" rifampicin resistance" as surrogate marker for rapid detection of MDR-TB using Real-Time PCR directly on FNAC samples of tuberculous lymphadenitis. British Journal of Medicine and Medical Research, 9(5).4.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 22(2), pp.195-201.

Ates, L.S., Houben, E.N. and Bitter, W., 2016. Type VII secretion: a highly versatile secretion system. Microbiology spectrum, 4(1).

Badje, A., Moh, R., Gabillard, D., Guéhi, C., Kabran, M., Ntakpé, J.B., Le Carrou, J., Kouame, G.M., Ouattara, E., Messou, E. and Anzian, A., 2017. Effect of isoniazid preventive therapy on risk of death in west African, HIV-infected adults with high

CD4 cell counts: long-term follow-up of the Temprano ANRS 12136 trial. The Lancet Global Health, 5(11), pp.e1080-e1089.

Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. and Martin, M.J., 2005. The universal protein resource (UniProt). Nucleic acids research, 33(suppl_1), pp.D154-D159.

Bansal, R., Sharma, D. and Singh, R., 2018. Tuberculosis and its treatment: an overview. Mini Reviews in Medicinal Chemistry, 18(1), pp.58-71.

Barry, C.E., Boshoff, H.I., Dartois, V., Dick, T., Ehrt, S., Flynn, J., Schnappinger, D., Wilkinson, R.J. and Young, D., 2009. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nature Reviews Microbiology, 7(12), pp.845-855.

Benkert, P., Biasini, M. and Schwede, T., 2010. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 27(3), pp.343-350.

Benkert, P., Künzli, M. and Schwede, T., 2009. QMEAN server for protein model quality estimation. Nucleic acids research, 37(suppl_2), pp.W510-W514.

Benkert, P., Tosatto, S.C. and Schomburg, D., 2008. QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Structure, Function, and Bioinformatics, 71(1), pp.261-277.

Berman, H.M., Kleywegt, G.J., Nakamura, H. and Markley, J.L., 2014. The Protein Data Bank archive as an open data resource. Journal of computer-aided molecular design, 28(10), pp.1009-1014.

Bernal, J.D. and Crowfoot, D., 1934. X-ray photographs of crystalline pepsin. Nature, 133(3369), p.794.

Bhembe, N.L., Nwodo, U.U., Govender, S., Hayes, C., Ndip, R.N., Okoh, A.I. and Green, E., 2014. Molecular detection and characterization of resistant genes in Mycobacterium tuberculosis complex from DNA isolated from tuberculosis patients in the Eastern Cape province South Africa. BMC infectious diseases, 14(1), p.479.

Binda, G., Domenichini, E., Gottardi, A., Orlandi, B., Ortelli, E., Pacini, B. and Fowst, G., 1971. Rifampicin, a general review. Arzneimittel-forschung, 21(12), pp.1907-1977.

Brito, J. A., & Archer, M. (2013). X-ray Crystallography. Practical Approaches to Biological Inorganic Chemistry, 217–255.

Brooks, B.R., Bruccoleri, R.E. and Olafson, B.D., 1983. States, dj; swaminathan, s.; karplus, m. J. comput. Chem, 4(2), pp.187-217.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K. and Parsons, L.M., 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proceedings of the national academy of Sciences, 99(6), pp.3684-3689.

Brown, D.K., Penkler, D.L., Sheik Amamuddy, O., Ross, C., Atilgan, A.R., Atilgan, C. and Tastan Bishop, Ö., 2017. MD-TASK: a software suite for analyzing molecular dynamics trajectories. Bioinformatics, 33(17), pp.2768-2771.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. and Darst, S.A., 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell, 104(6), pp.901-912.

Case, D.A., Cheatham III, T.E., Darden, T., Gohlke, H., Luo, R., Merz Jr, K.M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J., 2005. The Amber biomolecular simulation programs. Journal of computational chemistry, 26(16), pp.1668-1688.

Calmette, A., Guérin, C., Boquet, A. and Nègre, L., 1927. La vaccination préventive contre la tuberculose par le" BCG,". Masson et cie.

Centers for Disease Control and Prevention (CDC, 2006. Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs--worldwide, 2000-2004. MMWR. Morbidity and mortality weekly report, 55(11), p.301.

CDC [a]. How TB Spreads. CDC; 2016 [August 12 2016]. Available online: https://www.cdc.gov/tb/topic/basics/howtbspreads.htm.

CDC [b]. Chapter 2: Transmission and Pathogenesis of Tuberculosis [August 15 2016]. Available online: https://www.cdc.gov/tb/education/corecurr/pdf/chapter2.pdf.

CDC [c]. Tuberculin Skin Testing. CDC, 2016 [August 15 2016]. Available online: http://www.cdc.gov/tb/publications/factsheets/testing/skintesting.htm.

Chikaonda, T., Ketseoglou, I., Nguluwe, N., Krysiak, R., Thengolose, I., Nyakwawa, F., Rosenberg, N.E., Stanley, C., Mpunga, J., Hoffman, I.F. and Papathanasopoulos, M.A., 2017. Molecular characterisation of rifampicin-resistant Mycobacterium

tuberculosis strains from Malawi. African journal of laboratory medicine, 6(2), pp.1-7.

Churchyard, G.J. and Swindells, S., 2019. Controlling latent TB tuberculosis infection in high-burden countries: A neglected strategy to end TB. PLoS medicine, 16(4), p.e1002787.

Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.3. and Tekaia, F., 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature, 393(6685), pp.537-544.

Colditz, G.A., Brewer, T.F., Berkey, C.S., Wilson, M.E., Burdick, E., Fineberg, H.V. and Mosteller, F., 1994. Efficacy of BCG vaccine in the prevention of tuberculosis: meta-analysis of the published literature. Jama, 271(9), pp.698-702.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G. and Yeboah-Manu, D., 2013. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nature genetics, 45(10), p.1176.

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

Crippen, G.M. and Havel, T.F., 1988. Distance geometry and molecular conformation (Vol. 74). Taunton: Research Studies Press.

Crowley, L.V., 2013. An introduction to human disease: Pathology and pathophysiology correlations. Jones & Bartlett Publishers. pp390.

Cruz-Knight, W. and Blake-Gumbs, L., 2013. Tuberculosis: an overview. Primary Care: Clinics in Office Practice, 40(3), pp.743-756.

Daffé, M. and Draper, P., 1997. The envelope layers of mycobacteria with reference to their pathogenicity. In Advances in microbial physiology (Vol. 39, pp. 131-203). Academic Press.

da Silva, A.W.S. and Vranken, W.F., 2012. ACPYPE-Antechamber python parser interface. BMC research notes, 5(1), p.367.

Dessault Systemes. 2017. BIOVIA Discovery Studio® Visualizer Datasheet. Dessault Systemes. Accessed 17 November 2019. <https://www.3dsbiovia.com/products/datasheets/discovery-studio-visualizer.pdf>.

Dessault Systemes. 2019. Biovia Discovery Studio Comprehensive Modeling and Simulations for Life Sciences Datasheet. Dessault Systemes. Accessed 17 November 2019. <https://www.3dsbiovia.com/products/datasheets/biovia-discovery-studio-overview.pdf>.

Dheda, K., Gumbo, T., Maartens, G., Dooley, K.E., McNerney, R., Murray, M., Furin, J., Nardell, E.A., London, L., Lessem, E. and Theron, G., 2017. The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. The lancet Respiratory medicine, 5(4), pp.291-360.

Dookie, N., Sturm, A.W. and Moodley, P., 2016. Mechanisms of first-line antimicrobial resistance in multi-drug and extensively drug resistant strains of Mycobacterium tuberculosis in KwaZulu-Natal, South Africa. BMC infectious diseases, 16(1), p.609.

Eramian, D., Eswar, N., Shen, M.Y. and Sali, A., 2008. How well can the accuracy of comparative protein structure models be predicted?. Protein Science, 17(11), pp.1881-1893.

Eswar, N., Eramian, D., Webb, B., Shen, M.Y. and Sali, A., 2008. Protein structure modeling with MODELLER. In Structural proteomics (pp. 145-159). Humana Press.

Frieden, T.R., Sterling, T.R. and Munsiff, S.S., 2003. 1. Watt CJ, Dye C. Tuberculosis. Lancet, 362, pp.887-99. Wehrli, W., Handschin, J. and Wunderli, W., 1976. Interaction between rifampicin and DNA-dependent RNA polymerase of E. coli. Cold Spring Harbor Monograph Archive, 6, pp.397-412.

Global, W.H.O., 2017. tuberculosis report 2017. Geneva, Switzerland: World Health Organization.

Gordon, S.V. and Parish, T., 2018. Microbe profile: Mycobacterium tuberculosis: humanity's deadly microbial foe. Microbiology, 164(4), pp.437-439.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E. and Hutchison, G.R., 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of cheminformatics, 4(1), p.17.

Hartmann, G., 1967. The specific inhibition of DNA-directed RNA synthesis by rifamycin. Biochem. Biophys. Acta, 145, pp.843-844.19. Heil, A. and Zillig, W., 1970. Reconstitution of bacterial DNA-dependent RNA-polymerase from isolated subunits as a tool for the elucidation of the role of the subunits in transcription. FEBS letters, 11(3), pp.165-168.

Hatherley, R., Brown, D.K., Glenister, M. and Bishop, Ö.T., 2016. PRIMO: An interactive homology modeling pipeline. PloS one, 11(11), p.e0166698.

Havel, T.F. and Snow, M.E., 1991. A new method for building protein conformations from sequence alignments with homologues of known structure. Journal of molecular biology, 217(1), pp.1-7.

Havel, T.F., Kuntz, I.D. and Crippen, G.M., 1983. The theory and practice of distance geometry. Bulletin of Mathematical Biology, 45(5), pp.665-720.

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C., 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics, 65(3), pp.712-725.

Humphrey, W., Dalke, A. and Schulten, K., 1996. VMD: visual molecular dynamics. Journal of molecular graphics, 14(1), pp.33-38.

Illergård, K., Ardell, D.H. and Elofsson, A., 2009. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. Proteins: Structure, Function, and Bioinformatics, 77(3), pp.499-508.

International Union of Crystallography (1989) Acta Cryst A45:658.

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.

Johnston, D.E. and McClure, W.R., 1976. Abortive initiation of in vitro RNA synthesis on bacteriophage λ DNA. Cold Spring Harbor Monograph Archive, 6, pp.413-428.

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.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H. and Xu, J., 2012. Template-based protein structure modeling using the RaptorX web server. Nature protocols, 7(8), p.1511.

Kandasamy, R., Voysey, M., McQuaid, F., de Nie, K., Ryan, R., Orr, O., Uhlig, U., Sande, C., O'Connor, D. and Pollard, A.J., 2016. Non-specific immunological effects of selected routine childhood immunisations: systematic review. Bmj, 355, p.i5225.

Kapopoulou, A., Lew, J.M. and Cole, S.T., 2011. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis, 91(1), pp.8-13.

Keshavjee, S. and Farmer, P.E., 2012. Tuberculosis, drug resistance, and the history of modern medicine. New England Journal of Medicine, 367(10), pp.931-936.

Kessler, C., Huaifeng, M.I. and Hartmann, G.R., 1982. Competition of rifampicin with binding of substrate and RNA to RNA polymerase. European journal of biochemistry, 122(3), pp.515-518.

King, R.A., Markov, D., Sen, R., Severinov, K. and Weisberg, R.A., 2004. A conserved zinc binding domain in the largest subunit of DNA-dependent RNA polymerase modulates intrinsic transcription termination and antitermination but does not stabilize the elongation complex. Journal of molecular biology, 342(4), pp.1143-1154.

Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O. and Shaw, D.E., 2009. Long-timescale molecular dynamics simulations of protein structure and function. Current opinion in structural biology, 19(2), pp.120-127.

Kleywegt, G.J. and Jones, T.A., 1997. [11] Model building and refinement practice. In Methods in enzymology (Vol. 277, pp. 208-230). Academic Press.

Koch, A., Mizrahi, V. and Warner, D.F., 2014. The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin?. Emerging microbes & infections, 3(1), pp.1-11.

Lane, W.J. and Darst, S.A., 2010. Molecular evolution of multisubunit RNA polymerases: sequence analysis. Journal of molecular biology, 395(4), pp.671-685.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography, 26(2), pp.283-291.

Laskowski, R.A., MacArthur, M.W. and Thornton, J.M., 2006. PROCHECK: validation of protein-structure coordinates.

Lai, C., Xu, J., Tozawa, Y., Okamoto-Hosoya, Y., Yao, X. and Ochi, K., 2002. Genetic and physiological characterization of rpoB mutations that activate antibiotic production in Streptomyces lividans. Microbiology, 148(11), pp.3365-3373.

Lee, J.Y., 2015. Diagnosis and treatment of extrapulmonary tuberculosis. Tuberculosis and respiratory diseases, 78(2), pp.47-55.

Lin, W., Mandal, S., Degen, D., Liu, Y., Ebright, Y.W., Li, S., Feng, Y., Zhang, Y., Mandal, S., Jiang, Y. and Liu, S., 2017. Structural basis of Mycobacterium tuberculosis transcription and transcription inhibition. Molecular cell, 66(2), pp.169-179.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. and Shaw, D.E., 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Structure, Function, and Bioinformatics, 78(8), pp.1950-1958.

Long, R., 2000. Drug-resistant tuberculosis. Cmaj, 163(4), pp.425-428.

Lüthy, R., Bowie, J.U. and Eisenberg, D., 1992. Assessment of protein models with three-dimensional profiles. Nature, 356(6364), p.83.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C. and Stover, C.K., 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. Journal of bacteriology, 178(5), pp.1274-1282.

Mangtani, P., Abubakar, I., Ariti, C., Beynon, R., Pimpin, L., Fine, P.E., Rodrigues, L.C., Smith, P.G., Lipman, M., Whiting, P.F. and Sterne, J.A., 2013. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. Clinical infectious diseases, 58(4), pp.470-480.

Mangtani, P., Nguipdop-Djomo, P., Keogh, R.H., Trinder, L., Smith, P.G., Fine, P.E., Sterne, J., Abubakar, I., Vynnycky, E., Watson, J. and Elliman, D., 2017. Observational study to estimate the changes in the effectiveness of bacillus Calmette-Guerin (BCG) vaccination with time since vaccination for preventing tuberculosis in the UK. Health technology assessment, 21(39), pp.7-7.

McClure, W.R. and Cech, C.L., 1978. On the mechanism of rifampicin inhibition of RNA synthesis. Journal of Biological Chemistry, 253(24), pp.8949-8956.

Misura, K.M. and Baker, D., 2005. Progress and challenges in high-resolution refinement of protein structure models. Proteins: Structure, Function, and Bioinformatics, 59(1), pp.15-29.

Misura, K.M., Chivian, D., Rohl, C.A., Kim, D.E. and Baker, D., 2006. Physically realistic homology models built with ROSETTA can be more accurate than their templates. Proceedings of the National Academy of Sciences, 103(14), pp.5361-5366.

Molodtsov, V., Scharf, N.T., Stefan, M.A., Garcia, G.A. and Murakami, K.S., 2017. Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in Mycobacterium tuberculosis. Molecular microbiology, 103(6), pp.1034-1045.

Moodley, P., Shah, N.S., Tayob, N., Connolly, C., Zetola, N., Gandhi, N., Friedland, G. and Sturm, A.W., 2011. Spread of extensively drug-resistant tuberculosis in KwaZulu-Natal province, South Africa. PloS one, 6(5), p.e17513. 28. Munro, S.A., Lewin, S.A., Smith, H.J., Engel, M.E., Fretheim, A. and Volmink, J., 2007. Patient adherence to tuberculosis treatment: a systematic review of qualitative research. PLoS medicine, 4(7), p.e238.

Nathan, C.F. and Hibbs Jr, J.B., 1991. Role of nitric oxide synthesis in macrophage antimicrobial activity. Current opinion in immunology, 3(1), pp.65-70.

National Center for Biotechnology Information. PubChem Database. Rifampicin, CID=135398735, https://pubchem.ncbi.nlm.nih.gov/compound/Rifampicin (accessed on Nov. 14, 2019)

Nayak, S. and Acharjya, B., 2012. Mantoux test and its interpretation. Indian dermatology online journal, 3(1), p.2.

Nguipdop-Djomo, P., Heldal, E., Rodrigues, L.C., Abubakar, I. and Mangtani, P., 2016. Duration of BCG protection against tuberculosis and change in effectiveness with time since vaccination in Norway: a retrospective population-based cohort study. The Lancet infectious diseases, 16(2), pp.219-226.

Nieuwenhuizen, N.E. and Kaufmann, S.H., 2018. Next-Generation Vaccines Based on Bacille Calmette–Guérin. Frontiers in immunology, 9, p.121.

Notredame, C., Higgins, D.G. and Heringa, J., 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology, 302(1), pp.205-217.

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. and Hutchison, G.R., 2011. Open Babel: An open chemical toolbox. Journal of cheminformatics, 3(1), p.33.

Parish, T. and Brown, A.C. eds., 2009. Mycobacterium: genomics and molecular biology. Horizon Scientific Press. pp98.

Petrey, D. and Honig, B., 2005. Protein structure prediction: inroads to biology. Molecular cell, 20(6), pp.811-819.

Penkler, D.L., Atilgan, C. and Tastan Bishop, O., 2018. Allosteric Modulation of Human Hsp90α Conformational Dynamics. Journal of chemical information and modeling, 58(2), pp.383-404.

Piccaro, G., Giannoni, F., Filippini, P., Mustazzolu, A. and Fattorini, L., 2013. Activities of drug combinations against Mycobacterium tuberculosis grown in aerobic and hypoxic acidic conditions. Antimicrobial agents and chemotherapy, 57(3), pp.1428-1433.

Protein Data Bank Nat. New Biol., 233 (1971), p. 223.

Qazi, O., Rahman, H., Tahir, Z., Qasim, M., Khan, S., Anjum, A.A., Yaqub, T., Tayyab, M., Ali, N. and Firyal, S., 2014. Mutation pattern in rifampicin resistance determining region of rpoB gene in multidrug-resistant Mycobacterium tuberculosis isolates from Pakistan. International journal of mycobacteriology, 3(3), pp.173-177. 33.

Rabussay, D. and Zillig, W., 1969. A rifampicin resistent rna-polymerase from E. coli altered in the β-subunit. FEBS letters, 5(2), pp.104-106.

Ramaswamy, S. and Musser, J.M., 1998. Molecular genetic basis of antimicrobial agent resistance inMycobacterium tuberculosis: 1998 update. Tubercle and Lung disease, 79(1), pp.3-29.

Rienthong, D.R., 2017. Rapid detection for early appearance of rifampin and isoniazid resistance in Mycobacterium tuberculosis. Siriraj Medical Journal, 61(1), pp.49-55.

Rittershaus, E.S., Baek, S.H. and Sassetti, C.M., 2013. The normalcy of dormancy: common themes in microbial quiescence. Cell host & microbe, 13(6), pp.643-651.

Rodrigues, L.C., Diwan, V.K. and Wheeler, J.G., 1993. Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-analysis. International journal of epidemiology, 22(6), pp.1154-1158.

Roy, A., Eisenhut, M., Harris, R.J., Rodrigues, L.C., Sridhar, S., Habermann, S., Snell, L., Mangtani, P., Adetifa, I., Lalvani, A. and Abubakar, I., 2014. Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. Bmj, 349, p.g4643.

Sala, C., Dhar, N., Hartkoorn, R.C., Zhang, M., Ha, Y.H., Schneider, P. and Cole, S.T., 2010. Simple model for testing drugs against nonreplicating Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy, 54(10), pp.4150-4158.

Šali and T. L. Blundell, 1993. Comparative protein modelling by satisfaction of spatial restraints.J. Mol. Biol.234,779–815.

Sánchez, R. and Šali, A., 1997. Evaluation of comparative protein structure modeling by MODELLER-3. Proteins: Structure, Function, and Bioinformatics, 29(S1), pp.50-58.

Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B.K., Church, G.M. and Murray, M.B., 2009. Tuberculosis drug resistance mutation database. PLoS medicine, 6(2), p.e1000002.

Sanyanga, T.A., Nizami, B. and Tastan Bishop, Ö., 2019. Mechanism of Action of Non-Synonymous Single Nucleotide Variations Associated with α-Carbonic Anhydrase II Deficiency. Molecules, 24(21), p.3987.

Schulz, W. and Zillig, W., 1981. Rifampidn inhibition of RNA synthesis by destabilisation of DNA-RNA polymerase-oligonucleotide-complexes. Nucleic acids research, 9(24), pp.6889-6906.

Scitable. "Scitable by Nature Education". 2014 [July 6 2014], Available online: <http://www.nature.com/scitable/definition/single-nucleotide-polymorphism-snp-295>

Sensi, P., Margalith, P. and Timbal, M.T., 1959. Rifomycin, a new antibiotic; preliminary report. Il Farmaco; edizione scientifica, 14(2), p.146.38. Shah, N.S., Wright, A., Bai, G.H., Barrera, L., Boulahbal, F., Martín-Casabona, N., Drobniewski, F., Gilpin, C., Havelková, M., Lepe, R. and Lumb, R., 2007. Worldwide emergence of extensively drug-resistant tuberculosis. Emerging infectious diseases, 13(3), p.380.

Shen, M.Y. and Sali, A., 2006. Statistical potential for assessment and prediction of protein structures. Protein science, 15(11), pp.2507-2524.

Silva Miranda, M., Breiman, A., Allain, S., Deknuydt, F. and Altare, F., 2012. The tuberculous granuloma: an unsuccessful host defence mechanism providing a safety shelter for the bacteria?. Clinical and Developmental Immunology, 2012.

Söding, J., Biegert, A. and Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic acids research, 33(suppl_2), pp.W244-W248.

Srinivasan, N. and Blundell, T.L., 1993. An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. Protein Engineering, Design and Selection, 6(5), pp.501-512.

Srivastava, G., Tripathi, S., Kumar, A. and Sharma, A., 2018. Molecular insight into multiple RpoB clinical mutants of Mycobacterium tuberculosis: An attempt to probe structural variations in rifampicin binding site underlying drug resistance. International journal of biological macromolecules, 120, pp.2200-2214.

Talbot, E.A., Perkins, M.D., Silva, S.F.M. and Frothingham, R., 1997. Disseminated bacille Calmette-Guerin disease after vaccination: case report and review. Clinical Infectious Diseases, 24(6), pp.1139-1146.

Thillai, M., Pollock, K., Pareek, M. and Lalvani, A., 2014. Interferon-gamma release assays for tuberculosis: current and future applications. Expert Review of Respiratory Medicine, 8(1), pp.67-78.

Unissa, A.N., Hassan, S., Kumari, V.I., Revathy, R. and Hanna, L.E., 2016. Insights into RpoB clinical mutants in mediating rifampicin resistance in Mycobacterium tuberculosis. Journal of Molecular Graphics and Modelling, 67, pp.20-32.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J., 2005. GROMACS: fast, flexible, and free. Journal of computational chemistry, 26(16), pp.1701-1718.

van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P.K.H.C., Mark, A.E., Scott, W.R.P. and Tironi, I.G., 1996. Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH Zürich, Zürich, 86.

Wang, J., Wang, W., Kollman, P.A. and Case, D.A., 2006. Automatic atom type and bond type perception in molecular mechanical calculations. Journal of molecular graphics and modelling, 25(2), pp.247-260.

Webb, B. and Sali, A., 2014. Comparative protein structure modeling using MODELLER. Current protocols in bioinformatics, 47(1), pp.5-6.

Wright, A. and Zignol, M., 2008. Anti-tuberculosis drug resistance in the world: fourth global report: the world health organization/international union against tuberculosis and lung disease (who/union) global project on anti-tuberculosis drug resistance surveillance, 2002-2007. World Health Organization.

World Health Organization and Stop TB Initiative (World Health Organization), 2010. Treatment of tuberculosis: guidelines. World Health Organization.

World Health Organization. c1948. Tuberculosis fact sheet. In: Geneva (Switzerland): WHO global TB programme; 2014 Oct.

World Health Organization, 2017. Global tuberculosis report 2017. Geneva. World Health Organization.

World Health Organization. 2018. Global tuberculosis report 2018.

Wüthrich, K., 1986. NMR with proteins and nucleic acids. Europhysics News, 17(1), pp.11-13.

Zimmermann, W. and Wehrli, W., 1977. Rifampicin-resistance inE. coli: Comparison of microbiological and enzymatic properties. Cellular and Molecular Life Sciences, 33(1), pp.132-133.

Appendix A

A-1: Example of PIR file used for comparative modeling

>P1;5uhc

sequence:5uhc::::::::

ISQRPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVLHE

FT

TVPGVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHNPG

MH

IATLNDKGKLEVELVVERGRGYVPAVQNRASGAEIGRIPVDSIYSPVLKVTYKVDATRV
E

QRTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELN/

RPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVLHEFTT

VP

GVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHNPGMHI

A

TLNDKGKLEVELVVERGRGYVPAVQNRASGAEIGRIPVDSIYSPVLKVTYKVDATRVE

QR

TDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELNVEAEGI/

SNNSVPGAPNRVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVG
GLE

EVLYELSPIEDFSGSMSLSFSDPRFDDVKAPVDECKDKDMTYAAPLFVTAEFINNNTGE
I

KSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVYFDETIDKSTDKTLHSVKVIP
S

RGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSEQIVERFGFSEIMRSTLEKD
NT

VGTDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKRYDLARVGRYKVNKKLGLHVGE
P

ITSSTLTEEDVVATIEYLVRLHEGQTTMTVPGGVEVPVETDDIDHFGNRRLRTVGELIQN

QIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPVVAAIKEFFGTSQLSQFMDQNNPL
S

GLTHKRRLSALGPGGLSRERAGLEVRDVHPSHYGRMCPIETPEGPNIGLIGSLSVYAR
VN

PFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQANSPIDADGRFVEPRVLVRRKA
GE

VEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHDDANRALMGANMQRQAVPLVRSEA
PLVG

TGMELRAAIDAGDVVVAEESGVIEEVSADYITVMHDNGTRRTYRMRKFARSNHGTCA
NQC

PIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLVAIMPWEGHNYEDAIILSNRLVEEDV
L

TSIHIEEHEIDARDTKLGAEEITRDIPNISDEVLADLDERGIVRIGAEVRDGDILVGKVT

PKGETELTPEERLLRAIFGEKAREVRDTSLKVPHGESGKVIGIRVFSREDEDELPAGVN
E

LVRVYVAQKRKISDGDKLAGRHGNKGVIGKILPVEDMPFLADGTPVDIILNTHGVPRRM
N

IGQILETHLGWCAHSGWKVDAAKGVPDWAARLPDELLEAQPNAIVSTPVFDGAQEAEL
QG

LLSCTLPNRDGDVLVDADGKAMLFDGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARS
TG

PYSMITQQPLGGKAQFGGQRFGEMECWAMQAYGAAYTLQELLTIKSDDTVGRVKVYE
AIV

KGENIPEPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGEDE./

DVNFFDELRIGLATAEDIRQWSYGEVKKPETINYRTLKPEKDGLFCEKIFGPTRDWEC

YCGKYKRVRFKGIICERCGVEVTRAKVRRERMGHIELAAPVTHIWYFKGVPSRLGYLL
DL

APKDLEKIIYFAAYVITSVDEEMRHNELSTLEAEMAVERKAVEDQRDGELEARAQKLEA
D

LAELEAEGAKADARRKVRDGGEREMRQIRDRAQRELDRLEDIWSTFTKLAPKQLIVDE
NL

YRELVDRYGEYFTGAMGAESIQKLIENFDIDAEAESLRDVIRNGKGQKKLRALKRLKVV
A

AFQQSGNSPMGMVLDAVPVIPPELRPMVQLDGGRFATSDLNDLYRRVINRNNRLKRLI
DL

GAPEIIVNNEKRMLQESVDALFDNGRRGRPVTGPGNRPLKSLSDLLKGKQGRFRQNLL
GK

RVDYSGRSVIVVGPQLKLHQCGLPKLMALELFKPFVMKRLVDLNHAQNIKSAKRMVER
QR

PQVWDVLEEVIAEHPVLLNRAPTLHRLGIQAFEPMLVEGKAIQLHPLVCEAFNADFDGD
Q

MAVHLPLSAEAQAEARILMLSSNNILSPASGRPLAMPRLDMVTGLYYLTTEVPGDTGE
YQ

PASGDHPETGVYSSPAEAIMAADRGVLSVRAKIKVRLTQLRPPVEIEAELFGHSGWQP
GD

AWMAETTLGRVMFNELLPLGYPFVNKQMHKKVQAAIINDLAERYPMIVVAQTVDKLKD
AG

FYWATRSGVTVSMADVLVPPRKKEILDHYEERADKVEKQFQRGALNHDERNEALVEI
WKE

ATDEVGQALREHYPDDNPIITIVDSGATGNFTQTRTLAGMKGLVTNPKGEFIPRPVKSS
F

REGLTVLEYFINTHGARKGLADTALRTADSGYLTRRLVDVSQDVIVREHDCQTERGIVV
E

LAERAPDGTLIRDPYIETSAYARTLGTDAVDEAGNVIVERGQDLGDPEIDALLAAGITQV

KVRSVLTCATSTGVCATCYGRSMATGKLVDIGEAVGIVAAQSIGEPGTQLTGGGGLPR
VQE

LFEARVPRGKAPIADVTGRVRLEDGERFYKITIVPDDGGEEVVYDKISKRQRLRVFKHE
D

GSERVLSDGDHVEVGQQLMEGSADPHEVLRVQGPREVQIHLVREVQEVYRAQGVSIH
DKH

IEVIVRQMLRRVTIIDSGSTEFLPGSLIDRAEFEAENRRVVAEGGEPAAGRPVLMGITKA

SLATDSWLSAASFQETTRVLTDAAINCRSDKLNGLKENVIIGKLIPAGTGINRYRNIAVQ

PTEEARAA.../

GYDTPLGITNPPIDELLDRVSSKYALVIYAAKRARQINDYYNQLGEGILEYVGPLVEPGL

QEKPLSIALREIHADLLEHTE/

DESEALRQARKDAELTASADSVRAYLKQIGKVALLNAEEEVELAKRIEAGLYATQLMTE
L

SERGEKLPAAQRRDMMWICRDGDRAKNHLLEANLRLVVSLAKRYTGRGMAFLDLIQE
GNL

GLIRAVEKFDYTKGYKFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLGRIQREL

LQDLGREPTPEELAKEMDITPEKVLEIQQYAREPISLDQTIGDEGDSQLGDFIEDSEAVV

AVDAVSFTLLQDQLQSVLDTLSEREAGVVRLRFGLTDGQPRTLDEIGQVYGVTRERIR
QI

ESKTMSKLRHPSRSQVLRDYLD/

...../

...../

.../*

>P1;5uhc.pdb

structureX:5uhc.pdb:FIRST:A :LAST:I::::

ISQRPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVL

HEFTTVPGVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVH
N

PGMHIATLNDKGKLEVELVVERGRGYVPAVQNRASGAEIGRIPVDSIYSPVLKVTYKVD
A

TRVEQRTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELN/

RPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVL

HEFTTVPGVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVH
N

PGMHIATLNDKGKLEVELVVERGRGYVPAVQNRA--AEIGRIPVDSIYSPVLKVTYKVDA

TRVEQRTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELNVEAEGI/

SNNSVPGAPNRVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVG
GLE

EVLYELSPIEDFSGSMSLSFSDPRFDDVKAPVDECKDKDMTYAAPLFVTAEFINNNTGE
I

KSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVYFDETIDKSTDKTLHSVKVIP
S

RGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSEQIVERFGFSEIMRSTLEKD
NT

VGTDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKRYDLARVGRYKVNKKLGLHVGE
P

ITSSTLTEEDVVATIEYLVRLHEGQTTMTVPGGVEVPVETDDIDHFGNRRLRTVGELIQN

QIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPVVAAIKEFFGTSQLSQFMDQNNPL
S

GLTHKRRLSALGPGGLSRERAGLEVRDVHPSHYGRMCPIETPEGPNIGLIGSLSVYAR
VN

PFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQANSPIDADGRFVEPRVLVRRKA
GE

VEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHDDANRALMGANMQRQAVPLVRSEA
PLVG

TGMELRAAIDAGDVVVAEESGVIEEVSADYITVMHDNGTRRTYRMRKFARSNHGTCA
NQC

PIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLVAIMPWEGHNYEDAIILSNRLVEEDV
L

TSIHIEEHEIDARDTKLGAEEITRDIPNISDEVLADLDERGIVRIGAEVRDGDILVGKVT

PKGETELTPEERLLRAIFGEKAREVRDTSLKVPHGESGKVIGIRVFSREDEDELPAGVN
E

LVRVYVAQKRKISDGDKLAGRHGNKGVIGKILPVEDMPFLADGTPVDIILNTHGVPRRM
N

IGQILETHLGWCAHSGWKVDAAKGVPDWAARLPDELLEAQPNAIVSTPVFDGAQEAEL
QG

LLSCTLPNRDGDVLVDADGKAMLFDGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARS
TG

PYSMITQQPLGGKAQFGGQRFGEMECWAMQAYGAAYTLQELLTIKSDDTVGRVKVYE
AIV

KGENIPEPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGEDE./

DVNFFDELRIGLATAEDIRQWSYGEVKKPETINYRTLKPEKDGLFCEKIFGPTRDWEC

YCGKYKRVRFKGIICERCGVEVTRAKVRRERMGHIELAAPVTHIWYFKGVPSRLGYLL
DL

APKDLEKIIYFAAYVITSVDEEMRHNELSTLEAEMAVERKAVEDQRDGELEARAQKLEA
D

LAELEAEGAKADARRKVRDGGEREMRQIRDRAQRELDRLEDIWSTFTKLAPKQLIVDE
NL

YRELVDRYGEYFTGAMGAESIQKLIENFDIDAEAESLRDVIRNGKGQKKLRALKRLKVV
A

AFQQSGNSPMGMVLDAVPVIPPELRPMVQLDGGRFATSDLNDLYRRVINRNNRLKRLI
DL

GAPEIIVNNEKRMLQESVDALFDNGRRGRPVTGPGNRPLKSLSDLLKGKQGRFRQNLL
GK

RVDYSGRSVIVVGPQLKLHQCGLPKLMALELFKPFVMKRLVDLNHAQNIKSAKRMVER
QR

PQVWDVLEEVIAEHPVLLNRAPTLHRLGIQAFEPMLVEGKAIQLHPLVCEAFNADFDGD
Q

MAVHLPLSAEAQAEARILMLSSNNILSPASGRPLAMPRLDMVTGLYYLTTEVPGDTGE
YQ

PASGDHPETGVYSSPAEAIMAADRGVLSVRAKIKVRLTQLRPPVEIEAELFGHSGWQP
GD

AWMAETTLGRVMFNELLPLGYPFVNKQMHKKVQAAIINDLAERYPMIVVAQTVDKLKD
AG

FYWATRSGVTVSMADVLVPPRKKEILDHYEERADKVEKQFQRGALNHDERNEALVEI
WKE

ATDEVGQALREHYPDDNPIITIVDSGATGNFTQTRTLAGMKGLVTNPKGEFIPRPVKSS
F

REGLTVLEYFINTHGARKGLADTALRTADSGYLTRRLVDVSQDVIVREHDCQTERGIVV
E

LAERAPDGTLIRDPYIETSAYARTLGTDAVDEAGNVIVERGQDLGDPEIDALLAAGITQV

KVRSVLTCATSTGVCATCYGRSMATGKLVDIGEAVGIVAAQSIGEPGTQLT--
GGLPRVQE

LFEARVPRGKAPIADVTGRVRLEDGERFYKITIVPDDGGEEVVYDKISKRQRLRVFKHE
D

GSERVLSDGDHVEVGQQLMEGSADPHEVLRVQGPREVQIHLVREVQEVYRAQGVSIH
DKH

IEVIVRQMLRRVTIIDSGSTEFLPGSLIDRAEFEAENRRVVAEGGEPAAGRPVLMGITKA

SLATDSWLSAASFQETTRVLTDAAINCRSDKLNGLKENVIIGKLIPAGTGINRYRNIAVQ

PTEEARAA.../

GYDTPLGITNPPIDELLDRVSSKYALVIYAAKRARQINDYYNQLGEGILEYVGPLVEPGL

QEKPLSIALREIHADLLEHTE/

DESEALRQARKDAELTASADSVRAYLKQIGKVALLNAEEEVELAKRIEAGLYATQLMTE
L

SERGEKLPAAQRRDMMWICRDGDRAKNHLLEANLRLVVSLAKRYTGRGMAFLDLIQE
GNL

GLIRAVEKFDYTKGYKFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLGRIQREL

LQDLGREPTPEELAKEMDITPEKVLEIQQYAREPISLDQTIGDEGDSQLGDFIEDSEAVV

AVDAVSFTLLQDQLQSVLDTLSEREAGVVRLRFGLTDGQPRTLDEIGQVYGVTRERIR
QI

ESKTMSKLRHPSRSQVLRDYLD/

...../

...../

.../*

A-2: Example of MODELLER script used to build and rank models

```
# Homology modelling by the automodel class

from modeller import *
from modeller.automodel import *                    # Load the automodel class


log.verbose()                                        # request verbose output
env = environ()       # create a new MODELLER environment to build this model in

import sys
import os
from sys import argv

pir_file_name = sys.argv[1]
sequence_name = sys.argv[2]
top_models_dir = sys.argv[3]

# directories for input atom files
env.io.atom_files_directory = [' /Run/']

#read in HETATMS
env.io.hetatm = True
env.io.water = False

a = automodel(env,
              alnfile  = 'alignment2.pir',          # alignment filename
              knowns   = ('5uhc.pdb'),              # code of the template
```

```
            sequence = '5uhc', assess_methods=(assess.DOPE, assess.GA341))   #
        code of the target
a.starting_model= 1                 # index of the first model
a.ending_model  = 100               # index of the last model
                                    # (determines how many models to calculate)
a.final_malign3d = False # generate superimposed templates and model (*_fit.pdb files)
a.md_level = refine.very_slow       # No refinement of model
a.make()                            # do the actual homology modelling

# Get a list of all successfully built models from a.outputs
ok_models = [x for x in a.outputs if x['failure'] is None]

# Rank the models by DOPE score
key = 'DOPE score'
if sys.version_info[:2] == (2,3):
    # Python 2.3's sort doesn't have a 'key' argument
    ok_models.sort(lambda a,b: cmp(a[key], b[key]))
else:
    ok_models.sort(key=lambda a: a[key])

# Get top model
m = ok_models[0]
print("Top model: {0} (DOPE score {1})".format(m['name'], m[key]))
os.system("cp {0} {1}".format(m['name'], top_models_dir)) #copy the top model to that file
direct
```

A-3: Validation results for SNP models

Table 218: Validation information for the respective SNP models. The SNP models were ranked from the generated 100 models using MODELLER's GA341 and z-DOPE score after which they were evaluated using PROCHECK, VERIFY3D and QMEAN.

Models	Validation methods				
	PROCHECK (Ramachandran plot)	VERIFY3D (3D-1D score)	QMEAN (normalized QMEAN6 score)	MODELLER (GA341 score)	MODELLER (z-DOPE score)
D441Y	Favoured region: 94% ; Disallowed region: 0.0%	56.86	0.72	1.00	-0.53
D441V	Favoured region: 93.6% ; Disallowed region: 0.0%	59.25	0.71	1.00	-0.52
H451D	Favoured region: 93.9% ; Disallowed region: 0.1%	57.85	0.72	1.00	-0.54
H451N	Favoured region: 93.6% ; Disallowed region: 0.2%	58.51	0.71	1.00	-0.53
H451L	Favoured region: 93.7% ; Disallowed region: 0.0%	55.63	0.72	1.00	-.053

H451R	Favoured region: 93.7% ; Disallowed region: 0.0%	55.92	0.72	1.00	-0.52
S456W	Favoured region: 93.9% ; Disallowed region: 0.1%	67.92	0.71	1.00	-0.53
S456L	Favoured region: 94.1% ; Disallowed region: 0.0%	55.62	0.71	1.00	-0.52

Appendix B

B-1: Tables of Δ*L* and Δ*BC* holo and complex residues

Table 226: Δ*L* residues with values 3SD< = mean and 3SD >= mean for the holo dataset in accordance with the MODELLER structure sequence (actual numbering in Annex D1). The residues are colored based on their commonality among the listed SNP models. The blue, green, yellow, orange, and red indicate ΔL residues that are important in protein communication for 4, 5, 6, 7 and 8 SNP models, respectively. Residues that are bold signify commonality between the holo and complex systems, while those with a larger format signify residues that would be in contact with RIF (cutoff of 7 Å calculated by PyMol) in the complex system.

Model	Positive Δ*L* (mean >= 3 SD)	Negative Δ*L* (mean <= 3 SD)
D441Y	T13,V181,E182,K396,**D398**,E709,D2519,E2929,G3035,R3036,R3201,V3218,E3224,I3226,S3243,**Q3244**,V3245	T72,E73,**L76**,**N77**,E102,**M121**,**H122**,T125,**E281**,**G450**,**I451**,**R501**,D503,**F523**,D537,E1237,P1264,H1265,R1382,N1393,E1575,Y1637,**C1651**,**E1652**,**R1653**,**C1654**,**G1655**,**V1656**,K1808
D441V	I1,E182,L223,**Q239**,K396,**D398**,**A399**,E403,Q404,**I412**,V446,D2519,**I2644**,**S2645**,**K2646**,**E2894**,**G2895**,V3245	**E281**,**F282**,**I451**,N453,**G458**,R891,G893,L894,K1484,G1908,G1909,P2403,**H2655**,**S2659**,L2897,Q3111,D3112,G3114,R3115,E3131,**G3151**,**D3152**,**E3153**,**G3154**,**D3155**
H451R	V181,E182,V278,S374,T2469,D2519,**I2525**,**D2526**,**L2529**,A2530,E2620,D2621,G2622,**E2623**,**R2624**,F2625,Y2626,**I2644**,**S2645**,**K2646**,**R2647**,R2651,V2652,F2653,K2654,H2655,E2656,G2658,E2660,R2661,D2665,G3035,T3211,D3213,E3214,G3216,Q3217,R3227	L12,N15,R16,S17,**L41**,K171,**L219**,**N224**,**I451**,E1577,D1579,**N1581**,A1788,**G3207**,P3240
H451N	E182,K350,F1583,G1971,D2621,G2622,**E2623**,**R2624**,F2625,Y2626,**I2644**,**S2645**,**K2646**,**R2647**,P3209,R3	**I157**,**N224**,**R225**,**P226**,L228,E230,V241,**E243**,**P244**,E246,P247,G248,**A399**,**T400**,**R401**,**V402**,**E403**,**Q404**,**R40**

	210,T3211,E3214,I3215,G3216,Q3217,V3218,E3224,**Q3244**,V3245	**5**,**T406**,**D407**,**F408**,D409,K410,A448,**E449**,**G450**,**I451**,**R501**,G502,E1575,D1576,E1577,Y1637,C1638,**E1652**,D2371,I2896
H451D	T179,V181,E182,Q183,R184,T185,D309,S374,G375,D616,K617,S618,T619,D620,**G1574**,D2519,**D2526**,**I2644**,**S2645**,**K2646**,**R2647**,**A2845**	V285,P286,**G458**,E1747,A1750,Q1751,E1754,A1758,L1760,A1762,E1763,G1764,A1765,K1766,A1767,D1768,A1769,R1770,R1771,K1772,R1774,D1775,G1776,E1778,R1779,E1780,R1782,Q1783,R1785,D1786,N1815,I2896,L2897,**D3152**,**E3153**,**D3155**
H451L	E182,**D398**,**A399**,E403,L1245,**R2054**,**D2522**,A2527,L2528,**L2529**,A2530,G2532,I2533,**G2893**,**E2894**,**G2895**,E3224	**I451**,P1264,H1265,G1266,E1386,K1484,A1485,Y1637,**C1651**,**E1652**,**R1653**,**C1654**,**G1655**,**V1656**,E1657,R1858,G2622,R2624,Y2626,K2627,I2628,T2629,V2631,K2646,**E2656**,**D2657**,**G2658**,**S2659**,E2660,**R2661**,**V2662**,L2663,S2664,D2665
S456W	E182,**E449**,G450,I451,E709,**K1187**,H2229,I2492,D2519,L2520,G2521,**D2522**,P2523,E2524,**I2525**,**D2526**,A2527,L2528,**L2529**,A2530,A2531,G2532,**E2894**,D3213,E3224	**N224**,**R225**,**P226**,**T227**,E246,P247,**T400**,**R401**,**V402**,**Q404**,**R405**,**T406**,**D407**,**F408**,**F523**,T857,H1265,**L1571**,**R1572**,**E1573**,E1575,D1576,E1577,**I1650**,**C1651**,**E1652**,**C1654**,L1760,A1765,K1766,A1767,D1768,A1769,R1770,R1771,K1772,V1773,R1774,D1775,G1776,G1777,E1778,R1779,E1780,R1782,Q1783,D1786,**T3205**,**D3206**,**G3207**

S456L	E182,V278,K396,**D398**,**A399**,R1572,**A2844**,**A2845**,R3201,E3224	I1,S2,S452,N453,**V456**,**P457**,**F523**,T830,Q831,D832,E834,A835,R1382,N1393,D1579,V1580,**N1581**,L1746,R1749,A1750,K1752,L1753,E1754,A1755,D1756,L1757,A1758,E1759,L1760,E1761,A1762,E1763,G1764,A1765,K1766,A1767,D1768,A1769,R1770,R1771,K1772,V1773,R1774,D1775,G1776,G1777,E1778,R1779,E1780,M1781,R1782,Q1783,R1785,D1786,**R2661**,**V2662**,I2896

Table 234: ΔL residues with values 3SD <= mean and 3SD >= mean for the complex dataset in accordance with the MODELLER structure sequence. The residues are colored based on their commonality among the listed SNP models. The blue, green, yellow, orange, and red indicate ΔL residues that are important in communication for 4, 5, 6, 7 and 8 SNP models, respectively. Residues that are bold signify commonality between the holo and complex systems, while those with a larger format signify residues that would be in contact with RIF (cutoff of 7 Å calculated by PyMol) in the complex system.

Model	Positive ΔL (mean >= 3 SD)	Negative ΔL (mean <= 3 SD)
D441Y	**Q239**,V241,E243,L279,**D398**,**A399**,T400,T406,D407,K410,**I412**,D1579,N1581,F1582,Y1637,Y2497,A2498,**D2522**,**I2525**,**D2526**,**L2529**,A2530,L2892,**G2893**,M3126,D3127,I3128,T3129,P3130,E3131,K3132,E3135	**G450**,**I451**,**R501**,**F523**,**H2655**,**G3151**,**G3207**
D441V	D14,**D398**,**A399**,E470,K728,D1576,E1577,D1579,**R2054**,**I2644**,**S2645**,**K2646**,**R2647**,P2853,E3125,**Q3244**	**N224**,**E449**,**G450**,**I451**,**C1654**,**G1655**,**V1656**,P1970,E2493,T2494,F2653,K2654,**H2655**,**E2656**,**D2657**,**G3151**,**D3152**,**G3154**
H451R	V833,E834,K1766,D1768,A1769,K1772,V1773,G1776,E1780,Q1783,A2498,**S2645**,E3116,T3118,P3119,E3120,E3121,L3122,A3123,K3124,E3125,M3126,I3128,T3129,P3130	**L41**,S42,G46,A48,**L76**,**N77**,**M121**,**H122**,G143,Y144,V145,I165,Y166,S167,P168,V169,**R225**,**P226**,H280,**E281**,**F282**,**R401**,**E403**,**Q404**,**R405**,**T406**,**G450**,**I451**,**V456**,**P457**,**G458**,A459,**L1571**,**R1572**,**E1573**,**C1651**,**E1652**,**R1653**,**C1654**,**G1655**,V1686,F2653,K2654,**H2655**,**E2656**,**D2657**,**G2658**

H451N	D14,N15,E1203,**G1574**,N1581,Q1783,P2403,K2404,**I2644**,S2645,R3032,**Q3244**	R225,**E243**,**T400**,R401,**V402**,E403,**Q404**,**R405**,**T406**,D407,**E449**,G450,I451,H2229,K2654,**H2655**,E2656,**D2657**,**G2658**,**G3151**,**D3152**,G3154,**D3155**
H451D	A448,**E449**,G450,S1201,E1203,V1204,**R2054**,A2481,P2482,A2498,**A2844**,**A2845**,**G2893**	**I157**,**E281**,R401,**Q404**,**R405**,**T406**,D407,**F523**,G1648,C1651,**C1654**,G1655,**V1656**,R1965,P1966,V1967,T1968,G1969,P1970,G1971,N1972,R1973,P1974,L1975,**D2657**,Q3148,T3149,**G3151**,**D3152**,**E3153**,G3154,S3156,G3203,L3204,**G3207**,Q3208
H451L	L468,E1203,V1204	R225,L279,H280,**E281**,**F282**,T283,**T400**,R401,**V402**,E403,**Q404**,**R405**,**T406**,D407,**F408**,G450,I451,V1644,G1648,I1649,**I1650**,C1651,**E1652**,**R1653**,**C1654**,G1655,**V1656**,G1969,P1970,G1971,N1972,R1973,**R2661**,**V2662**,Q3148,T3149,**G3151**,**D3152**,L3204
S456W	L468,K713,Q717,N722,K728,T830,**K1187**,Y1637,C1638,A2498,R2499,L2779,**G2893**,**E2894**,**G2895**,E3125,M3126,**Q3244**	D309,M342,G450,I451,N454,**V456**,**P457**,A459,P460,K739,K742,Q859,G886,**L1571**,**R1572**,**N1581**,C1651,**R1653**,**C1654**,G1685,V1686,H2229,**H2655**,E2656,**D2657**,**G2658**,**S2659**,**R2661**,**V2662**,S2950,T3149,**G3151**,**D3152**,**E3153**,G3154,**D3155**,Q3208

S456L	A448,**E449**,**E2623**,**R2624**,**S2645**,S2659,**E2894**	**L219**,A220,R221,E222,L223,**N224**,**R225**,**P226**,**T227**,**P244**,**A399**,**T400**,**R401**,**V402**,**E403**,**Q404**,**R405**,**T406**,**D407**,**F408**,G1971,**G3151**,G3203,L3204,**T3205**,**D3206**,**G3207**,Q3208,P3209,R3210,T3211,L3212,D3213,E3214

Table 242: Δ*BC* residues with values 3SD <= mean and 3SD >= mean for the holo dataset in accordance with the MODELLER structure sequence. The residues are colored based on their prevalence among the listed SNP models. The blue, green, yellow, orange, and red indicate Δ*BC* residues that are important in communication for 4, 5, 6, 7 and 8 SNP models, respectively. Residues that are bold signify commonality between the holo and complex systems, while those with a larger format signify residues that would be in contact with RIF (cutoff of 7 Å calculated by PyMol) in the complex system.

Model	Positive Δ*BC* (mean >= 3 SD)	Negative Δ*BC* (mean <= 3 SD)
D441Y	826,**Q859**,**L860**,M864,N867,**V899**,**H900**,N1154, 301,D1305,L1309,L1323,D1327,**L1457**,H1459,K1464, 490,**G1493**,**E1496**,**Q1951**,V1954,**F1958**,**L1978**,**S1979**, 987,**R1988**,**L1994**,R2003,**S2004**,M2023,I2085,**Q2086**, 097,L2103,V2104,**R2148**,R2212,G2245,V2247,**N2250**, 288,N2428,T2429,R2433,**A2437**,**D2438**,L2441,R2442, 446,**L2592**,**H2714**,I2802,T2945,A2946,**R3032**	**D398**,**Q2011**,**T3034**
D441V	Y610,L699,I794,**Q859**,**L860**,L894,**V899**,**H900**,E914, N1034,**R1184**,E1326,**P1329**,S1439,E1441,K1464,**F1492**, **G1493**,**E1496**,G1909,Q1986,**R1988**,**F1989**,**N1992**,**L1994** ,**K1996**,I2085,**Q2086**,**E2094**,L2103,**R2148**,P2149,**N2250**, G2385,L2441,**H2714**,A2787,V2796,I2802,D2930,E2931,L 2935,T2945,**T3149**,D3152,E3153,D3155	**D398**,**A399**,I412,**R605**,**I629**,R706,E709,D792,D793,N7 99,L879,R1044,A1287,**K1316**,P1347,Q1490,**R1491**,**C1 497**,W1681,R1794,R1948,**G1987**,**R1990**,Q1991,**G199 5**,**L2092**,E2106,N2109,A2110,G2114,D2115,P2144,**L2 276**,M2559,G2580,T2775,**L2779**,**E2792**,T2794,**L2944**, R2952,Q3045
H451R	**L41**,**N799**,E826,**Q859**,D898,N1034,G1288,E1441,K14 64,**M1475**,**F1492**,**G1493**,**E1496**,R1689,P1902,**S1979**,Q1 986,G1987,**R1988**,**F1989**,**L1994**,I2085,**Q2086**,L2103,V21 04,P2149,**A2437**,L2441,**D2445**,**H2714**,A2787,V2796,I280 2,T2945,**L3029**,**R3032**,**T3149**,D3152,D3155	T31,N34,P607,R706,D793,**I809**,**R819**,**R822**,T857,**D8 65**,**V896**,T1343,V1346,N1351,G1437,G1440,**L1481**,I 1680,**Y1682**,**F1683**,I1798,V1954,G1969,**G1971**,L1975, L2150,**L2276**,**A2277**,E2345,**T2450**,**L2779**,**E2792**,R295 2,**T3034**
H451N	G248,**T406**,D407,F408,**S606**,Y610,V626,I794,D795,E826 ,**V899**,G915,**V1043**,**S1173**,**R1184**,E1237,R1301,E1441,K 1464,E1635,A1708,D1786,**L1981**,**R1988**,**N1992**,**K1996**, M2023,T2079,L2080,**E2094**,L2103,V2104,M2117,G2385,	Q404,V604,**R605**,**I629**,P710,G919,A1287,**D1305**,**I1322 **,**V1325**,P1347,N1351,**R1438**,V1547,**R1689**,L1694,L18 05,K1808,Q1809,G1969,**G1971**,**G1995**,N2109,**E2581**, L2592,Q2703,S2789,**E2792**,**T2798**,G2959,G3151

	Q2389,T2392,K2397,N2428,R2433,**T2494**,I2579,**H2714**,I2802,D2930,E2931,L2935,A2946,**R3032**	
H451D	Y610,V626,F638,I794,D795,**Q859**,L882,R897,**V899**,**H900**,N1034,**V1043**,N1154,**S1173**,D1327,E1441,H1459,K1464,**G1493**,**E1496**,Y1706,A1708,**Q1951**,Q1986,G1987,I2085,**Q2086**,**E2094**,L2103,V2104,**R2148**,P2191,**N2250**,**L2253**,G2385,Q2389,K2397,N2428,**A2437**,**H2714**,A2787,**F2790**,V2796,D2799,I2802,D2930,E2931,L2935,A2946,**G3151**,D3152,E3153,D3155	V604,**R605**,**I629**,P630,**W635**,D651,G708,**V823**,P846,**D865**,L879,R1348,G1437,**R1438**,G1440,**C1497**,E1551,W1681,Y1710,R1948,**L2092**,E2106,**L2142**,T2775,A2788,**E2792**,T2794,**T2798**,**L2944**,S2947
H451L	R38,Y610,V626,I794,**T857**,**S858**,P913,A1039,**V1043**,**H1265**,D1305,**I1322**,E1326,D1327,**P1329**,E1441,K1456,**L1457**,K1464,Q1486,**F1487**,**F1492**,**E1496**,A1708,A1709,D1797,**I1811**,**L1816**,**Q1951**,**S1979**,**L1981**,Q1986,G1987,**K1996**,L2080,I2085,**Q2086**,**E2094**,A2097,L2103,**S2146**,**R2148**,P2149,P2153,**N2250**,G2385,T2429,**A2437**,T2439,**H2714**,I2802,D2930,E2931,T2945	**D398**,**A399**,S606,D792,A849,**D865**,T874,L879,**L894**,T1106,V1166,N1351,**L1460**,E1551,**Y1682**,**L1994**,E2053,E2106,P2144,V2187,D2438,A2788,L2797,**T2798**,**E2933**,**R2936**,**R2939**,K2956,I3150,G3154
S456W	G248,**T406**,**S606**,Y610,V626,F638,I794,E826,**T857**,**S858**,**L860**,L882,D898,**H900**,P913,G915,**R1184**,E1441,K1464,**G1493**,E1635,G1987,**E2094**,L2103,V2104,G2385,L2441,**D2445**,**H2714**,A2787,I2802,D2930,E2931,L2935,T2945,A2946,**R3032**	**R605**,**I629**,P630,**W635**,D651,L705,R706,E709,**D865**,**V896**,A1026,T1186,K1187,**D1305**,**G1306**,**L1323**,G1437,**R1438**,G1440,**V1461**,**F1683**,I1798,D1813,L1816,N1992,K1996,**V1998**,**L2276**,S2778,**L2779**,**E2792**,**T2798**,R2952,**Y3033**,**T3034**,R3087
S456L	I629,L699,I794,V823,E826,**T857**,**S858**,L882,R897,**V899**,G915,N1154,E1326,**P1329**,E1441,K1464,I1465,**G1493**,**E1496**,G1987,I2085,**Q2086**,**E2094**,A2097,L2103,V2104,P2149,**N2250**,G2385,Q2389,K2397,N2428,R2433,**A2437**,D2930,E2931,L2935,T2945,A2946,**L3029**,**R3032**	**A399**,S606,**R703**,K704,R706,E709,R801,L802,A1026,**G1306**,K1308,**K1316**,V1346,P1347,N1351,**C1497**,I1798,L1805,D1813,L1982,**R2076**,G2084,D2113,G2114,D2115,G2147,A2270,**L2276**,K2434,**L2779**,**T2798**,K2956,G3151,D3152,G3154

Table 250: ΔBC residues with values 3SD <= mean and 3SD >= mean for the complex dataset in accordance with the MODELLER structure sequence. The residues are colored based on their prevalence among the listed SNP models. The blue, green, yellow, orange, and red indicate ΔBC residues that are important in communication for 4, 5, 6, 7 and 8 SNP models, respectively. Residues that are bold signify commonality between the holo and complex systems, while those with a larger format signify residues that would be in contact with RIF (cutoff of 7 Å calculated by PyMol) in the complex system.

Model	Positive Δ*BC* (mean >= 3 SD)	Negative Δ*BC* (mean <= 3 SD)
D441Y	S606,G608,W635,D651,L705,R706,E709,D792,H796,N811,S818,E821,**R822**,**T857**,**Q859**,**L860**,R1101,E1178,H1265,F1487,R1794,I1798,L1957,L1978,L1981,R1988,N1992,L1994,R1997,S2146,R2148,H2714,S2789,K2956,T3149,**G3151**,G3154,Q3157,L3178	D398,Y610,V626,I629,R703,I794,D795,I809,R845,L894,M1495,C1497,Q1789,D1793,D1797,G1987,R1990,Q2011,L2092,L2142,V2158,H2430,T2439,T2443,D2445,L2449,R2452,L2779,T2798,A2938,D2941,L2944,T2945,A2946,V3096,R3107
D441V	S606,G608,E709,D792,H796,N799,R801,S818,**T857**,**Q859**,**L860**,I910,I918,H1265,L1457,F1487,G1493,E1496,S1544,V1547,E1551,A1834,G1908,I1941,Q1951,F1958,N2075,A2270,K2434,D2438,D2445,T2494,L2592,K2956,T3149,**G3151**,G3154,Q3157	D398,A399,T406,R605,I629,R703,I794,D795,R819,R822,A850,**M864**,**D865**,A1062,N1154,R1221,G1288,H1458,L1460,V1461,D1462,C1497,R1988,R1990,K2049,L2092,P2254,A2444,S2446,R2452,I2644,H2717,L2779,D2941,L2944,D2949,T3034,Q3138,Y3139
H451R	L41,L76,R401,D792,D793,H796,N799,E821,**T857**,V1043,N1163,L1457,Q1479,E1496,V1547,I1798,I1811,L1816,A1834,Q1951,R1988,N1992,L1994,K1996,R1997,S2004,E2064,E2251,L2252,G2435,D2438,T3149,L3179	E60,F61,Y610,I794,D795,E826,**N867**,G893,L894,V896,N1034,A1298,K1316,L1323,E1326,D1327,L1460,S1474,I1476,E1494,Q1783,G1987,R1990,L2142,N2250,L2253,L2276,A2277,T2439,R2451,E2581,L2779,K2813,E2933,R2936,R2939,R3032
H451N	T400,R401,T406,S606,W635,R706,L808,S818,**T857**,**Q859**,**L860**,**L879**,I918,R1184,P1347,M1475,Q1479,F1487,G1493,E1496,I1798,A1834,L1994,E2064,N207	N255,D544,I794,D795,R819,R877,L894,D898,G934,D1305,L1323,D1327,R1438,K1464,I1465,G1482,G1483,A1485,E1494,C1497,G1987,R1990,L2080,M2248,P2

	5,L2083,Q2086,S2146,S2189,N2250,L2253,I2378,A2437,D2455,E2597,P2604,H2714,T3149,**G3151**,Q3157	254,L2436,T2439,R2442,S2679,E2701,L2779,R3032,Y3033
H451D	R401,D792,H796,N799,S818,E821,**R822**,**T857**,S858,**Q859**,**L860**,V899,H900,D1025,H1265,P1347,Q1479,F1487,A1834,Q1951,A1956,G1969,G1971,L1975,L1982,N1992,K1996,R1997,S2146,R2148,S2189,I2378,H2714,K2956,T3149,**G3151**,G3154,Q3157	N255,D612,I629,R703,I794,D795,F797,R800,R897,**P913**,G1288,D1305,L1323,D1327,R1349,R1438,L1460,K1464,I1465,R1491,G1493,P1542,R1689,F1989,L1994,G1995,R2076,A2097,N2140,I2141,L2142,S2143,A2145,V2285,A2432,T2439,S2446,T2450,R2452,L2779,T2798,A2948,D2949,I3091,V3096,I3099,G3103,R3107,Y3139
H451L	R401,D792,H796,N799,N811,S818,E821,**T857**,S858,**Q859**,**L860**,V899,I918,D1025,V1043,S1173,I1322,P1347,Q1479,F1487,F1492,E1496,A1834,Q1951,L1982,N1992,R1997,E2064,N2075,I2378,K2956,T3149,**G3151**,Q3157	V480,R605,I629,W635,R703,I794,D795,R800,V823,E826,A850,**M864**,**D865**,L894,G1266,G1306,K1316,K1321,L1323,D1327,R1438,R1491,C1497,M1500,Y1682,D1797,S1953,L1981,G1987,R1990,L2022,R2076,A2432,T2439,R2442,L2779,R2795,L3029,R3032
S456W	S606,E709,D792,D793,H796,N799,R801,L808,S818,E821,**R822**,**Q859**,**L860**,P1347,Q1479,G1493,M1495,E1496,A1834,S1979,L1982,F1989,L1994,E2064,S2146,S2189,H2714,K2956,T3149,**G3151**,G3154,Q3157,L3179	E516,R605,I629,K728,I794,D795,R800,I813,E826,G1288,H1458,H1459,K1464,I1465,L1481,C1497,P1542,F1683,L1701,A1708,D1797,G1971,L1978,L1981,V1998,W2060,A2432,R2442,R2452,L2779,E2792,L2944
S456L	T400,R401,Q404,T406,D792,H796,N799,S818,E821,**R822**,**T857**,S858,**Q859**,**L860**,I918,A1158,P1329,Q1951,G1969,G1971,L1982,E2064,E2094,S2189,L2253,	Y610,A634,R703,I794,D795,R845,**N867**,D898,N1163,G1288,R1301,D1305,I1322,L1323,V1325,D1327,I1465,I1704,I1705,V1998,R2003,L2025,L2080,A2097,L2252,R2302,H2430,T2439,S2446,L2449,L2779,L2944,V3096,N3100,G3103,R3107,Y3139

A2270,D2438,S2789,F2790,L3029,R3032,T3149,**G3151**,G3154,Q3157

Appendix C

C-1: Example of PyMol script used to identify interface residues

```
from pymol import cmd, stored

def interfaceResidues(cmpx, cA='c. A', cB='c. B', cutoff=1.0, selName="interface"):
    """
    interfaceResidues -- finds 'interface' residues between two chains in a complex.

    PARAMS
        cmpx
            The complex containing cA and cB

        cA
            The first chain in which we search for residues at an interface
            with cB

        cB
            The second chain in which we search for residues at an interface
```

with cA

cutoff

The difference in area OVER which residues are considered interface residues. Residues whose dASA from the complex to a single chain is greater than this cutoff are kept. Zero keeps all residues.

selName

The name of the selection to return.

RETURNS

** A selection of interface residues is created and named depending on what you passed into selName*

** An array of values is returned where each value is: (modelName, residueNumber, dASA)*

NOTES

If you have two chains that are not from the same PDB that you want to complex together, use the create command like:

```
        create myComplex, pdb1WithChainA or pdb2withChainX
    then pass myComplex to this script like:
        interfaceResidues myComlpex, c. A, c. X

    This script calculates the area of the complex as a whole.  Then,
    it separates the two chains that you pass in through the arguments
    cA and cB, alone.  Once it has this, it calculates the difference
    and any residues ABOVE the cutoff are called interface residues.

AUTHOR:
    Jason Vertrees, 2009.
"""
# Save user's settings, before setting dot_solvent
oldDS = cmd.get("dot_solvent")
cmd.set("dot_solvent", 1)

# set some string names for temporary objects/selections
tempC, selName1 = "tempComplex", selName+"1"
chA, chB = "chA", "chB"
```

```
# operate on a new object & turn off the original
cmd.create(tempC, cmpx)
cmd.disable(cmpx)

# remove cruft and inrrelevant chains
cmd.remove(tempC + " and not (polymer and (%s or %s))" % (cA, cB))

# get the area of the complete complex
cmd.get_area(tempC, load_b=1)
# copy the areas from the loaded b to the q, field.
cmd.alter(tempC, 'q=b')

# extract the two chains and calc. the new area
# note: the q fields are copied to the new objects
# chA and chB
cmd.extract(chA, tempC + " and (" + cA + ")")
cmd.extract(chB, tempC + " and (" + cB + ")")
cmd.get_area(chA, load_b=1)
cmd.get_area(chB, load_b=1)
```

```
# update the chain-only objects w/the difference
cmd.alter( "%s or %s" % (chA,chB), "b=b-q" )

# The calculations are done.  Now, all we need to
# do is to determine which residues are over the cutoff
# and save them.
stored.r, rVal, seen = [], [], []
cmd.iterate('%s or %s' % (chA, chB), 'stored.r.append((model,resi,b))')

cmd.enable(cmpx)
cmd.select(selName1, 'none')
for (model,resi,diff) in stored.r:
        key=resi+"-"+model
        if abs(diff)>=float(cutoff):
                if key in seen: continue
                else: seen.append(key)
                rVal.append( (model,resi,diff) )
                # expand the selection here; I chose to iterate over stored.r instead of
                # creating one large selection b/c if there are too many residues PyMOL
                # might crash on a very large selection.  This is pretty much guaranteed
```

```
                    # not to kill PyMOL; but, it might take a little longer to run.
                    cmd.select( selName1, selName1 + " or (%s and i. %s)" % (model,resi))

        # this is how you transfer a selection to another object.
        cmd.select(selName, cmpx + " in " + selName1)
        # clean up after ourselves
        cmd.delete(selName1)
        cmd.delete(chA)
        cmd.delete(chB)
        cmd.delete(tempC)
        # show the selection
        cmd.enable(selName)

        # reset users settings
        cmd.set("dot_solvent", oldDS)

        return rVal

cmd.extend("interfaceResidues", interfaceResidues)
```

C-2: DDRP interface residues identified through the PyMol script (C-1) using a cutoff of 6.7 Å

Table 258: DDRP holo interface residues identified using PyMol script (C-1) where each of the 6 chains were compared to each other with an applied cutoff distance of 6.7 Å.. The red colored residues indicate positive 3SD Δ*BC* residues. The blue colored residues indicate negative Δ*BC* residues. The bold Δ*BC* residues indicate residues that fall into both of the aforementioned categories.

Chain	Residues
A	I1, S2, Q3, T6, R16, L24, E25, P26, G27, F28, Y30, T31, N34, S35, L36, R37, R38, T39, S42, S43, P45, L58, H59, E60, F61, T63, P65, G66, V67, K68, D70, T72, E73, L76, K79, T125, N127, K129, R142, Y144, R151, I157, P161, D163, I165, K171, V172, T173, Y174, R180, E182, Q183, D186, E195, I200, R203, D204, L206, A207, S208, G210, K211, T212, L213, V214, E215, L216, F217, L219, A220, R221, N224
B	T227, S229, L245, E246, G248, F249, Y251, T252, L253, N255, S256, L257, R258, R259, T260, L262, A268, A269, E281, T293, E294, L297, N298, K300, E360, R361, G362, R363, Y365, P367, V369, Q370, R372, I381, P382, D384, I386, L391, K392, Y395, R401, Q404, T406, R424, A428, G431, L434, V435, L440, A441, R442, E443, L444, N445, V446, E447, A448, E449, I451
C	K540, F558, F577, D580, E696, L699, R703, R706, P707, G708, E709, P710, P711, K713, R822, E826, T829, Q839, I842, I844, R845, V848, E853, R889, **L894**, R897, D898, V899, P901, Y904, P909, I910, T912, I918, Q967, L984, R986, V992, P1007, M1010, L1021, H1023, D1024, D1025, A1026, N1027, L1030, V1043, R1044, E1076, V1077, S1078, A1079, D1080, K1098, F1099, C1111, P1112, D1115, A1116, I1147, P1149, E1151, G1152, H1153, Y1155, E1156, D1157, A1158, V1166, E1167, D1169, I1174, R1184, K1187, N1199, R1215, I1216, G1217, A1218, E1219, R1221, D1222, G1223, D1224, T1236, E1237, T1239, P1240, E1241, R1243, L1244, L1245, A1247, I1248, F1249, E1251, H1265, K1270, I1272, R1279, E1284, A1287, A1298, K1300, K1302, D1305, G1306, K1308, K1316, G1317, V1318, G1320, V1325, E1326, D1327, F1330, A1332, D1333, G1334, T1335, P1336, N1342, T1343, H1344, P1347, R1348, M1350, I1352, I1355, L1356, H1359, F1401, Q1405, E1406, Q1410, R1420, D1421, D1429, K1431, F1435, D1436, G1437, R1438, S1439, G1440, E1441, P1442, F1443, P1444, Y1445, P1446, T1448, V1461, D1462, K1464, I1465, H1466, A1467, R1468, S1469, T1470, P1472, Y1473, S1474, M1475, I1476, T1477, Q1478, Q1479, P1480, L1481, G1482, Q1486, F1487,

	G1489, Q1490, R1491, F1492, E1494, M1495, E1496, C1497, W1498, A1499, M1500, Q1501, A1502, Y1503, G1504, A1505, A1506, Y1507, T1508, L1509, Q1510, E1511, L1512, L1513, T1514, I1515, K1516, S1517, T1520, R1523, V1526, Y1527, I1530, V1531, K1532, G1533, E1534, N1535, I1536, E1538, P1539, G1540, I1541, P1542, E1543, S1544, F1545, V1547, L1548, K1550, E1551, L1552, Q1553, S1554, L1555, C1556, L1557, N1558, V1559, E1560, V1561, L1562, S1564, I1569, E1577
D	D1579, N1581, F1582, F1583, D1584, E1585, L1586, R1587, I1588, G1589, L1590, A1591, R1597, Q1598, W1599, S1600, Y1601, E1608, T1609, I1610, Y1612, R1613, L1615, E1618, E1635, K1642, R1643, V1644, R1645, A1661, K1662, R1665, E1666, M1668, H1679, W1681, K1703, A1708, R1779, R1782, Q1783, D1786, R1790, V1812, D1813, E1814, N1815, R1818, V1889, P1894, I1896, E1899, L1900, P1902, M1903, V1904, Q1905, L1906, D1907, G1908, R1910, F1911, A1912, T1913, S1914, D1915, D1918, R1921, R1922, N1925, R1926, R1929, R1932, L1933, L1936, G1937, A1938, P1939, I1941, I1942, N1945, E1946, R1948, M1949, E1952, F1958, R1963, P1970, G1971, R1973, K1976, L1978, L1981, L1982, Q1986, R1988, F1989, R1990, Q1991, N1992, L1993, **L1994**, G1995, **K1996**, R1997, V1998, D1999, Y2000, S2001, G2002, R2003, S2004, V2005, V2007, V2008, P2010, Q2011, K2013, H2015, K2021, M2023, E2026, P2030, F2031, M2033, K2034, R2035, Q2043, N2044, I2045, K2046, K2049, R2050, E2053, R2054, Q2055, P2057, W2060, E2064, E2065, V2066, A2068, E2069, N2075, A2077, P2078, T2079, L2080, R2082, I2085, E2089, L2092, E2094, G2095, K2096, A2097, Q2099, P2102, C2105, A2110, F2112, D2113, G2114, Q2116, A2118, H2120, P2122, L2123, S2124, A2125, E2126, A2127, Q2128, A2129, E2130, R2132, I2133, L2134, M2135, L2136, N2139, R2154, L2155, D2156, Y2163, G2180, D2181, P2183, E2184, V2187, Y2188, S2189, A2192, E2193, I2195, M2196, D2199, R2200, V2202, R2206, R2212, M2239, P2281, M2282, I2283, V2284, Q2287, K2291, F2297, Y2298, A2300, T2301, R2302, S2303, G2304, V2305, T2306, V2307, S2308, A2310, D2311, V2312, I2375, I2378, T2384, Q2389, T2392, L2393, K2397, R2410, F2416, R2417, E2418, G2419, L2420, T2421, V2422, L2423, Y2425, F2426, N2428, T2429, H2430, A2432, R2433, K2434, L2436, A2437, D2438, A2440, L2441, T2450, R2451, V2454, T2561, K2563, D2566, I2567, A2570, I2573, V2574, Q2577, W2784, L2785, L2797, L2812, K2813, V2816, I2817, I2818, G2819, L2821, I2822, G2825, T2826, R2830, Y2831, R2832, N2833, I2834, A2835, V2836, Q2837, P2838, T2839, A2842
E	G2849, Y2850, D2851, L2854, G2855, I2856, T2857, N2858, P2860, I2861, D2862, L2865, S2869, S2870, K2871, Y2872, A2873, L2874, V2875, I2876, A2879, K2880, A2882, R2883, Q2884, N2886, D2887, N2890, V2900, L2903, Q2909, K2911, L2913, S2914, L2917, D2923, L2924, L2925, E2926, H2927

F T2928, E2929, D2930, E2931, S2932, E2933, A2934, L2935, R2936, A2938, R2939, K2940, A2942, E2943, L2944, T2945, A2946, A2948, S2950, R2952, A2953, Y2954, K2956, Q2957, K2960, K3015, N3016, L3019, E3020, L3023, R3032, T3034, A3039, L3041, D3042, Q3045, E3046, N3048, L3049, I3052, Q3085, T3088, I3089, R3090, I3091, P3092, M3095, Q3106, R3107, L3110, Q3111, G3114, R3115, E3116, T3118, K3132, L3134, E3135, Q3138, Y3139, R3141, E3142, P3143, I3144, S3145, L3146, D3147, Q3148, **G3151**, **D3152**, E3153, D3155, Q3157, L3158, G3159, D3160, F3161, I3162, E3163, D3164, E3166, A3167, V3168, A3170, V3171, D3172, A3173, V3174, S3175, F3176, T3177, L3178, L3179, L3183, R3201, F3202, G3203, L3204, D3206, G3207, Q3208, P3209, M3234, L3237, S3241, Q3244, L3246, R3247, D3248, Y3249, L3250, D3251

Table 266: DDRP complex interface residues identified using PyMol script (C-1) where each of the 6 chains were compared to each other with an applied cutoff distance of 6.7 Å.. The red colored residues indicate positive 3SD ΔBC residues. The blue colored residues indicate negative ΔBC residues. The bold ΔBC residues indicate residues that fall into both of the aforementioned categories.

Chain	Residues
A	I1, S2, Q3, T6, R16, L24, E25, P26, G27, F28, Y30, T31, N34, S35, L36, R37, R38, T39, S42, S43, P45, L58, H59, E60, F61, T63, P65, G66, V67, K68, D70, T72, E73, L76, K79, T125, N127, K129, R142, Y144, R151, I157, P161, D163, I165, K171, V172, T173, Y174, R180, E182, Q183, D186, E195, I200, R203, D204, L206, A207, S208, G210, K211, T212, L213, V214, E215, L216, F217, L219, A220, R221, N224
B	T227, S229, L245, E246, G248, F249, Y251, T252, L253, N255, S256, L257, R258, R259, T260, L262, A268, A269, E281, T293, E294, L297, N298, K300, E360, R361, G362, R363, Y365, P367, V369, Q370, R372, I381, P382, D384, I386, L391, K392, Y395, R401, Q404, **T406**, R424, A428, G431, L434, V435, L440, A441, R442, E443, L444, N445, V446, E447, A448, E449, I451
C	K540, F558, F577, D580, E696, L699, R703, R706, P707, G708, E709, P710, P711, K713, **R822**, E826, T829, Q839, I842, I844, R845, V848, E853, R889, L894, R897, D898, V899, P901, Y904, P909, I910, T912, I918, Q967, L984, R986, V992, P1007, M1010, L1021, H1023, D1024, D1025, A1026, N1027, L1030, V1043, R1044, E1076, V1077, S1078, A1079, D1080, K1098, F1099, C1111, P1112, D1115, A1116, I1147, P1149, E1151, G1152, H1153, Y1155, E1156, D1157, A1158, V1166, E1167, D1169, I1174, R1184, K1187, N1199, R1215, I1216, G1217, A1218, E1219, R1221, D1222, G1223, D1224, T1236, E1237, T1239, P1240, E1241, R1243, L1244, L1245, A1247, I1248, F1249, E1251, H1265, K1270, I1272, R1279, E1284, A1287, A1298, K1300, K1302, D1305, G1306, K1308, K1316, G1317, V1318, G1320, V1325, E1326, D1327, F1330, A1332, D1333, G1334, T1335, P1336, N1342, T1343, H1344, P1347, R1348, M1350, I1352, I1355, L1356, H1359, F1401, Q1405, E1406, Q1410, R1420, D1421, D1429, K1431, F1435, D1436, G1437, R1438, S1439, G1440, E1441, P1442, F1443, P1444, Y1445, P1446, T1448, V1461, D1462, K1464, I1465, H1466, A1467, R1468, S1469, T1470, P1472, Y1473, S1474, M1475, I1476, T1477, Q1478, Q1479, P1480, L1481, G1482, Q1486, F1487, G1489, Q1490, R1491, F1492, E1494, **M1495**, E1496, C1497, W1498, A1499, M1500, Q1501, A1502, Y1503, G1504, A1505, A1506, Y1507, T1508, L1509, Q1510, E1511, L1512, L1513, T1514, I1515, K1516, S1517, T1520, R1523, V1526, Y1527, I1530, V1531, K1532, G1533, E1534, N1535, I1536, E1538, P1539, G1540, I1541, P1542, E1543, S1544, F1545, V1547, L1548, K1550, E1551, L1552, Q1553, S1554, L1555, C1556, L1557, N1558, V1559, E1560, V1561, L1562, S1564, I1569, E1577
D	D1579, N1581, F1582, F1583, D1584, E1585, L1586, R1587, I1588, G1589, L1590, A1591, R1597, Q1598, W1599, S1600, Y1601, E1608, T1609, I1610, Y1612, R1613, L1615, E1618, E1635, K1642, R1643, V1644,

	R1645, A1661, K1662, R1665, E1666, M1668, H1679, W1681, K1703, A1708, R1779, R1782, Q1783, D1786, R1790, V1812, D1813, E1814, N1815, R1818, V1889, P1894, I1896, E1899, L1900, P1902, M1903, V1904, Q1905, L1906, D1907, G1908, R1910, F1911, A1912, T1913, S1914, D1915, D1918, R1921, R1922, N1925, R1926, R1929, R1932, L1933, L1936, G1937, A1938, P1939, I1941, I1942, N1945, E1946, R1948, M1949, E1952, F1958, R1963, P1970, G1971, R1973, K1976, **L1978**, **L1981**, L1982, Q1986, **R1988**, **F1989**, R1990, Q1991, N1992, L1993, **L1994**, G1995, K1996, R1997, V1998, D1999, Y2000, S2001, G2002, R2003, S2004, V2005, V2007, V2008, P2010, Q2011, K2013, H2015, K2021, M2023, E2026, P2030, F2031, M2033, K2034, R2035, Q2043, N2044, I2045, K2046, K2049, R2050, E2053, R2054, Q2055, P2057, W2060, E2064, E2065, V2066, A2068, E2069, N2075, A2077, P2078, T2079, L2080, R2082, I2085, E2089, L2092, E2094, G2095, K2096, A2097, Q2099, P2102, C2105, A2110, F2112, D2113, G2114, Q2116, A2118, H2120, P2122, L2123, S2124, A2125, E2126, A2127, Q2128, A2129, E2130, R2132, I2133, L2134, M2135, L2136, N2139, R2154, L2155, D2156, Y2163, G2180, D2181, P2183, E2184, V2187, Y2188, S2189, A2192, E2193, I2195, M2196, D2199, R2200, V2202, R2206, R2212, M2239, P2281, M2282, I2283, V2284, Q2287, K2291, F2297, Y2298, A2300, T2301, R2302, S2303, G2304, V2305, T2306, V2307, S2308, A2310, D2311, V2312, I2375, I2378, T2384, Q2389, T2392, L2393, K2397, R2410, F2416, R2417, E2418, G2419, L2420, T2421, V2422, L2423, Y2425, F2426, N2428, T2429, H2430, A2432, R2433, K2434, L2436, A2437, D2438, A2440, L2441, T2450, R2451, V2454, T2561, K2563, D2566, I2567, A2570, I2573, V2574, Q2577, W2784, L2785, L2797, L2812, K2813, V2816, I2817, I2818, G2819, L2821, I2822, G2825, T2826, R2830, Y2831, R2832, N2833, I2834, A2835, V2836, Q2837, P2838, T2839, A2842
E	G2849, Y2850, D2851, L2854, G2855, I2856, T2857, N2858, P2860, I2861, D2862, L2865, S2869, S2870, K2871, Y2872, A2873, L2874, V2875, I2876, A2879, K2880, A2882, R2883, Q2884, N2886, D2887, N2890, V2900, L2903, Q2909, K2911, L2913, S2914, L2917, D2923, L2924, L2925, E2926, H2927

F	T2928, E2929, D2930, E2931, S2932, E2933, A2934, L2935, R2936, A2938, R2939, K2940, A2942, E2943, L2944, T2945, A2946, A2948, S2950, R2952, A2953, Y2954, K2956, Q2957, K2960, K3015, N3016, L3019, E3020, L3023, **R3032**, T3034, A3039, L3041, D3042, Q3045, E3046, N3048, L3049, I3052, Q3085, T3088, I3089, R3090, I3091, P3092, M3095, Q3106, R3107, L3110, Q3111, G3114, R3115, E3116, T3118, K3132, L3134, E3135, Q3138, Y3139, R3141, E3142, P3143, I3144, S3145, L3146, D3147, Q3148, G3151, D3152, E3153, D3155, Q3157, L3158, G3159, D3160, F3161, I3162, E3163, D3164, E3166, A3167, V3168, A3170, V3171, D3172, A3173, V3174, S3175, F3176, T3177, L3178, L3179, L3183, R3201, F3202, G3203, L3204, D3206, G3207, Q3208, P3209, M3234, L3237, S3241, Q3244, L3246, R3247, D3248, Y3249, L3250, D3251

C-3: MODELLER numbering vs universal *MTB* numbering

Table 9: MODELLER numbered residues (left) with corresponding universal *MTB* numbered residues (right) of the DDRP holoenzyme.

Chain A: I1, S2, Q3, R4, P5, T6, L7, S8, E9, D10, V11, L12, T13, D14, N15, R16, S17, Q18, F19, V20, I21, E22, P23, L24, E25, P26, G27, F28, G29, Y30, T31, L32, G33, N34, S35, L36, R37, R38, T39, L40, L41, S42, S43, I44, P45, G46, A47, A48, V49, T50, S51, I52, R53, I54, D55, G56, V57, L58, H59, E60, F61, T62, T63, V64, P65, G66, V67, K68, E69, D70, V71, T72, E73, I74, I75, L76, N77, L78, K79, S80, L81, V82, V83, S84, S85, E86, E87, D88, E89, P90, V91, T92, M93, Y94, L95, R96, K97, Q98, G99, P100, G101, E102, V103, T104, A105, G106, D107, I108, V109, P110, P111, A112, G113, V114, T115, V116, H117, N118, P119, G120, M121, H122, I123, A124, T125, L126, N127, D128, K129, G130, K131, L132, E133, V134, E135, L136, V137, V138, E139, R140, G141, R142, G143, Y144, V145, P146, A147, V148, Q149, N150, R151, A152, S153, G154, A155, E156, I157, G158, R159, I160, P161, V162, D163, S164, I165, Y166, S167, P168, V169, L170, K171, V172, T173, Y174, K175, V176, D177, A178, T179, R180, V181, E182, Q183, R184, T185, D186, F187, D188, K189, L190, I191, L192, D193, V194, E195, T196, K197, N198, S199, I200, S201, P202, R203, D204, A205, L206, A207, S208, A209, G210, K211, T212, L213, V214, E215, L216, F217, G218, L219, A220, R221, E222, L223, N224, R225, P226

Chain A: I3, S4, Q5, R6, P7, T8, L9, S10, E11, D12, V13, L14, T15, D16, N17, R18, S19, Q20, F21, V22, I23, E24, P25, L26, E27, P28, G29, F30, G31, Y32, T33, L34, G35, N36, S37, L38, R39, R40, T41, L42, L43, S44, S45, I46, P47, G48, A49, A50, V51, T52, S53, I54, R55, I56, D57, G58, V59, L60, H61, E62, F63, T64, T65, V66, P67, G68, V69, K70, E71, D72, V73, T74, E75, I76, I77, L78, N79, L80, K81, S82, L83, V84, V85, S86, S87, E88, E89, D90, E91, P92, V93, T94, M95, Y96, L97, R98, K99, Q100, G101, P102, G103, E104, V105, T106, A107, G108, D109, I110, V111, P112, P113, A114, G115, V116, T117, V118, H119, N120, P121, G122, M123, H124, I125, A126, T127, L128, N129, D130, K131, G132, K133, L134, E135, V136, E137, L138, V139, V140, E141, R142, G143, R144, G145, Y146, V147, P148, A149, V150, Q151, N152, R153, A154, X155, X156, A157, E158, I159, G160, R161, I162, P163, V164, D165, S166, I167, Y168, S169, P170, V171, L172, K173, V174, T175, Y176, K177, V178, D179, A180, T181, R182, V183, E184, Q185, R186, T187, D188, F189, D190, K191, L192, I193, L194, D195, V196, E197, T198, K199, N200, S201, I202, S203, P204, R205, D206, A207, L208, A209, S210, A211, G212, K213, T214, L215, V216, E217, L218, F219, G220, L221, A222, R223, E224, L225, N226

Chain B: T227, L228, S229, E230, D231, V232, L233, T234, D235, N236, R237, S238, Q239, F240, V241, I242, E243, P244, L245, E246, P247, G248, F249, G250, Y251, T252, L253, G254, N255, S256, L257, R258, R259, T260, L261, L262, S263, S264, I265, P266, G267, A268, A269, V270, T271, S272, I273, R274, I275, D276, G277, V278, L279, H280, E281, F282, T283, T284, V285, P286, G287, V288, K289, E290, D291, V292, T293, E294, I295, I296, L297, N298, L299, K300, S301, L302, V303, V304, S305, S306, E307, E308, D309, E310, P311, V312, T313, M314, Y315, L316, R317, K318, Q319, G320, P321, G322, E323, V324, T325, A326, G327, D328, I329, V330, P331, P332, A333, G334, V335, T336, V337, H338, N339, P340, G341, M342, H343, I344, A345, T346, L347, N348, D349, K350, G351, K352, L353, E354, V355, E356, L357, V358, V359, E360, R361, G362, R363, G364, Y365, V366, P367, A368, V369, Q370, N371, R372, A373, S374, G375, A376, E377, I378, G379, R380, I381, P382, V383, D384, S385, I386, Y387, S388, P389, V390, L391, K392, V393, T394, Y395, K396, V397, D398, A399, T400, R401, V402, E403, Q404, R405, T406, D407, F408, D409, K410, L411, I412, L413, D414, V415, E416, T417, K418, N419, S420, I421, S422, P423, R424, D425, A426, L427, A428, S429, A430, G431, K432, T433, L434, V435, E436, L437, F438, G439, L440, A441, R442, E443, L444, N445, V446, E447, A448, E449, G450, I451	**Chain B**: R6, P7, T8, L9, S10, E11, D12, V13, L14, T15, D16, N17, R18, S19, Q20, F21, V22, I23, E24, P25, L26, E27, P28, G29, F30, G31, Y32, T33, L34, G35, N36, S37, L38, R39, R40, T41, L42, L43, S44, S45, I46, P47, G48, A49, A50, V51, T52, S53, I54, R55, I56, D57, G58, V59, L60, H61, E62, F63, T64, T65, V66, P67, G68, V69, K70, E71, D72, V73, T74, E75, I76, I77, L78, N79, L80, K81, S82, L83, V84, V85, S86, S87, E88, E89, D90, E91, P92, V93, T94, M95, Y96, L97, R98, K99, Q100, G101, P102, G103, E104, V105, T106, A107, G108, D109, I110, V111, P112, P113, A114, G115, V116, T117, V118, H119, N120, P121, G122, M123, H124, I125, A126, T127, L128, N129, D130, K131, G132, K133, L134, E135, V136, E137, L138, V139, V140, E141, R142, G143, R144, G145, Y146, V147, P148, A149, V150, Q151, N152, R153, A154, S155, G156, A157, E158, I159, G160, R161, I162, P163, V164, D165, S166, I167, Y168, S169, P170, V171, L172, K173, V174, T175, Y176, K177, V178, D179, A180, T181, R182, V183, E184, Q185, R186, T187, D188, F189, D190, K191, L192, I193, L194, D195, V196, E197, T198, K199, N200, S201, I202, S203, P204, R205, D206, A207, L208, A209, S210, A211, G212, K213, T214, L215, V216, E217, L218, F219, G220, L221, A222, R223, E224, L225, N226, V227, E228, A229, E230, G231, I232
Chain C: S452, N453, N454, S455, V456, P457, G458, A459, P460, N461, R462, V463, S464, F465, A466, K467, L468, R469, E470, P471, L472, E473, V474, P475, G476, L477, L478, D479, V480, Q481, T482, D483, S484, F485, E486, W487, L488, I489, G490, S491, P492, R493, W494, R495, E496, S497, A498, A499, E500, R501, G502, D503,	**Chain C**: S28, N29, N30, S31, V32, P33, G34, A35, P36, N37, R38, V39, S40, F41, A42, K43, L44, R45, E46, P47, L48, E49, V50, P51, G52, L53, L54, D55, V56, Q57, T58, D59, S60, F61, E62, W63, L64, I65, G66, S67, P68, R69, W70, R71, E72, S73, A74, A75, E76, R77, G78, D79, V80, N81, P82, V83, G84, G85, L86, E87, E88, V89, L90, Y91, E92, L93, S94, P95, I96, E97,

V504, N505, P506, V507, G508, G509, L510, E511, E512, V513, L514, Y515, E516, L517, S518, P519, I520, E521, D522, F523, S524, G525, S526, M527, S528, L529, S530, F531, S532, D533, P534, R535, F536, D537, D538, V539, K540, A541, P542, V543, D544, E545, C546, K547, D548, K549, D550, M551, T552, Y553, A554, A555, P556, L557, F558, V559, T560, A561, E562, F563, I564, N565, N566, N567, T568, G569, E570, I571, K572, S573, Q574, T575, V576, F577, M578, G579, D580, F581, P582, M583, M584, T585, E586, K587, G588, T589, F590, I591, I592, N593, G594, T595, E596, R597, V598, V599, V600, S601, Q602, L603, V604, R605, S606, P607, G608, V609, Y610, F611, D612, E613, T614, I615, D616, K617, S618, T619, D620, K621, T622, L623, H624, S625, V626, K627, V628, I629, P630, S631, R632, G633, A634, W635, L636, E637, F638, D639, V640, D641, K642, R643, D644, T645, V646, G647, V648, R649, I650, D651, R652, K653, R654, R655, Q656, P657, V658, T659, V660, L661, L662, K663, A664, L665, G666, W667, T668, S669, E670, Q671, I672, V673, E674, R675, F676, G677, F678, S679, E680, I681, M682, R683, S684, T685, L686, E687, K688, D689, N690, T691, V692, G693, T694, D695, E696, A697, L698, L699, D700, I701, Y702, R703, K704, L705, R706, P707, G708, E709, P710, P711, T712, K713, E714, S715, A716, Q717, T718, L719, L720, E721, N722, L723, F724, F725, K726, E727, K728, R729, Y730, D731, L732, A733, R734, V735, G736, R737, Y738, K739, V740, N741, K742, K743, L744, G745, L746, H747, V748, G749, E750, P751, I752, T753, S754, S755, T756, L757, T758, E759, E760, D761, V762, V763, A764, T765, I766, E767, Y768, L769, V770, R771, L772, H773, E774, G775, Q776, T777, T778, M779, T780, V781, P782, G783, G784, V785, E786, V787, P788, V789, E790, T791, D792, D793, I794, D795, H796, F797, G798, N799, D98, F99, S100, G101, S102, M103, S104, L105, S106, F107, S108, D109, P110, R111, F112, D113, D114, V115, K116, A117, P118, V119, D120, E121, C122, K123, D124, K125, D126, M127, T128, Y129, A130, A131, P132, L133, F134, V135, T136, A137, E138, F139, I140, N141, N142, N143, T144, G145, E146, I147, K148, S149, Q150, T151, V152, F153, M154, G155, D156, F157, P158, M159, M160, T161, E162, K163, G164, T165, F166, I167, I168, N169, G170, T171, E172, R173, V174, V175, V176, S177, Q178, L179, V180, R181, S182, P183, G184, V185, Y186, F187, D188, E189, T190, I191, D192, K193, S194, T195, D196, K197, T198, L199, H200, S201, V202, K203, V204, I205, P206, S207, R208, G209, A210, W211, L212, E213, F214, D215, V216, D217, K218, R219, D220, T221, V222, G223, V224, R225, I226, D227, R228, K229, R230, R231, Q232, P233, V234, T235, V236, L237, L238, K239, A240, L241, G242, W243, T244, S245, E246, Q247, I248, V249, E250, R251, F252, G253, F254, S255, E256, I257, M258, R259, S260, T261, L262, E263, K264, D265, N266, T267, V268, G269, T270, D271, E272, A273, L274, L275, D276, I277, Y278, R279, K280, L281, R282, P283, G284, E285, P286, P287, T288, K289, E290, S291, A292, Q293, T294, L295, L296, E297, N298, L299, F300, F301, K302, E303, K304, R305, Y306, D307, L308, A309, R310, V311, G312, R313, Y314, K315, V316, N317, K318, K319, L320, G321, L322, H323, V324, G325, E326, P327, I328, T329, S330, S331, T332, L333, T334, E335, E336, D337, V338, V339, A340, T341, I342, E343, Y344, L345, V346, R347, L348, H349, E350, G351, Q352, T353, T354, M355, T356, V357, P358, G359, G360, V361, E362, V363, P364, V365, E366, T367, D368, D369, I370, D371, H372, F373, G374, N375, R376, R377, L378, R379, T380, V381, G382, E383, L384, I385, Q386, N387, Q388, I389, R390, V391, G392, M393, S394, R395, M396, E397, R398, V399, V400, R401,

R800, R801, L802, R803, T804, V805, G806, E807, L808, I809, Q810, N811, Q812, I813, R814, V815, G816, M817, S818, R819, M820, E821, R822, V823, V824, R825, E826, R827, M828, T829, T830, Q831, D832, V833, E834, A835, I836, T837, P838, Q839, T840, L841, I842, N843, I844, R845, P846, V847, V848, A849, A850, I851, K852, E853, F854, F855, G856, T857, S858, Q859, L860, S861, Q862, F863, M864, D865, Q866, N867, N868, P869, L870, S871, G872, L873, T874, H875, K876, R877, R878, L879, S880, A881, L882, G883, P884, G885, G886, L887, S888, R889, E890, R891, A892, G893, L894, E895, V896, R897, D898, V899, H900, P901, S902, H903, Y904, G905, R906, M907, C908, P909, I910, E911, T912, P913, E914, G915, P916, N917, I918, G919, L920, I921, G922, S923, L924, S925, V926, Y927, A928, R929, V930, N931, P932, F933, G934, F935, I936, E937, T938, P939, Y940, R941, K942, V943, V944, D945, G946, V947, V948, S949, D950, E951, I952, V953, Y954, L955, T956, A957, D958, E959, E960, D961, R962, H963, V964, V965, A966, Q967, A968, N969, S970, P971, I972, D973, A974, D975, G976, R977, F978, V979, E980, P981, R982, V983, L984, V985, R986, R987, K988, A989, G990, E991, V992, E993, Y994, V995, P996, S997, S998, E999, V1000, D1001, Y1002, M1003, D1004, V1005, S1006, P1007, R1008, Q1009, M1010, V1011, S1012, V1013, A1014, T1015, A1016, M1017, I1018, P1019, F1020, L1021, E1022, H1023, D1024, D1025, A1026, N1027, R1028, A1029, L1030, M1031, G1032, A1033, N1034, M1035, Q1036, R1037, Q1038, A1039, V1040, P1041, L1042, V1043, R1044, S1045, E1046, A1047, P1048, L1049, V1050, G1051, T1052, G1053, M1054, E1055, L1056, R1057, A1058, A1059, I1060, D1061, A1062, G1063, D1064, V1065, V1066, V1067, A1068, E1069, E1070, S1071, G1072, V1073, I1074, E402, R403, M404, T405, T406, Q407, D408, V409, E410, A411, I412, T413, P414, Q415, T416, L417, I418, N419, I420, R421, P422, V423, V424, A425, A426, I427, K428, E429, F430, F431, G432, T433, S434, Q435, L436, S437, Q438, F439, M440, D441, Q442, N443, N444, P445, L446, S447, G448, L449, T450, H451, K452, R453, R454, L455, S456, A457, L458, G459, P460, G461, G462, L463, S464, R465, E466, R467, A468, G469, L470, E471, V472, R473, D474, V475, H476, P477, S478, H479, Y480, G481, R482, M483, C484, P485, I486, E487, T488, P489, E490, G491, P492, N493, I494, G495, L496, I497, G498, S499, L500, S501, V502, Y503, A504, R505, V506, N507, P508, F509, G510, F511, I512, E513, T514, P515, Y516, R517, K518, V519, V520, D521, G522, V523, V524, S525, D526, E527, I528, V529, Y530, L531, T532, A533, D534, E535, E536, D537, R538, H539, V540, V541, A542, Q543, A544, N545, S546, P547, I548, D549, A550, D551, G552, R553, F554, V555, E556, P557, R558, V559, L560, V561, R562, R563, K564, A565, G566, E567, V568, E569, Y570, V571, P572, S573, S574, E575, V576, D577, Y578, M579, D580, V581, S582, P583, R584, Q585, M586, V587, S588, V589, A590, T591, A592, M593, I594, P595, F596, L597, E598, H599, D600, D601, A602, N603, R604, A605, L606, M607, G608, A609, N610, M611, Q612, R613, Q614, A615, V616, P617, L618, V619, R620, S621, E622, A623, P624, L625, V626, G627, T628, G629, M630, E631, L632, R633, A634, A635, I636, D637, A638, G639, D640, V641, V642, V643, A644, E645, E646, S647, G648, V649, I650, E651, E652, V653, S654, A655, D656, Y657, I658, T659, V660, M661, H662, D663, N664, G665, T666, R667, R668, T669, Y670, R671, M672, R673, K674, F675, A676, R677, S678, N679, H680, G681, T682, C683, A684, N685, Q686, C687, P688, I689, V690, D691, A692, G693, D694, R695, V696, E697, A698, G699, Q700,

E1075, E1076, V1077, S1078, A1079, D1080, Y1081, I1082, T1083, V1084, M1085, H1086, D1087, N1088, G1089, T1090, R1091, R1092, T1093, Y1094, R1095, M1096, R1097, K1098, F1099, A1100, R1101, S1102, N1103, H1104, G1105, T1106, C1107, A1108, N1109, Q1110, C1111, P1112, I1113, V1114, D1115, A1116, G1117, D1118, R1119, V1120, E1121, A1122, G1123, Q1124, V1125, I1126, A1127, D1128, G1129, P1130, C1131, T1132, D1133, D1134, G1135, E1136, M1137, A1138, L1139, G1140, K1141, N1142, L1143, L1144, V1145, A1146, I1147, M1148, P1149, W1150, E1151, G1152, H1153, N1154, Y1155, E1156, D1157, A1158, I1159, I1160, L1161, S1162, N1163, R1164, L1165, V1166, E1167, E1168, D1169, V1170, L1171, T1172, S1173, I1174, H1175, I1176, E1177, E1178, H1179, E1180, I1181, D1182, A1183, R1184, D1185, T1186, K1187, L1188, G1189, A1190, E1191, E1192, I1193, T1194, R1195, D1196, I1197, P1198, N1199, I1200, S1201, D1202, E1203, V1204, L1205, A1206, D1207, L1208, D1209, E1210, R1211, G1212, I1213, V1214, R1215, I1216, G1217, A1218, E1219, V1220, R1221, D1222, G1223, D1224, I1225, L1226, V1227, G1228, K1229, V1230, T1231, P1232, K1233, G1234, E1235, T1236, E1237, L1238, T1239, P1240, E1241, E1242, R1243, L1244, L1245, R1246, A1247, I1248, F1249, G1250, E1251, K1252, A1253, R1254, E1255, V1256, R1257, D1258, T1259, S1260, L1261, K1262, V1263, P1264, H1265, G1266, E1267, S1268, G1269, K1270, V1271, I1272, G1273, I1274, R1275, V1276, F1277, S1278, R1279, E1280, D1281, E1282, D1283, E1284, L1285, P1286, A1287, G1288, V1289, N1290, E1291, L1292, V1293, R1294, V1295, Y1296, V1297, A1298, Q1299, K1300, R1301, K1302, I1303, S1304, D1305, V701, I702, A703, D704, G705, P706, C707, T708, D709, D710, G711, E712, M713, A714, L715, G716, K717, N718, L719, L720, V721, A722, I723, M724, P725, W726, E727, G728, H729, N730, Y731, E732, D733, A734, I735, I736, L737, S738, N739, R740, L741, V742, E743, E744, D745, V746, L747, T748, S749, I750, H751, I752, E753, E754, H755, E756, I757, D758, A759, R760, D761, T762, K763, L764, G765, A766, E767, E768, I769, T770, R771, D772, I773, P774, N775, I776, S777, D778, E779, V780, L781, A782, D783, L784, D785, E786, R787, G788, I789, V790, R791, I792, G793, A794, E795, V796, R797, D798, G799, D800, I801, L802, V803, G804, K805, V806, T807, P808, K809, G810, E811, T812, E813, L814, T815, P816, E817, E818, R819, L820, L821, R822, A823, I824, F825, G826, E827, K828, A829, R830, E831, V832, R833, D834, T835, S836, L837, K838, V839, P840, H841, G842, E843, S844, G845, K846, V847, I848, G849, I850, R851, V852, F853, S854, R855, E856, D857, E858, D859, E860, L861, P862, A863, G864, V865, N866, E867, L868, V869, R870, V871, Y872, V873, A874, Q875, K876, R877, K878, I879, S880, D881, G882, D883, K884, L885, A886, G887, R888, H889, G890, N891, K892, G893, V894, I895, G896, K897, I898, L899, P900, V901, E902, D903, M904, P905, F906, L907, A908, D909, G910, T911, P912, V913, D914, I915, I916, L917, N918, T919, H920, G921, V922, P923, R924, R925, M926, N927, I928, G929, Q930, I931, L932, E933, T934, H935, L936, G937, W938, C939, A940, H941, S942, G943, W944, K945, V946, D947, A948, A949, K950, G951, V952, P953, D954, W955, A956, A957, R958, L959, P960, D961, E962, L963, L964, E965, A966, Q967, P968, N969, A970, I971, V972, S973, T974, P975, V976, F977, D978, G979, A980, Q981, E982, A983, E984, L985, Q986, G987, L988, L989, S990, C991, T992, L993, P994, N995, R996, D997, G998, D999, V1000, L1001, V1002, D1003,

G1306, D1307, K1308, L1309, A1310, G1311, R1312, H1313, G1314, N1315, K1316, G1317, V1318, I1319, G1320, K1321, I1322, L1323, P1324, V1325, E1326, D1327, M1328, P1329, F1330, L1331, A1332, D1333, G1334, T1335, P1336, V1337, D1338, I1339, I1340, L1341, N1342, T1343, H1344, G1345, V1346, P1347, R1348, R1349, M1350, N1351, I1352, G1353, Q1354, I1355, L1356, E1357, T1358, H1359, L1360, G1361, W1362, C1363, A1364, H1365, S1366, G1367, W1368, K1369, V1370, D1371, A1372, A1373, K1374, G1375, V1376, P1377, D1378, W1379, A1380, A1381, R1382, L1383, P1384, D1385, E1386, L1387, L1388, E1389, A1390, Q1391, P1392, N1393, A1394, I1395, V1396, S1397, T1398, P1399, V1400, F1401, D1402, G1403, A1404, Q1405, E1406, A1407, E1408, L1409, Q1410, G1411, L1412, L1413, S1414, C1415, T1416, L1417, P1418, N1419, R1420, D1421, G1422, D1423, V1424, L1425, V1426, D1427, A1428, D1429, G1430, K1431, A1432, M1433, L1434, F1435, D1436, G1437, R1438, S1439, G1440, E1441, P1442, F1443, P1444, Y1445, P1446, V1447, T1448, V1449, G1450, Y1451, M1452, Y1453, I1454, M1455, K1456, L1457, H1458, H1459, L1460, V1461, D1462, D1463, K1464, I1465, H1466, A1467, R1468, S1469, T1470, G1471, P1472, Y1473, S1474, M1475, I1476, T1477, Q1478, Q1479, P1480, L1481, G1482, G1483, K1484, A1485, Q1486, F1487, G1488, G1489, Q1490, R1491, F1492, G1493, E1494, M1495, E1496, C1497, W1498, A1499, M1500, Q1501, A1502, Y1503, G1504, A1505, A1506, Y1507, T1508, L1509, Q1510, E1511, L1512, L1513, T1514, I1515, K1516, S1517, D1518, D1519, T1520, V1521, G1522, R1523, V1524, K1525, V1526, Y1527, E1528, A1529, I1530, V1531, K1532, G1533, E1534, N1535, I1536, A1004, D1005, G1006, K1007, A1008, M1009, L1010, F1011, D1012, G1013, R1014, S1015, G1016, E1017, P1018, F1019, P1020, Y1021, P1022, V1023, T1024, V1025, G1026, Y1027, M1028, Y1029, I1030, M1031, K1032, L1033, H1034, H1035, L1036, V1037, D1038, D1039, K1040, I1041, H1042, A1043, R1044, S1045, T1046, G1047, P1048, Y1049, S1050, M1051, I1052, T1053, Q1054, Q1055, P1056, L1057, G1058, G1059, K1060, A1061, Q1062, F1063, G1064, G1065, Q1066, R1067, F1068, G1069, E1070, M1071, E1072, C1073, W1074, A1075, M1076, Q1077, A1078, Y1079, G1080, A1081, A1082, Y1083, T1084, L1085, Q1086, E1087, L1088, L1089, T1090, I1091, K1092, S1093, D1094, D1095, T1096, V1097, G1098, R1099, V1100, K1101, V1102, Y1103, E1104, A1105, I1106, V1107, K1108, G1109, E1110, N1111, I1112, P1113, E1114, P1115, G1116, I1117, P1118, E1119, S1120, F1121, K1122, V1123, L1124, L1125, K1126, E1127, L1128, Q1129, S1130, L1131, C1132, L1133, N1134, V1135, E1136, V1137, L1138, S1139, S1140, D1141, G1142, A1143, A1144, I1145, E1146, L1147, R1148, E1149, G1150, E1151, D1152, E1153

P1537, E1538, P1539, G1540, I1541, P1542, E1543, S1544, F1545, K1546, V1547, L1548, L1549, K1550, E1551, L1552, Q1553, S1554, L1555, C1556, L1557, N1558, V1559, E1560, V1561, L1562, S1563, S1564, D1565, G1566, A1567, A1568, I1569, E1570, L1571, R1572, E1573, G1574, E1575, D1576, E1577

Chain D: D1579, V1580, N1581, F1582, F1583, D1584, E1585, L1586, R1587, I1588, G1589, L1590, A1591, T1592, A1593, E1594, D1595, I1596, R1597, Q1598, W1599, S1600, Y1601, G1602, E1603, V1604, K1605, K1606, P1607, E1608, T1609, I1610, N1611, Y1612, R1613, T1614, L1615, K1616, P1617, E1618, K1619, D1620, G1621, L1622, F1623, C1624, E1625, K1626, I1627, F1628, G1629, P1630, T1631, R1632, D1633, W1634, E1635, C1636, Y1637, C1638, G1639, K1640, Y1641, K1642, R1643, V1644, R1645, F1646, K1647, G1648, I1649, I1650, C1651, E1652, R1653, C1654, G1655, V1656, E1657, V1658, T1659, R1660, A1661, K1662, V1663, R1664, R1665, E1666, R1667, M1668, G1669, H1670, I1671, E1672, L1673, A1674, A1675, P1676, V1677, T1678, H1679, I1680, W1681, Y1682, F1683, K1684, G1685, V1686, P1687, S1688, R1689, L1690, G1691, Y1692, L1693, L1694, D1695, L1696, A1697, P1698, K1699, D1700, L1701, E1702, K1703, I1704, I1705, Y1706, F1707, A1708, A1709, Y1710, V1711, I1712, T1713, S1714, V1715, D1716, E1717, E1718, M1719, R1720, H1721, N1722, E1723, L1724, S1725, T1726, L1727, E1728, A1729, E1730, M1731, A1732, V1733, E1734, R1735, K1736, A1737, V1738, E1739, D1740, Q1741, R1742, D1743, G1744, E1745, L1746, E1747, A1748, R1749, A1750, Q1751, K1752, L1753, E1754, A1755, D1756, L1757, A1758, E1759,	**Chain D**: D3, V4, N5, F6, F7, D8, E9, L10, R11, I12, G13, L14, A15, T16, A17, E18, D19, I20, R21, Q22, W23, S24, Y25, G26, E27, V28, K29, K30, P31, E32, T33, I34, N35, Y36, R37, T38, L39, K40, P41, E42, K43, D44, G45, L46, F47, C48, E49, K50, I51, F52, G53, P54, T55, R56, D57, W58, E59, C60, Y61, C62, G63, K64, Y65, K66, R67, V68, R69, F70, K71, G72, I73, I74, C75, E76, R77, C78, G79, V80, E81, V82, T83, R84, A85, K86, V87, R88, R89, E90, R91, M92, G93, H94, I95, E96, L97, A98, A99, P100, V101, T102, H103, I104, W105, Y106, F107, K108, G109, V110, P111, S112, R113, L114, G115, Y116, L117, L118, D119, L120, A121, P122, K123, D124, L125, E126, K127, I128, I129, Y130, F131, A132, A133, Y134, V135, I136, T137, S138, V139, D140, E141, E142, M143, R144, H145, N146, E147, L148, S149, T150, L151, E152, A153, E154, M155, A156, V157, E158, R159, K160, A161, V162, E163, D164, Q165, R166, D167, G168, E169, L170, E171, A172, R173, A174, Q175, K176, L177, E178, A179, D180, L181, A182, E183, L184, E185, A186, E187, G188, A189, K190, A191, D192, A193, R194, R195, K196, V197, R198, D199, G200, G201, E202, R203, E204, M205, R206, Q207, I208, R209, D210, R211, A212, Q213, R214, E215, L216, D217, R218, L219, E220, D221, I222, W223, S224, T225, F226, T227, K228, L229, A230, P231, K232, Q233, L234, I235, V236, D237, E238, N239, L240, Y241, R242, E243, L244, V245, D246, R247, Y248, G249, E250, Y251, F252, T253, G254, A255, M256, G257, A258, E259, S260, I261, Q262, K263,

L1760, E1761, A1762, E1763, G1764, A1765, K1766, L264, I265, E266, N267, F268, D269, I270, D271, A272, E273,
A1767, D1768, A1769, R1770, R1771, K1772, V1773, A274, E275, S276, L277, R278, D279, V280, I281, R282,
R1774, D1775, G1776, G1777, E1778, R1779, E1780, N283, G284, K285, G286, Q287, K288, K289, L290, R291,
M1781, R1782, Q1783, I1784, R1785, D1786, R1787, A292, L293, K294, R295, L296, K297, V298, V299, A300,
A1788, Q1789, R1790, E1791, L1792, D1793, R1794, A301, F302, Q303, Q304, S305, G306, N307, S308, P309,
L1795, E1796, D1797, I1798, W1799, S1800, T1801, M310, G311, M312, V313, L314, D315, A316, V317, P318,
F1802, T1803, K1804, L1805, A1806, P1807, K1808, V319, I320, P321, P322, E323, L324, R325, P326, M327,
Q1809, L1810, I1811, V1812, D1813, E1814, N1815, V328, Q329, L330, D331, G332, G333, R334, F335, A336,
L1816, Y1817, R1818, E1819, L1820, V1821, D1822, T337, S338, D339, L340, N341, D342, L343, Y344, R345,
R1823, Y1824, G1825, E1826, Y1827, F1828, T1829, R346, V347, I348, N349, R350, N351, N352, R353, L354,
G1830, A1831, M1832, G1833, A1834, E1835, S1836, K355, R356, L357, I358, D359, L360, G361, A362, P363,
I1837, Q1838, K1839, L1840, I1841, E1842, N1843, E364, I365, I366, V367, N368, N369, E370, K371, R372, M373,
F1844, D1845, I1846, D1847, A1848, E1849, A1850, L374, Q375, E376, S377, V378, D379, A380, L381, F382,
E1851, S1852, L1853, R1854, D1855, V1856, I1857, D383, N384, G385, R386, R387, G388, R389, P390, V391,
R1858, N1859, G1860, K1861, G1862, Q1863, K1864, T392, G393, P394, G395, N396, R397, P398, L399, K400,
K1865, L1866, R1867, A1868, L1869, K1870, R1871, S401, L402, S403, D404, L405, L406, K407, G408, K409,
L1872, K1873, V1874, V1875, A1876, A1877, F1878, Q410, G411, R412, F413, R414, Q415, N416, L417, L418,
Q1879, Q1880, S1881, G1882, N1883, S1884, P1885, G419, K420, R421, V422, D423, Y424, S425, G426, R427,
M1886, G1887, M1888, V1889, L1890, D1891, A1892, S428, V429, I430, V431, V432, G433, P434, Q435, L436,
V1893, P1894, V1895, I1896, P1897, P1898, E1899, K437, L438, H439, Q440, C441, G442, L443, P444, K445,
L1900, R1901, P1902, M1903, V1904, Q1905, L1906, L446, M447, A448, L449, E450, L451, F452, K453, P454,
D1907, G1908, G1909, R1910, F1911, A1912, T1913, F455, V456, M457, K458, R459, L460, V461, D462, L463,
S1914, D1915, L1916, N1917, D1918, L1919, Y1920, N464, H465, A466, Q467, N468, I469, K470, S471, A472,
R1921, R1922, V1923, I1924, N1925, R1926, N1927, K473, R474, M475, V476, E477, R478, Q479, R480, P481,
N1928, R1929, L1930, K1931, R1932, L1933, I1934, Q482, V483, W484, D485, V486, L487, E488, E489, V490,
D1935, L1936, G1937, A1938, P1939, E1940, I1941, I491, A492, E493, H494, P495, V496, L497, L498, N499,
I1942, V1943, N1944, N1945, E1946, K1947, R1948, R500, A501, P502, T503, L504, H505, R506, L507, G508,
M1949, L1950, Q1951, E1952, S1953, V1954, D1955, I509, Q510, A511, F512, E513, P514, M515, L516, V517,
A1956, L1957, F1958, D1959, N1960, G1961, R1962, E518, G519, K520, A521, I522, Q523, L524, H525, P526,
R1963, G1964, R1965, P1966, V1967, T1968, G1969, L527, V528, C529, E530, A531, F532, N533, A534, D535,
P1970, G1971, N1972, R1973, P1974, L1975, K1976, F536, D537, G538, D539, Q540, M541, A542, V543, H544,
S1977, L1978, S1979, D1980, L1981, L1982, K1983, L545, P546, L547, S548, A549, E550, A551, Q552, A553,
G1984, K1985, Q1986, G1987, R1988, F1989, R1990, E554, A555, R556, I557, L558, M559, L560, S561, S562,

Q1991, N1992, L1993, L1994, G1995, K1996, R1997, V1998, D1999, Y2000, S2001, G2002, R2003, S2004, V2005, I2006, V2007, V2008, G2009, P2010, Q2011, L2012, K2013, L2014, H2015, Q2016, C2017, G2018, L2019, P2020, K2021, L2022, M2023, A2024, L2025, E2026, L2027, F2028, K2029, P2030, F2031, V2032, M2033, K2034, R2035, L2036, V2037, D2038, L2039, N2040, H2041, A2042, Q2043, N2044, I2045, K2046, S2047, A2048, K2049, R2050, M2051, V2052, E2053, R2054, Q2055, R2056, P2057, Q2058, V2059, W2060, D2061, V2062, L2063, E2064, E2065, V2066, I2067, A2068, E2069, H2070, P2071, V2072, L2073, L2074, N2075, R2076, A2077, P2078, T2079, L2080, H2081, R2082, L2083, G2084, I2085, Q2086, A2087, F2088, E2089, P2090, M2091, L2092, V2093, E2094, G2095, K2096, A2097, I2098, Q2099, L2100, H2101, P2102, L2103, V2104, C2105, E2106, A2107, F2108, N2109, A2110, D2111, F2112, D2113, G2114, D2115, Q2116, M2117, A2118, V2119, H2120, L2121, P2122, L2123, S2124, A2125, E2126, A2127, Q2128, A2129, E2130, A2131, R2132, I2133, L2134, M2135, L2136, S2137, S2138, N2139, N2140, I2141, L2142, S2143, P2144, A2145, S2146, G2147, R2148, P2149, L2150, A2151, M2152, P2153, R2154, L2155, D2156, M2157, V2158, T2159, G2160, L2161, Y2162, Y2163, L2164, T2165, T2166, E2167, V2168, P2169, G2170, D2171, T2172, G2173, E2174, Y2175, Q2176, P2177, A2178, S2179, G2180, D2181, H2182, P2183, E2184, T2185, G2186, V2187, Y2188, S2189, S2190, P2191, A2192, E2193, A2194, I2195, M2196, A2197, A2198, D2199, R2200, G2201, V2202, L2203, S2204, V2205, R2206, A2207, K2208, I2209, K2210, V2211, R2212, L2213, T2214, Q2215, L2216, R2217, P2218, P2219, V2220, E2221, N563, N564, I565, L566, S567, P568, A569, S570, G571, R572, P573, L574, A575, M576, P577, R578, L579, D580, M581, V582, T583, G584, L585, Y586, Y587, L588, T589, T590, E591, V592, P593, G594, D595, T596, G597, E598, Y599, Q600, P601, A602, S603, G604, D605, H606, P607, E608, T609, G610, V611, Y612, S613, S614, P615, A616, E617, A618, I619, M620, A621, A622, D623, R624, G625, V626, L627, S628, V629, R630, A631, K632, I633, K634, V635, R636, L637, T638, Q639, L640, R641, P642, P643, V644, E645, I646, E647, A648, E649, L650, F651, G652, H653, S654, G655, W656, Q657, P658, G659, D660, A661, W662, M663, A664, E665, T666, T667, L668, G669, R670, V671, M672, F673, N674, E675, L676, L677, P678, L679, G680, Y681, P682, F683, V684, N685, K686, Q687, M688, H689, K690, K691, V692, Q693, A694, A695, I696, I697, N698, D699, L700, A701, E702, R703, Y704, P705, M706, I707, V708, V709, A710, Q711, T712, V713, D714, K715, L716, K717, D718, A719, G720, F721, Y722, W723, A724, T725, R726, S727, G728, V729, T730, V731, S732, M733, A734, D735, V736, L737, V738, P739, P740, R741, K742, K743, E744, I745, L746, D747, H748, Y749, E750, E751, R752, A753, D754, K755, V756, E757, K758, Q759, F760, Q761, R762, G763, A764, L765, N766, H767, D768, E769, R770, N771, E772, A773, L774, V775, E776, I777, W778, K779, E780, A781, T782, D783, E784, V785, G786, Q787, A788, L789, R790, E791, H792, Y793, P794, D795, D796, N797, P798, I799, I800, T801, I802, V803, D804, S805, G806, A807, T808, G809, N810, F811, T812, Q813, T814, R815, T816, L817, A818, G819, M820, K821, G822, L823, V824, T825, N826, P827, K828, G829, E830, F831, I832, P833, R834, P835, V836, K837, S838, S839, F840, R841, E842, G843, L844, T845, V846, L847, E848, Y849, F850, I851, N852, T853, H854, G855, A856, R857, K858, G859, L860, A861, D862,

I2222, E2223, A2224, E2225, L2226, F2227, G2228, H2229, S2230, G2231, W2232, Q2233, P2234, G2235, D2236, A2237, W2238, M2239, A2240, E2241, T2242, T2243, L2244, G2245, R2246, V2247, M2248, F2249, N2250, E2251, L2252, L2253, P2254, L2255, G2256, Y2257, P2258, F2259, V2260, N2261, K2262, Q2263, M2264, H2265, K2266, K2267, V2268, Q2269, A2270, A2271, I2272, I2273, N2274, D2275, L2276, A2277, E2278, R2279, Y2280, P2281, M2282, I2283, V2284, V2285, A2286, Q2287, T2288, V2289, D2290, K2291, L2292, K2293, D2294, A2295, G2296, F2297, Y2298, W2299, A2300, T2301, R2302, S2303, G2304, V2305, T2306, V2307, S2308, M2309, A2310, D2311, V2312, L2313, V2314, P2315, P2316, R2317, K2318, K2319, E2320, I2321, L2322, D2323, H2324, Y2325, E2326, E2327, R2328, A2329, D2330, K2331, V2332, E2333, K2334, Q2335, F2336, Q2337, R2338, G2339, A2340, L2341, N2342, H2343, D2344, E2345, R2346, N2347, E2348, A2349, L2350, V2351, E2352, I2353, W2354, K2355, E2356, A2357, T2358, D2359, E2360, V2361, G2362, Q2363, A2364, L2365, R2366, E2367, H2368, Y2369, P2370, D2371, D2372, N2373, P2374, I2375, I2376, T2377, I2378, V2379, D2380, S2381, G2382, A2383, T2384, G2385, N2386, F2387, T2388, Q2389, T2390, R2391, T2392, L2393, A2394, G2395, M2396, K2397, G2398, L2399, V2400, T2401, N2402, P2403, K2404, G2405, E2406, F2407, I2408, P2409, R2410, P2411, V2412, K2413, S2414, S2415, F2416, R2417, E2418, G2419, L2420, T2421, V2422, L2423, E2424, Y2425, F2426, I2427, N2428, T2429, H2430, G2431, A2432, R2433, K2434, G2435, L2436, A2437, D2438, T2439, A2440, L2441, R2442, T2443, A2444, D2445, S2446, G2447, Y2448, L2449, T2450, R2451, R2452, T863, A864, L865, R866, T867, A868, D869, S870, G871, Y872, L873, T874, R875, R876, L877, V878, D879, V880, S881, Q882, D883, V884, I885, V886, R887, E888, H889, D890, C891, Q892, T893, E894, R895, G896, I897, V898, V899, E900, L901, A902, E903, R904, A905, P906, D907, G908, T909, L910, I911, R912, D913, P914, Y915, I916, E917, T918, S919, A920, Y921, A922, R923, T924, L925, G926, T927, D928, A929, V930, D931, E932, A933, G934, N935, V936, I937, V938, E939, R940, G941, Q942, D943, L944, G945, D946, P947, E948, I949, D950, A951, L952, L953, A954, A955, G956, I957, T958, Q959, V960, K961, V962, R963, S964, V965, L966, T967, C968, A969, T970, S971, T972, G973, V974, C975, A976, T977, C978, Y979, G980, R981, S982, M983, A984, T985, G986, K987, L988, V989, D990, I991, G992, E993, A994, V995, G996, I997, V998, A999, A1000, Q1001, S1002, I1003, G1004, E1005, P1006, G1007, T1008, Q1009, L1010, T1011, X1012, X1013, G1026, G1027, L1028, P1029, R1030, V1031, Q1032, E1033, L1034, F1035, E1036, A1037, R1038, V1039, P1040, R1041, G1042, K1043, A1044, P1045, I1046, A1047, D1048, V1049, T1050, G1051, R1052, V1053, R1054, L1055, E1056, D1057, G1058, E1059, R1060, F1061, Y1062, K1063, I1064, T1065, I1066, V1067, P1068, D1069, D1070, G1071, G1072, E1073, E1074, V1075, V1076, Y1077, D1078, K1079, I1080, S1081, K1082, R1083, Q1084, R1085, L1086, R1087, V1088, F1089, K1090, H1091, E1092, D1093, G1094, S1095, E1096, R1097, V1098, L1099, S1100, D1101, G1102, D1103, H1104, V1105, E1106, V1107, G1108, Q1109, Q1110, L1111, M1112, E1113, G1114, S1115, A1116, D1117, P1118, H1119, E1120, V1121, L1122, R1123, V1124, Q1125, G1126, P1127, R1128, E1129, V1130, Q1131, I1132, H1133, L1134, V1135, R1136, E1137, V1138, Q1139, E1140, V1141, Y1142, R1143, A1144, Q1145, G1146, V1147, S1148, I1149, H1150, D1151, K1152, H1153, I1154, E1155,

L2453, V2454, D2455, V2456, S2457, Q2458, D2459, V2460, I2461, V2462, R2463, E2464, H2465, D2466, C2467, Q2468, T2469, E2470, R2471, G2472, I2473, V2474, V2475, E2476, L2477, A2478, E2479, R2480, A2481, P2482, D2483, G2484, T2485, L2486, I2487, R2488, D2489, P2490, Y2491, I2492, E2493, T2494, S2495, A2496, Y2497, A2498, R2499, T2500, L2501, G2502, T2503, D2504, A2505, V2506, D2507, E2508, A2509, G2510, N2511, V2512, I2513, V2514, E2515, R2516, G2517, Q2518, D2519, L2520, G2521, D2522, P2523, E2524, I2525, D2526, A2527, L2528, L2529, A2530, A2531, G2532, I2533, T2534, Q2535, V2536, K2537, V2538, R2539, S2540, V2541, L2542, T2543, C2544, A2545, T2546, S2547, T2548, G2549, V2550, C2551, A2552, T2553, C2554, Y2555, G2556, R2557, S2558, M2559, A2560, T2561, G2562, K2563, L2564, V2565, D2566, I2567, G2568, E2569, A2570, V2571, G2572, I2573, V2574, A2575, A2576, Q2577, S2578, I2579, G2580, E2581, P2582, G2583, T2584, Q2585, L2586, T2587, G2588, G2589, G2590, G2591, L2592, P2593, R2594, V2595, Q2596, E2597, L2598, F2599, E2600, A2601, R2602, V2603, P2604, R2605, G2606, K2607, A2608, P2609, I2610, A2611, D2612, V2613, T2614, G2615, R2616, V2617, R2618, L2619, E2620, D2621, G2622, E2623, R2624, F2625, Y2626, K2627, I2628, T2629, I2630, V2631, P2632, D2633, D2634, G2635, G2636, E2637, E2638, V2639, V2640, Y2641, D2642, K2643, I2644, S2645, K2646, R2647, Q2648, R2649, L2650, R2651, V2652, F2653, K2654, H2655, E2656, D2657, G2658, S2659, E2660, R2661, V2662, L2663, S2664, D2665, G2666, D2667, H2668, V2669, E2670, V2671, G2672, Q2673, Q2674, L2675, M2676, E2677, G2678, S2679, A2680, D2681, P2682, H2683, V1156, I1157, V1158, R1159, Q1160, M1161, L1162, R1163, R1164, V1165, T1166, I1167, I1168, D1169, S1170, G1171, S1172, T1173, E1174, F1175, L1176, P1177, G1178, S1179, L1180, I1181, D1182, R1183, A1184, E1185, F1186, E1187, A1188, E1189, N1190, R1191, R1192, V1193, V1194, A1195, E1196, G1197, G1198, E1199, P1200, A1201, A1202, G1203, R1204, P1205, V1206, L1207, M1208, G1209, I1210, T1211, K1212, A1213, S1214, L1215, A1216, T1217, D1218, S1219, W1220, L1221, S1222, A1223, A1224, S1225, F1226, Q1227, E1228, T1229, T1230, R1231, V1232, L1233, T1234, D1235, A1236, A1237, I1238, N1239, C1240, R1241, S1242, D1243, K1244, L1245, N1246, G1247, L1248, K1249, E1250, N1251, V1252, I1253, I1254, G1255, K1256, L1257, I1258, P1259, A1260, G1261, T1262, G1263, I1264, N1265, R1266, Y1267, R1268, N1269, I1270, A1271, V1272, Q1273, P1274, T1275, E1276, E1277, A1278, R1279, A1280, A1281

E2684, V2685, L2686, R2687, V2688, Q2689, G2690, P2691, R2692, E2693, V2694, Q2695, I2696, H2697, L2698, V2699, R2700, E2701, V2702, Q2703, E2704, V2705, Y2706, R2707, A2708, Q2709, G2710, V2711, S2712, I2713, H2714, D2715, K2716, H2717, I2718, E2719, V2720, I2721, V2722, R2723, Q2724, M2725, L2726, R2727, R2728, V2729, T2730, I2731, I2732, D2733, S2734, G2735, S2736, T2737, E2738, F2739, L2740, P2741, G2742, S2743, L2744, I2745, D2746, R2747, A2748, E2749, F2750, E2751, A2752, E2753, N2754, R2755, R2756, V2757, V2758, A2759, E2760, G2761, G2762, E2763, P2764, A2765, A2766, G2767, R2768, P2769, V2770, L2771, M2772, G2773, I2774, T2775, K2776, A2777, S2778, L2779, A2780, T2781, D2782, S2783, W2784, L2785, S2786, A2787, A2788, S2789, F2790, Q2791, E2792, T2793, T2794, R2795, V2796, L2797, T2798, D2799, A2800, A2801, I2802, N2803, C2804, R2805, S2806, D2807, K2808, L2809, N2810, G2811, L2812, K2813, E2814, N2815, V2816, I2817, I2818, G2819, K2820, L2821, I2822, P2823, A2824, G2825, T2826, G2827, I2828, N2829, R2830, Y2831, R2832, N2833, I2834, A2835, V2836, Q2837, P2838, T2839, E2840, E2841, A2842, R2843, A2844, A2845

Chain E: G2849, Y2850, D2851, T2852, P2853, L2854, G2855, I2856, T2857, N2858, P2859, P2860, I2861, D2862, E2863, L2864, L2865, D2866, R2867, V2868, S2869, S2870, K2871, Y2872, A2873, L2874, V2875, I2876, Y2877, A2878, A2879, K2880, R2881, A2882, R2883, Q2884, I2885, N2886, D2887, Y2888, Y2889, N2890, Q2891, L2892, G2893, E2894, G2895, I2896, L2897, E2898, Y2899, V2900, G2901, P2902, L2903,

Chain E: G28, Y29, D30, T31, P32, L33, G34, I35, T36, N37, P38, P39, I40, D41, E42, L43, L44, D45, R46, V47, S48, S49, K50, Y51, A52, L53, V54, I55, Y56, A57, A58, K59, R60, A61, R62, Q63, I64, N65, D66, Y67, Y68, N69, Q70, L71, G72, E73, G74, I75, L76, E77, Y78, V79, G80, P81, L82, V83, E84, P85, G86, L87, Q88, E89, K90, P91, L92, S93, I94, A95, L96, R97, E98, I99, H100, A101, D102, L103, L104, E105, H106, T107, E108

V2904, E2905, P2906, G2907, L2908, Q2909, E2910, K2911, P2912, L2913, S2914, I2915, A2916, L2917, R2918, E2919, I2920, H2921, A2922, D2923, L2924, L2925, E2926, H2927, T2928, E2929

Chain F: D2930, E2931, S2932, E2933, A2934, L2935, R2936, Q2937, A2938, R2939, K2940, D2941, A2942, E2943, L2944, T2945, A2946, S2947, A2948, D2949, S2950, V2951, R2952, A2953, Y2954, L2955, K2956, Q2957, I2958, G2959, K2960, V2961, A2962, L2963, L2964, N2965, A2966, E2967, E2968, E2969, V2970, E2971, L2972, A2973, K2974, R2975, I2976, E2977, A2978, G2979, L2980, Y2981, A2982, T2983, Q2984, L2985, M2986, T2987, E2988, L2989, S2990, E2991, R2992, G2993, E2994, K2995, L2996, P2997, A2998, A2999, Q3000, R3001, R3002, D3003, M3004, M3005, W3006, I3007, C3008, R3009, D3010, G3011, D3012, R3013, A3014, K3015, N3016, H3017, L3018, L3019, E3020, A3021, N3022, L3023, R3024, L3025, V3026, V3027, S3028, L3029, A3030, K3031, R3032, Y3033, T3034, G3035, R3036, G3037, M3038, A3039, F3040, L3041, D3042, L3043, I3044, Q3045, E3046, G3047, N3048, L3049, G3050, L3051, I3052, R3053, A3054, V3055, E3056, K3057, F3058, D3059, Y3060, T3061, K3062, G3063, Y3064, K3065, F3066, S3067, T3068, Y3069, A3070, T3071, W3072, W3073, I3074, R3075, Q3076, A3077, I3078, T3079, R3080, A3081, M3082, A3083, D3084, Q3085, A3086, R3087, T3088, I3089, R3090, I3091, P3092, V3093, H3094, M3095, V3096, E3097, V3098, I3099, N3100, K3101, L3102, G3103, R3104, I3105, Q3106, R3107, E3108, L3109, L3110, Q3111, D3112, L3113, G3114, R3115, E3116, P3117,

Chain F: D207, E208, S209, E210, A211, L212, R213, Q214, A215, R216, K217, D218, A219, E220, L221, T222, A223, S224, A225, D226, S227, V228, R229, A230, Y231, L232, K233, Q234, I235, G236, K237, V238, A239, L240, L241, N242, A243, E244, E245, E246, V247, E248, L249, A250, K251, R252, I253, E254, A255, G256, L257, Y258, A259, T260, Q261, L262, M263, T264, E265, L266, S267, E268, R269, G270, E271, K272, L273, P274, A275, A276, Q277, R278, R279, D280, M281, M282, W283, I284, C285, R286, D287, G288, D289, R290, A291, K292, N293, H294, L295, L296, E297, A298, N299, L300, R301, L302, V303, V304, S305, L306, A307, K308, R309, Y310, T311, G312, R313, G314, M315, A316, F317, L318, D319, L320, I321, Q322, E323, G324, N325, L326, G327, L328, I329, R330, A331, V332, E333, K334, F335, D336, Y337, T338, K339, G340, Y341, K342, F343, S344, T345, Y346, A347, T348, W349, W350, I351, R352, Q353, A354, I355, T356, R357, A358, M359, A360, D361, Q362, A363, R364, T365, I366, R367, I368, P369, V370, H371, M372, V373, E374, V375, I376, N377, K378, L379, G380, R381, I382, Q383, R384, E385, L386, L387, Q388, D389, L390, G391, R392, E393, P394, T395, P396, E397, E398, L399, A400, K401, E402, M403, D404, I405, T406, P407, E408, K409, V410, L411, E412, I413, Q414, Q415, Y416, A417, R418, E419, P420, I421, S422, L423, D424, Q425, T426, I427, G428, D429, E430, G431, D432, S433, Q434, L435, G436, D437, F438, I439, E440, D441, S442, E443, A444, V445, V446, A447, V448, D449, A450,

T3118, P3119, E3120, E3121, L3122, A3123, K3124, E3125, M3126, D3127, I3128, T3129, P3130, E3131, K3132, V3133, L3134, E3135, I3136, Q3137, Q3138, Y3139, A3140, R3141, E3142, P3143, I3144, S3145, L3146, D3147, Q3148, T3149, I3150, G3151, D3152, E3153, G3154, D3155, S3156, Q3157, L3158, G3159, D3160, F3161, I3162, E3163, D3164, S3165, E3166, A3167, V3168, V3169, A3170, V3171, D3172, A3173, V3174, S3175, F3176, T3177, L3178, L3179, Q3180, D3181, Q3182, L3183, Q3184, S3185, V3186, L3187, D3188, T3189, L3190, S3191, E3192, R3193, E3194, A3195, G3196, V3197, V3198, R3199, L3200, R3201, F3202, G3203, L3204, T3205, D3206, G3207, Q3208, P3209, R3210, T3211, L3212, D3213, E3214, I3215, G3216, Q3217, V3218, Y3219, G3220, V3221, T3222, R3223, E3224, R3225, I3226, R3227, Q3228, I3229, E3230, S3231, K3232, T3233, M3234, S3235, K3236, L3237, R3238, H3239, P3240, S3241, R3242, S3243, Q3244, V3245, L3246, R3247, D3248, Y3249, L3250, D3251

V451, S452, F453, T454, L455, L456, Q457, D458, Q459, L460, Q461, S462, V463, L464, D465, T466, L467, S468, E469, R470, E471, A472, G473, V474, V475, R476, L477, R478, F479, G480, L481, T482, D483, G484, Q485, P486, R487, T488, L489, D490, E491, I492, G493, Q494, V495, Y496, G497, V498, T499, R500, E501, R502, I503, R504, Q505, I506, E507, S508, K509, T510, M511, S512, K513, L514, R515, H516, P517, S518, R519, S520, Q521, V522, L523, R524, D525, Y526, L527, D528