Structural dynamic investigation of the mutation-induced resistance mechanisms of *Mycobacterium tuberculosis* DNA-directed RNA polymerase against Rifampicin

A thesis submitted in fulfilment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

IN BIOINFORMATICS

Rhodes University, South Africa

Department of Biochemistry, Microbiology and Bioinformatics

Faculty of Science

by

Mokgerwa Zacharia Monama

May 2024

Abstract

Emerging resistant strains of *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis (TB) disease, continue to plague mankind and reduce the efficacies of antitubercular therapies that have been an effective defence against TB for decades. More specifically, mutations located in the β subunit of the multisubunit *Mtb* RNA replicative machinery, RNA polymerase (RNAP), have been well established as the reason behind resistance to the first-line antitubercular drug rifampicin (RIF), which has resulted in therapeutic failure in several clinical cases. Additionally, elusive details pertaining to the underlying mechanisms associated with RIF resistance due to the presence of *Mtb*-RNAP-β mutations, have resulted in setbacks in the development of novel and effective drugs that might be able to curb the ongoing threat. Hence, in this investigation, we attempted to resolve the involved *Mtb*-RNAP structural events at the molecular level to discern potentially important details regarding the nine clinically relevant *Mtb*-RNAP-β missense mutations under investigation.

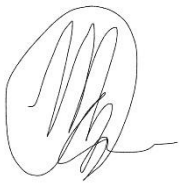
Hence, for the first time, we conducted an *in-silico* RIF resistance investigation using the *Mtb*-RNAP complex. To accomplish the set-out task, we first employed the use of more traditional post-MD analytical approaches such as root mean square deviation, root mean square fluctuation, radius of gyration, center of mass distance analyses, hydrogen bond occupancies, and binding free energy calculations, to conduct a global analysis of the mutated *Mtb*-RNAP proteins referencing RIF efficacy. Our findings revealed that the mutations may have a perturbation effect resulting in the disruption of essential structural dynamics attributed to the protein's catalytic functions. This was for instance observed for the βfork loop 2 domain, the β'zinc-binding domain, the β' trigger loop domain, and the β'jaw domain, which happen to be in line with previously reported experiments detailing changes in RNAP processivity. Complementarily,

some of the mutations more specifically perturbed the RIF binding pocket (RIF-BP) which observably led to the reorientation of RIF from the native or active orientation needed to obstruct the processive addition of nucleoside triphosphates to the growing RNA transcript. The mutation-induced repositioning from the active RIF orientation was also reflected through the loss of essential interactions between RIF and the RIF-BP along with the loss of binding affinities captured for a majority of the mutant proteins.

In conjunction with traditional analytical approaches, we further employed computational alanine scanning, weighted contact map analyses, and dynamic residue network (DRN) analyses, a novel approach that delineates residue-residue communication pathways through several metrics, to further elucidate how a set of clinically relevant mutations affect *Mtb*-RNAP function. With that, we were able to observe several key changes in residue importance and interactions that may be instrumental in bringing about RIF resistance and the compensatory conformational changes we observed among the *mt* systems through global analysis. Furthermore, we identified persistent hubs that may be particularly important in maintaining transcriptional activities in the presence and absence of the investigated mutations and RIF that could serve as potential resistance markers for future therapeutic investigations. We believe these findings will significantly aid future efforts in the discovery of new treatment options with the potential to overcome antitubercular resistance.

Declaration

I, Mokgerwa Zacharia Monama, hereby declare that the contents of this thesis are my original work. I further declare that the thesis being submitted for the degree of Doctor of Philosophy at Rhodes University has not been previously submitted for any degree or examination at any other institution.



27/05/2024

Signature **Date**

Acknowledgments

- Firstly, I would like to give my sincere thanks and appreciation for the continuous advice, support, and motivation that my supervisor Prof. Özlem Tastan Bishop gave me throughout my PhD journey. Without this incredible opportunity I would not be where I am in life, and I cannot even begin to express how truly thankful I am.
- Secondly, I would like to extend my thanks to my co-supervisor Dr. Fisayo Olotu whose assistance, dedication, and encouragement inspired me to progress as a researcher.
- Thirdly, I would like to thank Dr. Thommas Musyoka, Dr. Olivier Sheik Amamuddy, and Dr. Taremekedzwa Allan Sanyanga for their valuable insight throughout my PhD.
- I would also like to give a big thank you to the Centre for High-Performance Computing (CHPC) in Cape Town, South Africa, for their computational resources.
- Finally, I would like to thank the National Research Foundation (NRF) for their generous funding of my PhD degree (Grant number 121287).

Dedication

I dedicate this thesis to my family.

Table of contents

List of figures

List of tables

List of supplementary figures

List of supplementary tables

List of supplementary text

Abbreviations

AM	Alveolar macrophage
BC	*Betweenness centrality*
CASM	Computational alanine scanning
CC	*Closeness centrality*
CoM	Center of mass
DC	*Degree centrality*
DNA	Deoxyribonucleic acid
ED	Essential dynamics
FL2	Fork loop 2
NRF2	Erythroid factor 2-related factor 2
HIV	Human immunodeficiency virus
KC	*Katz centrality*
MD	Molecular dynamics
MDR-TB	Multi-drug resistant tuberculosis
mt	Mutant
Mtb	*Mycobacterium tuberculosis*
PCA	Principal component analysis
RIF	Rifampicin
Rg	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
RNAP	RNA polymerase

ROS	Reactive oxygen species
RR-TB	Rifampicin resistant tuberculosis
TB	Tuberculosis
WHO	World Health Organization
wt	Wild type
XDR-TB	Extremely drug-resistant TB
ZBD	N-terminal zinc-binding domain

Research outputs

Monama, M.Z., Olotu, F. and Tastan Bishop, Ö., 2023. Investigation of Multi-Subunit Mycobacterium tuberculosis DNA-Directed RNA Polymerase and Its Rifampicin Resistant Mutants. International Journal of Molecular Sciences, 24(4), p.3313.

Thesis overview

The research detailed in this thesis explores the complex relationship between β subunit-located clinically relevant missense mutations of the *Mtb*-RNAP proteins and the resulting structural dynamics in an attempt to elucidate their mechanisms of RIF resistance at the molecular level. The thesis consists of a total of 6 chapters. **Chapter 1** describes the background of the *Mtb* organism, its prevalence based on the latest statistics, the symptoms and available treatment options for those suffering from TB, and details covering the ongoing threat of resistance. Additionally, this chapter delves into the first-line drug RIF, its target, *Mtb*-RNAP, and associated resistance factors. **Chapter 2** explores the knowledge gap that inspired the study, including the study's aims and objectives, and hypotheses that informed the subsequent analyses. **Chapter 3** explores the various concepts and methodologies, inclusive of the online databases, servers, and tools, employed throughout the study to tackle the aim. **Chapter 4** involves the application of traditional post-MD analytical approaches to conduct the global analysis of the RIF-bound and RIF-unbound *wt* (wild-type) and *mt* (mutant) *Mtb*-RNAP systems to investigate the dynamic structure-associated changes. **Chapter 5** describes the use of local analytical approaches to explore the residue-residue communication paths and additional residue-level effects induced by the mutations and/or RIF on the *wt* and *mt Mtb*-RNAP proteins. **Chapter 6** serves as a summary of the findings made throughout the *in-silico* investigation.

Chapter 1
Study Background

Chapter overview

The following chapter covers background information about tuberculosis (TB) disease and details current TB projections, detection techniques, treatment plans, preventative measures, novel antitubercular drugs, as well as antibiotic resistance and its many implications for human health. Furthermore, the focus of our study on the antitubercular drug, rifampicin, and its target, the bacterial RNA polymerase (RNAP), has been included herein.

1.1. Introduction

1.1.1. Tuberculosis

Across history, major infectious diseases affecting human health have been observed to correlatively increase with advancements in agriculture (Dobson and Robin Carper, 1996; Hardin and Diamond, 1997; Wolfe, Dunavan and Diamond, 2007). This is mainly because the increased control over food production within a given population allows that population to grow immensely in numbers, giving rise to the possibility of pathogens from various sources spreading and perhaps even becoming endemic in the resulting populace (Wolfe, Dunavan and Diamond, 2007). TB disease, which is responsible for over a billion lives lost to date, is no exception to this established trend (Paulson, 2013; Barberis *et al.*, 2017; Khanna, Saha and Ahmad, 2023). Moreover, the disease has plagued humanity for thousands of years, with mentions of tell-tale symptoms appearing in multiple historical records in ancient civilizations (Comas *et al.*, 2013; Barberis *et al.*, 2017). TB is primarily spread through the inhalation of respiratory droplets (expelled through a cough) harbouring the pathogenic Actinomycete, *Mycobacterium tuberculosis* (*Mtb*) (Barberis *et al.*, 2017). Although TB

is one of the oldest diseases on record with the origins of the causative agent believed to date back to over 150 million years ago, *Mtb* was only successfully cultivated for the first time under laboratory settings in 1882 by Robert Koch, with the help of Paul Ehrlich (Hayman, 1984; Gradmann, 2001). Infection by *Mtb* can result in either the active or transmissible form of the disease (active TB), which entails the active growth of infecting *Mtb* bacilli and requires immediate medical treatment; or the latent and asymptomatic form of the disease (latent TB), which may affect ~1/4 of the world's population (serving as a reservoir for future infections), play a role in comorbidity, and is generally characterized by slow-growing or non-replicating *Mtb* inside granulomas (O'Garra *et al.*, 2013; Liu *et al.*, 2020). The disease usually manifests in the lungs (pulmonary TB). However, sites of infection can include, but are not limited to, pleura, bones and joints, and the central nervous system (Golden and Vikram, 2005).

1.1.2. Statistical information about the prevalence of TB

Although the prevalence of TB has seen a slow decline over recent years in some high-burden regions (WHO, 2021), the devastating impact of the COVID-19 pandemic and its disruption of TB-related healthcare services resulted in reduced diagnoses and reports of affected patients, likely resulting in increases in local transmission due to undiagnosed and untreated cases (WHO, 2023). Contrastingly, the establishment of lockdowns which restricted human contact worldwide, may have had the opposite effect, making it particularly difficult to estimate the number of TB-related incidents between 2020 and 2021, especially when considering the limitations of applied predictive models (WHO, 2023). Regardless, in 2022 alone, TB was ranked second only to COVID-19 as the leading cause of death by a single infectious agent globally. It was estimated that TB caused the deaths of over 1.3 million individuals worldwide, almost doubling that of HIV (0.63 million). Among the individuals that accounted for

the estimated deaths, the non-HIV related fatalities comprised ~587,000 (52%) men, ~365000 (32%) women, and ~183,000 (16%) children, whereas the HIV-infected fatalities comprised ~78,000 (47%) men, ~58,000 (35%) women, and 16,000 (18%) children. Additionally, the global incidence of TB infections rose by ~300,000 thousand in 2022 from the previous year (WHO, 2023).

1.1.3. Mtb life cycle

As previously stated, pathogenic *Mtb* generally infects a new host through cough-generated aerosols produced by an already infected person (**Figure 1.1**) (Barberis *et al.*, 2017). Furthermore, given *Mtb*'s capacity to evade the innate immune response, the bacterium is especially transmissible with a shockingly low infectious dose of ~3 bacilli (Ehrt and Schnappinger, 2009; Corleis *et al.*, 2012; Donald *et al.*, 2018). It's also worth noting that infections tend to be contained (latent TB) by an effective adaptive immune response but may lead to either the active form of the disease or to *Mtb* clearance and subclinical TB disease (Henry Boom, Schaible and Achkar, 2021). Although specific molecular details of how *Mtb* can overcome the host's immune defences from early to late stages remain elusive, some critical details about its life cycle have been delineated through a mouse model (Flynn, 2006; Rothchild *et al.*, 2019; Chandra, Grigsby and Philips, 2022). For instance, it is clear that once in the lungs, alveolar macrophages (AMs) are the first phagocytic cells to be infected by *Mtb* and tend to be especially targeted for intracellular growth, highlighting a key attribute in the bacterium's virulence (Cohen *et al.*, 2018; Rothchild *et al.*, 2019; Chandra, Grigsby and Philips, 2022). AMs are the most numerous macrophages in the lungs and they serve as sentinels that constantly sample foreign airway-related particles to avoid any unnecessary pro-inflammatory responses, making them a critical line of defence against airborne antigens, crucial for inducing an adaptive immune response

when necessary (Hussell and Bell, 2014). In contrast however, once the *Mtb* bacilli are phagocytosed by the AMs, it triggers the up-regulation of a cell-preservation transcriptional antioxidant, the nuclear erythroid factor 2-related factor 2 (NRF2), consequently allowing for further progression of the infection and informing the characteristic delayed response to the ensuing bacterial burden (Rothchild *et al.*, 2019). Within ~2 weeks of the initial infection, the *Mtb*-infected AMs translocate to the pulmonary interstitium as mediated by the signalling of the host's interleukin-1 (IL-1) on epithelial cells and the bacteria's ESX-1 type VII secretion system (Cohen *et al.*, 2018). Once there, the *Mtb* causes necrosis of the infected cells, inevitably releasing the bacilli and resulting in the further recruitment of immune cells such as dendritic cells, monocytes, neutrophils, and additional macrophages, leaving more immune cells infected by *Mtb* (Cohen *et al.*, 2018). Though the neutrophils attempt to control the bacterial burden by releasing reactive oxygen species (ROS), *Mtb* is able to neutralize these molecules through a catalase-peroxidase (KatG) and superoxide dismutase (SOD), as well as inhibit host cell apoptosis and the production of ROS using the type I NADH dehydrogenase (NuoG) (Stallings and Glickman, 2010; Corleis *et al.*, 2012; Bruns and Stenger, 2014). *Mtb*'s ability to survive in such diverse cellular environments that would destroy most pathogens is the direct result of its co-evolution with humans, making it especially adept at evading the host's natural defences (M Cristina *et al.*, 2005). For instance, lysosomal trafficking pathways and acidic phagosomal pH can be respectively avoided through the inefficient effects of human interferon-γ (at least partly) on macrophages and by the *Mtb*-released terpene, 1-tuberculosinyladenosine (Baker, Dechow and Abramovitch, 2019; Ghanavi *et al.*, 2021). As the disease progresses, immune cells including AMs and monocyte-derived macrophages launch a more aggressive pro-inflammatory response applying

additional stress on the bacilli (Lee *et al.*, 2020). Matured dendritic cells that have engulfed the bacilli move to the draining lymph nodes where *Mtb* antigens are used to prime naive T cells ($CD4^+$ and $CD8^+$), effectively initiating an adaptive immune response (Bodnar, Serbina and Flynn, 2001). Inflammatory chemokines produced by the *Mtb*-infected cells lead to the coalescence of additionally recruited immune cells around the infected phagocytes, forming a mass of encapsulating cells called a granuloma (Pagán and Ramakrishnan, 2018). If *Mtb* is able to cause sufficient granuloma necrosis, the structure may eventually break open and release several thousand infectious bacilli into the airways forming aerosols that will eventually be released to infect new hosts (Russell, 2007; Pagán and Ramakrishnan, 2018).

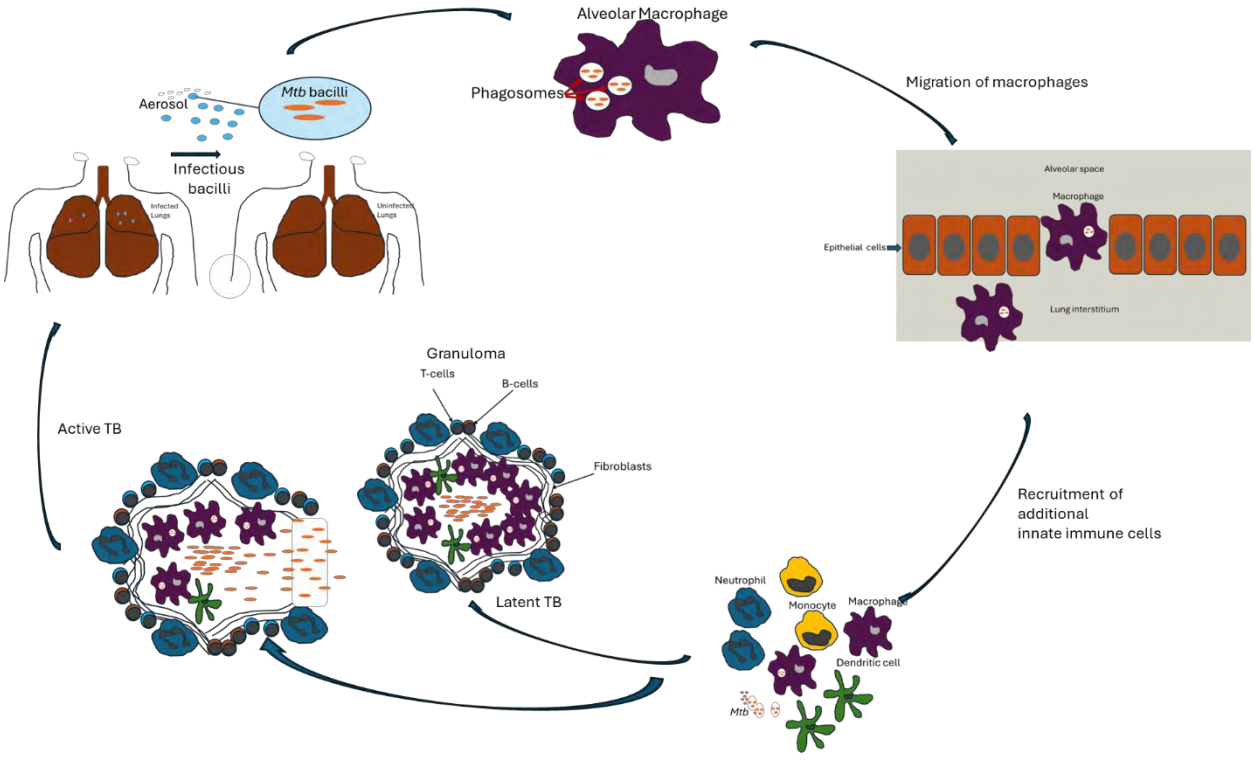


Figure 1.1. Summary of the *Mtb* life cycle.

1.1.4. Detection of TB

Chest radiographs, medical history, and the assessment of symptoms of pulmonary TB which include a chronic cough, coughing up blood (haemoptysis), fatigue, fever or chills, night sweats, shortness of breath, tightness of the chest and weight loss are generally used by clinicians to screen for suspected TB disease (Assefa *et al.*, 2019). However, extrapulmonary TB may present with different symptoms depending on the site of infection (Golden and Vikram, 2005). Early detection of TB along with the accuracy of diagnostic tools used to determine the type and/or severity of the infection can therefore immensely contribute to the quality of a patient's medical treatment plan (Acharya *et al.*, 2020). In this regard, the World Health Organization (WHO) endorses diagnostic approaches such as the rapid molecular test Xpert MTB/RIF® assay which produces results within ~2 hours of testing (Steingart *et al.*, 2014); the sputum smear microscopy test developed over a century ago whose results can be ready on the same day; and a liquid culture-based method which is the most preferred method in terms of accuracy and entails the confirmation of *Mtb* presence or absence through a cultured sample in selective media (World Health Organization, 2018; Acharya *et al.*, 2020). Besides testing for *Mtb* infections, some diagnostic tools including the Xpert MTB/RIF® assay and the culture-based method further test for drug resistance, a growing threat in the treatment of TB (World Health Organization, 2018; Acharya *et al.*, 2020).

1.1.5. The antitubercular drug, rifampicin; its receptor, Mtb-RNAP; and receptor-associated mutations

1.1.5.1. Rifampicin

Rifampicin (RIF) (**Figure 1.2**), also referred to as rifampin, is a highly potent and cyclic semisynthetic drug derived from rifamycins and composes of an ansa bridge, a

naphthol ring and a methylated tail (Sensi, 1983). RIF's antitubercular activities and dose-dependent bactericidal effects were confirmed through several *in vitro* and *in vivo* investigations in the late 1960s to early 1970s during which several patients experiencing chronic resistance to other drugs were treated successfully (Grobbelaar *et al.*, 2019). The first-line drug has therefore been used to treat TB since the late 20th century and has shown broad-spectrum antimicrobial activities, making it an especially effective therapy against a range of other infections (Forrest and Tamura, 2010; Piccaro *et al.*, 2013; Dartois and Rubin, 2022).

RIF's mode of action involves the specific binding of RIF to the β subunit interface ~12 Å away from the bacterial DNA-directed RNA polymerase (RNAP) active site to induce a steric occlusion effect on 2-3 nucleotide long nascent RNAs on the 5' end, effectively inhibiting the transcription process and killing the infecting bacilli (Campbell *et al.*, 2001; Lin *et al.*, 2017). Additionally, the protein-drug interaction dynamics were recently assessed using the crystallized structures and certain RNAP residues such as the β residues Q435, Q438, F439, R454, and S456 have been linked with RIF's active orientation which is particularly important given how the drug performs its inhibitory action (Lin *et al.*, 2017). An important aspect of RIF's activities is that it is capable of killing slowly-replicating and non-replicating bacteria, a fundamental property important in *Mtb*'s virulence, making RIF's sterilizing activity essential for shortening TB treatment (Xie, Siddiqi and Rubin, 2005; Sala et al., 2010; Piccaro et al., 2013; Koch, Mizrahi and Warner, 2014).

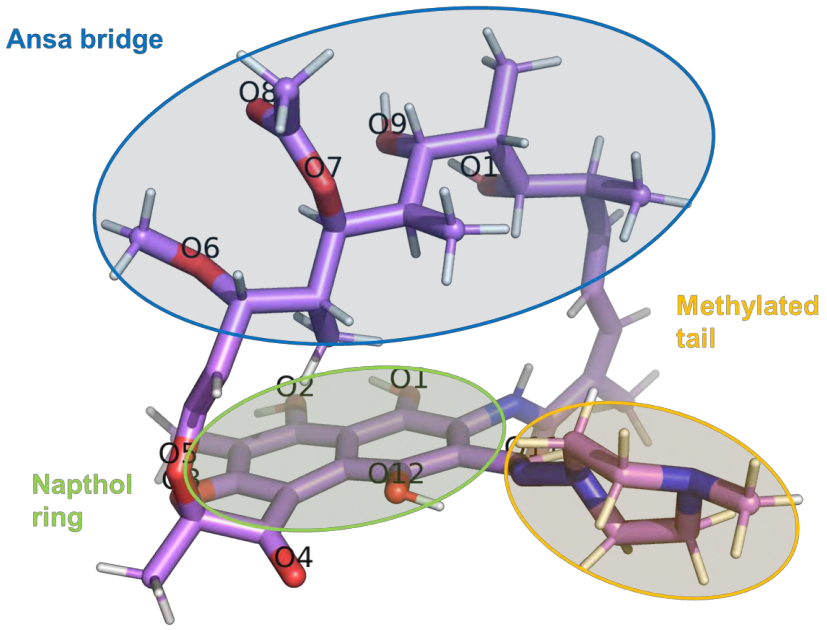


Figure 1.2. 3D structural representation of rifampicin (RIF). The main regions of the drug (in licorice representation) are encircled and annotated. The oxygen atoms (red) are further annotated. Nitrogen and hydrogen atoms are colored blue and white respectively.

1.1.5.2. *Mtb* RNAP, the RIF drug target

Mtb RNAP is a large multisubunit enzyme (exceeding 3000 amino acid residues) (**Figure 1.3A**) that plays a central role in gene expression, making it an attractive drug target. This has made elucidation of RNAP's complex structural dynamics (in relation to its function) a topic of great interest, necessitating additional investigations (Chen, Boyaci and Campbell, 2021). Structurally, the enzyme's catalytic core is made up of an α homodimer (αI and αII), β, β', and ω subunits encoded by *rpoA*, *rpoB*, *rpoC*, and *rpoZ* genes respectively (Lane and Darst, 2010b). However, initiating transcription requires further incorporation of a σ factor to complete the holoenzyme assembly and form the closed complex (RPc) capable of promoter-DNA recognition, followed by the transition to the transcriptionally competent open complex (RPo) obtained through a series of isomerization steps that promote DNA unwinding and formation of a bubble

~12 nucleotides long (Travers and Burgess, 1969; Chamberlin and J., 1976; Mooney, Darst and Landick, 2005). Among the bacterium's repertoire of σ factors, it usually has at least one primary σ factor that transcribes housekeeping genes essential for cellular viability under normal conditions, while the expression of the rest of the σ factors is regulated by specific stimuli in response to various environmental factors (Paget and Helmann, 2003; Flentie, Garner and Stallings, 2016). Additionally, these σ factors (also referred to as σ^{70} factors) are generally separated into four groups based on phylogeny and structure, with group 1 having all four domains (Campbell *et al.*, 2002; Paget and Helmann, 2003). *Mtb* in particular has an impressive repertoire of 10 alternative σ factors with σ^{A} (encoded by *rpoD*) as the primary σ factor (Rodrigue *et al.*, 2006). RNAP's β and β' subunits not only form the composite pincers of the characteristic crab-claw shape; they also encapsulate the DNA-binding channel (primary channel) through which DNA can reach the active site to produce nascent RNA (Basu et al., 2014; Sutherland and Murakami, 2018). The formed transcription bubble triggers DNA-template-directed RNA synthesis which results in the generation of a covalent bond between the 3'OH group of the nascent RNA and α-phosphate of the incoming NTP substrate (which enters the active site through the secondary channel), causing the addition of a nucleoside diphosphate (NDP) to the growing RNA and the release of a pyrophosphate by-product (Brueckner, Ortiz and Cramer, 2009; Stephanie, Tambunan and Siahaan, 2022). To accomplish promoter DNA escape and transition into the elongation complex, DNA scrunching occurs by dragging more of the template DNA into the transcription bubble whilst maintaining promoter interactions (Winkelman and Gourse, 2017). This causes tension to build up until the promoter is ultimately released and the σ subunit dissociates unless it induces abortive initiation, in which case the RNAP would successively revert to the RPo state, releasing short

RNA molecules until a successful transition to the elongation complex occurs (Winkelman and Gourse, 2017). The elongation phase involves growing the RNA chain through maintained DNA melting and template-directed NTP base pairing in the 3' to 5' direction to form a DNA-RNA hybrid, whereas the termination phase can occur either intrinsically or through a factor-dependent manner when RNAP reaches a termination sequence (Vassylyev, 2009; Roberts, 2019). RNAP structure is ubiquitously conserved over all the domains of life, with some discrepancies found in domain-specific sequence insertions (Ebright, 2000; Darst, 2001; Cramer, 2002; Lane and Darst, 2010a). It's worth noting that most biochemical studies surrounding the dynamics of bacterial RNAP function were generally undertaken using the evolutionarily distinct *Escherichia coli's* (*E. coli*) RNAP as a model of inference (Boyaci, Saecker and Campbell, 2020). However, recent studies have changed this narrative in an attempt to expand on the species-specific structural dynamics of *Mtb* RNAP (Boyaci, Saecker and Campbell, 2020). For instance, the essential transcription factors CarD and RbpA are unique to mycobacteria and absent in *E. coli* (Boyaci, Saecker and Campbell, 2020).

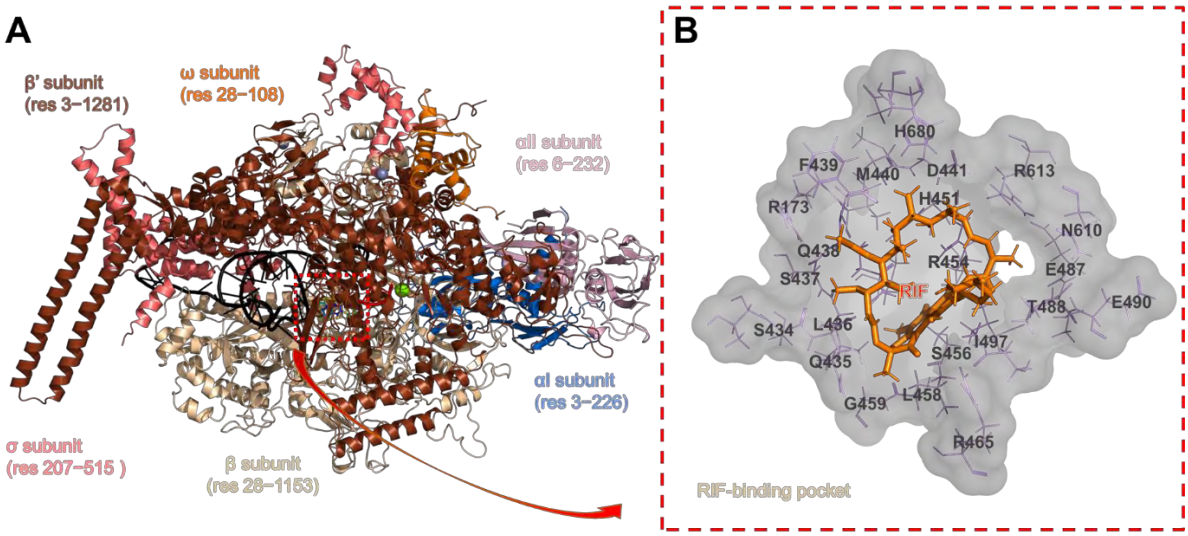


Figure 1.3. Structural representation of multisubunit *Mtb*-RNAP and RIF interactions. (A) The 3D structure of *Mtb*-RNAP is represented as a cartoon and colored blue, green, purple, brown, orange, and teal for the αI, αII, β, β', ω, and σ subunits respectively. (B) A 3D representation of RIF (orange in stick representation) with the associated RIF-binding pocket (RIF-BP) residues. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

1.1.5.3. *Mtb*-RNAP and its functional domains

As mentioned above, the main function of template DNA-directed RNAP is the catalytic addition of an incoming NTP to the 3'OH end of the growing nascent RNA (Brueckner, Ortiz and Cramer, 2009). This would then be followed by the translocation of the RNA-DNA assembly to accommodate successive NTP addition at the insertion site, a process referred to as the nucleotide addition cycle (Brueckner, Ortiz and Cramer, 2009). To accomplish this, the RNAP functional domains, not limited to those outlined in **Figure 1.4**, work in concert to ensure transcriptional fidelity and proficiency.

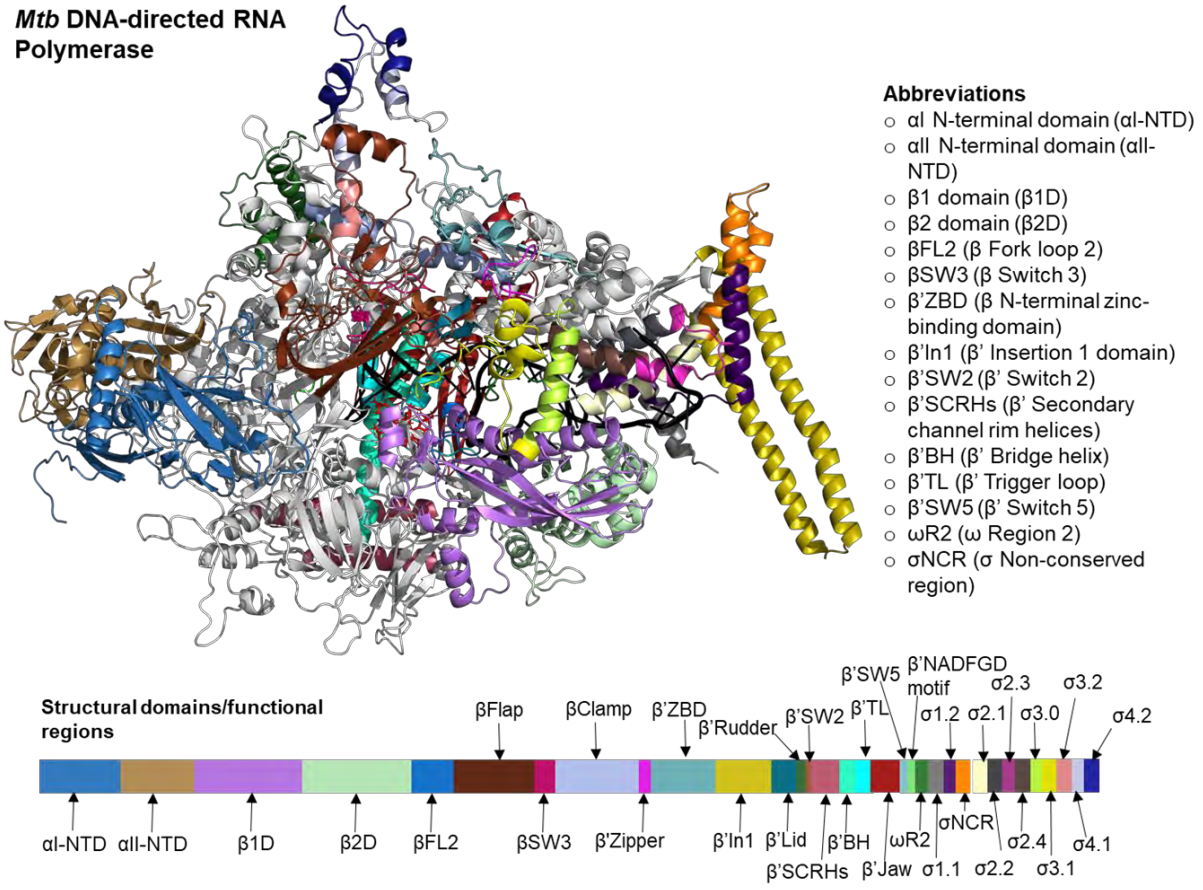


Figure 1.4. 3D structure of mapped *Mtb*-RNAP structural domains. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

1.1.5.3.1. α N-terminal domains

As previously mentioned, *Mtb*-RNAP consists of two α-NTDs (residues 1–347), namely, the αI-NTD and αII-NTD. These homodimers are responsible for initiating RNAP assembly and forming the platform with which the much larger β and β' subunits interact during the assembly process (Wu and Oen, 1977; Zhang and Darst, 1998). Furthermore, the αR40 and αL43 along with αG103 and αK191 were determined to be key interacting residues with the β and β' subunits for the purpose of maintaining a stable assembly (Zhang and Darst, 1998). The α-NTD has been divided into 2 subdomains, namely, domain 1 which consists of two regions (henceforth referred to

as R1 and R2; residues 1–47 and 170–235 respectively), and the transitional region, referred to as domain 2 (residues 48–170) (Zhang and Darst, 1998).

1.1.5.3.2. β1 and β2 domains

Some of the larger *Mtb*-RNAP domains, the β1 domain (β1D) (residues 56–169 and 385–437) and β2 domain (β2D) (residues 181–370), otherwise known as the protrusion and lobe, fold over the RNA-DNA assembly, structurally forming a channel within the active centre's transcription bubble that guides single-stranded non-template DNA (Lane and Darst, 2010b). The β1D also forms part of the RIF-binding pocket (RIF-BP) and therefore plays a role in *Mtb*'s drug resistance, while the β2D forms additional interactions with downstream DNA (Lane and Darst, 2010b; Monama, Olotu and Tastan Bishop, 2023).

1.1.5.3.3. Fork loop 2 domain

The fork loop 2 (βFL2) (residues 459–473) is a highly mobile domain that serves to maintain the transcription bubble by preventing the reannealing of the separated strands of DNA and interfering with the downstream edge of the double-stranded DNA (Gnatt *et al.*, 2001; Meyer *et al.*, 2009; Lane and Darst, 2010b). Furthermore, βFL2 has been shown to exhibit multiple conformations which may be associated with the different energetic stages within the elongation process (Meyer *et al.*, 2009).

1.1.5.3.4. βFlap domain

The βflap domain (residues 752–874), also referred to as the "wall" in eukaryotic RNAP II, forms key structural elements of the active site and plays a critical role in all phases of the transcription process (Yang *et al.*, 2009; Lane and Darst, 2010b). The βflap domain also forms part of the RNA exit channel; the channel through which nascent RNA leaves RNAP during the elongation phase (Yang *et al.*, 2009; Lane and

Darst, 2010b). Furthermore, the βflap consists of the βflap tip (residues 810–831), which plays an essential role in transcriptional pausing through interactions with the RNA pause hairpin (nascent RNA structural element that forms in the RNA exit channel) (Toulokhonov and Landick, 2003); and the βflap tip helix (residues 816–827), which is particularly conserved in bacteria and interacts with the σ4 domain to spatially position it to correctly interact with the –10 and –35 promoter elements during initiation (Kuznedelov *et al.*, 2002; Lane and Darst, 2010b).

1.1.5.3.5. Switch regions

The switch region serves as a hinge that allows the necessary conformational changes needed to accommodate DNA into the RNAP active center cleft (Molodtsov *et al.*, 2015). The switches are further segmented into 5 subregions, namely switch 1 (β'SW1) (residues 1208-1230), β'switch 2 (β'SW2) (residues 405–423), βswitch 3 (βSW3) (residues 1046–1067), switch 4 (βSW4) (residues 1089–1098) and switch 5 (β'SW5) (residues 1257–1259), with β'SW2, βSW3 and β'SW5 indicating particular importance in RNAP function (Lane and Darst, 2010b; Molodtsov *et al.*, 2015).

1.1.5.3.6. Clamp module

The clamp domain (β residues 1093–1136, and β' residues 5–137, 254–418 and 1222–1245) mediates promoter recognition and DNA melting by switching between open and closed conformations respectively to accommodate the loading of DNA into the active center cleft during the initiation phase and to keep it loaded for subsequent stages of the transcription process (Feklistov *et al.*, 2017; Morichaud *et al.*, 2023).

1.1.5.3.7. β'Zipper domain

The β'zipper domain (residues 32–44) is a highly conserved loop among bacterial RNAPs and forms part of the RNA exit channel (Borukhov and Nudler, 2008). The

β'zipper has been found to form important interactions with the DNA segment between the –10 and –35 promoter elements, otherwise referred to as the promoter spacer, contributing to the formation of the RPc and potentially the RPo as well (Yuzenkova *et al.*, 2011).

1.1.5.3.8. β'ZBD

The β' N-terminal zinc-binding domain (β'ZBD) (residues 24–98) plays several roles during each phase of the process of transcription, some of which include assisting in the formation of a stable RNAP holoenzyme assembly with the σ^{54} during intuition, forming important interactions with the promoter spacer in σ^{38}, interactions with transcription regulators such as RbpA, regulation of transcriptional pausing, along with termination and antitermination, among other functions (Hu and Liu, 2022).

1.1.5.3.9. β'In1 domain

The taxon-specific β' *Mtb* sequence insertion (residues 140–229) structurally forms an α-helical coiled-coil. It is more commonly referred to as the β' insertion 1 domain (β'In1) and extends from the clamp domain. The β'In1 has been associated with the role of a gate meant to securely trap DNA into the active center cleft (Lane and Darst, 2010b; Lin *et al.*, 2017).

1.1.5.3.10. β'Lid

The β'lid domain (residues 326–338) is an evolutionarily conserved loop element that extends into the primary channel and forms part of the edge of the RNA exit channel (Borukhov and Nudler, 2008). The β'lid not only contributes to the correct positioning of a portion of the sigma subunit but also plays a role in the separation of the RNA-DNA assembly (Naryshkina, Kuznedelov and Severinov, 2006; Borukhov and Nudler, 2008).

1.1.5.3.11. β'Rudder

The β'rudder domain (residues 382–401) is a structure that is part of the clamp module wherein its function has been reported as contributing towards the stability of the elongation complexes by interacting with DNA at the upstream edge, RNA-DNA assembly and aiding in sigma binding to the RNAP core (Borukhov and Nudler, 2008; Lane and Darst, 2010b).

1.1.5.3.12. β'Bridge helix

The β'bridge helix (β'BH) domain (residues 847–881), otherwise known as the β'F helix, is a helical structure that spans the active center cleft and plays a direct role in catalysis by shifting between straight and bent conformations to accommodate NTP binding and/or RNAP translocation (Tuske *et al.*, 2005). There have also been suggestions that it plays a role in the coordination of surrounding domain conformational changes through its N-terminal region (Weinzierl, 2010).

1.1.5.3.13. β'Trigger loop

The β'trigger loop (β'TL) domain (residues 994–1037) is a helix-turn-helix structure wherein the two helixes are represented by residues 994–1010 and 1027–1037 whilst the connecting loop is represented by residues 1011–1026 (Lane and Darst, 2010b). The β'TL has been shown to participate in catalytic activities involving RNAP translocation and NTP substrate complementarity, and also switches between open and closed conformations allowing for the entry and obstruction of the NTP substrate respectively (Meji, Nudler and Bustamante, 2015; Mazumder *et al.*, 2020).

1.1.5.3.14. ω subunit and the R2 segment

The ω subunit serves in roles such as RNAP assembly and structural integrity (Kurkela *et al.*, 2021). Ω R2 (residues 65–94) in particular is a relatively conserved region of the ω subunit and is important in RNAP assembly (Mao *et al.*, 2018).

1.1.5.3.15. σ^A domain 1 and the non-conserved region (NCR)

Domain 1 of the σ^A subunit (residues 1–259) is composed of the subdomains 1.1, (residues 1–224) which serves to prevent the premature binding of DNA to free σ^A and contributes to RPo formation, and 1.2 (residues 226–259), which forms strong interactions with the β' subunit and may also contribute to RPo formation (Sachdeva *et al.*, 2010). The non-conserved region (NCR) (residues 260–291) is a loop region connecting domain 1 to domain 2 and has been associated with contributing to DNA unwinding around the transcription start-site (Borukhov and Severinov, 2002; Narayanan *et al.*, 2018).

1.1.5.3.16. σ^A domain 2

Domain 2 of the σ^A subunit (residues 292–368) consists of subdomain 2.1 (residues 292–312) which participates in core enzyme interactions, subdomain 2.2 (residues 313–331) which forms interactions with the β' subunit, along with the subdomains 2.3 (residues 332–349) and 2.4 (residues 350–368) which form interactions with the –10 promoter element and participate in recognition of the promoter respectively (Borukhov and Severinov, 2002; Narayanan *et al.*, 2018).

1.1.5.3.17. σ^A domain 3

Domain 3 of the σ^A subunit (residues 369–455) is composed of subdomains 3.0 (residues 369–388), 3.1 (residues 389–431), and 3.2 (residues 432–455). Subdomain 3.0 participates in recognition of the extended –10 promoter element, while

subdomains 3.1 and 3.2 (otherwise known as the σ-finger) play roles in RNAP assembly and NTP-binding respectively (Campbell *et al.*, 2002; Zhang *et al.*, 2012).

1.1.5.3.18. σ^A domain 4

The σ^A subunit further consists of domain 4 (residues 456–515) which can be additionally divided into subdomains 4.1 (residues 456–481) and 4.2 (residues 487–515) which play essential roles in binding transcriptional activators and recognition of the –35 promoter element during transcription initiation (Sachdeva *et al.*, 2010).

1.1.5.4. RIF-resistance mutations

Although RIF showed great efficacy in monotherapies during its early period of application starting in the 1960s, combination therapies have since been favoured particularly due to several instances of RIF-resistant (RIF^R) strains identified and reported through *in vitro* studies and among patients (Goldstein, 2014). RIF-resistance mutations tend to occur as non-synonymous single nucleotide polymorphisms (nsSNPs) within an ~81 bp region of the RNAP β subunit (inclusive of the RIF-BP) called the RIF-resistance determining region (RRDR) (Zaw, Emran and Lin, 2018). The RRDR consists of 4 cluster regions, namely, the RRDR-N (residues 167-172), RRDR-I (residues 423-457), RRDR-II (residues 481-495), and RRDR-III (residues 604-611), with the most common mutations D441V, H451Y, and S456L occurring in cluster I (Molodtsov *et al.*, 2017). While most mutations associated with RIF resistance occur within the RRDR, mutations outside the RRDR have been reported, although they tend to show low levels of RIF resistance (Siu *et al.*, 2011; Zaw, Emran and Lin, 2018; Shea *et al.*, 2021). According to prior investigations, given the rarity of RIF-resistance mutations compared to isoniazid-resistance mutations, the presence of RIF^R strains within a given sample tends to serve as a surrogate marker of isoniazid

resistance (Drobniewski and Wilson, 1998; Yam *et al.*, 2004). To help curb the ongoing trend of drug resistance, improving drug efficacy through modification has been necessitated. However, previous studies have suggested that modification of the naphthol ring at the O1 and O2 positions along with the hydroxyl groups at positions 9 and 10 may be detrimental for RIF's activities against RNAP (Campbell *et al.*, 2001).

1.1.6. Other established antitubercular drugs and their mechanisms of action

1.1.6.1. Isoniazid

Isonicotinic acid hydrazide, or isoniazid (INH), is a highly effective synthetic drug with a simple chemical structure consisting of a hydrazide group and a pyridine ring (Muschenheim, 1952; Sacchettini and Blanchard, 1996). To acquire the active form of INH whose mode of action includes forming toxic free radicals that eradicate susceptible infecting bacilli, the pro-drug form of INH is activated by the catalase-peroxidase enzyme, KatG (Johnsson and Schultz, 1994). In spite of INH's structural simplicity, its bacteriostatic and bactericidal effects against *Mtb* are elusively complex and involve multiple pathways responsible for macromolecular synthesis (Unissa *et al.*, 2016). Primarily, INH's selectivity and specificity for mycobacterial species is based on its inhibition of *Mtb* mycolic acid (MA) synthesis. The cell wall's mycolyl-arabinogalactan-peptidoglycan component highly contributes to *Mtb*'s survival, virulence, and resistance to antitubercular drugs, making it a clear drug target (Takayama, Wang and David, 1972; Takayama, Armstrong and David, 1974; Davidson and Takayama, 1979; Abrahams and Besra, 2018).

1.1.6.2. Ethambutol

Ethambutol (EMB) is an adjunctive synthetic drug derived from ethylenediamine and composes of ethane-1,2-diamine with each nitrogen attached to a 1-hydroxy-butan-2-

yl group (Place and Thomas, 1963; Lee *et al.*, 2003). Besides the mycolyl-arabinogalactan-peptidoglycan component, the *Mtb* cell wall further constitutes the virulent glycolipid, lipoarabinomannan (LAM), which plays a crucial role in the manipulation of the endocytic pathway and phagocytic competencies (Vergne, Gilleron and Nigou, 2014). Although not fully delineated, it has been shown that EMB's mode of action involves the competitive inhibition of D-arabinofuranose substrate binding to the arabinosyltransferase B (EmbB) and EmbC active site; effectively preventing the addition of the D-arabinofuranose units to the growing chains of arabinogalactan and LAM, and depriving the *Mtb* bacilli of an essential cell wall component necessary for their growth and virulence (Tan *et al.*, 2020; Zhang *et al.*, 2020).

1.1.6.3. Pyrazinamide

Pyrazine-2-carboxamide or pyrazinamide (PZA) is a synthetic analogue of nicotinamide which is composed of a pyrazine ring connected to an amide group (Yeager, Munroe and Dessau, 1952; Ying Zhang *et al.*, 2014). PZA exists as a prodrug that is first converted to the active pyrazinoic acid (POA) form by the *Mtb*'s pyrazinamidase after entering the bacilli through passive diffusion (Scorpio and Zhang, 1996). Although the mode of action appears to be complex and continues to be investigated (Ying Zhang *et al.*, 2014; Sun *et al.*, 2020; Santucci *et al.*, 2021); the most widely accepted model involves the extracellular diffusion of POA where it is protonated (HPOA) under acidic conditions and readily passes through *Mtb*'s cellular envelope, where it collects intracellularly (Zhang *et al.*, 1999, 2003; Ying Zhang *et al.*, 2014; Sun *et al.*, 2020). As time passes this inevitably kills the bacilli through several mechanisms including membrane potential disruption (Zhang *et al.*, 1999, 2003; Ying Zhang *et al.*, 2014; Sun *et al.*, 2020). It's worth noting that the drug is uniquely effective

against non-replicating and slowly replicating *Mtb* bacilli, with limited to no effect on actively replicating bacilli (Zhang, Permar and Sun, 2002). It has also been particularly credited for its role in shortening TB combination therapies (Fox, Ellard and Mitchison, 1999).

1.1.6.4. Fluoroquinolones

Fluoroquinolones (FQs) are synthetic drugs generally characterized by having a substituted fluorine at the carbon 6 of a heterocyclic quinoline group which is further bonded to a carboxyl group (Bush *et al.*, 2020). However, the bactericidal efficacy and broad-spectrum capacity of FQ have been improved through several modifications of these structures, resulting in four generations of highly active compounds (Appelbaum and Hunter, 2000; Emmerson and Jones, 2003; Bolon, 2011; Bush *et al.*, 2020). FQs halt DNA synthesis by specifically targeting and inhibiting the structurally and functionally related type II topoisomerases, DNA gyrase and topoisomerase IV, the complexes responsible for the topological regulation of DNA (Levine, Hiasa and Marians, 1998).

1.1.6.5. Bedaquiline

Bedaquiline (BDQ) is derived from diarylquinolines and can be characterized by its quinolone group bonded to a dimethylamino moiety with a hydroxyl group (Andries *et al.*, 2005). BDQ is a relatively novel antitubercular drug whose mode of action has been linked with the inhibition of *Mtb*'s F-ATP synthase, the enzyme responsible for producing the energy-storing/supplying molecule, adenosine triphosphate (ATP) (Koul *et al.*, 2007; Sarathy, Gruber and Dick, 2019). Interfering with the production of ATP has been shown to lead to cell death in both growing and non-replicating bacilli through a complex series of follow-up events that affect multiple metabolic pathways such as

DNA synthesis, protein synthesis and nitrogen metabolism (Koul *et al.*, 2014; Sarathy, Gruber and Dick, 2019; Wang *et al.*, 2019).

1.1.6.6. Linezolid

Linezolid (LNZ) is a synthetic antitubercular drug derived from oxazolidinone (Brickner *et al.*, 1996). The 5-acylaminomethyl group has been specifically credited as essential for its bacteriostatic activity while a fluorine atom and the adjacent morpholino group have been shown to increase the activity (Livermore, 2003; Watkins, Lemonovich, and File Jr, 2012; Poce *et al.*, 2014). LNZ inhibits protein synthesis by specifically binding to the 50S ribosomal subunit, effectively disrupting the incorporation of the 30S ribosomal subunit and preventing the formation of a competent protein synthesis initiation complex needed for further translation of mRNA (Livermore, 2003).

1.1.7. WHO-recommended TB treatment regimen

The current WHO-recommended treatment plan for drug-susceptible pulmonary TB disease has an ~85% success rate and involves the use of the first-line drugs INH, RIF, EMB and PZA over the course of 6 months (WHO, 2021). However, a longer duration of treatment (between ~9 and 20 months) using a different combination of drugs with varying adverse reactions or side effects and reduced success rates is required if one is diagnosed with INH-resistant TB; RIF-resistant TB (RR-TB); multidrug-resistant TB (MDR-TB) which is resistant to both INH and RIF; pre-extensively-drug-resistant TB (pre-XDR TB), which would be resistant to RIF and an FQ (a class of second-line antitubercular drugs); or extensively-drug-resistant TB (XDR-TB) which has resistance against RIF, an FQ and either BDQ or LNZ (WHO, 2021). Regardless, additional antitubercular drugs have therefore been discovered and continue to be investigated in an attempt to curb the growing trend of drug

resistance (Butler, Blaskovich and Cooper, 2013; Hoagland *et al.*, 2016; Bahuguna and Rawat, 2020).

1.1.8. Preventing TB

Immunocompromised persons living with HIV/AIDS, including individuals at particular risk of developing active TB e.g., household contacts of clinically confirmed PTB patients, are especially recommended to take TB preventative treatment (TPT) (WHO, 2021). Although INH-resistant *Mtb* strains continue to pose a challenge in treating latent TB, INH-preventative treatment (IPT) is the most common TPT and involves taking INH for a specified period of time to prevent the onset of active TB disease (Jenkins, Zignol and Cohen, 2011; Surie *et al.*, 2019). This however requires a comprehensive TB diagnosis beforehand to make sure that the individual does not have active TB in order to avoid undertreating them (Surie *et al.*, 2019). Another means of preventative care involves the use of one of the most widely used vaccines first developed in the early 20th century, the *Mycobacterium bovis* bacille Calmette–Guérin (BCG) vaccine, whose use has particularly been recommended for infants and children by WHO (Andersen and Doherty, 2005; WHO, 2021). Although BCG has notable success rates in preventing severe TB disease among vaccinated children, there is an evident decline in efficacy among adults with pre-vaccine sensitization to environmental mycobacteria or *Mtb*, making current vaccines under clinical trials potentially critical for future TB preventative care (Andersen and Doherty, 2005; Martin *et al.*, 2020).

1.1.9. Discovery of novel antitubercular drugs and their mechanisms of action

In the fight to eradicate TB, several drugs with a range of targets such as the *Mtb* cell wall and *Mtb*'s various biosynthetic pathways have been discovered and continue to

be investigated for their effectiveness (Chauhan *et al.*, 2021). Among these targets, the disruption of *Mtb*-RNAP activities through the administration of an inhibitor such as RIF also remains a viable and attractive treatment plan for which a considerable amount of research continues to be undertaken (Stephanie, Tambunan and Siahaan, 2022). Regardless, the following outlines some of the more novel *Mtb*-RNAP inhibitors.

1.1.9.1. D-AAP1

The small D-AAP1 molecule which belongs to the group of N-α-aroyl-N-aryl-phenylalaninamides (AAPs) was first reported in 2017 and allosterically inhibits *Mtb*-RNAP activities by binding to the N-terminus of the catalytic β'BH, preventing it from adopting the necessary conformational changes to allow transcription to proceed (Lin *et al.*, 2017; Agoni, Ramharack and Soliman, 2018a).

1.1.9.2. Sorangicin A (Sor)

Sorangicin A (Sor) is an antibiotic that binds to the same binding pocket as RIF and functions with a similar mechanism of action. However, the drug has displayed more torsional flexibility, allowing it to be less affected by physical changes in the binding site (Lilic *et al.*, 2020). Furthermore, Sor's mechanism of action has been shown to include preventing promoter DNA that is partially unwound from being fully unwound and preventing the template DNA from entering the active site (Lilic *et al.*, 2020).

1.1.9.3. GE23077 (GE)

The antibiotic GE23077 (GE) is a cyclic peptide that is effective against both Gram-negative and Gram-positive RNAPs, a contrasting effect to that of human RNAP I, II and III (Ciciliato *et al.*, 2004; Yu Zhang *et al.*, 2014). Furthermore, GE's mode of action involves blocking the initiation of the transcription process by sterically occluding NTPs

from binding to the i and i+1 sites, along with the catalytic Mg^{2+} (MgI) (Yu Zhang *et al.*, 2014).

Chapter 2

Study knowledge gap, aims and objectives, and hypotheses

Chapter overview

This section covers the identified knowledge gap that inspired and necessitated this study, the associated aims and objectives, and the formulated hypotheses that informed the subsequent analyses.

2.1. Knowledge gap

Although some progress has been made toward curbing the spread of tuberculosis on a global scale, the threat to human health remains (Global TB report, 2020). For this reason, research into the RIF resistance effects of clinically relevant β subunit RRDR mutations such as D441V, H451Y and S456L continues to be conducted to fully elucidate the molecular mechanisms of resistance for the purpose of producing more effective drug therapies (Kumar and Jena, 2014; Nusrath Unissa *et al.*, 2016; Agoni, Ramharack and Soliman, 2018a; Srivastava *et al.*, 2018). However, no study has thus far taken a holistic structural dynamical approach to deciphering resistance mechanisms of these mutations. Furthermore, structure-function implications of the mutations outside the RRDR (non-RDRR) still require much investigation. Hence, the identified research gaps entail the elucidation of clinically relevant RRDR and non-RRDR mutations on multisubunit *Mtb*-RNAP structure-function dynamics regarding RIF potency to help improve future drug therapy investigations and avoid similar therapeutic setbacks.

2.2. Research aim and objectives

2.2.1. Aim

- The fundamental aim of the current study was to build on knowledge regarding the molecular mechanisms of RIF resistance caused by the presence of nsSNPs located in the β subunit of *Mtb*'s DNA-directed RNA replicative machinery through the application of traditional post-MD analyses and the novel dynamic residue network (DRN) analyses on the RIF-bound and unbound *wt* and *mt* systems. In essence, to reveal critical structure-function dynamics and local communication paths that have remained elusive due to the limitation incurred by investigating mutational effects on the β subunit alone.

2.2.2. Objectives

1) To perform a literature search and identify clinically relevant RIF-resistance mutations, along with the identification and retrieval of adequate 3D structural templates of *Mtb*-RNAP.
2) To generate and select acceptable models of the RIF-bound and unbound *wt* and *mt Mtb*-RNAP proteins through homology modelling.
3) To validate the quality of the homology models.
4) To perform MD simulations of the RIF-bound and unbound *wt* and *mt* systems.
5) To perform post-MD processing and successive analyses of equilibrated timeframes using global and local metrics on the *wt* and *mt Mtb*-RNAPs to better understand the mechanism of RIF resistance at the molecular level.

2.3. Hypotheses

- The RRDR SNPs cause local perturbations that destabilize the RIF-binding pocket resulting in a reduced affinity of RIF to the *Mtb*-RNAP β subunit, further causing reduced RIF efficacy.

- The non-RRDR SNPs cause local perturbations in functional domains associated with the RIF-binding pocket resulting in a modified binding mode by RIF, thus reducing RIF efficacy.
- The SNPs cause local effects that extend globally which inevitably results in a shift in the way nucleic acids interact with the *Mtb*-RNAP protein, further disrupting the steric mechanism needed to prevent RNA elongation.

Chapter 3

Computational tools and concepts and applied

Chapter 3 serves as a permitted partial reproduction for the following publication:

Monama, M.Z., Olotu, F. and Tastan Bishop, Ö., 2023. Investigation of Multi-Subunit Mycobacterium tuberculosis DNA-Directed RNA Polymerase and Its Rifampicin Resistant Mutants. International Journal of Molecular Sciences, 24(4), p.3313.

Authors contributions were as follows:

The formal analysis, investigation, methodology, validation, visualization, writing the original draft, and review editing were conducted by Mr Monama MZ. Contributions such as formal analysis, project administration, writing the original draft, and review-editing were conducted by Dr Olotu F. Finally, Prof Tastan Bishop Ö provided and conducted (where applicable) conceptualization, funding acquisition, project administration, resources, visualization, and review editing.

Chapter overview

This chapter covers the various concepts and methodologies applied to carry out the global and local investigations into the molecular basis of *Mtb*-RNAP RIF-resistance.

3.1. Theoretical concepts utilized in the study

The following details concepts that were used to compile the study's objectives along with those used to analyse and discuss the results. The topics covered herein include sequence alignment, homology modelling, molecular dynamics (MD) simulations, and post-MD analytical approaches such as root mean squared deviation (RMSD), root mean squared fluctuation (RMSF), radius of gyration (Rg/RoG), center of mass (CoM) distance analysis, essential dynamics (ED), binding free energy, and dynamic residue network analysis (DRN).

3.1.1. Sequence alignment

Sequence analysis of genomic data in the form of DNA and RNA remains a great and powerful resource in determining the evolutionary relationships between all living organisms over time (Gostling, 2002). Furthermore, due to continuing advancements in sequencing technologies, massive amounts of genomic and RNA data have been made accessible (Liu *et al.*, 2012). There are several peer-reviewed databases that host this wealth of genomic data, for example, the National Center for Biotechnology Institute's (NCBI's) GenBank, the DNA Databank of Japan (DDBJ), and the European Bioinformatics Institute (EBI), which form part of an international collaboration and have greatly impacted research in the field of molecular biology (Shumway, Cochrane and Sugawara, 2009; Liu *et al.*, 2012). These great efforts have allowed researchers to decipher the sequences of several proteins, which further allowed them to determine protein homology (i.e., degree of evolutionary relatedness) and the association of some diseases with specific changes to the transcriptomes. This is all based on the understanding that, due to evolution, genetic information in the form of DNA generally changes over time from the common ancestor, the consequences of which are reflected in the resulting protein sequences and structures (Gostling, 2002). Hence, to determine how much a sequence of interest diverges from a set of target sequences, researchers have made use of multiple sequence alignment (MSA) algorithms. Within the context of protein sequence comparisons, the application of MSA methods can be particularly informative by revealing highly conserved regions that directly translate to maintained protein functionalities. Multiple MSA methods exist today spanning both global and local approaches (Thompson *et al.*, 2011). Global methods entail the alignment and comparison of whole sequences in an end-to-end manner whereas local approaches identify matching sub-sequence regions within the

set of sequences being compared (Thompson *et al.*, 2011). Additionally, given the shortcomings of the aforementioned methods in some test cases, relatively newer iterative MSA methods make use of refinement strategies such as Hidden Markov Models and Genetic algorithms to improve the accuracy of the alignment of more related sequences (Notredame and Higgins, 1996; Eddy, 1998; Thompson *et al.*, 2011).

3.1.2. Protein structure prediction

Although nuclear magnetic resonance (NMR) and X-ray crystallography are the major approaches that have been used to experimentally determine macromolecular structures in great atomistic detail, there are clear limiting factors such as high costs, the time requirements, and the inability to accurately determine particularly mobile/disordered structures (Krishnan and Rupp, 2012). This ultimately results in an ever-increasing gap between experimentally determined protein sequences and structures, as clearly indicated as recently as 2022, wherein UniProt housed non-redundant amino acid sequences amounting to ~204 million whereas the protein data bank (PDB) only had ~190,000 high-quality 3D protein structures (Burley *et al.*, 2019; Bateman *et al.*, 2021; Varadi *et al.*, 2022). To remedy this, one of the most accurate means of protein structure prediction, homology modelling, otherwise referred to as comparative modelling, has been used to predict protein structures. This generally involves using a protein sequence of interest and the target sequence(s) of one or more proteins with experimentally determined 3D structures (templates) preferably displaying a high degree of sequence identity (at least 30–40% identity) and/or structural similarity. Homology modelling is founded on the observation that the 3D structures of proteins are more likely to be conserved in nature even in the presence of medium to small changes in protein sequence (Lesk and Chothia, 1986). An

effective alternative method is the fairly accurate threading or fold recognition method which, unlike homology modelling, makes use of an energy function to predict protein folding (Agnihotry *et al.*, 2021). Also, although the rates of success of *ab initio* methods (which involve the prediction of 3D protein structures from scratch) remain low, they continue to be researched and improved (T. Wang *et al.*, 2019; David *et al.*, 2022; Lee, Su and Tseng, 2022). Regardless, there are still measures that should be taken to reduce the chance of introducing potential errors when building a model through homology modelling, one of which is taking into consideration the level of sequence and structural conservation within protein families (Larsson *et al.*, 2008; Cavasotto and Phatak, 2009).

3.1.3. Molecular dynamics (MD) simulations

In the context of molecular biology, classical molecular dynamics (MD) simulations involve the prediction of the functional motions of proteins or other biomolecules by exploring their possible interatomic interactions within a solvated 3D environment over time, through the application of Newton's equations of motion to each atom (Karplus and McCammon, 2002; Hollingsworth and Dror, 2018). The potential energy function of the simulated system is generally described using molecular mechanics (MM) force fields, which are mathematical expressions derived from experimental observations that are used to calculate the energy of the system with reference to the position of each particle (González, 2011). In essence, MD calculations make use of classical MM physics principles, wherein chemical bonds are treated as springs and force field parameters are used to describe bonded terms which account for chemical bond stretching or bending and torsional angles, and non-bonded terms that cater for charged/electrostatic interactions (via Coulomb's law) and van der Waals interactions captured using the Leonard-Jones potential (**Figure 3.1**).

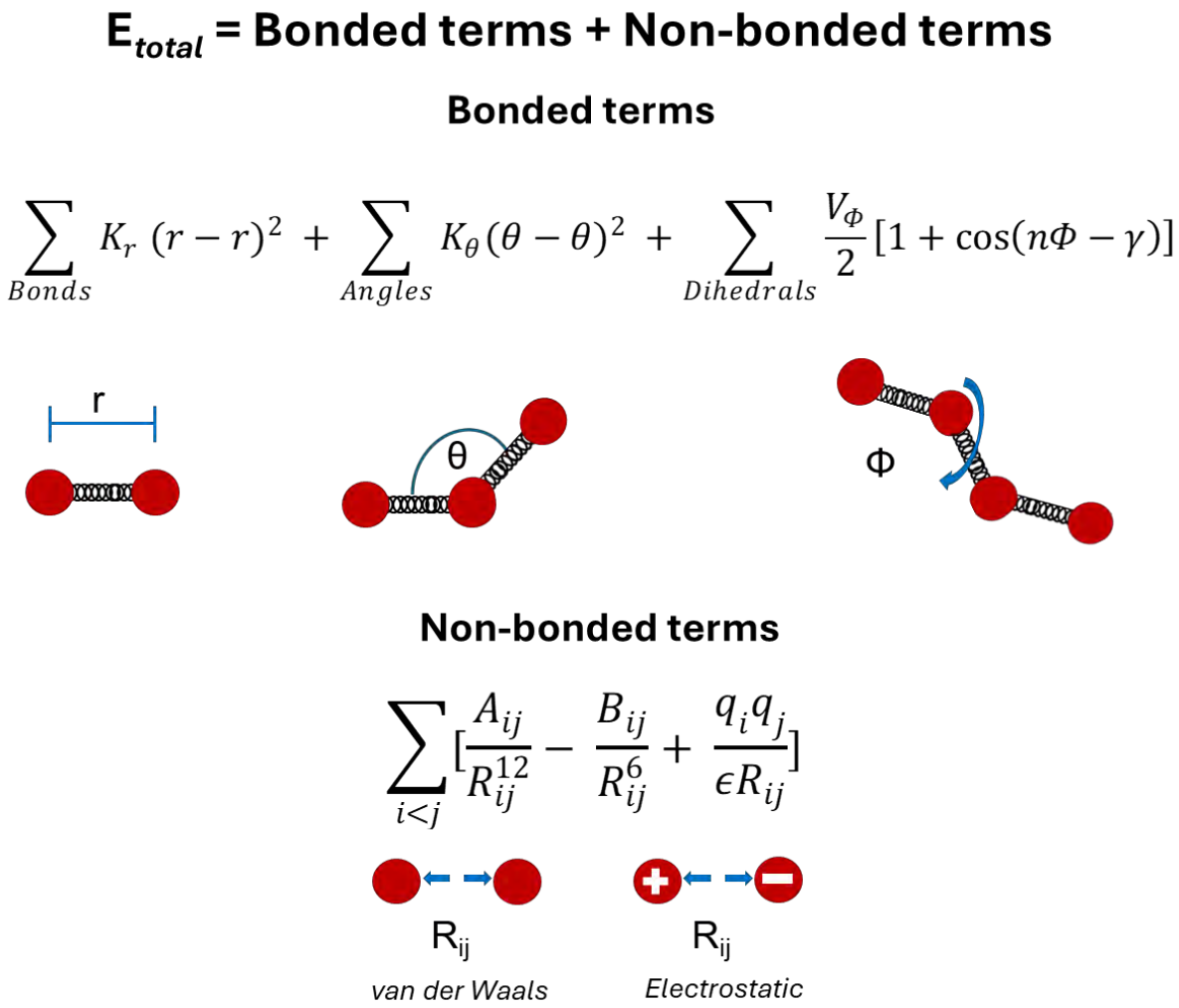


Figure 3.1. An example of the potential energy function used to approximate the atomic forces that determine molecular motions in MD simulation as indicated by (Durrant and McCammon, 2011).

That said, it is worth noting that although the accuracy of MD simulations continues to improve through rigorous research, limitations remain (Lindorff-Larsen *et al.*, 2012). For instance, the electronic effects in MM are only accounted for through approximations determined using either specific experimental measurements or extensive quantum mechanical calculations, which primarily describe the low-energy or ground state of covalent bonds. Therefore, although MDs are less computationally expensive than the alternative, that is, quantum mechanical (QM) approaches,

covalent bonds cannot be converted to the excited state in MM calculations (i.e., no breaking or formation of covalent bonds throughout the simulation), making them less accurate. Another instance would be in the form of physiochemical properties of a protein of interest, e.g. protonation state, which cannot be altered over the course of the simulation. Therefore, the protonation state of a protein must be inspected and, if necessary, corrected in order to accurately sample the relevant MD data. By extension, the results in MD studies will be greatly affected by the level of accuracy in the protein model. Hence, the protein model in use must have all the relevant bonds needed to address the research question prior to the start of the simulation. Regardless, MD simulations surpass the limitations of current physical experiments. That is, through detailed knowledge and direct control of the parameters associated with the simulated system (i.e., temperature, pressure, protein protonation state, etc.), and by capturing the spatial arrangement and velocity of each atom at each point in time in the presence of contributing forces, an overall trajectory describing probable biomolecular motions can be approximated (Hollingsworth and Dror, 2018).

3.1.4. Post-MD analysis

Performing MD simulations can result in large quantities of data being generated which can make it difficult to analyse and extract useful information. The level of difficulty may vary depending on the investigative question you want to answer. Hence, it is always advisable to interpret MD data with reference to as much of the updated and relevant literature for the system in question (Hollingsworth and Dror, 2018). Methods for analysing MD data usually consist of visualization of the simulation (or aspects of the simulation), which was done here primarily using VMD (Humphrey, Dalke and Schulten, 1996), PyMOL (DeLano, 2002; Rosignoli and Paiardini, 2022) and Discovery Studio Visualizer software (BIOVIA, 2012); along with quantitative analysis

approaches (e.g., GROMACS tools and MDM-TASK tools) which imply the generation of additional data that allows the researcher to probe an aspect of the protein's dynamics, allowing them to observe the evolution of a specific structural component to gain more information (Hollingsworth and Dror, 2018).

3.1.4.1. Root mean square deviation (RMSD)

Root mean square deviation (RMSD) can be used to quantify the differences in backbones between two superimposed protein structures. When used dynamically it can give details about protein stability by comparing the conformational evolution of the protein from the initial timeframe to its conformation at the final timeframe. The following equation can be applied through the GROMACS *gmx rms* tool:

$$RMSD(t_1, t_2) = \left[\frac{1}{M}\sum_{i=1}^{N} m_i ||r_i(t_1) - r_i(t_2)||^2\right]^{\frac{1}{2}}$$

wherein $M = \sum_{i=1}^{N} m_i$ and $r_i(t)$ represent the position of an atom i at time t. This can be used for least-square fitting the structure of interest to the reference structure followed by subsequent estimation of RMSD (Abraham, van der Spoel and Lindahl, 2016).

3.1.4.2. Root mean square fluctuation (RMSF)

Root mean square fluctuation (RMSF), unlike RMSD, does not convey global protein stability; instead, it generally makes use of the more stable backbone Cα atom as a representative to more accurately estimate the flexibility of the protein at the residue level over time. RMSF can be calculated using the following equation via the *gmx rmsf* tool:

$$\rho_i^{RMSF} = \sqrt{\langle (r_i - \langle r_i \rangle)^2}$$

wherein the atomic coordinates and average position of the ensemble are represented by i and $\langle r_i \rangle$ respectively (Beckstein, 2017).

3.1.4.3. Radius of gyration (Rg/RoG)

The radius of gyration (Rg/RoG) is a metric that is used to deduce the level of packing for a given protein which therefore relays an estimate of the amount of compactness (Abraham, van der Spoel and Lindahl, 2016). The following equation can be used to calculate Rg through the *gmx gyrate* tool:

$$R_g = \left(\frac{\sum_i ||r_i||^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}}$$

wherein m_i and r_i represent the mass of the atom i and the relative position of the atom with respect to the molecule's center of mass (Abraham, van der Spoel and Lindahl, 2016).

3.1.4.4. Solvent accessible surface area (SASA)

The determination of how much a biomolecule's surface area is exposed to solvent molecules can be approximated using the *gmx sasa* tool which applies the double cubic lattice method (DCLM) (Eisenhaber *et al.*, 1995).

3.1.4.5. Dynamic cross-correlation (DCC)

Dynamic cross-correlation (DCC) can be used to approximate correlated residue motions over the course of an MD trajectory using the MDM-TASK-web calc_correlation.py script (Brown *et al.*, 2017; Sheik Amamuddy *et al.*, 2021). The script uses the following equation:

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{\sqrt{\langle \Delta r_j^2 \rangle} \cdot \sqrt{\langle \Delta r_i^2 \rangle}}$$

where r_i and $\langle \ \rangle$ represents the average distance of displaced atom i from its original position and the trajectory time respectively.

3.1.4.6. Center of Mass (CoM)

The *gmx distance* tool can be used to conduct a center of mass (CoM) analysis wherein the average change in distance over time between two specified points in a biomolecular system can be estimated using the captured distances between the CoM of each point per MD fame (Abraham, van der Spoel and Lindahl, 2016).

3.1.4.7. Hydrogen bonds

In the context of drug discovery and optimization, there are specific non-covalent interactions that contribute to the required interaction geometries and binding affinity between a ligand and a feasible protein receptor (Bissantz, Kuhn and Stahl, 2010; Zhou, Huang and Tian, 2012). Hydrogen bonds (H-bonds), which are generally described as non-covalent bonds that form between a hydrogen atom covalently bonded to a highly electronegative atom and further non-covalently bonded to an additional highly electronegative atom, form part of the essential non-covalent bonds needed to establish protein-ligand complexes (Bissantz, Kuhn and Stahl, 2010). It's also important to note that crystal structures with static interactions only represent a snapshot of the likely interactions that occur between the structural ensembles of the protein and ligand (Bissantz, Kuhn and Stahl, 2010). Hence, applying a more dynamic approach may lead to a more accurate description of the resulting non-covalent interactions. The *gmx hbond* tool in particular provides an option to estimate the protein-ligand H-bond interactions that occur throughout a simulation (Abraham, van der Spoel and Lindahl, 2016).

3.1.4.8. Essential dynamics

MD simulations generate multidimensional structural ensembles in the form of atomic trajectories that capture a protein's likely conformations best described by vector spaces that encompass several dimensions equivalent to the degrees of freedom (DOF) in the phase space (David and Jacobs, 2014). Therefore, protein function can be elucidated by observing those sampled conformations. However, to be certain that conclusions are accurate, the sampled conformations must be accessible to the protein (David and Jacobs, 2014). Principal component analysis (PCA) makes use of the atomic coordinates of the protein from each frame of the MD to generate a covariance matrix, or a correlation matrix if normalization of the results is necessary to prevent skewed results (David and Jacobs, 2014). That said, essential dynamics (ED) is used to describe protein dynamics through the application of PCA which in essence is a statistical approach that reduces the number of dimensions of multivariate datasets (David and Jacobs, 2014). Due to the likely association of the low-frequency modes, otherwise referred to as principal modes of collective motion, with the functionally relevant protein motions, the dynamics of the "essential subspace" in which these modes are found were referred to as essential dynamics, to emphasize their functional relevance (Amadei, Linssen and Berendsen, 1993). Similar to normal mode analysis (NMA), the idea is that the normal modes indicating the largest fluctuations, i.e. low-frequency modes, are associated with functional motions of the protein (Hayward and De Groot, 2008). This is evidenced by the observation that protein dynamics can be described by only a small set of DOFs (Hayward and De Groot, 2008).

3.1.4.9. Binding free energy

Free energy is a thermodynamic concept that governs a system's internal energy and capacity to do work while also giving details about the direction in which thermodynamic processes proceed and the state in which they may remain. Free energy is therefore regarded as the driving force behind molecular processes such as molecular associations and protein folding (Wang *et al.*, 2019). Since MD simulations have continued to improve in terms of accuracy, MDs have become a particularly useful tool in explaining the free energies of systems of interest, allowing researchers to explore otherwise hidden biomolecular states, among other applications (Christ *et al.*, 2009). One such application is the estimation of binding free energy (ΔG_{bind}) which conveys the binding strength or affinity of a ligand for a specific receptor (usually a protein) (Genheden and Ryde, 2015). Even though there is a wide range of accurate approaches for determining protein-ligand binding affinities, e.g. the computationally expensive alchemical methods such as free energy perturbation (FEP), end-point methods like the molecular mechanics Poisson−Boltzmann surface area or MM/PBSA, and the molecular mechanics generalized Born surface area or MM/GBSA are among the most widely used given their relative accuracy and lower computational requirements (Wang *et al.*, 2019; Wang *et al.*, 2022). That said, the MM/PB(GB)SA methods aren't without limitations. For instance, the determination of conformational entropy is a major source of some error and happens to make up the bulk of the computational cost, albeit some remedies have been developed (Genheden *et al.*, 2012; Duan, Liu and Zhang, 2016; Sun *et al.*, 2018). It's also worth noting that although the GB model, which is a simpler approximation of the PB model, is comparatively less accurate, it remains a viable alternative given its lower computational costs (Feig *et al.*, 2004; Wang *et al.*, 2019). Nonetheless, using the MM/PB(GB)SA method, the

equation used to determine the binding free energy for the protein-ligand complex is as follows:

$$G_{bind} = G_{RL} - G_R - G_L$$

where the R, L and RL represent the protein receptor, the ligand, and the protein-ligand complex respectively. The equation can be further expanded as follows:

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

where ΔH represents a change in enthalpy, ΔE_{MM} represents the gas phase MM energy, ΔG_{sol} represents the solvation-free energy, and $-T\Delta S$ represents the conformational entropy once the ligand is bound.

The following represent additional expansions of some of the above-mentioned energy terms:

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW}$$

where ΔE_{int} represents the changes in internal energies (encompassing bond, angle, and dihedral energies); and ΔE_{ele} and ΔE_{vdW} represent electrostatic and van der Waals energies respectively, and

$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA}$$

where $\Delta G_{PB/GB}$ and ΔG_{SA} represent the polar contribution (determined either through the GB or PB model, with GB making use of an analytical expression that expedites the calculation) and the non-polar contributions (determined through a solvent-accessible surface area (SASA) calculation) between the solute and continuum solvent respectively (Wang *et al.*, 2019).

3.1.4.10. Graph theory and centrality metrics

Graph theory has shown several applications in biology, including but not limited to, knowledge representation, e.g. metabolic and biochemical pathways represented in a “bubble-arrow” layout, identification of potential drug-target interactions, deducing either protein or gene function(s), and formulation of potential treatments to combat diseases (Huber *et al.*, 2007; Pavlopoulos *et al.*, 2011). Graph theory has therefore shown great promise in describing and analysing the interactions and relationships involved in biological systems (Huber *et al.*, 2007; Pavlopoulos *et al.*, 2011). A graph can also be used to describe a protein’s residue interaction network (RIN). This is based on the established notion that non-covalent interactions contribute to structural stability, govern the folding process, and are responsible for the bulk of the energy required to form high-ordered assemblies, hence making molecular communication possible (Del Sol *et al.*, 2006; Dill and MacCallum, 2012). One can therefore generate RINs using the backbone atoms (Cβ of the residue or Cα for glycine) wherein the node represents the residues while the edges represent the non-covalent bonds between each residue pair within the bounds of a specified distance (Atilgan, Akan and Baysal, 2004; Brown *et al.*, 2017). Moreover, using the frames generated through MDs, a researcher can compile the relevant RINs to determine the dynamic residue network (DRN) of their protein through a server such as MDM-TASK (Sheik Amamuddy *et al.*, 2021) which may prove especially beneficial if analysed using the appropriate averaged network centrality metric(s). Centrality is a measure of how important a residue is within the network. *Betweenness centrality* (*BC*) determines the frequency of shortest path transversals between connected residue pairs through a particular node, thereby relaying residue centrality in protein communication (Karabekmez and Kirdar, 2016; Brown *et al.*, 2017). *Closeness centrality* (*CC*) reveals the most

immediately accessible residues within the network by identifying nodes with the closest proximity to all other nodes within the network. This may convey the speed at which information is transferred between residues (Goldbeck, 2013; Okeke *et al.*, 2021). *Degree centrality* (*DC*) relays residue centrality by identifying nodes with the most connections within the network (Hanneman and Riddle, 2005; Zhan, Gurung and Parsa, 2017). *Eigenvector centrality* (*EC*) identifies highly influential residues by considering immediately connected high-degree neighbouring nodes in the network (Ruhnau, 2000; Goldbeck, 2013; Sheik Amamuddy *et al.*, 2021). *Katz centrality* (*KC*), much like *EC*, determines highly influential residues in the same manner as *EC* but also includes nodes non-immediately connected through the high-degree neighboring nodes (Koschützki and Schreiber, 2008; Zhan, Gurung and Parsa, 2017).

3.2. Methodologies

3.2.1. Mutation data and protein structure retrieval

We began by retrieving most of the clinically relevant missense mutation data that was designated as high confidence from the MUBII-TB-DB (Flandrois, Lina and Dumitrescu, 2014). High-confidence mutations are those mutation sites that tend to be well-reported in literature (Sandgren *et al.*, 2009; Flandrois, Lina and Dumitrescu, 2014). Of the identified high-confidence RRDR mutations, D441V, H451D, H451L, H451N, H451R, H451Y and S456L were selected for further study. Additionally, the non-RRDR high-confidence mutation D551E (Miotto *et al.*, 2017) along with the novel non-RRDR I65T mutation (Sandgren *et al.*, 2009) were also considered based on their distal positions from the RIF-BP, which allowed for a broader perspective for the investigation. We then retrieved the 3D *Mtb*-RNAP crystal structure (PDB:5UHC) (Lin *et al.*, 2017) from the Protein Data Bank. Along with the 5UHC structure, we retrieved the 3D structure of *Thermus thermophilus* (*Tth*) RNAP (PDB:1ZYR) (Tuske *et al.*,

2005) and used it as a partial template in order to build the missing residues to the *Mtb*-RNAP protein. For the purposes of this study, the final *Mtb*-RNAP structure consisted of the αI- and αII-NTDs, β, β', ω, and σ subunits, including DNA, RNA, and RIF.

3.2.2. Building the wildtype (wt) and mutant (mt) Mtb-RNAPs

To construct the *wt* and *mt Mtb*-RNAP models, we began by building the missing regions in the crystallized *Mtb*-RNAP (5UHC) structure (which consisted of the αII-NTD (residues 156-157) and the β'TL (residues 1012-1025) domain) utilizing the automodel classs of the MODELLER software (v9.15) (University of California, San Francisco, CA, USA) (Šali and Blundell, 1993) (refer to **Supplementary Text 1.1-1.2**). In that regard, the homologous *Tth*-RNAP was used as the partial template to complete the missing β'TL domain. We then proceeded to produce the *mt Mtb*-RNAPs for our investigation by incorporating the relevant point substitutions in the target sequence. To have more accurate builds of the *wt* and *mt MTB*-RNAPs, we applied a very slow refinement to the 3D structures producing a total of 100 models each, respectively. The models were then ranked based on the calculated normalized DOPE score (z-DOPE score) and the best structural models were selected from each set of models. This comparative modelling approach has been found to be efficient and less likely to introduce errors to the generated models, as previously reported (Tastan Bishop and Kroon, 2011). Finally, the quality of the selected models was validated using the PROCHECK (University of California, Los Angeles, CA, USA) (Laskowski *et al.*, 1993) and QMEAN (University of Basel, Basel, Switzerland) (Benkert, Biasini and Schwede, 2011) servers through the respective use of a Ramachandran plot and the normalized QMEAN6 scoring metric as indicated in **Supplementary Table S1**. Also, it's worth

noting that, to obtain the RIF-unbound *wt* and *mt* *Mtb*-RNAPs, RIF was simply removed from the respective structural models prior to running MD simulations.

3.2.3. Molecular dynamics (MD) simulations

The GROMACS 2016.1 software (Van der Spoel *et al.*, 2005) was used to conduct 700ns MD simulations on the *Mtb*-RNAP protein multisubunit studied herein. Specifically, these entail the unbound *wt*s and 9 *mt* (I65T, D441V, H451D, H451L, H451N, H451R, H451Y, S456L, and D551E) proteins as well as their respective RIF-bound forms, making a total of 22 systems. Simulations were performed on the Center for High Performance Computing (CHPC) clusters in Cape Town. Each *Mtb*-RNAP (both unbound and RIF-bound) system employed for our simulation study uniformly consisted of the αI-NTD and αII-NTD, β and β' subunits, ω and σ subunits, and DNA and RNA molecules. The simulations were carried out using Amber ff99SB-ILDN (Lindorff-Larsen *et al.*, 2010) with a triclinic box and a clearance space of 1.6 nm. The system was solvated using the rigid extended simple point charge (SPC/E) water model, and the protein's overall charge was neutralized with Na^+ and Cl^- counterions. RIF's topology parameters were generated using the Antechamber Python Parser Interface (ACPYPE) 0.1.1 tool (Sousa Da Silva and Vranken, 2012) following valency corrections using Discovery studio 4.5 (*BIOVIA Discovery Studio Visualizer - Dassault Systèmes*, 2021). The systems were minimized using up to 50,000 steps of steepest-descent and terminated when the maximum force was <1000 kJ mol^{-1} nm^{-1}. The short-range cut-offs for the non-bonded interactions were set to 1.3 nm. The long-range interactions were treated using the Particle Mesh Ewald (PME) algorithm by applying a grid spacing of 0.16 nm and a cut-off distance of 1 nm while using the LINCS method to restrain the bond length. The systems were then temperature and pressure equilibrated in the NVT ensemble with V-rescale coupling set to 300 K, and

an NPT ensemble with Parrinello-Rahman barostat set to 1 bar for 100 ps, respectively. Moreover, a duplicate run was performed with the *wt* RNAP multisubunit as the control sample to ensure a high degree of correctiveness and reproducibility. This was also important to ascertain that the conformational changes observed in the mutants post-equilibration were largely due to mutational effects.

3.2.4. Computational resources

Before proceeding with MD simulations for the 22 systems, a scaling test was required to determine if access to more computational resources on CHPC was needed. Five separate scaling tests were run beginning with 2 nodes (48 CPUs) over 48 hours, followed by an increment of 2 nodes per test. Since the test revealed usage of ~25 ns/day with the efficient use of 10 nodes, usage of additional resources was approved, after which 30 nodes were used for each of the successive 100 ns MD runs per system. Following MD simulations, a total of ~30TB of data was generated with approximately 7185043 CPU time used to run the MDs. Finally, the data was transferred from CHPC to a local machine and then to a Rhodes University cluster using Globus (Foster, 2011).

3.2.5. Post-MD preparations

Resulting trajectories were corrected for periodic boundary conditions using *gmx trjconv* while conformational analyses were performed with the *gmx rms*, *gmx rmsf*, *gmx gyrate*, *gmx sasa*, and *gmx hbond* tools which, respectively, were used to calculate RMSD, Cα-RMSF, Rg, SASA, and quantify H-bond interactions. Additionally, the *readHBmap.py* script (https://github.com/quytruong1808/vilas/blob/master/vilas/analyzer/readHBmap.py) was used for H-bond analyses. All versus all Cα-RMSD was calculated using a previously applied in-house script (Barozi *et al.*, 2022) that implemented the *pytraj* package (Roe and Cheatham, 2013) at a step size of 20. The

script processed each of the trajectories and then calculated the Cα-RMSD via comparison of each MD frame to all other frames and itself across the simulation time to further corroborate the stability of the control systems. Additionally, to analyse events at the RIF-BP, the *gmx select* tool was used to map out *wt* β subunit residues within 5Å of RIF, among which were R173, R176, F439, Q435, Q438, D441, H451, R454, S456, and R465, for the initial frame of the simulation. Also, the MDM-TASK-web calc_correlation.py script was used to calculate DCC using a step size of 5 and a cut-off distance 6.7 Å, to determine the RIF-BP's correlated residue motions. Motional deviations of RIF across the *wt* and *mt* systems relative to the RIF-BP were also determined via CoM analysis using the *gmx distance* tool. The VMD software (v1.9.3) (Humphrey, Dalke and Schulten, 1996) was used for structural visualization of the MDs. Graphical analyses and visualizations were further done using RStudio and a combination of Python packages, namely, pandas (McKinney, 2011), seaborn (Waskom, 2021), NumPy (Van Der Walt, Colbert and Varoquaux, 2011), and Matplotlib (Hunter, 2007).

3.2.6. Binding free energy and residue decomposition calculations

The binding affinity of RIF for the *wt* and *mt* β subunits was evaluated through Molecular Mechanics/Generalized-Born Surface Area (MM/GBSA) calculations using the gmx_MM/PB(GB)SA tool (Valdés-Tresanco *et al.*, 2021). The MM/GBSA calculations were done with the final 10 ns (5000 timeframes) of the trajectories and made use of the Amber ff99SB-ILDN (Lindorff-Larsen *et al.*, 2010) force field. These equilibrated timeframes were selected for the energy calculations to minimize possible entropical effects.

3.2.7. Comparative essential dynamics analysis

ED calculations with common principal components (PCs) (also referred to as orthogonal eigenvectors) attained through a single decomposition of multiple MD trajectories were conducted using the comparative ED MDM-TASK-web tool (Brown *et al.*, 2017; Sheik Amamuddy *et al.*, 2021) (https://github.com/RUBi-ZA/MODE-TASK/tree/mdm-task-web). The tool was used to process unbound and RIF-bound *wt* and *mt Mtb*-RNAP trajectories to produce decomposed conformational data in scatter plot representation with common PCs. Given that PC1 and PC2 capture the most variance, they were used in the analysis. The following Cα residue selections were made to improve resolution of the *Mtb*-RNAP ED results: '(chainid 0 and residue 7 to 221) or (chainid 1 and residue 228 to 444) or (chainid 2 and residue 482 to 1562) or (chainid 3 and residue 1584 to 2850) or (chainid 4 and residue 2873 to 2940) or (chainid 5 and residue 2945 to 3262)'. These selections correspond to chain A residues 9 to 223, chain B residues 9 to 225, chain C residues 58 to 1138, chain D residues 8 to 1278, chain E residues 40 to 107, and chain F residues 210 to 527, respectively. A step size of 5 was further used in the *Mtb*-RNAP EDs while additional RIF-BP ED calculations were conducted using a step size of 1.

3.2.8. Dynamic residue network calculations

Unlike RINs which derive their data from a single static structure, DRNs average the residue interactions from the multiple frames produced by MDs to decipher important contacts and relay important communication paths (refer to **Chapter 4**).

The MDM-TASK-web server (Sheik Amamuddy *et al.*, 2021) scripts (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) were therefore used to determine changes in communication paths in the presence of missense mutations

through DRN analysis of equilibrated 450 ns trajectories based on the RNAP Cα RMSD results of the individual subunits (with respect to their initial structure). DRN was conducted by applying a cut-off distance of 6.7 Å and step size of 50 frames from each sample of 225000 frames using the following averaged network graph metrics: *betweenness centrality* (*BC*), *closeness centrality* (*CC*), *degree centrality* (*DC*), *eigenvector centrality* (*EC*) and *Katz centrality* (*KC*). Heatmaps representing the raw DRN data for the β and β' subunits of the *wt* controls and *mt*s were then captured and compared (**Supplementary Figures S2.1-S2.5**). Additionally, the *Mtb*-RNAP β-β' complex DRN value differences were identified between the *wt*1 and *mt* residues and *wt*2 and *mt* residues through normalized heatmap representations, also referred to as the normalized delta heatmaps, which were generated and compared to identify any large discrepancies between *wt*1-*mt* and *wt*2-*mt* comparisons. Furthermore, the metrics were applied to global analysis, where the *Mtb*-RNAP β-β' complex Cβ atoms (Cα for the glycine residue) of all the RIF-bound or unbound systems were considered per metric, and local analysis, wherein Cβ atoms of the individual β-β' complexes per system were considered per metric. The global and local top 5% DRN residues of the β and β' subunits of the *wt*s were then identified for each metric and then represented as a heatmap and mapped to the relevant protein structures for further comparison (**Supplementary Figures S2.6-S2.10**). Finally, the global top 5% DRN residues of the *wt*1 β and β' subunits were compared to those of the *mt*s (**Supplementary Figures S2.16-S2.20**) while excluding the residues showing significant differences (mean ± 2 SD) between the *wt* controls per metric (**Supplementary Figures S2.11-S2.15**).

3.2.9. Weighted contact maps

Weighted contact maps are a dynamic approach to the more traditional contact maps, wherein the frequency of inter-residue contacts across MD frames is normalized.

Here, we determined weighted contact maps using the MDM-TASK-web server scripts (Sheik Amamuddy *et al.*, 2021) for the mutation sites. As with the DRN parameters, the calculations were conducted using a cut-off distance of 6.7 Å and a step size of 5 frames for the equilibrated trajectories.

Chapter 4

Global analysis of *wt* and *mt Mtb*-RNAPs in the presence and absence of RIF

Chapter 3 serves as a permitted partial reproduction for the following publication:

Monama, M.Z., Olotu, F. and Tastan Bishop, Ö., 2023. Investigation of Multi-Subunit Mycobacterium tuberculosis DNA-Directed RNA Polymerase and Its Rifampicin Resistant Mutants. International Journal of Molecular Sciences, 24(4), p.3313.

Authors contributions were as follows:

The formal analysis, investigation, methodology, validation, visualization, writing the original draft, and review editing were conducted by Mr Monama MZ. Contributions such as formal analysis, project administration, writing the original draft, and review-editing were conducted by Dr Olotu F. Finally, Prof Tastan Bishop Ö provided and conducted (where applicable) conceptualization, funding acquisition, project administration, resources, visualization, and review editing.

Chapter overview

This chapter describes the global analysis of the post-MD results to assess the *wt* and *mt Mtb*-RNAP systems and uncover details surrounding the RIF resistance mechanisms.

4.1. Introduction

For RIF to effectively achieve its antitubercular activities, it specifically targets the bacterial RNAP β subunit which allows it to prevent the elongation of RNA transcripts of 2-3 nucleotides in length, consequently trapping the enzyme in a state of abortive transcription (Campbell *et al.*, 2001; Lin *et al.*, 2017). In recent years, however, RIF's therapeutic efficacy has been challenged by several clinically reported RRDR

missense mutations which have negatively affected RIF's specificity and affinity for its binding site (Goldstein, 2014; Koch, Mizrahi and Warner, 2014).

Among the reported missense mutations, the most commonly reported *Mtb*-RNAP-β mutation, S456L, not only happens to be located in the cluster I region of the RRDR, but is also proximally located to the βFL2 domain (residues 459-473). Moreover, previous studies suggest that the S456L mutation may impact RIF activity through a disordered βFL2 while indicating a low fitness cost (Molodtsov *et al.*, 2017). Further details regarding this observation were kinetically expounded by (Stefan, Ugur and Garcia, 2018) whereby the enzymatic activities of S456L, D441V, and H451Y which are some of the most commonly occurring mutations, were delineated. A number of reports sourced from populations worldwide have also highlighted the clinical relevance of the H451D, H451L, H451N, and H451R mutations (Zaw, Emran and Lin, 2018).

A number of previous *in silico* studies have investigated RIF's binding to the *Mtb*-RNAP β subunit in the presence of mutations and have since provided some mechanistic insights into their perturbation effects (Singh *et al.*, 2017; Srivastava *et al.*, 2018; Zhang *et al.*, 2019; Amusengeri, Khan and Tastan Bishop, 2022). Some of these studies revealed that RIF's binding efficacy, and by extension its activity, may be adversely affected by mutations such as D441V, H451D, H451R, and H451Y, through the alteration of critical RIF-BP interactions (Zhang *et al.*, 2019; Amusengeri, Khan and Tastan Bishop, 2022). Also, although several RRDR and distal mutations have been reported, a number of clinical studies detailing RIF's reduced efficacy have been closely associated with cluster I mutations, particularly those positioned at D441, H451, and S456 (Zaw, Emran and Lin, 2018). To complement these findings, the current study focused on investigating the mechanistic effects of the high incidence

and high-confidence mutations D441V, H451D, H451L, H451N, H451R, H451Y, S456L and D551E (Flandrois, Lina and Dumitrescu, 2014; Miotto *et al.*, 2017) on the structural dynamics of Mtb-RNAP along with RIF's inhibitory activities. Also under investigation was the reported clinical mutation I65T (Shea *et al.*, 2021) which, like D551E, is a distal mutation.

We, therefore, for the first time, elucidated some of the underlying RIF-resistance mechanisms as mediated by the aforementioned missense mutations using the multisubunit *Mtb*-RNAP which is composed of the αI-NTD and αII-NTD, and the β, β', ω and σ subunits along with DNA, RNA, and RIF molecules.

4.2. Consequences of high incidence mutations on structural–functional stability

Nine clinically relevant and high incidence *Mtb*-RNAP-rpoB mutations, namely, I65T, D441V, H451D, H451L, H451N, H451R, H451Y, S456L, and D551E, were collected from various literature sources (Flandrois, Lina and Dumitrescu, 2014; Miotto *et al.*, 2017; Shea *et al.*, 2021) to investigate their impact on the functional dynamics of *Mtb*-RNAP systems in the presence and absence of RIF. Following the construction and validation of the RIF-bound and RIF-unbound *wt* and *mt Mtb*-RNAP structures through a tested and validated homology modelling approach (Bishop and Kroon, 2011), we proceeded to simulate then examine the introduced mutational and RIF effects using the Cα-RMSD, Cα-RMSF, and Rg metrics. Additionally, the novel comparative ED approach was utilized to investigate the global motional effects of *Mtb*-RNAP in response to the presence of the investigated mutations and RIF (Sheik Amamuddy *et al.*, 2021).

To determine the equilibrated timeframes for the simulated 700 ns *wt* and *mt Mtb*-RNAP systems, Cα-RMSD (with respect to the initial frame) line plots along with all

versus all Cα-RMSD heatmaps were constructed and then assessed. With that, equilibrated timeframes were consistently found to be between ~250–700 ns, as seen in **Figure 4.1A–D**.

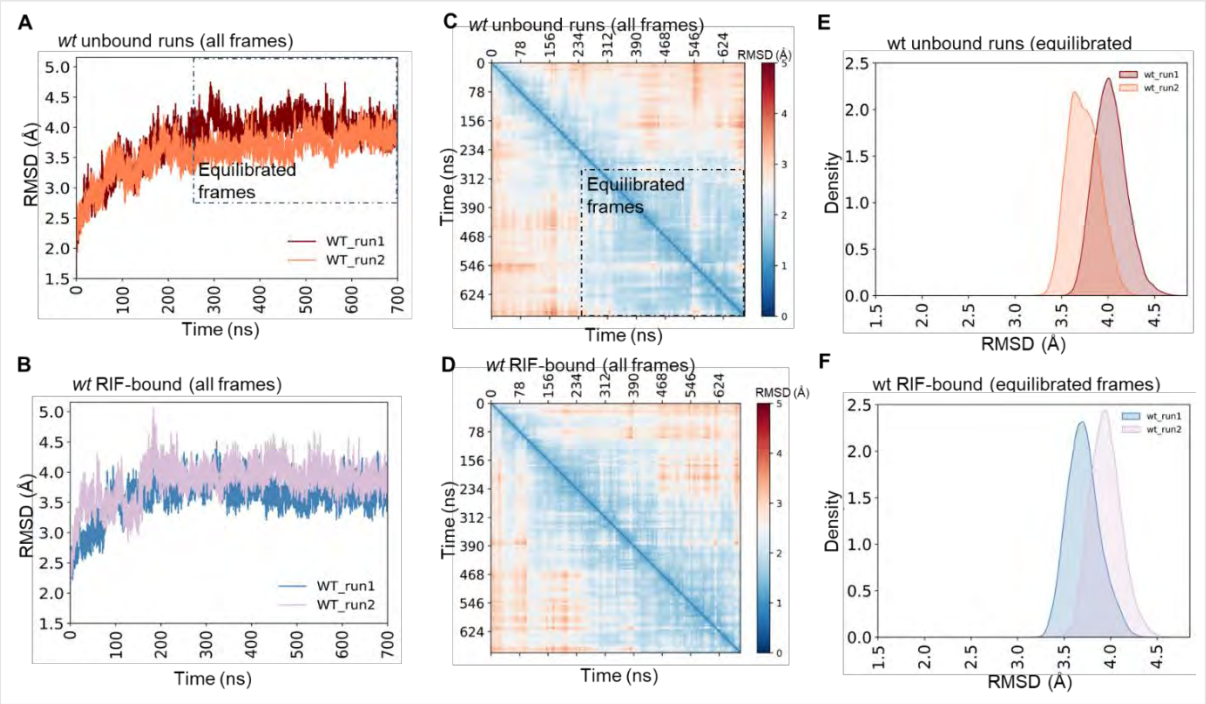


Figure 4.1. Graphical representation of the structural agreement between the *wt* standards for the equilibrated timeframes. (A-B) Shows the Cα-RMSD (Å) line plots for *wt* run 1 (unbound, colored maroon; RIF-bound, colored blue) and *wt* run 2 (unbound, colored orange; RIF-bound, colored pink), with reference to their initial structures. Heatmaps representing the determined all-versus-all Cα-RMSD are respectively displayed for the (C) unbound and (D) RIF-bound *wt* run1 proteins. The coloring scheme of the heatmaps ranges from red (representing conformational dissimilarity) and blue (representing conformational similarity). KDE plots are indicated for the (E) unbound and (F) RIF-bound duplicate *wt* runs with a similar color scheme to that of the Cα-RMSD line plots.

Therefore, to minimize noise and prevent capturing potentially erroneous results, only the equilibrated timeframes (250–700 ns) of each trajectory were used in the investigation unless otherwise specified.

It is important to note that, due to the computational resources required to simulate the *Mtb*-RNAP protein (~3000 residues),i.e. the 30 nodes (720 CPUs) per 100 ns (per 48 hours) and over 7 million in CPU time used, only duplicate *wt* runs were conducted. Our analysis of the differences between the *wt* standards (duplicate MD runs) through Cα-RMSD revealed average differences of 0.28 Å and 0.24 Å for the RIF-unbound and bound *wt* systems respectively (**Figure 4.1E, F**). Therefore, the observed consistency in conformational dynamics between the *wt* standards, which suggested higher confidence in the experimental approach, further implied that the changes observed in the *mt* systems were most likely a result of the introduced mutations (**Figures 4.2A, B, and Supplementary Figures S1.1A, B**). The detected conformational variabilities among the *mt* systems relative to the *wt* standards therefore served as indicators for the significance of the mutational impact on *Mtb*-RNAP stability.

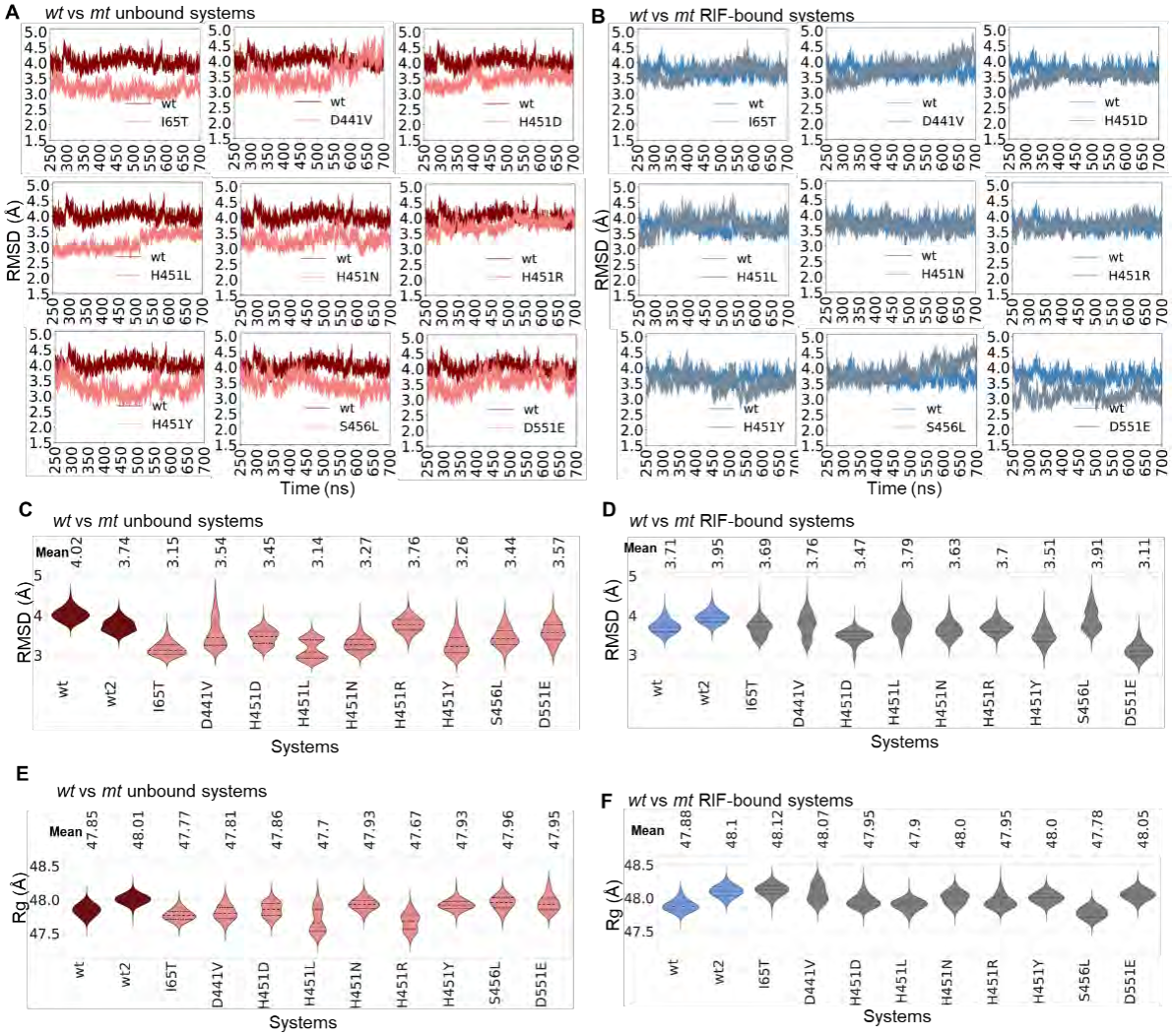


Figure 4.2. Line and violin plots that represent the *Mtb*-RNAP structural dynamics in response to mutations and/or RIF binding. (A) RIF-unbound and (B) RIF-bound Cα-RMSD line plots (relative to the initial structure) for the *wt* (unbound, colored maroon; RIF-bound, colored blue) and *mt* (unbound, colored orange; RIF-bound, colored pink) *Mtb*-RNAP proteins. Similarly, (C) RIF-unbound and (D) RIF-bound Cα-RMSD distribution violin plots (with mean values indicated above the plots) for the *wt* and *mt* *Mtb*-RNAP proteins are shown respectively with the same color scheme as the line plots. The all-atom Rg distribution violin plots for the (E) RIF-unbound and (F) bound proteins are also represented the same as the Cα-RMSD violins. The respective dotted lines in the violin plots represent the 25th, 50th, and 75th quartiles from top to bottom. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

As done by (Sheik Amamuddy, Verkhivker and Tastan Bishop, 2020; Okeke *et al.*, 2021), Cα-RMSD line and violin plots including all-atom Rg violin plots were utilized to explore the structure-function implications of the mutations and RIF binding. Pertaining

to the violin plots, the peaks represented the most sampled conformation whilst the width represented the extent to which the conformation was sampled per mode.

Based on **Figure 4.2A**, the RIF-unbound *mt Mtb*-RNAPs indicated lower Cα-RMSDs relative to the *wt*. The same was observed for *wt*2 as shown in **Supplementary Figure S1.2A**. Furthermore, when collectively looking at the RIF-unbound standard control runs (*wt*1 and *wt*2) violins (**Figure 4.2C**) we found that both displayed more stable conformational distributions compared to most of the *mt*s which appeared to sample multiple conformations. This was particularly evident for the H451L, D441V, D551E and H451D systems. Additionally, we observed notable changes wherein most of the RIF-unbound *mt*s displayed decreases in the calculated average Cα-RMSD estimations relative to the *wt* controls. Overall, this may suggest an attempt to maintain a level of functional integrity through compensatory motions.

In the presence of RIF, we noted that the *wt Mtb*-RNAP controls appeared to become more stable. However, most of the *mt* systems displayed several conformational changes over time and indicated relatively lower Cα-RMSD values as seen in **Figures 4.2B and S1.2B**. The changes in relative stability among the *mt*s were particularly evident in the D551E, H451Y, S456L, I65T, and D441V systems. Complementarily, when looking at the Cα-RMSD violin distribution plots for the RIF-bound systems (**Figure 4.2D**), we found that several of the *mt*s indicated changes in conformational sampling when compared to the *wt* controls. That said, the *mt*s that showed the most noteworthy changes, that is, *mt* proteins displaying multimodal distributions, were the S456L, H451L, D441V, I65T, H451Y, and H451D systems. Additionally, besides the D551E and H451D systems which exhibited notable reductions in average Cα-RMSD values, most of the RIF-bound *mt*s displayed minimal differences in average Cα-RMSDs relative to either one of the *wt* controls. Nonetheless, most of the RIF-bound

mt proteins displayed increases in terms of conformational sampling when compared to the *wt* controls. This may serve as an indicator giving insight into the extent of perturbation.

In summary, the observed stabilizing effects of RIF on the dynamics of the *wt Mtb*-RNAP proteins throughout the simulation may indicate an inhibitory pattern as reported previously (Lin *et al.*, 2017). Contrastingly, the observed decreases in Cα-RMSDs along the structural perturbations shown in the majority of *mt* proteins may corroborate the mutant-induced RNAP inefficiencies that have been experimentally determined and reported (Stefan, Ugur and Garcia, 2018).

Looking at the all-atom Rg estimations, we found that the mutational effect on protein compaction was fairly minimal as indicated for most of the *mt* systems (**Figures 4.2E**). That is, most of the *mt* proteins indicated similar Rg distributions to those of the *wt* controls in the RIF-unbound forms. It was however worth noting that the H451R and H451L proteins exhibited some compactional effect given that both *mt*s displayed larger sampling with multimodal distributions. In the presence of RIF, however, the changes in Rg for the *mt* proteins could be considered inconsequential given the minimal differences relative to the *wt* controls. Additionally, the captured average Rg estimations showed minimal differences across the RIF-unbound and bound proteins. Our analysis of comparative ED results of the RIF-unbound and RIF-bound *wt Mtb*-RNAPs revealed differences in the occupied phase space across PC2, showing a 10.93% and 12.15% captured variance, respectively (**Figure 4.3**).

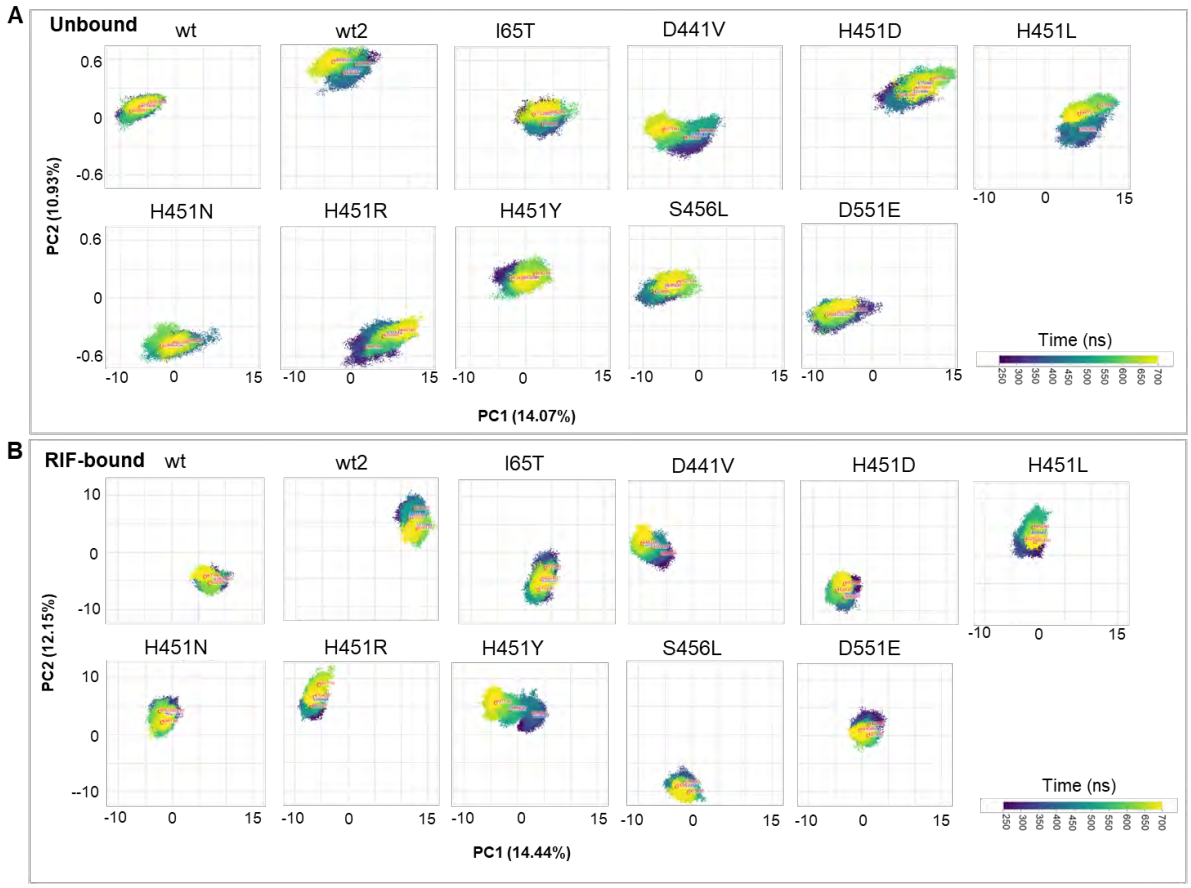


Figure 4.3. Comparative ED plots of the *wt* and *mt Mtb*-RNAPs in the (A) unbound and (B) RIF-bound forms. The dots in each of the plots represent a specific conformation. The dots are also colored based on the time at which each conformation was sampled, which ranges from dark to yellow. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

More similarities were however indicated along PC1 for the RIF-unbound and bound *wt* systems, displaying a captured variance of 14.07% and 14.44%, respectively. The tool employed herein to conduct this analysis addresses a potential issue that may occur when utilizing separate MD simulations to apply the more general ED approach, i.e., the generation of unique sets of eigenvector-eigenvalue pairs. To circumvent this, the tool initially aligns the trajectories, which may admittedly cause the generation of lower PC values but will allow for an explanation of the total variance by yielding a common set of eigenvector-eigenvalue pairs.

As informed by **Figures 4.4 and 4.5**, the observed differences pertaining to conformational distributions between the *wt* controls for the respective RIF-bound and unbound forms may likely be attributable to the discrepancies in either or both the conformational sampling and residue fluctuations detected in some of the subunits.

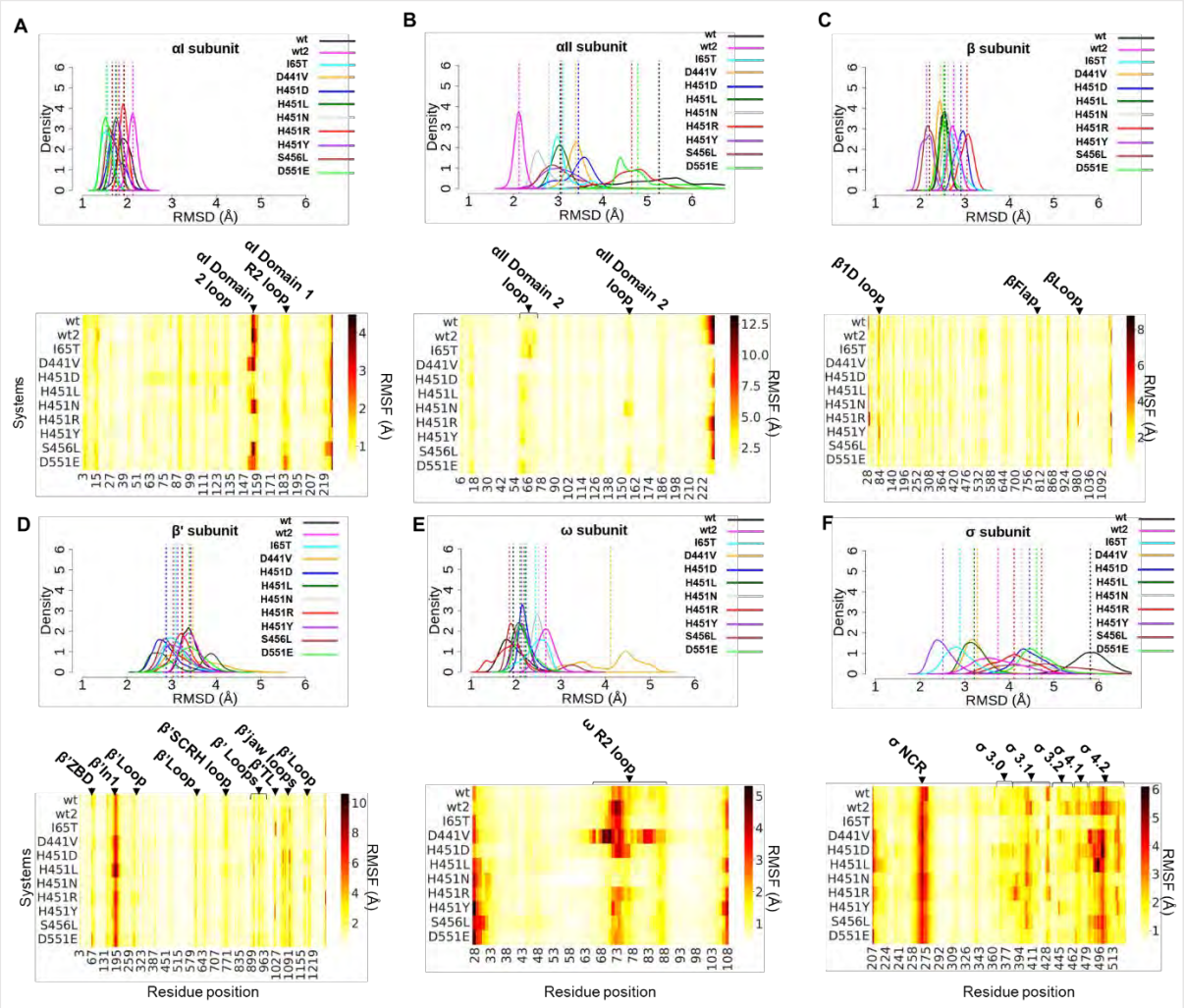


Figure 4.4. Graphical analysis of the *wt* and *mt Mtb*-RNAP subunits of the RIF-unbound form. (A-F, Top) Shows the Cα-RMSD (Å) KDE plots which represent the estimated conformational sampling for the *wt* and *mt Mtb*-RNAP subunits with reference to their corresponding initial structures. (A-F, Bottom) Shows the Cα-RMSF (Å) heatmap representations for the subunits of the respective proteins, detailing the level of fluctuations with the color scheme ranging from yellow (low-level residue fluctuations) to black (higher-level residue fluctuations). The annotated sections represent subunit regions contributing the most to subunit Cα-deviations. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

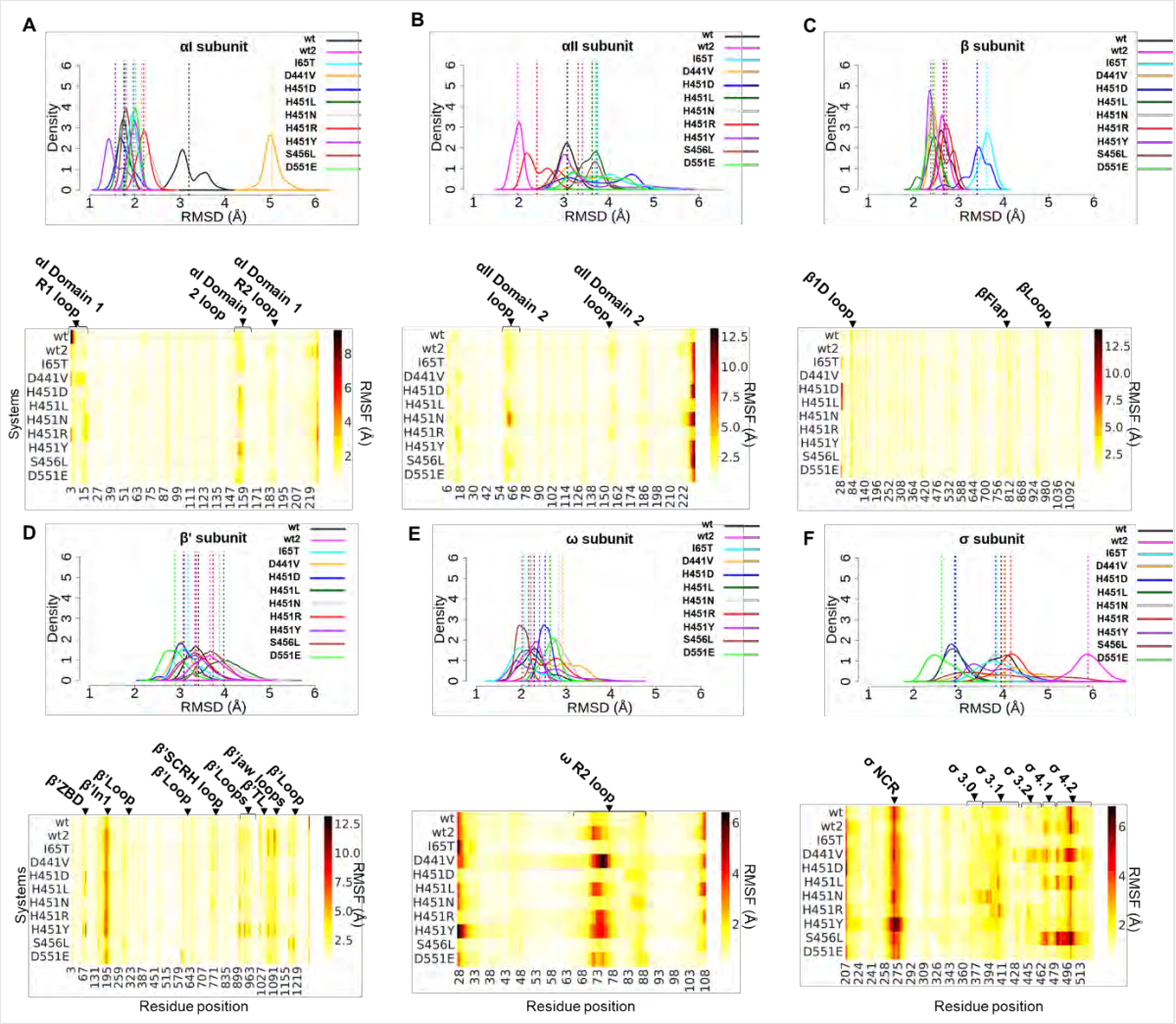


Figure 4.5. Graphical analysis of the wt and mt Mtb-RNAP subunits of the RIF-bound form. (A-F, Top) Shows the Cα-RMSD (Å) KDE plots which represent the estimated conformational sampling for the wt and mt Mtb-RNAP subunits with reference to their corresponding initial structures. (A-F, Bottom) Shows the Cα-RMSF (Å) heatmap representations for the subunits of the respective proteins, detailing the level of fluctuations with the color scheme ranging from yellow (low-level residue fluctuations) to black (higher-level residue fluctuations). The annotated sections represent subunit regions contributing the most to subunit Cα-deviations. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

Hence, the sum of the sampled *Mtb*-RNAP conformations may have contributed to the differences in distributions. Nonetheless, in agreement with the Cα-RMSD violin plots, we observed relatively larger discrepancies along PC1 and PC2 coupled with varying conformational distributions for a majority of the *mt* proteins. The most notable of these were the unbound H451D, H451L, and H451R systems, which displayed some of the

largest changes in occupancy along PC1 as compared to the *wt* controls. Additionally, pertaining to the RIF-bound *mt* systems, the H451N, H451R, D551E, and D441V proteins showed the largest differences in occupancies along PC2 (relative to the controls) as clearly shown in **Figure 4.3A**.

It's also worth noting that all the *mt*s, albeit to a lesser extent for D551E, unambiguously displayed differences along PC1, with the H451Y system in particular displaying a relatively large conformational distribution (**Figure 4.3B**).

We then explored the perturbation effects on the *Mtb*-RNAP subunits in response to the presence and absence of the mutations and/or RIF. These were quantified using subunit Cα-RMSD and Cα-RMSF calculations which were used to measure the respective stabilities and level of residue fluctuations of the *mt* αI-NTD, αII-NTD, β, β', ω, and σ *Mtb*-RNAP subunits (relative to the *wt*). Looking at the Cα-RMSD plots (**Figures 4.4 and 4.5**), we found that the most consistently stable distributions for the *wt* controls were observed in the αI-NTD, ω subunit along with the large β and β' subunits which compose of the RIF-BP and the functionally important DNA/RNA binding regions. Regarding the αI-NTD, we mainly found consistencies for the *wt* controls among the RIF-unbound forms wherein D551E showed the largest discrepancy in conformational sampling relative to the *wt* standards. We were also able to determine slight changes in Cα-RMSD conformational distributions for the β and β' subunits of the *mt* RIF-bound systems. Large changes in conformational distributions (relative to the *wt* controls) were however noted for the RIF-unbound β subunits of the H451R, H451D, S456L, and H451Y proteins. Additionally, although several of the RIF-unbound *mt*s indicated changes in the stability of the catalytic β' subunit when compared to the standards, the most notable changes were observed in the D441V and H451L proteins. When looking at the conformational distributions of

the RIF-bound systems, we found that the I65T and H451D (multimodally distributed) β subunits displayed large changes relative to the *wt* standards, whereas the S456L and H451L β subunits also presented with multimodal distributions which may imply mutation-induced changes that affected the structural integrity of the region. Given the relative agreement in Cα-RMSD among the ω subunits of the *wt* controls for the RIF-bound systems, it's worth noting that conformational variations were observed in the ω subunit for a majority of the *mt*s. The D441V and H451Y proteins, in particular, displayed multimodal distributions which were in contrast to the *wt*s. Also, although only minor changes were observed in the αI-NTD Cα-RMSD distributions of the RIF-unbound *mt* systems, that of the D441V protein particularly registered the largest changes in Cα-RMSD among all the tested *mt*s, potentially highlighting a unique RIF-binding effect. Overall, these detected perturbational effects may collectively impact the catalytic processes of the *Mtb*-RNAP assembly, which would have great implications for the dynamics of the RIF binding site and, by extension, RIF efficacy. Finally, considering the consistencies observed for the *wt* controls, all subsequent structural-dynamic analyses were conducted using *wt*1 (the initial *wt* run) with particular emphasis on the β and β' subunits and their constituents, given the relative agreement across both RIF-bound and RIF-unbound forms of the *wt* controls.

To begin, we conducted solvent accessible surface area (SASA) calculations to approximate the extent to which the *wt* and *mt* β subunits were exposed to the surrounding solvent over the course of the equilibrated timeframes. SASA is an important assessment which has been previously applied to reveal details pertaining to protein folding and stability (Savojardo *et al.*, 2021). Based on **Figure 4.6A-B**, we found that a majority of the unbound *mt*s and all of the RIF-bound *mt*s registered similar SASA values compared to the corresponding *wt*.

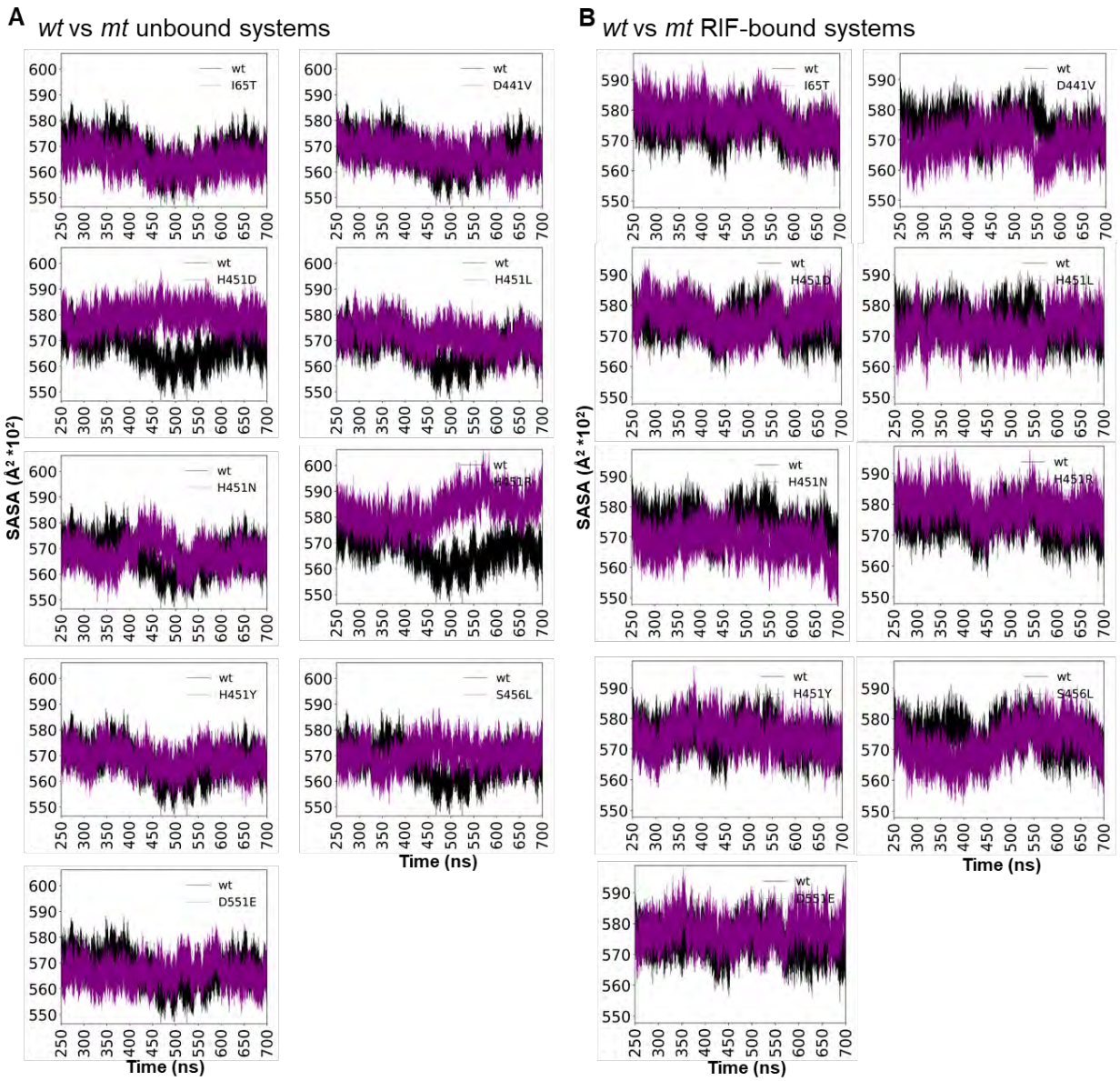


Figure 4.6. Graphical representation of the solvent accessible surface area ($Å^2*10^2$) for the *wt* (coloured black) and *mt* (coloured purple) *Mtb*-RNAP β subunits for the (A) unbound systems and the (B) RIF-bound systems over time.

It was worth noting, however, that the unbound H451D and H451R systems in particular showed increased SASA values relative to the *wt*, suggesting more unstable β subunits, which was in agreement with Cα-RMSD observations captured in **Figure 4.4**.

4.3. Mutation-induced destabilization of key functional *Mtb*-RNAP domains

At the core of *Mtb*-RNAP's catalytic activities, as facilitated by functionally critical structural domains, is the processive addition of nucleoside triphosphates (NTPs) to a growing RNA strand to produce a nascent RNA molecule; a process referred to as the nucleotide addition cycle (Basu *et al.*, 2014). Mutational effects along with RIF binding have been shown to alter or disrupt the nucleotide addition cycle by interfering with the inherent dynamics surrounding catalysis (Campbell *et al.*, 2001; Lin *et al.*, 2017; Stefan, Ugur and Garcia, 2018). Herein, we investigated the complementary effects associated with functional domain dynamics in an attempt to elucidate some of the elusive details not covered in the current literature. Therefore, we proceeded to employ Cα-RMSD and Cα-RMSF calculations to analyse and explore the varying stabilities involving specific β and β' functional domains (**Figure 1.3**), with emphasis on those particularly involved in enzymatic catalysis and function. Therefore, as shown in **Figures 4.7-4.13**, the βFL2 domain (residues 459–473), β'ZBD domain (residues 24–98), β'In1 domain (residues 140–228), and the β'jaw domain (residues 1039–1218) were investigated. This further included the catalytic β'NADFDGD motif (residues 533–539), the β'BH domain (residues 847–889), and the β'TL domain (residues 994–1037).

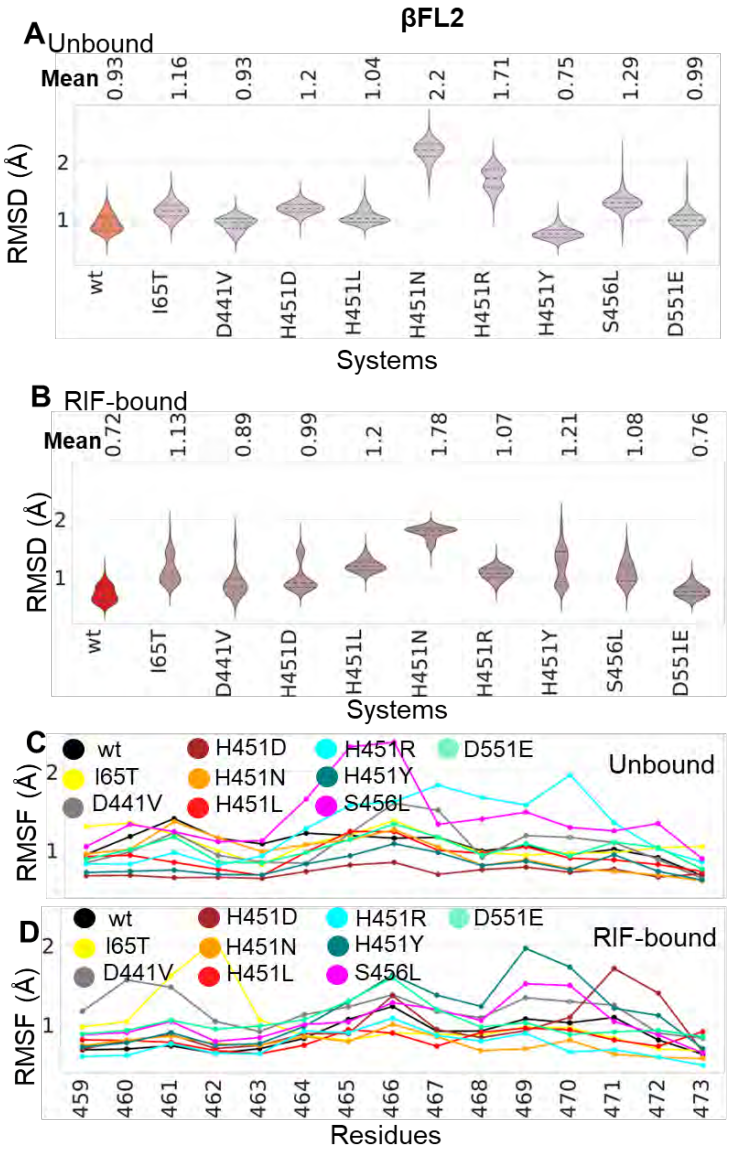


Figure 4.7. Structural dynamic analysis of the βfork loop2 (βFL2) of the *wt* and *mt Mtb*-RNAP proteins. Cα-RMSD estimations (with respect to the initial structures) were represented as violin distribution plots for the (A) RIF-unbound and (B) RIF-bound *wt* (RIF-unbound colored orange; RIF-bound colored red) and *mt* (RIF-unbound colored pink; RIF-bound colored brown) *Mtb*-RNAP structural domains. The Cα-RMSD mean values were shown above the respective plots. The dotted lines within each of the violin distributions represented the 25th, 50th, and 75th quartiles. Cα-RMSF line plots were also shown for the (C) RIF-unbound and (D) RIF-bound *wt* and *mt Mtb*-RNAP structural domains and colored as indicated by the keys. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

Based on the Cα-RMSD estimation results (**Figure 4.7A**), we found that the *wt* βFL2 domain displayed consistently lower Cα-RMSDs in the RIF-unbound and bound forms.

However, the β'In1 domain expressed similar conformational behaviour among the *wt* and *mt*s, while further displaying similar residue-level fluctuations (**Figure 4.8**).

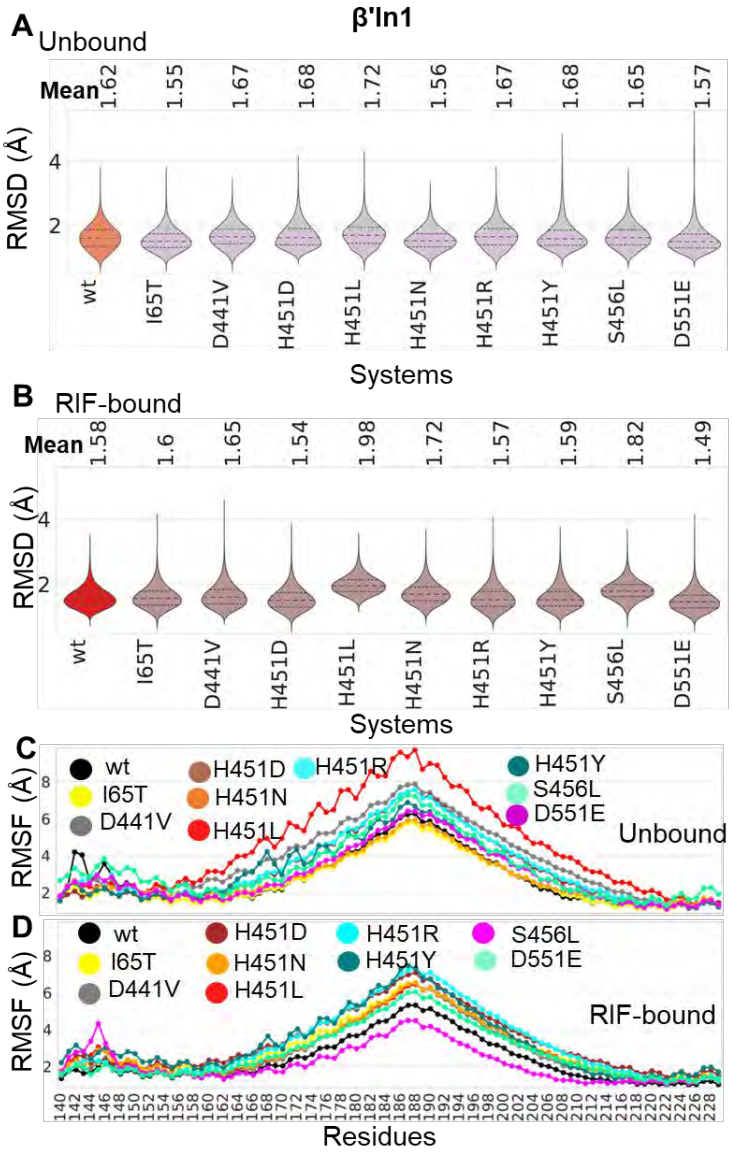


Figure 4.8. Structural dynamic analysis of the β'In1 domain of the *wt* and *mt Mtb*-RNAP proteins. Cα-RMSD estimations (with respect to the initial structures) were represented as violin distribution plots for the (A) RIF-unbound and (B) RIF-bound *wt*s (RIF-unbound colored orange; RIF-bound colored red) and *mt* (RIF-unbound colored pink; RIF-bound colored brown) *Mtb*-RNAP structural domains. The Cα-RMSD mean values were shown above the respective plots. The dotted lines within each of the violin distributions represented the 25th, 50th, and 75th quartiles. Cα-RMSF line plots were also shown for the (C) RIF-unbound and (D) RIF-bound *wt* and *mt Mtb*-RNAP structural domains and colored as indicated by the keys.

In contrast, the majority of *mt* protein β'ZBDs, β'TL domains, β'NADFDGD motifs, β'BH domains, and β'jaw domains exhibited greater differences in Cα-RMSDs and conformational distributions relative to the *wt*. This observation was complemented with domain Cα-RMSF results wherein instances of correlation between the extent of changes in residue fluctuations and Cα-RMSD could be identified. However, relative to the *wt*, the functional domains of the *mt* proteins displayed varying changes as reflected by the Cα-RMSD and Cα-RMSF plots. This might therefore reflect some of the mutational perturbation effects on the stability and flexibility of the inspected functional regions, particularly when considering the consequential effect on RIF binding and efficacy. To start with, as indicated by the large differences in Cα-RMSDs captured in **Figure 4.7A**, we found that most of the *mt* βFL2 domains of the RIF-unbound systems displayed relatively higher instabilities as compared to the *wt*. This was particularly evident in the H451D, S456L, H451R (bimodal), and H451N proteins. Regarding the RIF-bound systems, we found that the *mt*s again exhibited large changes in stability as shown by the higher Cα-RMSDs, especially H451N. Additionally, the βFL2 domains of the D441V, H451D, S456L, and H451Y proteins displayed increased conformational sampling as seen in **Figure 4.7B**. Exploring the Cα-RMSF results showed that the instabilities explained through Cα-RMSD for the RIF-unbound S456L and H451R systems along with the RIF-bound I65T and H451Y systems, may in part be informed by the registered increases in residue fluctuations shown in **Figure 4.7(C and D)**, with absolutely conserved residue βR465 (Lane and Darst, 2010b) being of particular interest among the RIF-unbound proteins. As depicted in **Figure 4.9A**, we deduced that all the *mt*s of the RIF-unbound proteins exhibited a decrease in β'ZBD's Cα-RMSD, whereas the D551E, H451R, S456L, H451Y, H451D, and H451N systems indicated multimodal distributions.

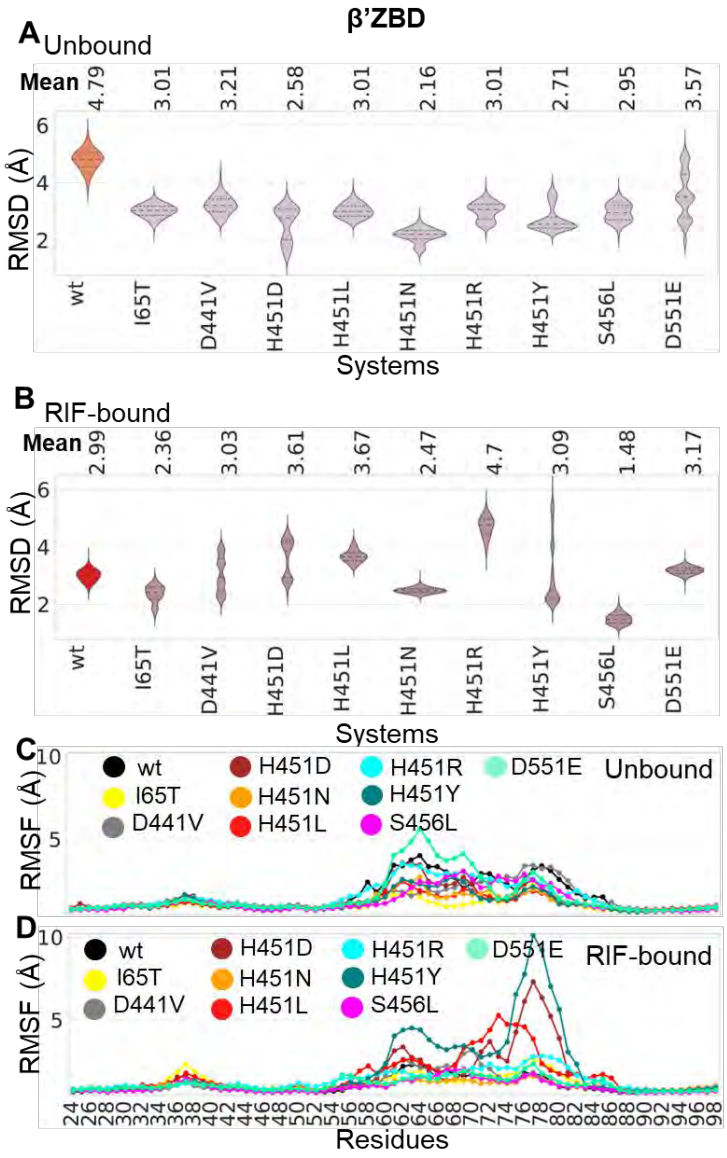


Figure 4.9. Structural dynamic analysis of the N-terminal zinc-binding domain (β'ZBD) domain of the *wt* and *mt Mtb*-RNAP proteins. Cα-RMSD estimations (with respect to the initial structures) were represented as violin distribution plots for the (A) RIF-unbound and (B) RIF-bound *wt* (RIF-unbound colored orange; RIF-bound colored red) and *mt* (RIF-unbound colored pink; RIF-bound colored brown) *Mtb*-RNAP structural domains. The Cα-RMSD mean values were shown above the respective plots. The dotted lines within each of the violin distributions represented the 25th, 50th, and 75th quartiles. Cα-RMSF line plots were also shown for the (C) RIF-unbound and (D) RIF-bound *wt* and *mt Mtb*-RNAP structural domains and colored as indicated by the keys. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

Furthermore, as indicated in **Figure 4.9B**, large discrepancies in β'ZBD Cα-RMSDs among the RIF-bound proteins were noted, with H451Y, H451R, H451L, H451D,

H451N, I65T, and S456L systems displaying especially large changes. Complementarily, large residue fluctuations were detected between β' residues 54–84 which encompassed the zinc-coordinating β'C60, β'C62, β'C75, and β'C78 residues that play a role in folding of the β'ZBD (Hu and Liu, 2022). The detected instabilities were particularly evident in the H451Y, H451D, and H451L proteins, which may have greatly contributed to the extent of conformational sampling (**Figure 4.9D**). Looking at the results shown in **Figure 4.10A**, we observed that the catalytic NADFDGD motif displayed similar stabilities amongst the RIF-unbound *wt* and *mt*s as indicated by the similarities in Cα-RMSDs and conformational distributions. However, notable exceptions included the D441V and D551E proteins, which exhibited higher Cα-RMSDs relative to the *wt*. We additionally observed that the increased instabilities were not further supplemented by any particularly large deviations in residue fluctuations as portrayed in **Figure 4.10C**. Interestingly, the RIF-bound *mt* proteins, particularly I65T, H451N, H451R, and S456L, displayed more stable NADFDGD motifs as compared to the *wt* which displayed one of the most unstable motif structures, suggesting a change in catalytic proficiency. Also, although the majority of residues did not exhibit great changes in fluctuations for most of the proteins, the S456L protein in particular displayed larger fluctuations for residue β'ASN533 (**Figure 4.10D**).

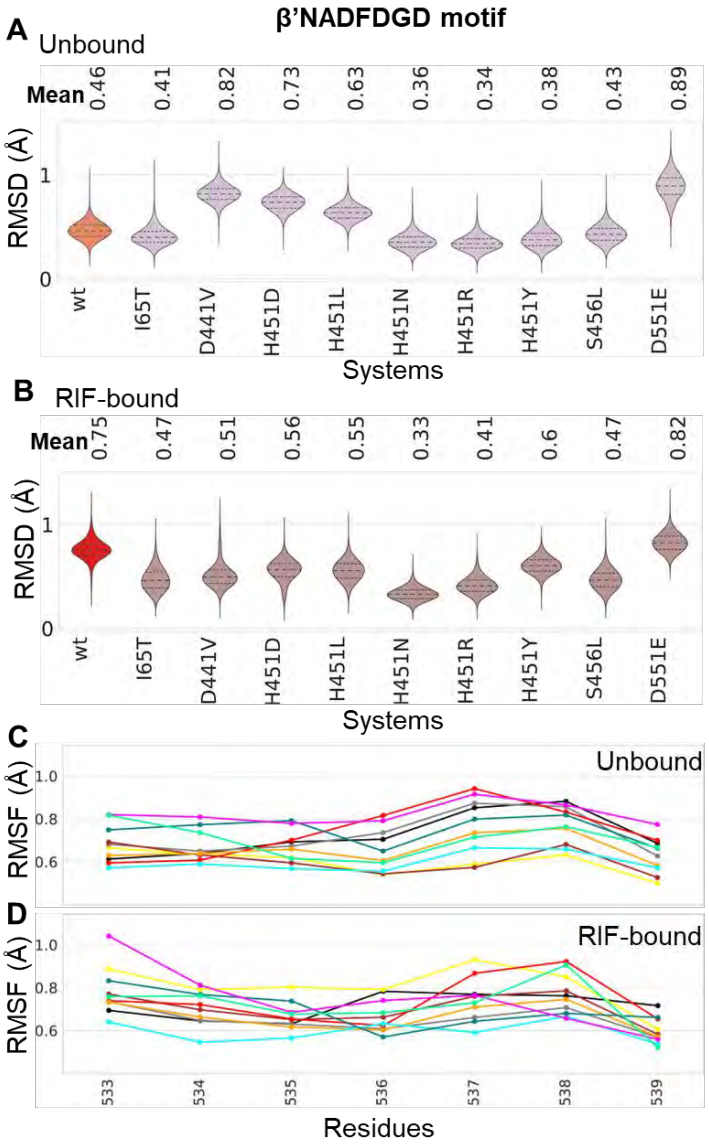


Figure 4.10. Structural dynamic analysis of the β'NADFDGD motif of the *wt* and *mt* *Mtb*-RNAP proteins. Cα-RMSD estimations (with respect to the initial structures) were represented as violin distribution plots for the (A) RIF-unbound and (B) RIF-bound *wt* (RIF-unbound colored orange; RIF-bound colored red) and *mt* (RIF-unbound colored pink; RIF-bound colored brown) *Mtb*-RNAP structural domains. The Cα-RMSD mean values were shown above the respective plots. The dotted lines within each of the violin distributions represented the 25th, 50th, and 75th quartiles. Cα-RMSF line plots were also shown for the (C) RIF-unbound and (D) RIF-bound *wt* and *mt* *Mtb*-RNAP structural domains and colored as indicated by the keys.

Based on the results depicted in **Figure 4.11A**, we found that the RIF-unbound *wt* generally exhibited greater β'BH stability as compared to the *mt*s, with D441V, H451N,

S456L, and D551E proteins indicating the greatest differences in Cα-RMSDs that could not be explained by the Cα-RMSF results (**Figure 4.11C**).

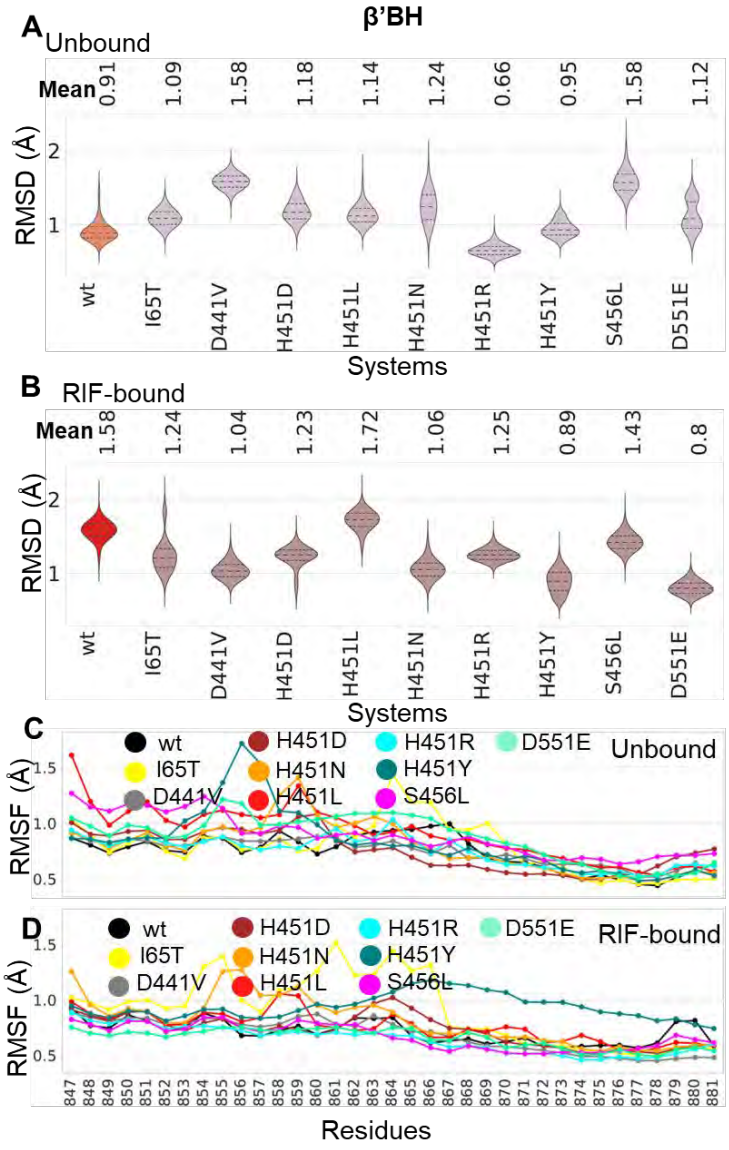


Figure 4.11. Structural dynamic analysis of the β'bridge helix (β'BH) domain of the *wt* and *mt Mtb*-RNAP proteins. Cα-RMSD estimations (with respect to the initial structures) were represented as violin distribution plots for the (A) RIF-unbound and (B) RIF-bound *wt* (RIF-unbound colored orange; RIF-bound colored red) and *mt* (RIF-unbound colored pink; RIF-bound colored brown) *Mtb*-RNAP structural domains. The Cα-RMSD mean values were shown above the respective plots. The dotted lines within each of the violin distributions represented the 25th, 50th, and 75th quartiles. Cα-RMSF line plots were also shown for the (C) RIF-unbound and (D) RIF-bound *wt* and *mt Mtb*-RNAP structural domains and colored as indicated by the keys.

Our observations pertaining to the RIF-bound systems, however, indicated a contrasting effect wherein most of the *mt*s displayed greater stabilities as compared to the *wt* (**Figure 4.11B**). The proteins that displayed particularly great differences in estimated β'BH Cα-RMSDs relative to the *wt* included the D441V, H451Y, and D551E proteins. Also of note is the multimodal distribution with increased conformational sampling displayed by the I65T protein, which further expressed enhanced residue-level fluctuations at β' residues 853–866 of the β'BH domain. Additionally, the H451Y protein interestingly displayed increased fluctuations for β' residues ~860–881, which included the hinge residues 868–870 that play a role in β'BH bending and straightening (Tuske *et al.*, 2005). Regarding the conformational sampling of the catalytic β'TL domain, we found that the RIF-unbound *mt* proteins (except D441V and S456L) exhibited multimodal Cα-RMSD distributions as can be seen in **Figure 4.12A**, with I65T, H451R, H451Y, D441V, and H451L proteins displaying the highest differences in Cα-RMSDs compared to the *wt* control. Based on our observations from **Figure 4.12B**, most of the RIF-bound *mt*s displayed great changes in the β'TL domain's Cα-RMSDs relative to the *wt* control while additionally displaying unimodal distributions. The *mt* proteins that showed particularly noteworthy differences in Cα-RMSDs in this regard were the I65T, H451N, H451Y, H451R, H451D, D551E, and D441V proteins. Among these proteins, large residue fluctuations were especially observed for the I65T and H451Y proteins between β' residues 1006–1028 (also similarly observed among the RIF-unbound forms (**Figure 4.12B–D**), which encompasses the β'R1012 and β'H1016 residues that have been reportedly deduced to orient incoming NTP phosphates (Mishanina *et al.*, 2017). Another functional domain indicating high differences in Cα-RMSDs for the *mt*s (relative to the *wt* control) was the β'jaw domain.

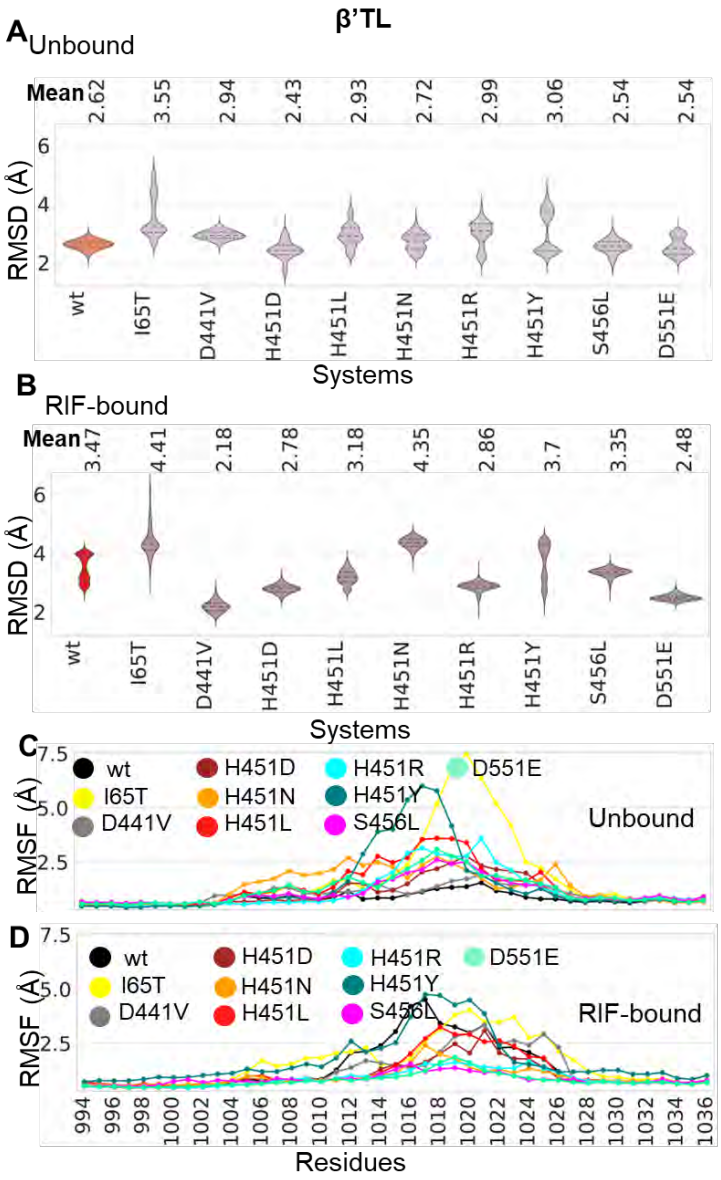


Figure 4.12. Structural dynamic analysis of the β'trigger loop (β'TL) domain of the *wt* and *mt* *Mtb*-RNAP proteins. Cα-RMSD estimations (with respect to the initial structures) were represented as violin distribution plots for the (A) RIF-unbound and (B) RIF-bound *wt* (RIF-unbound colored orange; RIF-bound colored red) and *mt* (RIF-unbound colored pink; RIF-bound colored brown) *Mtb*-RNAP structural domains. The Cα-RMSD mean values were shown above the respective plots. The dotted lines within each of the violin distributions represented the 25th, 50th, and 75th quartiles. Cα-RMSF line plots were also shown for the (C) RIF-unbound and (D) RIF-bound *wt* and *mt* *Mtb*-RNAP structural domains and colored as indicated by the keys. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

Of the *mt* proteins that showed Cα-RMSD differences compared to the *wt*, the I65T, D551E, S456L, H451Y, D441V, H451D, and H451L proteins were the most notable,

as can be seen in **Figure 4.13A**. From **Figure 4.13B**, we were able to identify clear increases in β'jaw domain Cα-RMSDs which were particularly pronounced in the RIF-bound H451L, H451Y, D551E, H451D, and D441V proteins.

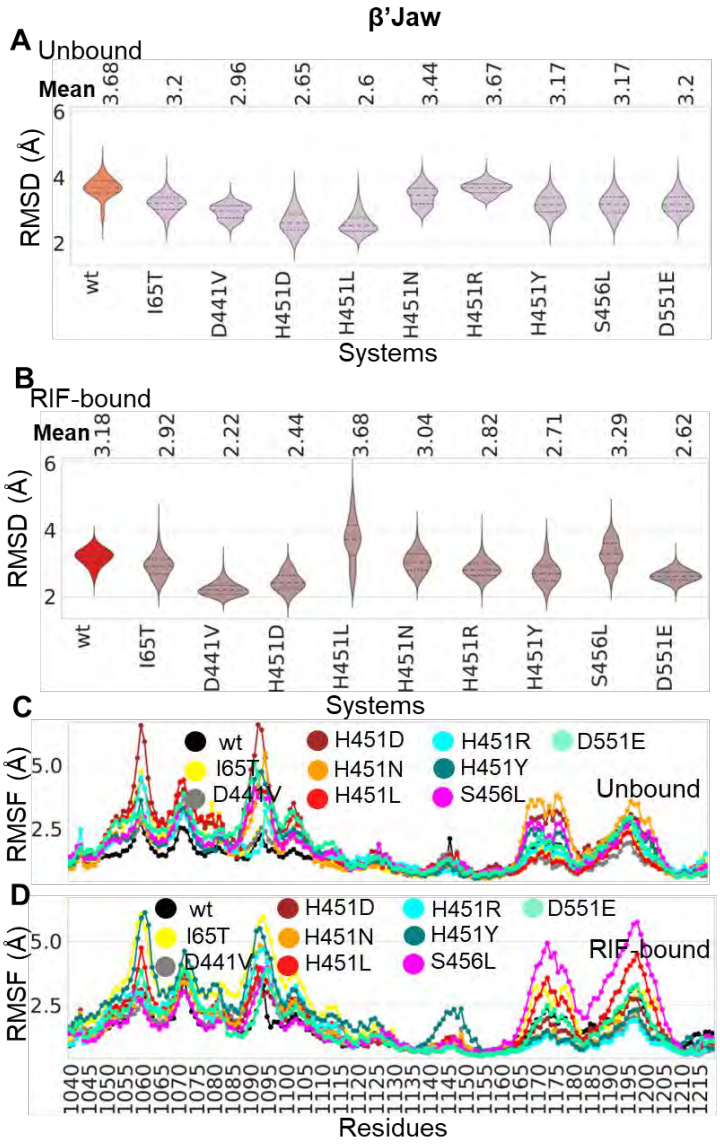


Figure 4.13. Structural dynamic analysis of the β'jaw domain of the *wt* and *mt Mtb*-RNAP proteins. Cα-RMSD estimations (with respect to the initial structures) were represented as violin distribution plots for the (A) RIF-unbound and (B) RIF-bound *wt* (RIF-unbound colored orange; RIF-bound colored red) and *mt* (RIF-unbound colored pink; RIF-bound colored brown) *Mtb*-RNAP structural domains. The Cα-RMSD mean values were shown above the respective plots. The dotted lines within each of the violin distributions represented the 25th, 50th, and 75th quartiles. Cα-RMSF line plots were also shown for the (C) RIF-unbound and (D) RIF-bound *wt* and *mt Mtb*-RNAP structural domains and colored as indicated by the keys. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

To further complement these observations, we again took a look at residue-level flexibility in the β'jaw domain for the RIF-bound and unbound forms as shown in **Figures 4.13C and D** and found changes in local flexibility in both cases. The H451L and S456L proteins displayed particularly high differences in flexibility for the bound forms, which may suggest some impact on critical interactions that occur downstream of DNA (Saecker, Record and Dehaseth, 2011).

Collectively, our analysis of the extensive changes in stabilities indicates a series of mutation- and RIF-induced perturbations on the functional domains and on *Mtb*-RNAP as a whole, which of course may potentially impact some core enzymatic functions. The β'TL and β'ZBD, for instance, play respective roles in the catalysis of transcription (Wang *et al.*, 2006) and the termination and antitermination of the transcription process (King *et al.*, 2004). Previous reports covering the diminished enzymatic capacity of the missense-mutated RNAPs to perform catalytic functions (Stefan, Ugur and Garcia, 2018) make our findings particularly interesting and may suggest a connection between our observations and the documented fitness costs (Lane and Darst, 2010b; Boyaci, Saecker and Campbell, 2020).

4.4. Mutational disruption of the RIF-BP, RIF 'active orientation' and binding affinity

To further understand the underlying mechanisms involved in how clinically relevant mutations affect RIF efficacy, we conducted the following investigation into the structural dynamic effects on the RIF-BP along with ligand dynamics and binding affinity. To begin with, dynamic cross-correlation (DCC) analysis was conducted on the RIF-BP to examine the extent of concerted/correlated atomic motions in the same direction throughout the course of the simulations. The structural dynamics of the RIF-

BP were then determined using Cα-RMSD and all-atom Rg estimations which aided in describing perturbation effects on the pocket for the *wt* and *mt* proteins. We additionally calculated center of mass (CoM) estimations to measure the distance changes between the RIF-BP and RIF.

Our analysis of **Figure 4.14** revealed that most of the RIF-unbound RIF-BP residue motions for the *wt* were correlated, suggesting limited conformational sampling (Sanyanga and Tastan Bishop, 2020). However, although the *mt*s tended to exhibit correlated motions, the I65T, D441V, H451D, H451Y, and S456L systems in particular, displayed a number of residues with less correlated motions, especially the Q442-R465 pair. Albeit it to a lesser extent, a similar trend wherein the *mt* RIF-BP residues displayed less correlated motions was observed for the RIF-bound systems. Although in this case, the I65T, H451D, H451L, H451N, H451Y, S456L, and D551E systems displayed the most notable correlation value differences relative to the *wt* (**Figure 4.15**).

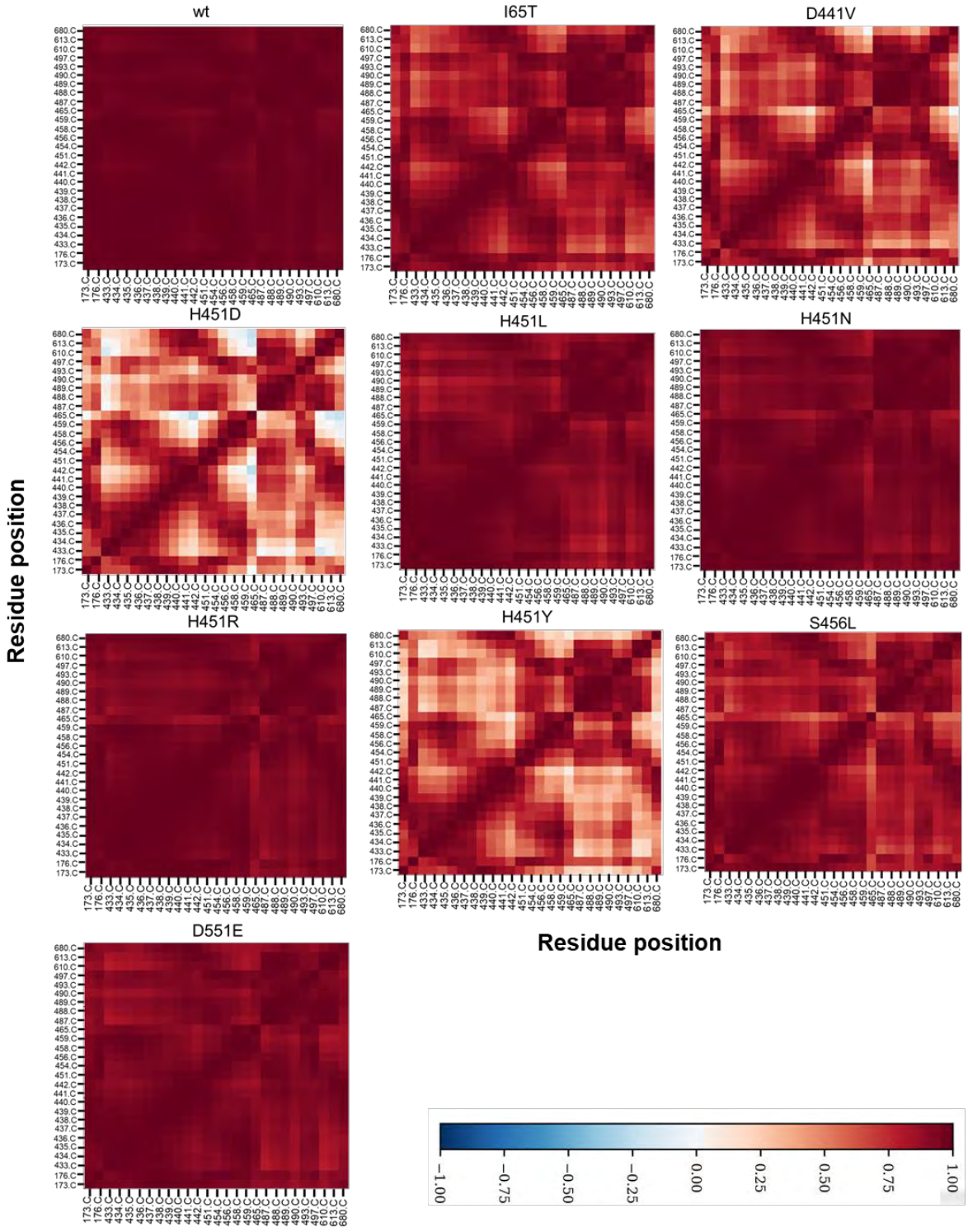


Figure 4.14. Heatmap representation of the dynamic cross-correlation (DCC) for the RIF-unbound *wt* and *mt* RIF-BPs of *Mtb*-RNAP. A DCC value of -1 indicated anticorrelated residue motions, 0 suggested a lack of correlation while a value of 1 was indicative of correlated motions.

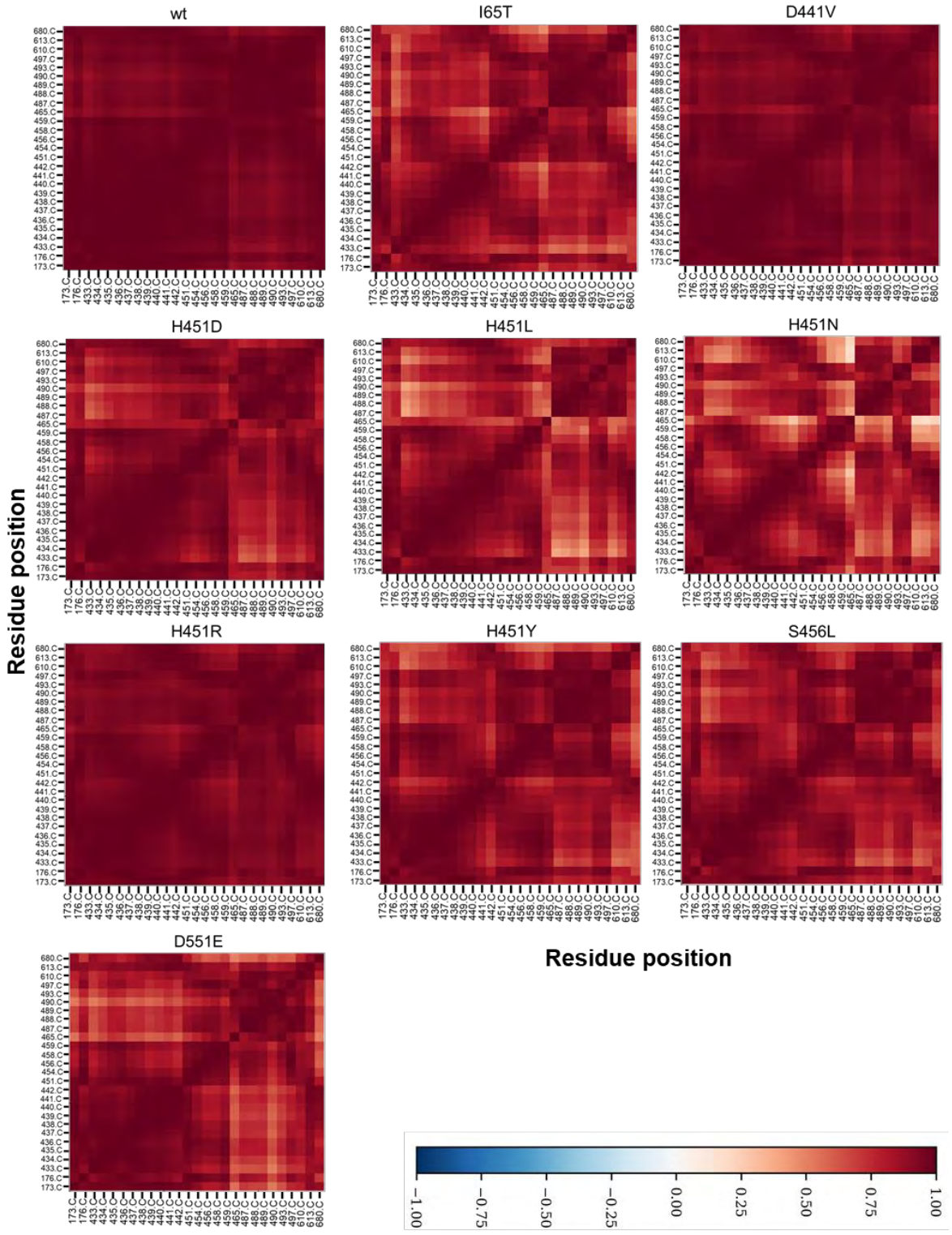


Figure 4.15. Heatmap representation of the dynamic cross-correlation (DCC) for the RIF-bound *wt* and *mt* RIF-BPs of *Mtb*-RNAP. A DCC value of -1 indicated anticorrelated residue motions, 0 suggested a lack of correlation while a value of 1 was indicative of correlated motions.

Based on **Figure 4.16A**, we found that the unbound RIF-BPs of the *mt* systems displayed varying magnitudes of instability when compared to the *wt* control. RIF-BP instabilities were particularly evident in the H451D, H451N, H451Y, S456L, and D551E proteins which was in agreement with our DCC results.

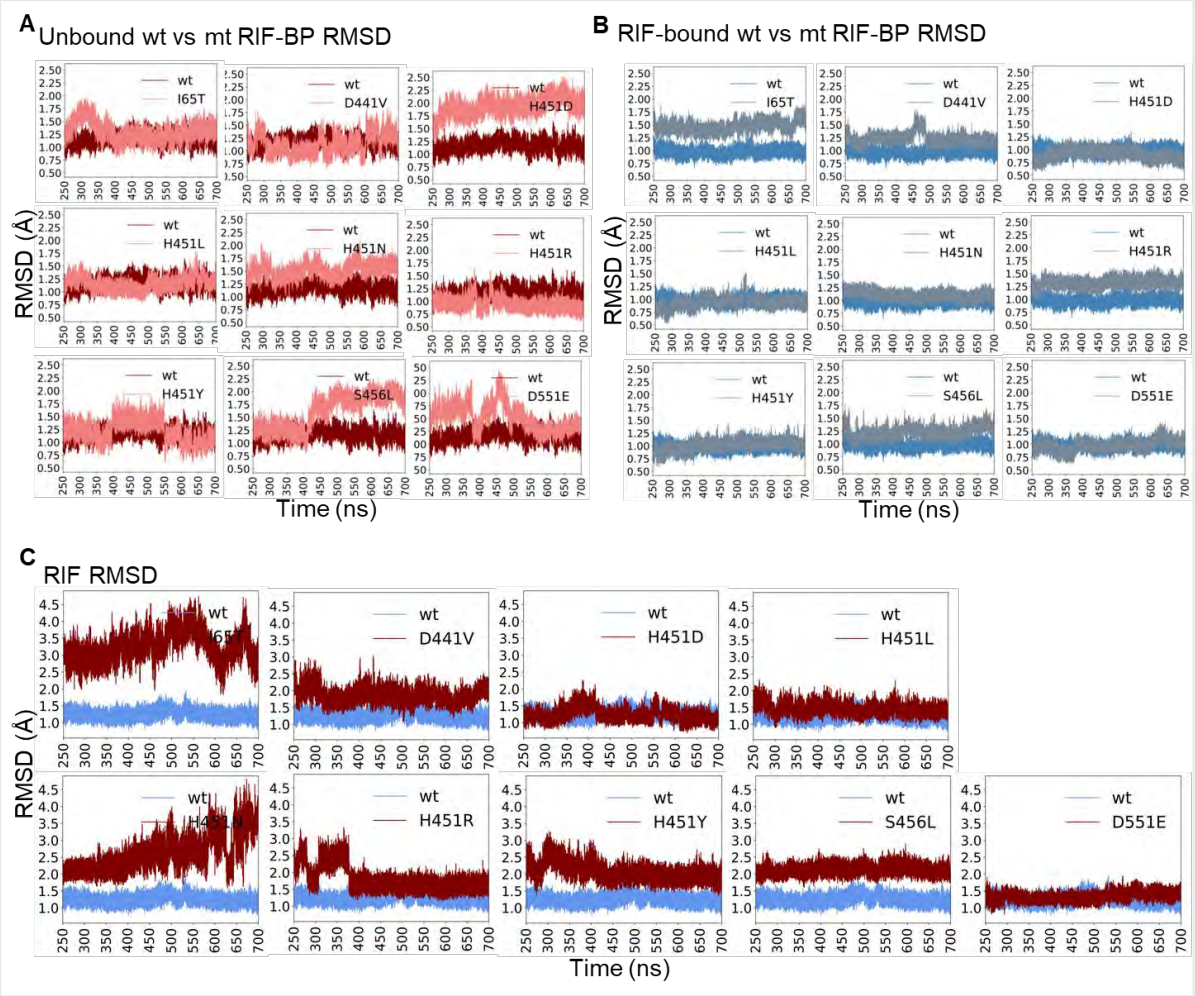


Figure 4.16. Graphical representation of the stability of RIF and the RIF-BP over time. Cα-RMSD (Å) estimation line plots of the RIF-BP are shown for the wt (unbound, colored maroon; RIF-bound, colored blue) and mts (unbound, colored orange; RIF-bound, colored pink) with reference to their initial structures for the (A) unbound and (B) RIF-bound states. (C) Shows the RMSD line plots of RIF (with reference to the RIF-BP) for the *wt* (colored blue) and *mt*s (colored maroon) over time.

Looking at the Cα-RMSD conformational distributions depicted in **Figure 4.17C**, we observed that the *wt* control exhibited a unimodal distribution while most of the *mt* systems displayed multiple equilibria which were most distinctly noted for the S456L, D551E, H451Y, D441V, and I65T proteins. Taken together, we were able to determine possible correlations between the mutational and perturbation effects on the RIF-BP, which highlighted a set of key residues such as R173, T433, Q438, F439, D441, P489, N493, and H680 that are needed to establish essential interactions to permit a stable RIF binding (Lin *et al.*, 2017; Srivastava *et al.*, 2018; Zhang *et al.*, 2019) (**Figure 4.17G**).

The presence of RIF appeared to stabilize the *wt* and *mt* RIF-BPs as suggested by the relatively lower Cα-RMSD estimations. RIF-BP perturbations were however consistent in the I65T, D441V, H451R, and S456L proteins, despite their tendency towards unimodal distributions (**Figure 4.17C**). Based on the all-atom Rg distribution plots in **Figure 4.17D**, we found that the mutational effects on the compact state of the RIF-BP were considerable, especially when considering the RIF-unbound forms. That is, the S456L, H451Y, H451R, H451L, and H451N proteins displayed relatively higher Rg values and unimodal distributions as compared to the *wt* control. However, the I65T, D551E, and H451D RIF-BPs displayed similar distributions to the *wt* protein, suggesting a minimal impact on compaction of the pocket. When considering the flexibility of the pocket in the presence of RIF, we found that the majority of the investigated proteins, with the exception of the H451N protein, exhibited varying residue fluctuations relative to the *wt*.

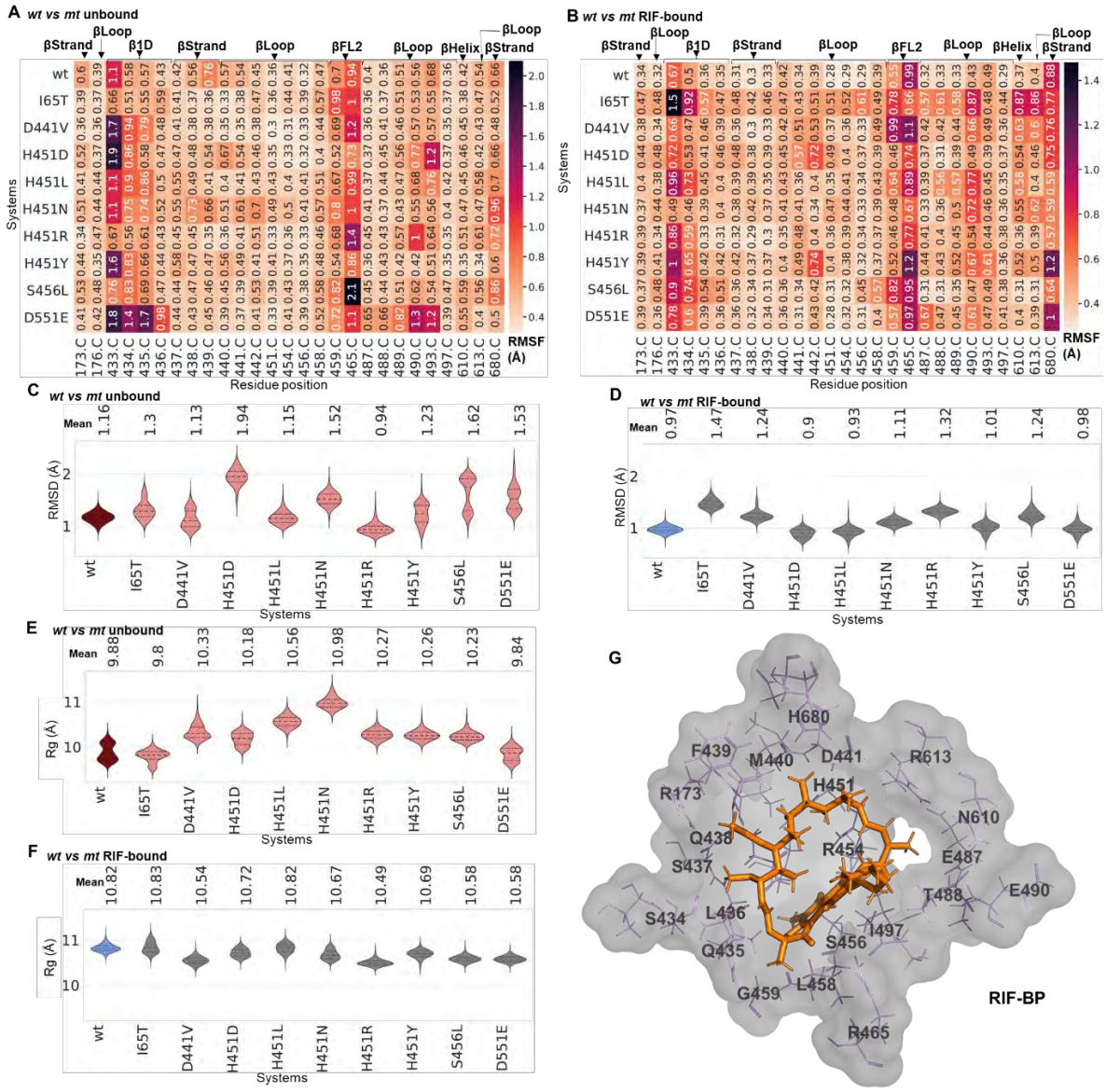


Figure 4.17. Structural-dynamic investigation of the *wt* and *mt Mtb*-RNAP RIF-BPs. Annotated heatmaps representing the Cα-RMSF estimations are shown for the (A) RIF-unbound and (B) bound *wt* and *mt* RIF-BPs. The heatmaps are colored from orange to dark purple and represent a range of low-to-high fluctuating regions respectively. Violin distribution plots representing the Cα-RMSD are depicted for the (C) unbound and (D) RIF-bound *wt* (unbound, colored maroon; RIF-bound, colored blue) and *mt*s (unbound, colored orange; RIF-bound, colored pink). All-atom Rg violin distribution plots are similarly illustrated for the I RIF-unbound and (F) bound proteins. The mean values are indicated above the violins whereas the dotted lines denote the 25th, 50th, and 75th quartiles. (G) Shows the 3D structure of the residue labelled RIF-BP (surface represented and colored grey) with the bound RIF drug (stick represented and colored orange). Adapted from (Monama, Olotu and Tastan Bishop, 2023).

Among the identified proteins displaying large average Cα-RMSF values, were the *wt* and similarly the I65T and D551E proteins, which happen to be distal mutations (**Figure 4.17E**). Altogether, this may indicate implemented changes in the RIF-BPs needed to accommodate RIF. Structural discrepancies in the *mt* RIF-BPs can be largely attributed to the fluctuation changes in the β1D R173, T433, and S434, β loop F439, and βFL2 R465 pocket residues as shown in the corresponding RMSF heatmaps (**Figure 4.17A**). Residues that consistently showed notable differences in fluctuation among the RIF-bound *mt* proteins, were the β residues T433, S434, R465, E490, and H680, as indicated in **Figure 4.17B**. Missense mutations such as D441V and H451Y that involve changes in the ionic state of the side chain, along with mutations involving changes in the size of the side chain, S456L for instance, are factors that may greatly influence the stable interaction network of the pocket, as suggested by observations made from **Figure 4.17C**. It's worth noting that the I65T protein which exhibited one of the more highly unstable RIF bindings, displayed particularly large fluctuations at the β1D residues T433 and S434. The D551E protein, however, displayed minimal changes at the same site, which appeared to result in a relatively stable yet altered RIF binding. Complementary to our RIF-BP results, we found that the ED results revealed increased RIF-BP sampling in various phase spaces for the *mt*s, additionally indicating a total explained variance of 80.18% and 5.42% along PC1 and PC2, respectively (**Figure 4.18A**). Much like the Cα-RMSD results, the *mt*s appeared to display increased conformational sampling relative to the *wt* while indicating differences in their phase space occupancies, as can be seen along both PC1 (78.56%) and PC2 (5.47%) (**Figure 4.18B**). It's also worth noting that, although the RIF-BPs displayed similarities along PC1 among the RIF-unbound and the bound *mt*s, larger differences were identified along PC2.

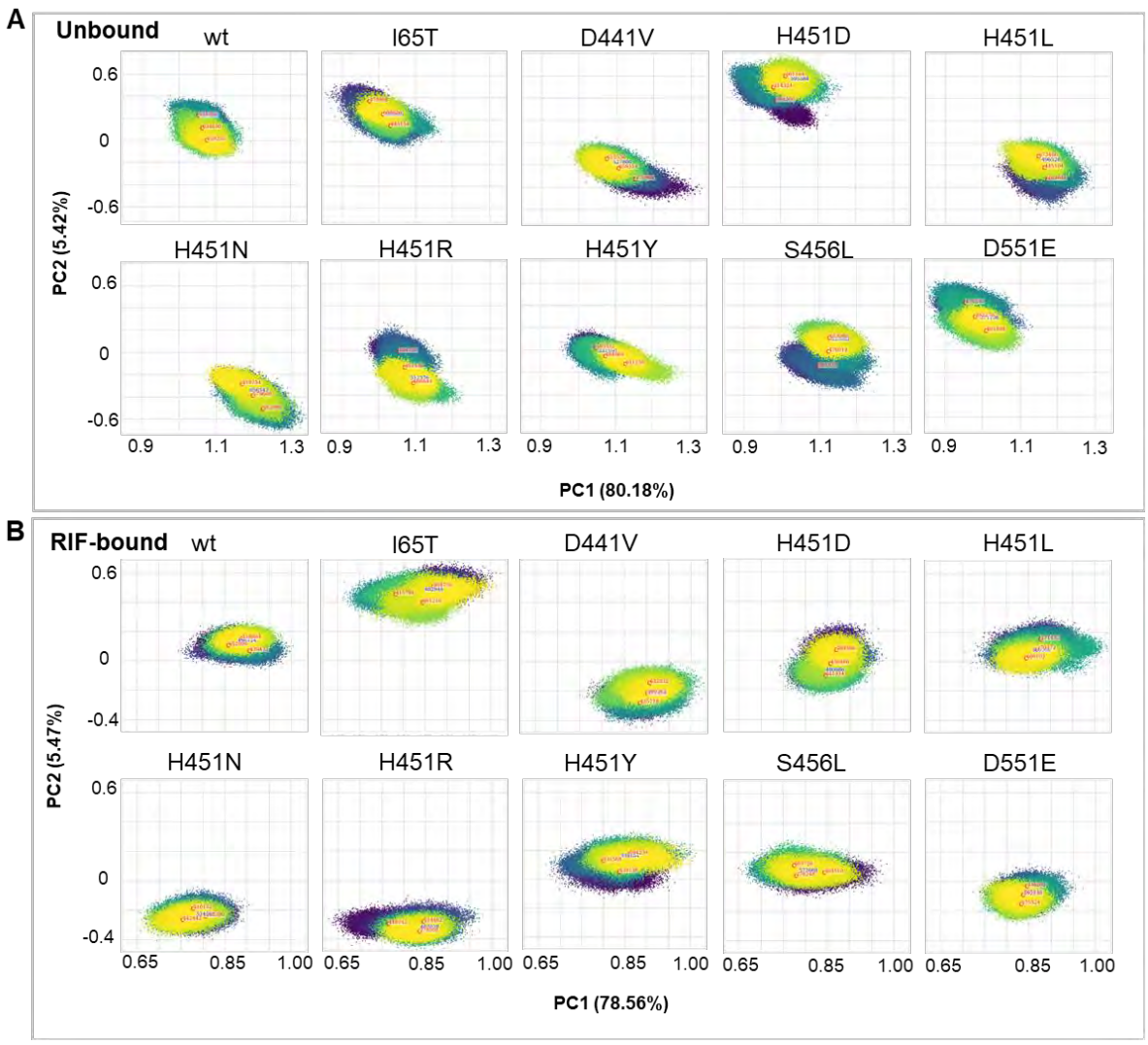


Figure 4 18. Comparative ED plots of the *wt* and *mt* RIF-BPs in the (A) unbound and (B) RIF-bound forms. The dots in each of the plots represent a specific conformation. The dots are also colored based on the time at which each conformation was sampled, which ranges from dark to yellow.

When we investigated the stability of RIF through RMSD (referencing the RIF-BP structure), we found that the characteristic mutation-induced perturbation effects consequently affected the stability of RIF, as depicted in **Figures 4.18B and C**. Looking at the ligand RMSD plots, we deduced that RIF exhibited especially heightened instability for the H451N, I65T, H451Y, S456L, H451D, D441V, and H451R proteins (**Figure 4.18C**). Noticeable changes in the RIF of the H451R protein in particular could in part be associated with the arginine substituent which is

characterized by its longer side chain than histidine, thus reducing RIF's ability to fit in the pocket, and may therefore explain the detected multimodal equilibria. A similar effect may also account for the high RMSDs detected for the S456L system's RIF which acquired a longer substituent, hence causing RIF displacement. In addition, specific substitutions of histidine in the H451R and H451N mutations may affect the formation of the ring-stabilizing π interactions that form with RIF, which may in part account for the unstable conformations observed in both *mt* proteins. Contrastingly, RIF bound to the *wt* protein displayed much more stability as indicated by the detected RMSDs, which may likely underlie a stable binding mode at the RIF-BP. Also, the stability of RIF for the H451L and D551E systems however appeared largely unaffected within the pocket, which may suggest either subtle or minimal overall effects on RIF efficacy.

The mutational impact on RIF binding and stability can be additionally explained by some of the drastic changes in hydrogen bond (H-bond) occupancies which informed the variations in RIF's active orientation, as shown in **Figure 4.19** and **Figure 4.20**. As previously mentioned, RIF's efficacy primarily hinges on a specific orientation within the pocket to exert the steric occlusion effects meant to prevent the extension of nascent RNA (Lin *et al.*, 2017). Hence, the way these mutations alter the active RIF pose may be an instrumental part of the RIF resistance mechanism. Looking at **Figure 4.20B**, we noted that RIF displayed a stable RMSD of 1.24 Å, whereas for most of the *mt*s it became increasingly unstable, displaying mean RMSDs ranging from 0.1 Å to 2.0 Å; a measure of the extent to which RIF was displaced from its initial pose. These deductions were further informed by the CoM calculation wherein we observed RIF's motions away from its binding pocket (**Figure 4.19B**).

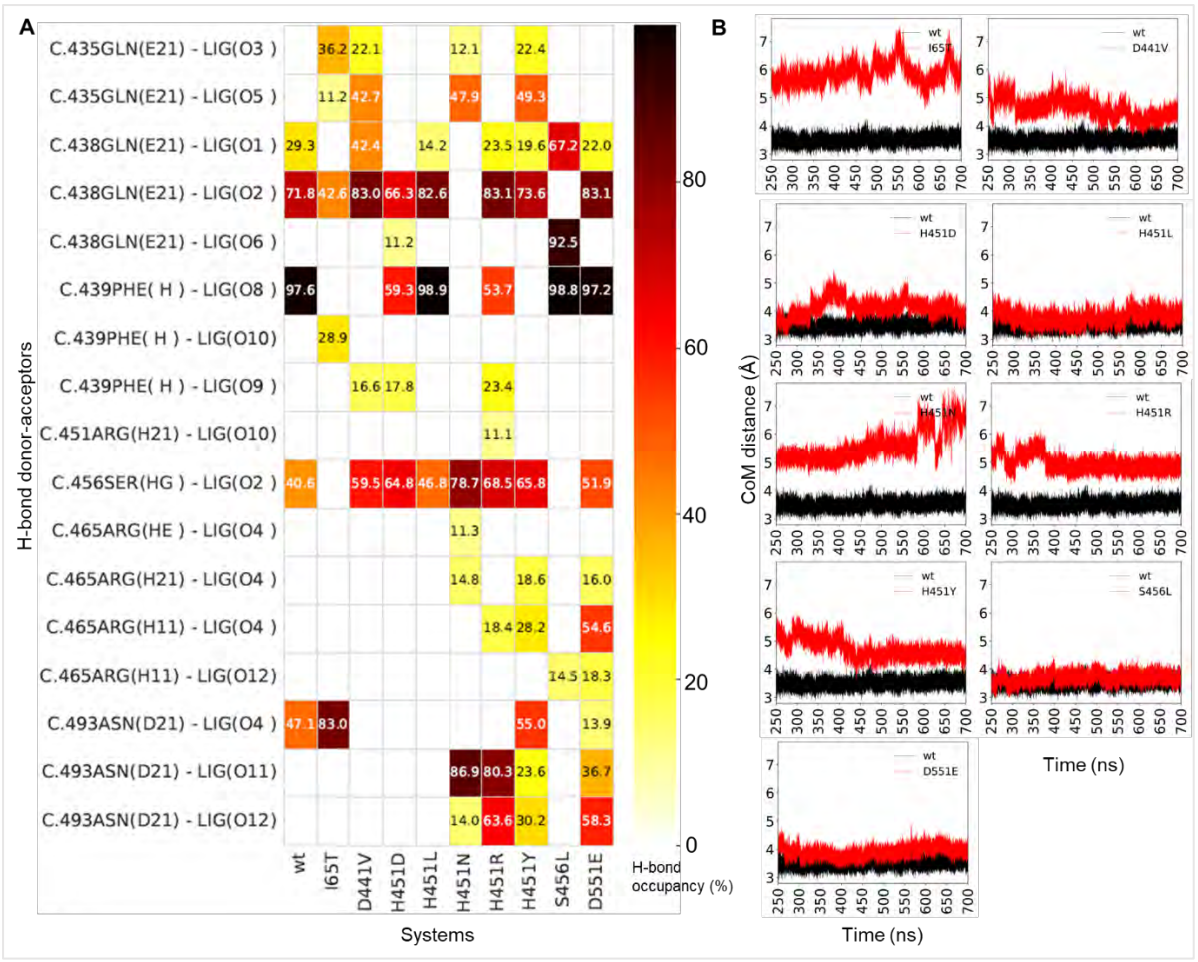


Figure 4.19. Graphical analysis of the RIF's H-bond interactions with the RIF-BP along with the displacement from the RIF-BP. (A) Shows the H-bond occupancy profile of the *wt* and *mt* bound RIF ligands for the equilibrated timeframes. (B) Shows the CoM plots which represent the changes in distance between RIF and the RIF-BP throughout the equilibrated timeframes. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

Visualization of the 3D structure of RIF through the course of the simulation captured through time-based snapshots (**Figure 4.20A**) further corroborated these observations.

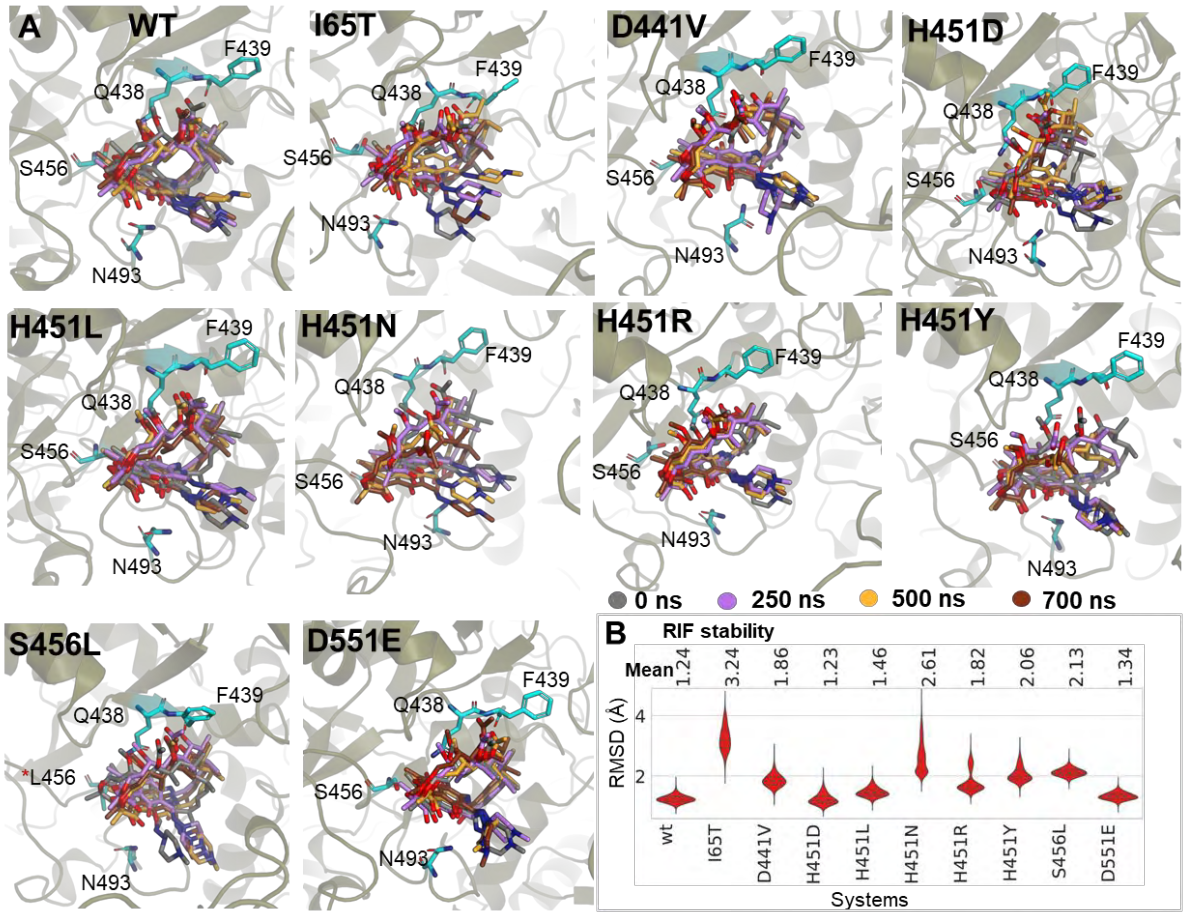


Figure 4.20. Structural and graphical representation of the mutant-induced destabilization of RIF. (A) Shows the 3D depiction of the degree to which RIF shifted from its initial active pose when in the presence of the investigated RIF resistance mutations. The various poses of RIF (stick representation) at the timeframes of 0 ns (colored grey), 250 ns (colored purple), 500 ns (colored orange), and 700 ns (colored brown) were respectively indicated. (B) Violin distribution plot representations of the RMSD for the *wt* and *mt* system RIFs (with respect to the RIF-BP) are shown for the equilibrated timeframes while the determined average RMSD values are displayed above each violin. Violin plots dotted lines represent the 25th, 50th, and 75th quartiles. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

As shown in the *wt*, RIF's orientation only slightly moved at the 500 ns timeframe, whereas the poses at the final timeframe of 700 ns were similar to those of the pre- and post-equilibrated timeframes of 0 ns and 250 ns respectively. The *mt* systems however showed varying degrees of difference in orientations at different timeframes relative to the *wt*. That is, while the active RIF pose appeared to be more conserved in the *wt* control, more noticeable differences were exhibited in the majority of the *mt*

systems. The consequence of these differences in RIF pose could be additionally captured through the resulting modifications of important interactions, i.e., the lowered frequency of H-bond occupancies (**Figure 4.19A**). The H-bonds of particular interest included those of O1 and O2 with βGLN438, O8 with βPHE439, O2 with βSER456, and finally those of O4 with βASN493 (Lin *et al.*, 2017; Srivastava *et al.*, 2018; Zhang *et al.*, 2019).

Using the low-energy minima structures extracted from the comparative ED estimations, RIF's molecular interaction profiles were analysed for the *wt* and *mt* systems. Based on **Figure 4.19A**, we observed that the *wt* system's RIF maintained a total of 6 H-Bonds with β residues GLN438, PHE439, SER456, and ASN493 which play a pivotal role in maintaining the pose and stability of RIF within its pocket. Contrastingly, a majority of the *mt*s exhibited differences in RIF bonding profiles, displaying either a complete loss of interactions with D429, for instance, or changes in interaction types, e.g. with Q438 for the H451R system (**Supplementary Figure S1.2**). Per-residue energy decomposition calculations were further conducted to reveal the mutational impact on relative energetic contributions of residues and how that translates to altered RIF stability and binding affinity. As shown in **Figure 4.21**, βPHE439's electrostatic contributions (ΔE_{ele}) decreased to −5.66 kcal/mol for the I65T protein, to −6.19 kcal/mol for the D441V protein, and to −3.34 kcal/mol and −5.48 kcal/mol for the H451D and H451Y proteins respectively, as compared to the *wt* control which displayed an ΔE_{ele} of −7.48 kcal/mol. Also, we noted decreases in the $_{mt}$βPHE439 van der Waals contributions (ΔE_{vdw}) to −1.98 kcal/mol for the I65T protein, to −2.36 kcal/mol for the H451N protein, and to −2.66 kcal/mol and −2.65 kcal/mol for the H451Y and D551E proteins respectively, relative to the ΔE_{vdw} of −3.49 kcal/mol for the $_{wt}$βPHE439.

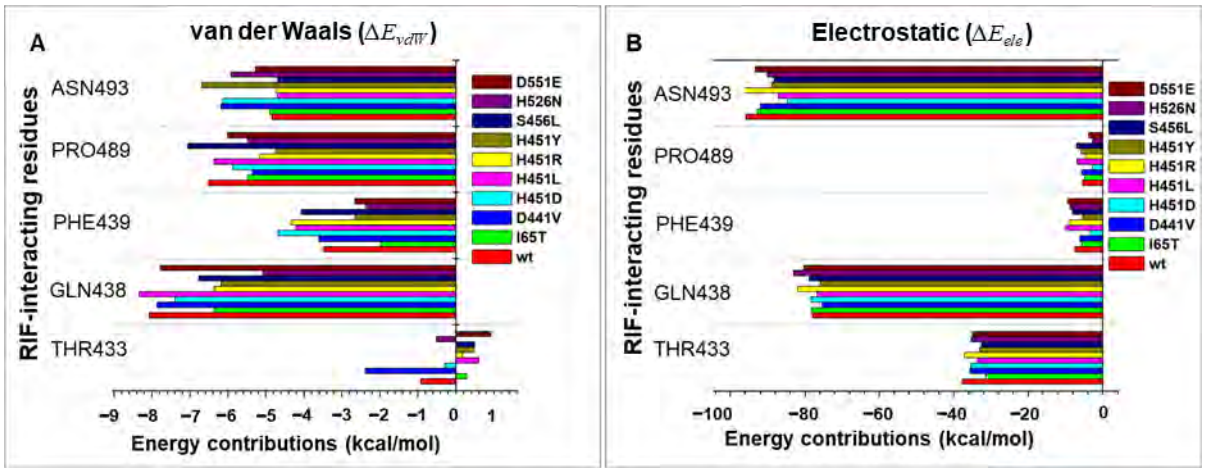


Figure 4.21. Energetic residue decomposition plots for the residues interacting with the RIF molecule for the *wt* and *mt*s. (A) Shows the energetic van der Waals energy contributions while (B) depicts the electrostatic energy contributions of the β residues ASN493, PRO489, PHE439, GLN438, and THR433. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

Furthermore, notable reductions in ΔE_{vdw} were detected for the βGLN438 residue, with energy differences of −1.71 kcal/mol for the I65T protein, −0.68 kcal/mol for the H451D protein, −2.98 kcal/mol for the H451N protein, −1.70 kcal/mol for the H451R protein, and −1.89 kcal/mol and −1.30 kcal/mol for the H451Y and S456L proteins respectively. Reductions in $_{\beta SER456}\Delta E_{vdw}$ from −5.33 kcal/mol for the *wt* were particularly noted for the I65T protein (−4.4 kcal/mol), whereas $_{\beta SER456}\Delta E_{ele}$ indicated lowered differences of −1.53 kcal/mol (I65T), −1.05 kcal/mol (D441V), −1.43 kcal/mol (H451D), −0.82 kcal/mol (H451L), −0.37 kcal/mol (H451N), −0.65 kcal/mol (H451R), −1.5 kcal/mol (H451Y), and −1.26 kcal/mol (D551E). Similarly, the $_{\beta ANS493}\Delta E_{ele}$ which displayed an energetic contribution of −95.76 kcal/mol for the *wt* observed a reduction in the H451D protein with an $_{\beta ANS493}\Delta E_{ele}$ of −84.67 kcal/mol. Additional $_{\beta ANS493}\Delta E_{ele}$ differences were also identified for the I65T protein (−3.04 kcal/mol), the D441V protein (−3.77 kcal/mol), the H451L protein (−8.62 kcal/mol), the H451N protein (−5.66 kcal/mol), the H451Y protein (−6.87 kcal/mol), the S456L protein (−7.49 kcal/mol), and the D551E

protein (−2.53 kcal/mol). We further noted complementary decreases in estimated ΔE_{ele} and ΔE_{vdw} contributions for the βTHR433 and βPRO489 which form part of the RIF-BP. Overall, our observations may be telling of the mutation-induced perturbations that result in reduced energetic contributions from residues that play a crucial role in RIF engaging with its pocket, which translates to decreases in RIF binding affinity, consequently leading to a reduced propensity to inhibit *Mtb*-RNAP as experimentally reported. Our deductions were further supported through binding energy estimations which showed a ΔG_{bind} of −43.58 kcal/mol for the $_{wt}$*Mtb*-RNAP RIF, whereas most of the *mt*s (excluding the H451R, S456L, and D551E proteins) indicated reduced binding affinities as seen in **Table 4.1**. The relatively higher ΔG_{bind} estimations were likely a result of the dynamic reorientation of the RIF molecules away from the active pose and towards residues with more favourable interactions such as the βARG465 residue wherein we observed a strong H-bond which happened to be complementary to the bi-phenyl ring placement (**Supplementary Figure S1.2G**, **I**, **J**). As mentioned earlier, relative to the *wt*, the side-chain substitution involving arginine in H451R may also have contributed to the notably high ΔG_{bind}. Nonetheless, we could safely presume that the reorientation of RIF from the active pose across the three *mt*s, but not limited to these three *mt* proteins (refer to **Figure 4.20**), is adequately capable of limiting RIF's propensity to obstruct *Mtb*-RNAPs functional capacity of extending RNA and enacts a level of resistance despite the detected increase in binding affinity.

Table 4.1. Respective energetic contributions and the binding energy for the RIF drug towards the *wt* and *mt Mtb*-RNAP structures.

Systems	ΔE_{vdW} (kcal/mol)	ΔE_{ele} (kcal/mol)	ΔE_{gb} (kcal/mol)	ΔE_{surf} (kcal/mol)	ΔG_{gas} (kcal/mol)	ΔG_{solv} (kcal/mol)	ΔG_{bind} (kcal/mol)
wt	−62.9 ± 3.3	−32.5 ± 3.4	59.0 ± 1.0	−7.2 ± 0.07	−95.4 ± 4.8	51.8 ± 1.1	−43.6 ± 4.9
I65T	−40.6 ± 3.5	−26.9 ± 5.7	52.3 ± 4.0	−5.5 ± 0.04	−67.6 ± 6.7	46.8 ± 4.0	−20.8 ± 7.8
D441V	−58.0 ± 3.3	−15.3 ± 2.6	46.1 ± 1.3	−7.0 ± 0.05	−73.2 ± 4.2	39.1 ± 1.3	−34.1 ± 4.4
H451D	−63.0 ± 3.3	−24.5 ± 4.8	55.6 ± 1.9	−7.6 ± 0.05	−87.6 ± 5.8	48.0 ± 1.9	−39.5 ± 6.1
H451L	−57.2 ± 3.5	−44.3 ± 2.3	66.3 ± 0.8	−6.9 ±0.06	−101.5 ± 4.2	59.4 ± 0.8	−42.1 ± 4.2
H451N	−45.5 ± 3.0	−30.3 ± 4.6	49.2 ± 2.8	−5.2 ± 0.06	−75.8 ± 5.6	44.0 ± 2.8	−31.8 ± 6.2
H451R	−60.0 ± 3.3	−65.3 ± 3.0	82.1 ± 1.4	−7.3 ± 0.04	−125.3 ± 4.5	74.8 ± 1.4	−50.5 ± 4.7
H451Y	−55.4 ± 3.0	−30.5 ± 1.9	56.3 ± 0.9	−6.9 ± 0.05	−85.9 ± 3.6	49.5 ± 0.9	−36.4 ± 3.7
S456L	−64.2 ± 3.5	−43.2 ± 3.7	71.4 ± 1.7	−7.7 ± 0.01	−107.3 ± 5.1	63.7 ± 1.7	−43.6 ± 5.4
D551E	−67.4 ± 3.3	−55.3 ± 3.6	77.3 ± 1.9	−8.8 ± 0.1	−122.7 ± 4.9	68.6 ± 1.9	−54.1 ± 5.2

ΔE_{vdW}—van der Waals energy, ΔE_{ele}—electrostatic energy, ΔE_{gb}—polar contribution to solvation free energy, ΔE_{surf}—non-polar contribution to solvation free energy, ΔG_{gas}—gas phase energy, ΔG_{solv}—solvation free energy and ΔG_{bind}—total binding energy.

4.5. Mutation-associated destabilization of catalytic nucleic acids cross-link to active RIF's disorientation

Dynamic alterations of DNA and RNA interactions with missense mutated *Mtb*-RNAPs may adversely affect how RIF binds to the β subunit and interacts with the 5′ end of RNA, along with how it prevents the transcription process by disrupting the formation of 2-3 nucleotide long RNA transcripts through steric occlusion (Lin *et al.*, 2017). Details surrounding these potential effects were thus investigated to reveal additional mechanistic effects in response to the presence of the investigated mutations and RIF. For the purposes of this investigation, RNA and DNA moiety (first 4nt from the 3' end of template DNA) was primarily conducted using H-bond analysis and the RMSD metric.

Based on the RNA results, as portrayed in **Figure 4.22A**, we found that the *wt*'s RNA displayed increased instability as indicated by the higher average RMSD (5.1 Å) compared to its initial structure. In the presence of RIF, however, the RNA of the *wt* became more stable (RMSD of 3.14 Å) (**Figures 4.22B**).

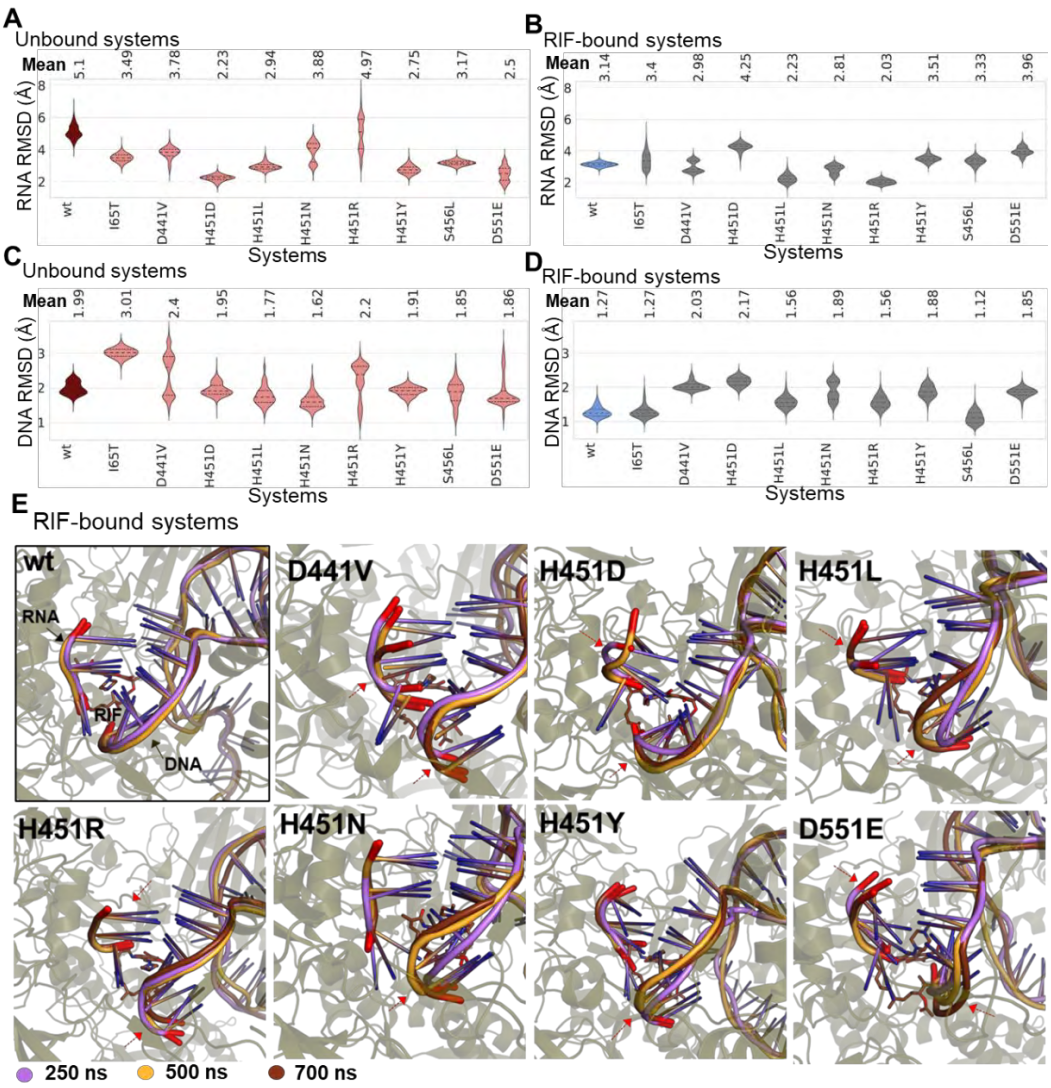


Figure 4.22. Illustration of the mutational and RIF binding perturbation effects on the catalytic nucleic acids. (A) RMSD distribution violins for the RNA molecules of the (A) RIF-unbound and (B) bound *mt*s (colored pink) and *wt* (colored maroon). The RMSD distribution violins of the 4nt long DNA moieties are shown for the (C) RIF-unbound and (D) bound *mt* (colored pink) and *wt* (colored maroon). Mean values are displayed above the respective plots. The dotted lines on the violins represent the 25th, 50^{th,} and 75th percentiles. (E) Shows the superimposed snapshots of the RNA and DNA moiety 3D structures at the 250ns (colored purple), 500ns (colored orange), and 750 ns (colored brown) timeframes. The most distinct changes in structural positions are indicated using red arrows. Adapted from (Monama, Olotu and Tastan Bishop, 2023).

These findings were also complemented by **Figures 4.23 and 4.24**, wherein we observed a greater number of low occupancy H-bonds in the RIF-unbound *wt* as compared to its RIF-bound counterpart. This further suggested an increase in RNA

stability or motional restriction of the RNA molecule for the RIF-bound *wt* (**Figure 4.22E**), a factor which might play a pivotal role in establishing RIF's inhibitory activities.

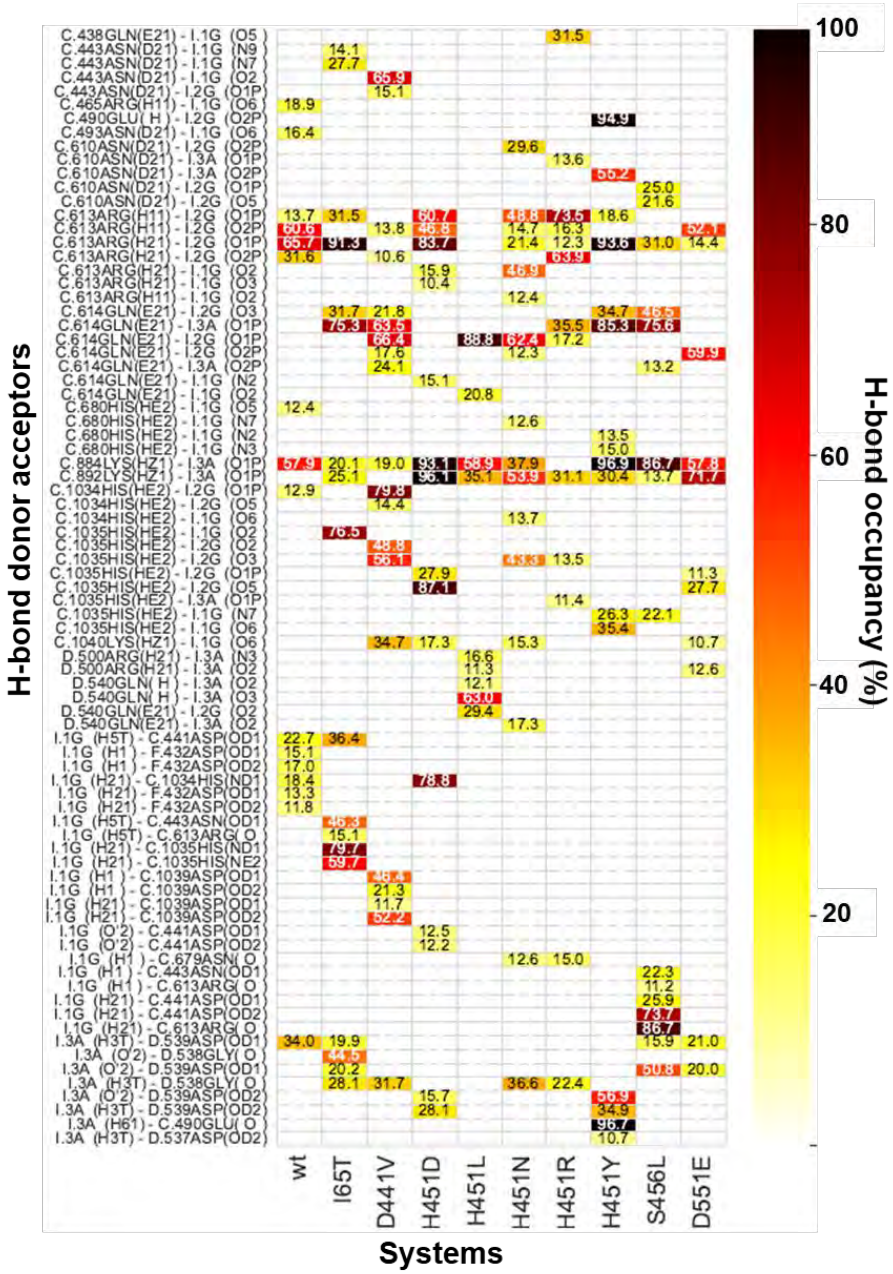


Figure 4.23. Heatmap representation of the RNA molecule's H-bond interactions with *Mtb*-RNAP. The heatmap shows the H-bond occupancy profile of the *wt* and *mt* RIF-unbound systems.

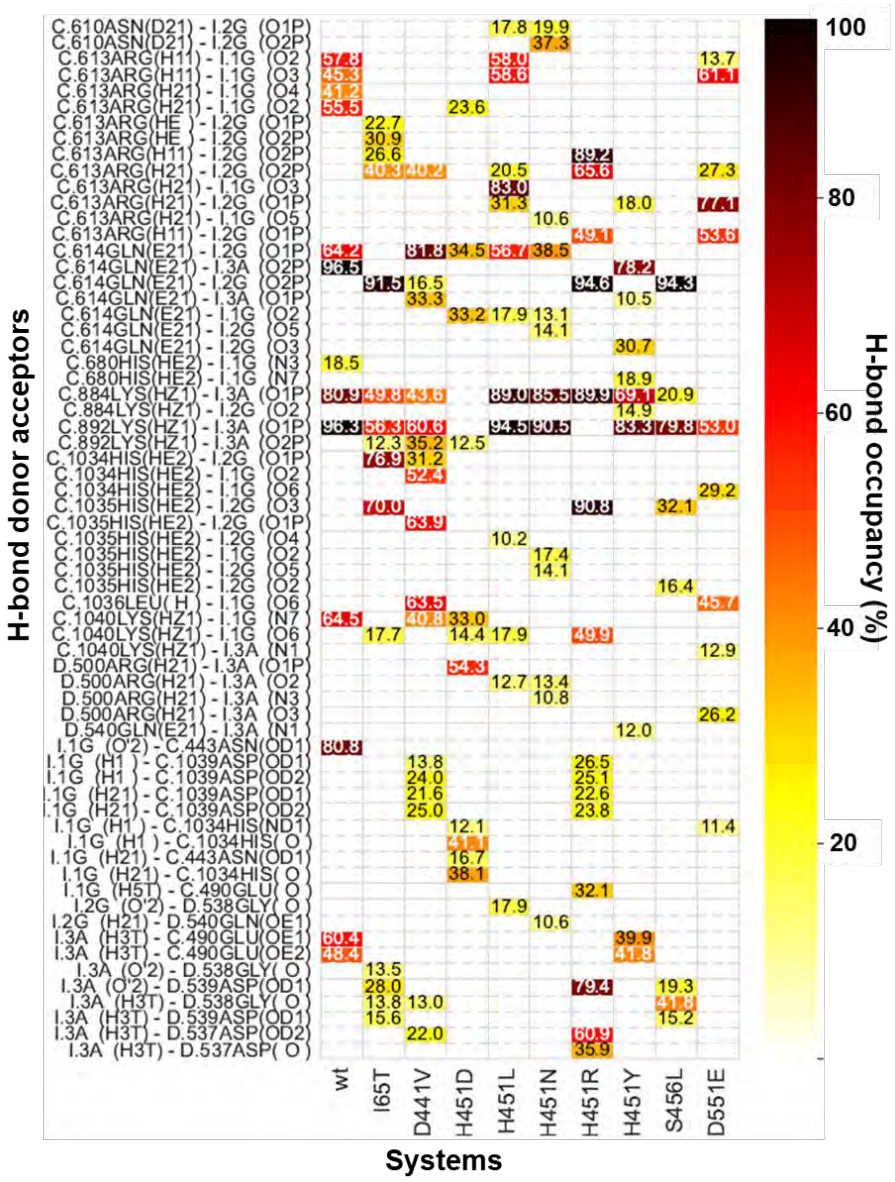


Figure 4.24. Heatmap representation of the RNA molecule's H-bond interactions with *Mtb*-RNAP. The heatmap shows the H-bond occupancy profile of the *wt* and *mt* RIF-bound systems.

The unbound and RIF-bound *mt* systems, however, displayed varying changes in H-bond profiles relative to the *wt*, thus implying a mutation-induced perturbation effect resulting in a shift from the original RNA position. A similar trend was observed when looking at the DNA moiety, that is, the RMSDs and reduced low occupancy H-bonds

(**Figures 4.25 and 4.26**) in the *wt* suggested a motional restriction that was induced as a consequence of RIF binding.

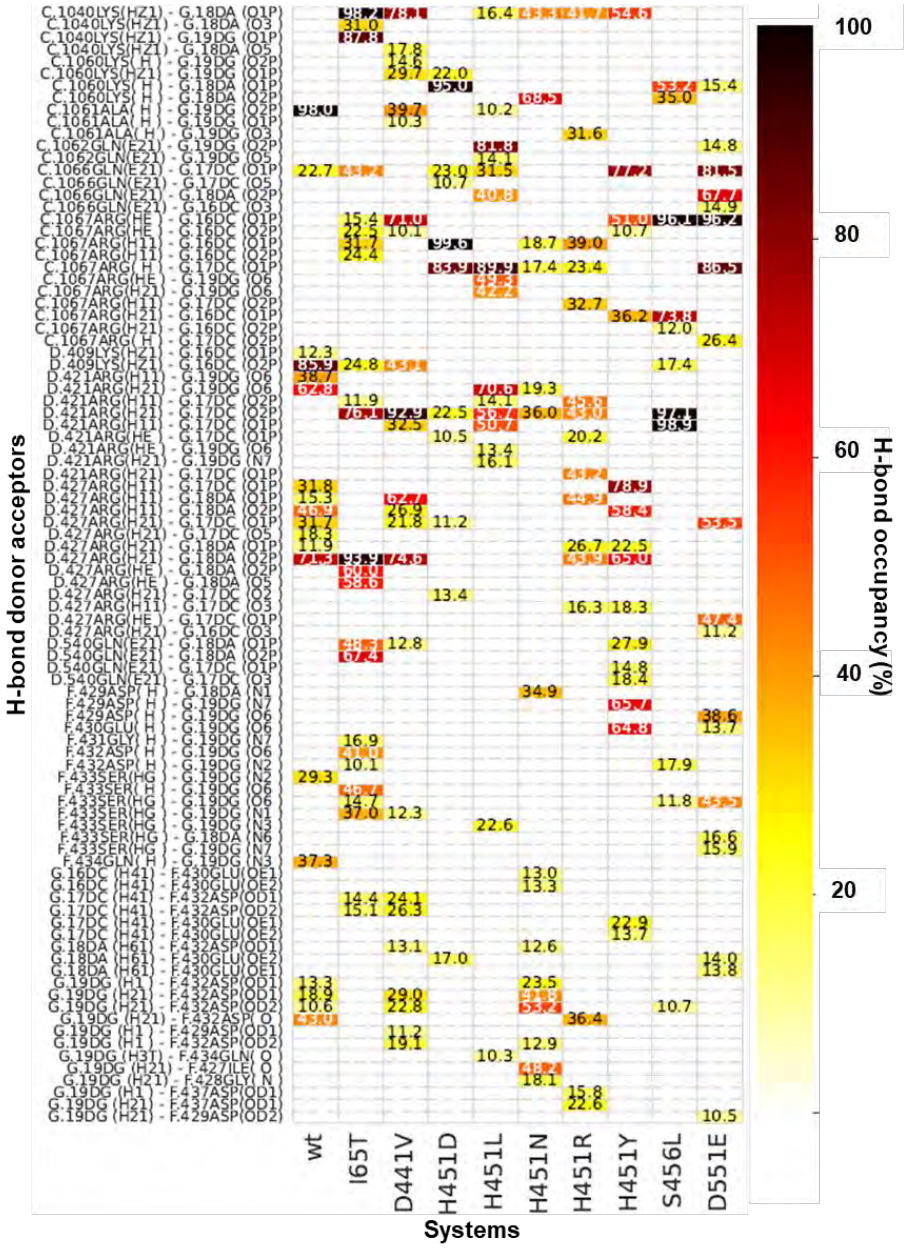


Figure 4.25. Heatmap representation of the DNA moiety's H-bond interactions with *Mtb*-RNAP. The heatmap shows the H-bond occupancy profile of the *wt* and *mt* RIF-unbound systems.

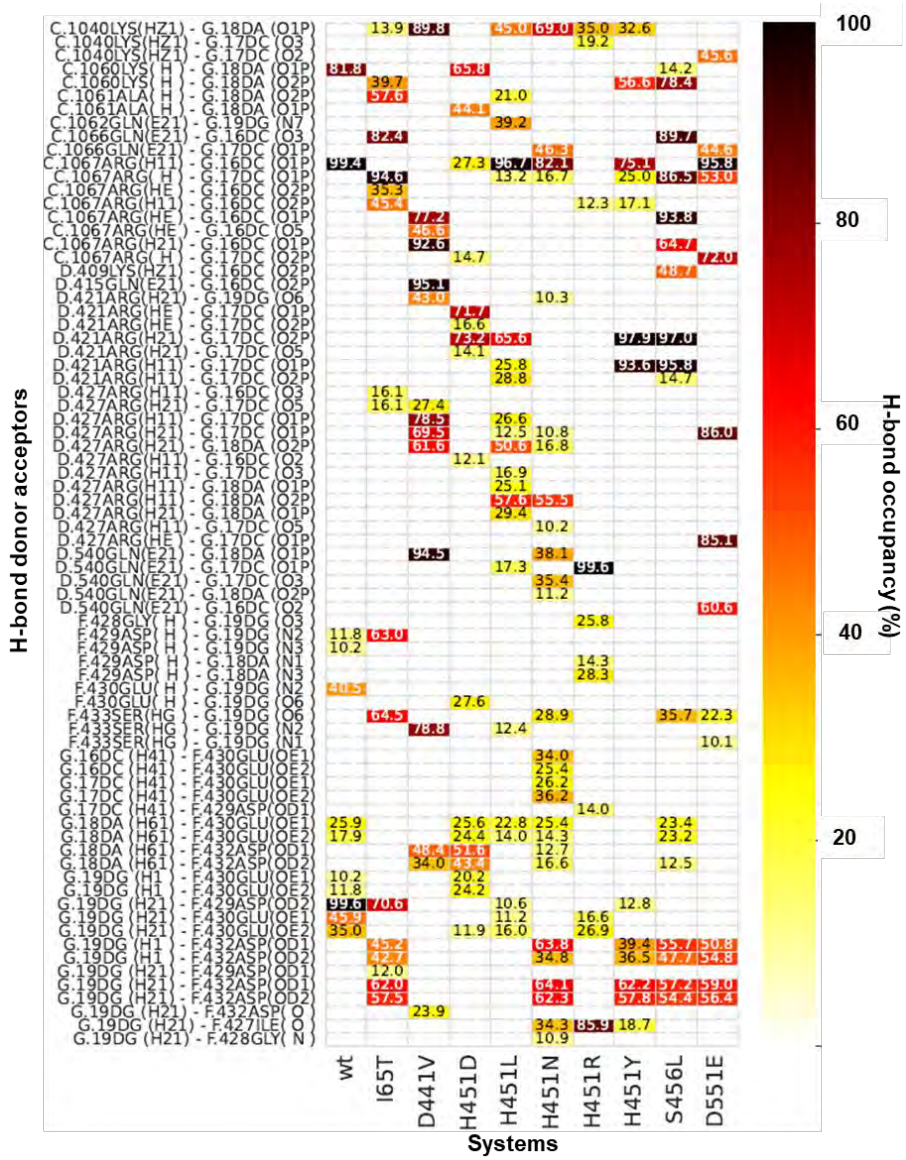


Figure 4.26. Heatmap representation of the DNA moiety's H-bond interactions with *Mtb*-RNAP. The heatmap shows the H-bond occupancy profile of the *wt* and *mt* RIF-bound systems.

Complementarily, as indicated in **Figure 4.22E**, we further observed consistencies in the superimposed snapshots of RNA and DNA for the specified timeframes for the *wt*. Regarding the unbound *mt* systems (excluding H451R) however, we noted lowered RNA RMSDs relative to the *wt*. This was further informed by the fewer low occupancy

H-bonds as compared to the *wt*, indicating more stable interactions in the *mt*s (**Figure 4.23A**). Additionally, most of the DNA moieties of the RIF-unbound *mt*s displayed large discrepancies in RMSDs and H-bond interactions, with H451N, H451L, S456L D551E, H451R, and D441V proteins displaying particularly large deviations. It is also worth noting some of the consistently occurring high occupancy H-bond interactions between G.19DG (19 deoxyguanosine) and F.ASP432 (σASP432) among the I65T, H451N, H451Y, S456L and D551E systems (**Figure 4.24B**), may be essential in the resistance mechanism of RIF in these cases. It is, however, worth mentioning that although no singular explanation such as the uniqueness of residue size and/or hydropathy could be associated with the aforementioned systems' behaviour given the complexity of the dynamics involved, further local effects detailing differences in mutation site contact frequencies and overall residue-residue communication pathways are illuminated in **Chapter 5**. Regardless, the additional observations made for the RNAs may collectively imply a mutation-and-RIF-based contribution towards *Mtb*-RNAP processive inefficiencies. Also, we notably observed that the majority of the *mt*s exhibited unstable RNAs as indicated by the RMSD distributions illustrated in **Figures 4.22A and 4.22B**. Furthermore, although a majority of the *mt* systems indicated lowered RMSDs for the DNA moiety as depicted in **Figure 4.22C**, the RIF-bound *mt* systems generally displayed greater instabilities (**Figure 4.22D**).

4.6. Conclusion

Throughout the investigation, we made a number of key observations. **I)** The *mt Mtb*-RNAPs generally displayed increased conformational sampling coupled with increased flexibility in several loop regions throughout the protein, indicating that the mutations may have caused perturbation effects resulting in accumulative global motions. The was however no large changes to the β subunit's exposure to the

surrounding solvent molecules suggesting minimal effects towards RIF binding in the *mt* proteins. Hence, the observed changes due to the accumulative perturbation effects were further linked to the altered flexibilities of the structural domains. This may be particularly important for the functioning of the *wt* protein, e.g., in protein stability, its capacity to terminate the transcription process, along with overall transcriptional processivity which tends to be diminished among the *mt* proteins, as reported in the literature (Stefan, Ugur and Garcia, 2018). The heightened motions that were observed may therefore form part of a compensatory mechanism needed to establish RIF resistance. This was further informed by the less correlated motions exhibited by the *mt* RIF-BPs as compared to the correlated motions of the *wt* in both unbound and RIF-bound forms. **II)** The native or active RIF pose clearly shifted by varying degrees in the *mt* proteins, even those *mt*s that displayed increased binding affinity, thereby suggesting that RIF efficacy especially relies on the RIF active pose to incite its inhibitory activities. **III)** We additionally found that, besides the RIF proximally located missense mutations such as D441V, which appears to affect RIF binding or pose through direct destabilization of the RIF-BP, more distal mutations such as the I65T can still perturb the effective binding of RIF by introducing instability to the β1D for instance, which happens to form part of the RIF-BP. This highlights the need to include distal RIF resistance mutations within the drug development pipeline given that they may prove more disruptive than some RIF-BP mutations. **IV)** Finally, it was of particular interest to note that part of the RIF resistance mechanism may involve the destabilization of catalytic nucleic acid molecules (DNA and/or RNA) through disruption of stabilizing H-bond interactions in order to overcome the inhibitory effects of RIF. Overall, the findings not only confirmed our hypotheses but also revealed a different perspective on the structural and functional effects of the investigated

mutations on the *Mtb*-RNAP protein and may prove beneficial for future drug discovery ventures.

Chapter 5
Local analysis of RIF-unbound and bound *wt* and *mt Mtb*-RNAPs

Chapter overview

This chapter involves the elucidation of the post-MD residue level interactions and communication paths through the application of local analytical approaches on the *wt* and *mt Mtb*-RNAP proteins.

5.1. Introduction

Although drugs that target protein functional sites (i.e., orthosteric drugs) have been the main source of bioactive modulators used to treat human diseases in recent history, allosteric drugs which affect functional sites by binding to more distal sites have been gaining more recognition given their higher selectivity, among other advantages (Tastan Bishop, Musyoka and Barozi, 2022). Interestingly, among the drug interventions for *Mtb*-RNAP missense mutations that have been investigated, the novel drug D-AAP1 (derived from N-α-aroyl-N-aryl-phenylalaninamides (AAPs)) which binds to the N-terminus of the β'bridge helix (β'BH) domain and the secondary channel rim helices has been shown to induce allosteric inhibition (Lin *et al.*, 2017; Agoni, Ramharack and Soliman, 2018a). Furthermore, when D-AAP1 is used in combination with RIF, it allows RIF to be effective even in the presence of some RIF-resistant mutations, thus further highlighting the great potential of allosteric drugs (Lin *et al.*, 2017; Agoni, Ramharack and Soliman, 2018a).

We therefore continued our previous work wherein, and we employed a structure-based computational pipeline involving post-MD analyses to delineate the global effects of the clinically relevant RRDR mutations D441V, H451D, H451L, H451N, H451R, H451Y, and S456L and non-RRDR mutations I65T and D551E on *Mtb*-RNAP

and *Mtb*-RNAP-RIF dynamics (Monama, Olotu and Tastan Bishop, 2023). In this study, we forewent the limitations of traditional post-MD analyses and described the effects of missense mutations on allostery and protein communication by employing dynamic residue network analysis (DRN) (Brown *et al.*, 2017; Sheik Amamuddy *et al.*, 2021), a novel local analysis approach based on graph theory which has proven its effectiveness in describing protein dynamics in previous studies (Penkler, Atilgan and Tastan Bishop, 2018; Sanyanga, Nizami and Tastan Bishop, 2019; Okeke *et al.*, 2021; Sheik Amamuddy *et al.*, 2021).

5.2. Dynamic residue network (DRN) analysis

In this study, for the first time, we attempted to gain insights into the molecular basis of RIF resistance of *mt Mtb*-RNAPs (with reference to the *wt*) in the presence and absence of RIF through the application of the averaged DRN metrics, *BC* (*betweenness centrality*), *CC* (*closeness centrality*), *EC* (*eigenvector centrality*), *DC* (*degree centrality*), and *KC* (*Katz centrality*). Centrality in this context refers to how important/central residues are within a specified residue network. That said, highly central residues, henceforth referred to as hubs, were identified for each metric using a global top 5 % cutoff for the *wt* and *mt Mtb*-RNAP DRN results (refer to **Chapter 4** for details). Also, the β and β' subunits (**Figure 5.1**) were the focus of the current study given the relative agreement in conformational distributions observed in the previous chapter for the two controls in both the RIF-unbound and bound forms (Monama, Olotu and Tastan Bishop, 2023). Hence, for the purposes of this investigation, a 5% global cutoff was applied for the β and β' subunit DRN results using the MDM-TASK-web script (Sheik Amamuddy *et al.*, 2021). To further explain the observed metric-specific effects, we introduced the term 'persistent hubs' which is defined as residue hubs that are present amongst all the compared *wt* and *mt* systems (Okeke *et al.*, 2021; Sheik

Amamuddy *et al.*, 2021). We further used the term ‘absolutely conserved’ to define residues that have been reported to be conserved among the several classes of RNAPs as indicated by Lane and Darst (2010) through their multiple sequence alignment experiments (Lane and Darst, 2010b). It’s also important to note that, before we proceeded to analyze the mutation/ligand-induced hubs, we first compared the two *wt* controls for each metric to determine the extent of their differences at the residue level. We also used the term “unique hubs” to indicate communication hubs that were unique to either the *wt* or the *mt* systems.

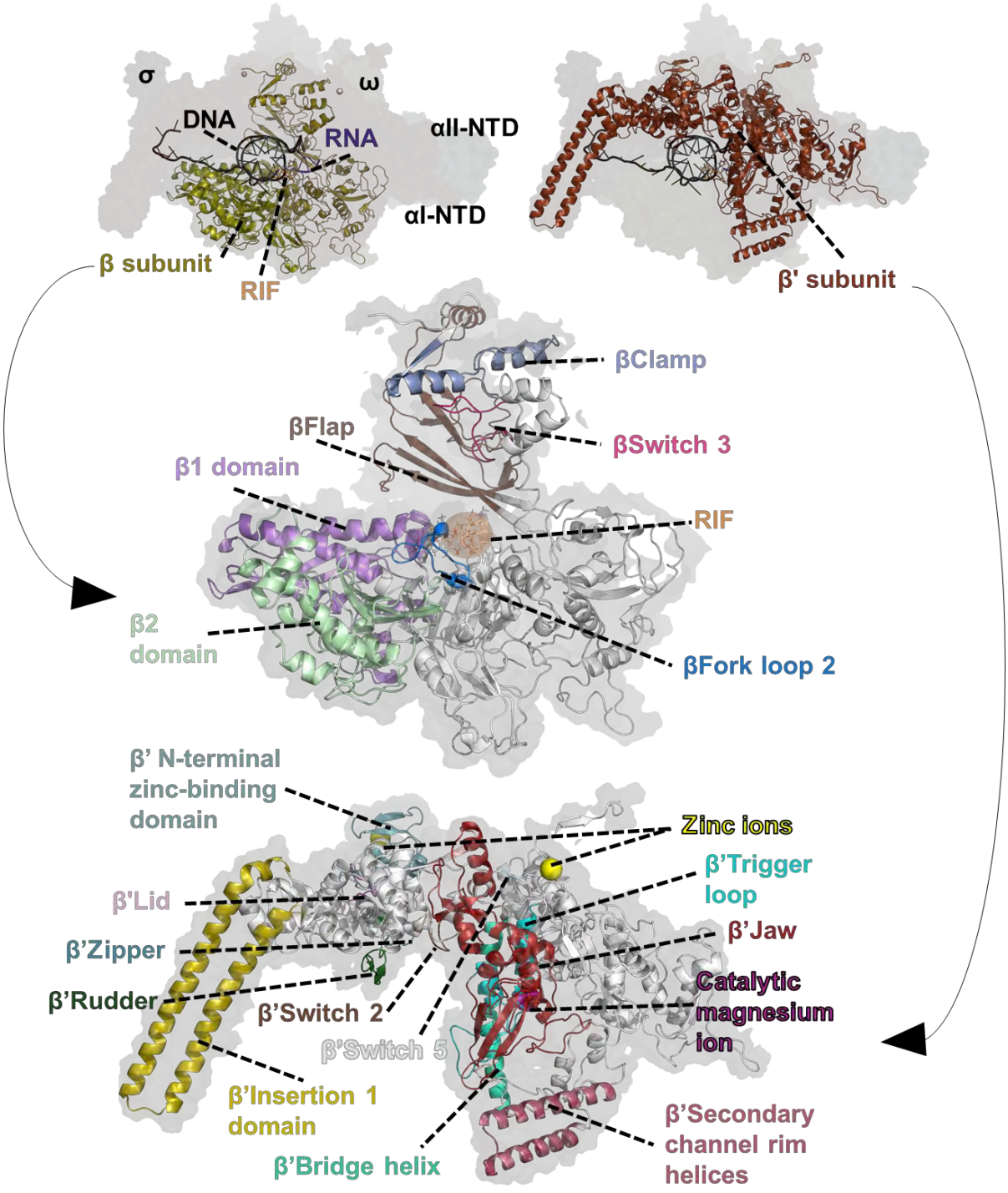


Figure 5.1. 3D representation of RIF-bound multisubunit *Mtb*-RNAP and the annotated functional domains of the β (olive) and β' (brown) subunits.

We initially presented the raw DRN data of all the systems along with the *wt1–mt* and *wt2–mt* delta results as heatmaps and found an acceptable amount of agreement, as seen in **Supplementary Figures S2.1-S2.5**. We then calculated and determined the global and local top 5% DRN residues of the β-β' complex for the *wt* controls using the MDM-TASK-web script (Sheik Amamuddy *et al.*, 2021) and mapped them to the *Mtb*-RNAP structure (**Supplementary Figures S2.6-S2.10**). We then proceeded to analyze our results using *wt*1 as the reference once a visual confirmation of high similarity between the global and local top 5% DRN mapped structures of the *wt* controls was confirmed through PyMol (Delano, 2002). Also, although the primary control in the current study was the same control we used in our previous investigation (designated henceforth as either *wt* or *wt1*), to more accurately compare the hub statuses between the *wt* and *mt* systems, we calculated the normalized delta *wt* (*wt1–wt2*) values per subunit for each metric and identified residues indicating significant differences (mean ± 2 SD) as shown in **Supplementary Figures S2.11-S2.15**, to better account for any discrepancies in the DRN results (**Supplementary Figures S2.16-S2.20**). In other words, the identified discrepancies between *wt1* and *wt2* entail a margin of lower confidence for those hubs used for *wt-mt* comparisons. However, it is worth noting that a majority of the significantly deviating residues were identified as loop residues, which was not surprising given their innate flexibility and role in conformational variability (Mitusińska, Skalski and Góra, 2020). Also, we identified the *mt* systems that displayed the greatest change overall per metric by inspecting *wt1–mt* and *wt2–mt* delta values for the β and β' DRN results.

5.2.1. Averaged betweenness centrality (BC) highlighted key residues that may play an instrumental role in both maintaining and deviating from Mtb-RNAP's normal functions due to mutations and RIF

Proteins are complex molecular structures that can be described with graph theory as a network of contacting nodes (residues) connected through edges (non-covalent interactions). Within this context, the *BC* network metric is determined by calculating the frequency of shortest-path transversals through a particular residue within a protein of interest (Karabekmez and Kirdar, 2016; Brown *et al.*, 2017). *BC* can therefore be used to determine residue hubs that play an important role in protein communication based on the assertion that residue-residue communication favors shorter paths (Penkler, Atilgan and Tastan Bishop, 2018). To that end, *BC* has been used in prior studies to identify functionally important residues with particular interest in changes in residue importance caused by the presence of mutations and/or ligand binding in several protein systems (Penkler, Atilgan and Tastan Bishop, 2018; Sanyanga, Nizami and Tastan Bishop, 2019; Okeke *et al.*, 2021; Sheik Amamuddy *et al.*, 2021).

Essentially, we visually observed that the *BC* hubs mapped onto the *wt Mtb*-RNAP structures were distributed over the length of the protein and associated with various functional domains (**Figure 5.2A-B**). We then used the PyMol interface residues script (https://pymolwiki.org/index. php/InterfaceResidues) applying a threshold distance of 4 Å to confirm that the hubs were mainly located along the interfaces (αI-β, αI-β', αII-β, αII-β', β-β', β-ω, β'-ω, ω-σ, β-σ, and β'-σ) of the *Mtb*-RNAP subunits (**Supplementary Figure S2.21**).

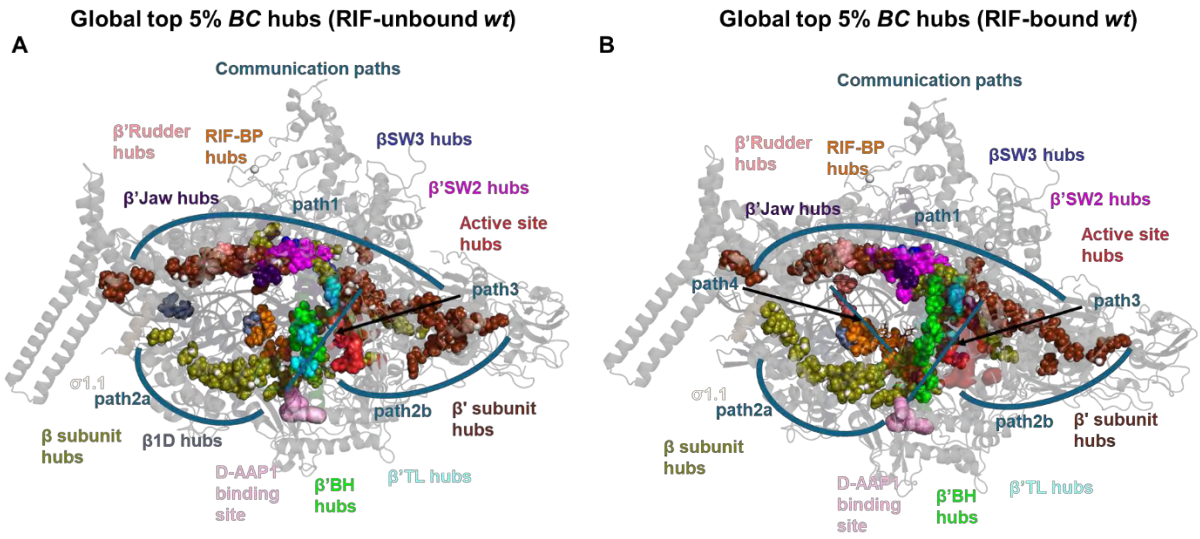


Figure 5.2. Distribution and paths of global 5% *BC* hubs and their structural associations for the (A) RIF-unbound and (B) bound *wt Mtb*-RNAPs.

When comparing the *mt*s to the corresponding *wt* system, it was also clear that the *mt* systems displayed fewer communication hubs that spanned from one end of the protein to the other (especially around path2a) (**Figure 5.3A-B**), the most notable of which being the H451D and H451N unbound proteins, and the D441V and S456L RIF-bound proteins. More notable, however, was the possible functional implication(s) due to the absence of path4 in all of the RIF-bound *mt*s. Path4 was formed by hubs that connect the β'rudder domain (residues 382–401) from path1 to the RIF-BP (**Figure 5.2B**). Additionally, the *mt* systems' unique hubs were dispersed throughout the protein structures with a greater concentration of these hubs appearing in the RIF-bound proteins. Among the *mt*s, the D441V and H451N unbound proteins, and the I65T, D441V, H451D, and S456L RIF-bound proteins were found to be of particular note. Our findings therefore suggested a mutation-induced effect at the residue level. This could additionally inform the drastic conformational changes we observed in our previous findings, particularly given the importance of interface residues in maintaining

stability in quaternary structures (Janin, Bahadur and Chakrabarti, 2008; Monama, Olotu and Tastan Bishop, 2023).

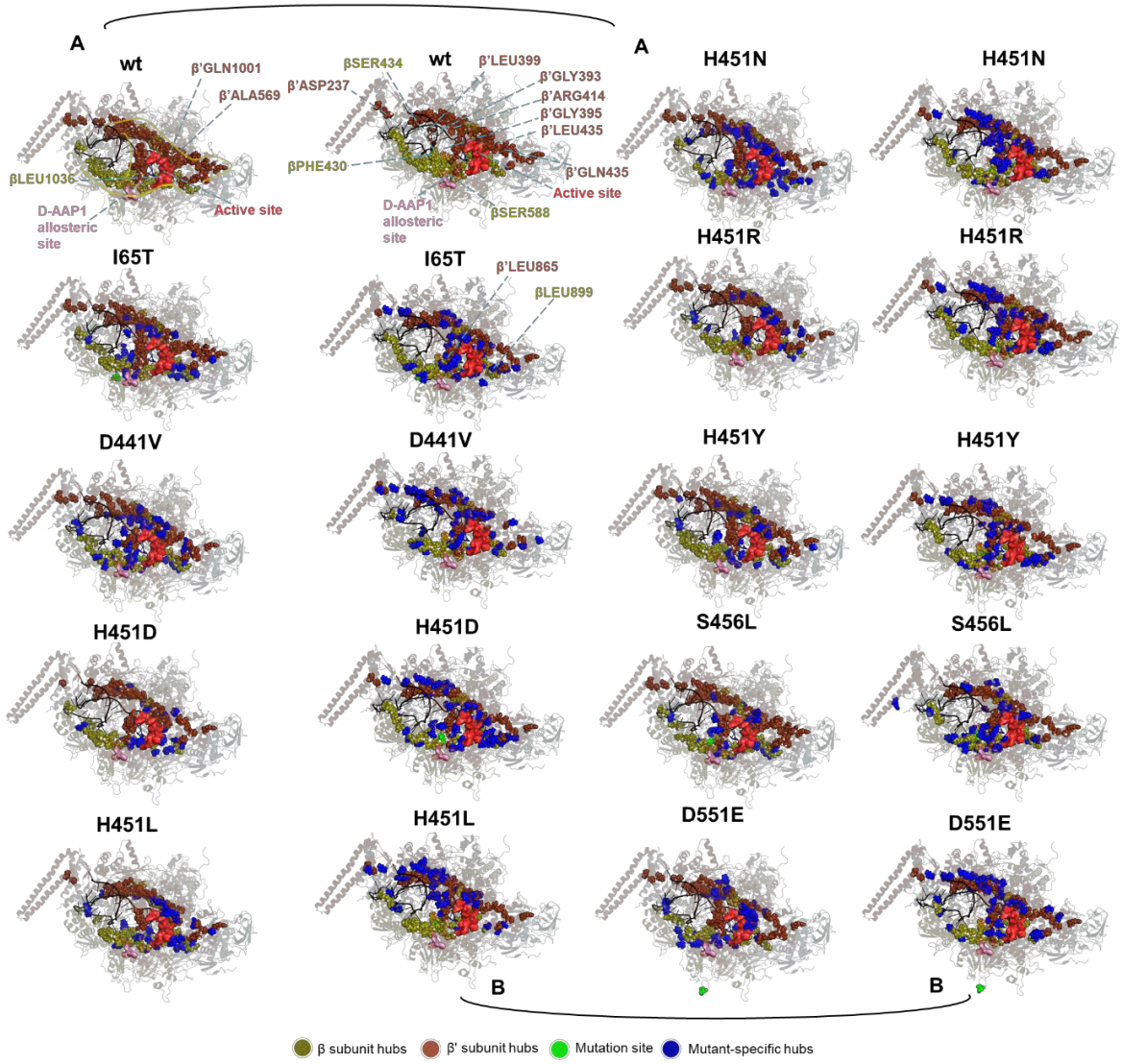


Figure 5.3. 3D *Mtb*-RNAP structures of the (A) RIF-unbound and (B) bound *wt* and representative *mt*s with mapped global top 5% *BC* hubs for the β and β' subunits. Hubs are shown in sphere representation and are colored olive and brown for the respective β and β' subunits. The blue- and purple-colored spheres represent the mutated site and *mt*-specific hubs respectively. Also, the active site (sphere representation), D-AAP1 allosteric site (surface representation), DNA (cartoon representation), RNA (cartoon representation), and RIF (stick representation) are each colored red, pink, black, dark blue, and orange. The residue annotations on the *wt* and *mt* systems represent hubs that are unique to the *wt* and all the *mt*s respectively for the bound and unbound states.

Interestingly, the RIF-BP *BC* hubs PHE430, SER434, and LEU436 (**Figure 5.4A**) (discussed further below) were present in the RIF-bound *wt* system but predominantly missing in the corresponding *mt* systems, suggesting that they may be instrumental in altering RIF's active orientation. We further observed a notable communication path leading from the known D-AAP1 allosteric site through to the active site (between path2a and path3), a trend that persisted for most of the employed DRN metrics**.**

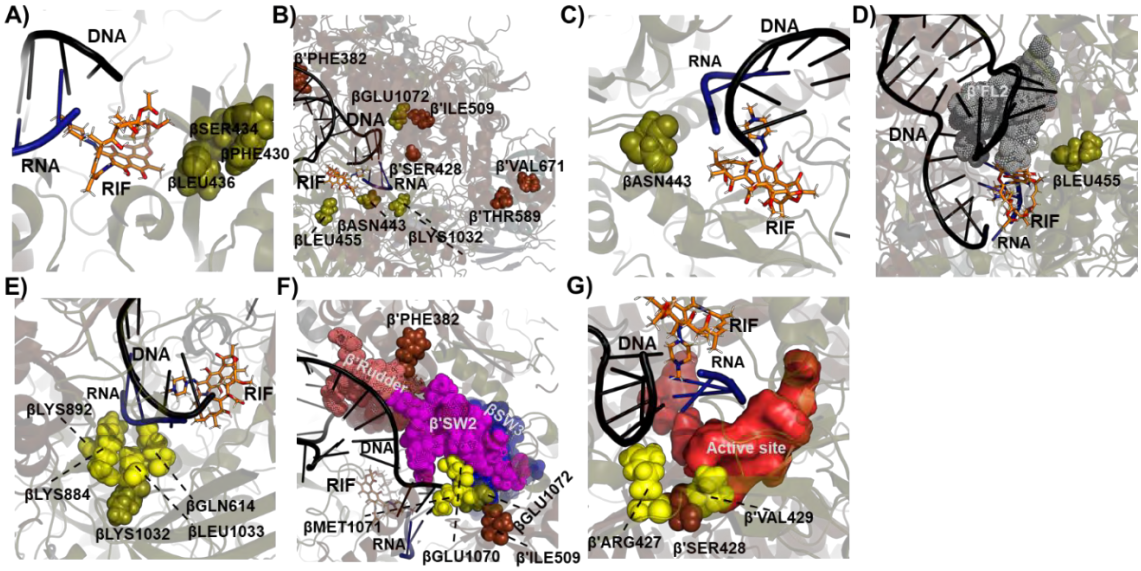


Figure 5.4. Illustration of the identified persistent hubs (sphere representation) colored in olive and brown for the *Mtb*-RNAP β and β' subunits respectively. Associated functionally relevant residue contacts are represented as yellow spheres while functional sites are indicated in dots and surface representations. The nucleic acids (DNA and RNA colored black and blue respectively) and RIF (orange) are represented as cartoons and sticks respectively. (A) Shows the annotated RIF-BP hubs βSER436, βPHE430, and βLEU456 relative to RIF, RNA, and DNA. (B) Indicates all the identified β-β' complex *wt* and *mt*s persistent hubs across both RIF-bound and unbound states. (C) Displays the annotated RIF-BP hub βASN443 relative to RIF, RNA, and DNA. (D) Shows the annotated βLEU455 hub and its relative position to the βFL2 functional domain (grey). (E) βLYS1032 and its contacting residues, βGLN614, βLYS884, βLYS892, and βLEU1033, are shown to contact RNA. (F) Indicates the β'PHE382 hub and associated β'rudder domain (salmon) along with βGLU1072 and β'ILE509 hubs and their contacting residues βGLU1070 and βMET1071 which further contact the β'SW2 (magenta) and βSW3 (blue) domains. (G) Shows β'SER428 with the β'ARG427 and β'VAL429 contacting residues, and their respective associations with DNA and the active site (red).

Based on **Supplementary Table S2.1**, the β residues ASN443, LEU455, LYS1032, and GLU1072, and β' residues PHE382, SER428, ILE509, LEU588, and VAL671 were identified as persistent *BC* hubs (**Figure 5.4B**) present in both the RIF-unbound and RIF-bound forms, further implying their functional importance, albeit some discrepancies were noted for βLEU455, βGLU1072, β'PHE382, β'SER428, and β'ILE509 (**Supplementary Figure S2.11**). βASN443, as captured in **Figure 5.4C**, forms part of the RIF-BP and was found to greatly fluctuate (relative to the *wt* in the presence and absence of the ligand) for the I65T system and may have contributed to the detected destabilization of RIF (Monama, Olotu and Tastan Bishop, 2023). As seen in **Figure 5.4D**, βLEU455 is located near the start of the βFL2 domain (residues 459–473) which is responsible for maintaining the transcription bubble and preventing the reassociation of ssDNA to dsDNA. βLYS1032 is adjacently positioned to absolutely conserved residue βLEU1033 which forms critical interactions with absolutely conserved residues βQ614, βK884, and βK892 (active site residue) (**Figure 5.4E**) allowing them to interact with the RNA transcript (Lane and Darst, 2010b). βGLU1072 lies directly adjacent to the residues βGLU1070 and absolutely conserved βMET1071 which interact with the switch 2 domain (β'SW2) (residues 405–423) and DNA respectively. Furthermore, βGLU1072 is also proximally located to the switch 3 (βSW3) domain (residues 1046–1067). β'SW2 serves as a hinge that connects the clamp domain (β residues 1093–1136; and β' residues 5–137, 254–418, and 1222–1245) to the *Mtb*-RNAP complex thus maintaining important clamp motional dynamics, while βSW3 forms interactions with the DNA-RNA assembly. β'PHE382 forms part of the β'rudder domain (**Figure 5.4F**) which serves to prevent the reassociation of the nascent RNA transcript by interacting with single-stranded DNA upstream of the DNA-RNA assembly (Naji *et al.*, 2008). β'SER428 is located between β'ARG427 and

β'Val429 which form interactions with DNA (**Figure 5.4G**) and contribute to the active site's structural core, respectively. Collectively, these residues may play an essential role in maintaining *Mtb*-RNAP fidelity and transcriptional capacity in the presence of missense mutations, therefore it might be therapeutically beneficial to identify a compound that could effectively interfere with the centrality of these essential hubs.

To explore the mutational/RIF-binding effects on intraprotein communication, we determined unique residue hubs that were absent in the *wt* but present in all the *mt*s and vice versa. Regarding the former criterion, only βLEU899 and the β'BH domain (residues 847–881) residue β'LEU865 were observed for the bound systems (**Figure 5.5A**).

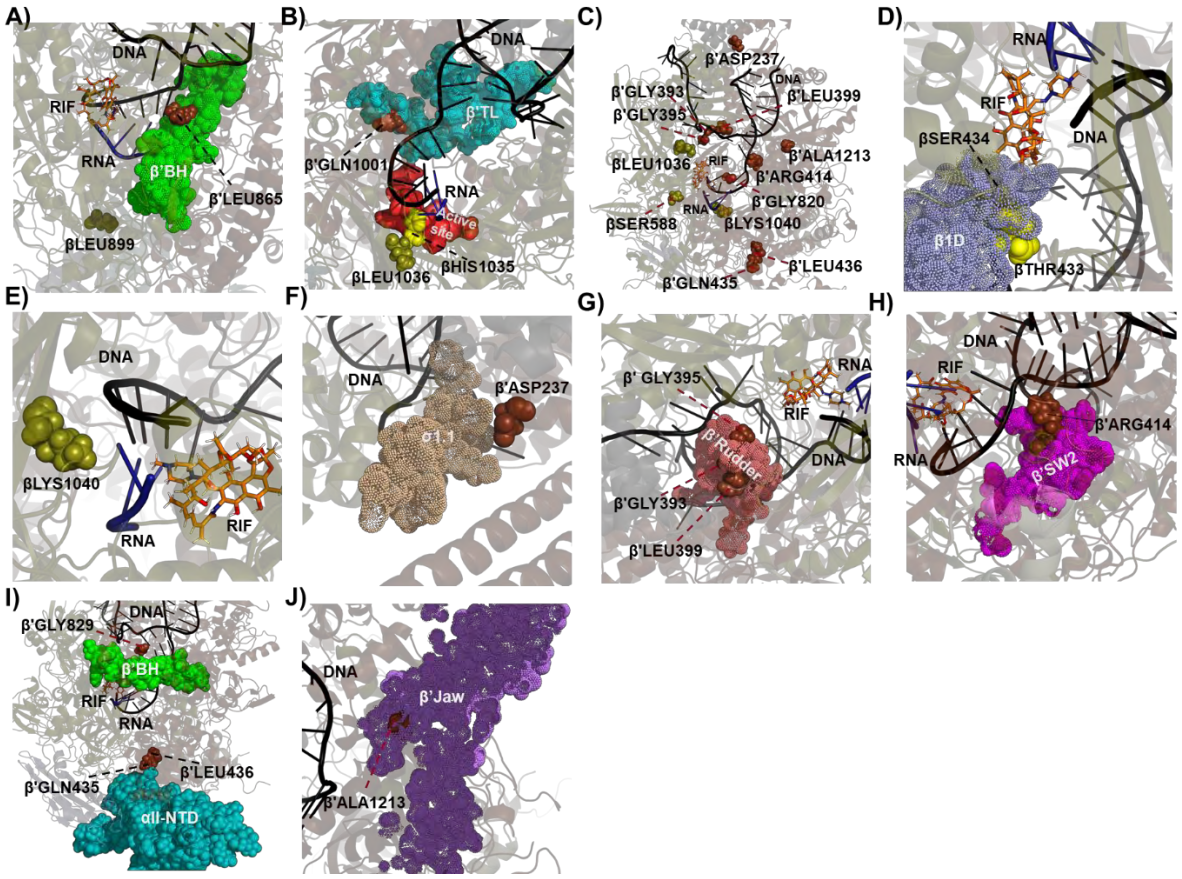


Figure 5.5. Illustration of the identified hubs (sphere representation) that either gained or lost centrality in the *mt* systems for the *Mtb*-RNAP β and β' subunits colored olive and brown respectively. Associated functionally relevant residue contacts are represented as yellow spheres while functional sites are indicated in dots and surface representations. The nucleic acids (DNA and RNA colored black and blue respectively) and RIF (orange) are represented as cartoons and sticks respectively. (A) Displays the β'LEU865 hub and its association with the β'BH domain and βLEU899's relative position. (B) Shows the β'GLN1001 hub and its association with the β'TL domain (teal) along with the βLEU1036 hub and its contacting residue βHIS1035 which are proximally located to the active site (red). (C) Displays the hubs that showed consistent loss of centrality across all of the *mt* systems. (D) The RIF-BP βSER434 hub which forms part of the β1D (light purple) is shown with its contacting residue βTHR433. (E) Shows the relative position of the βLYS1040 hub to DNA and RNA. (F) Shows the β'ASP237 hub and its close proximity to the σ1.1 domain (light brown). (G) Displays the β'GLY393, β'GLY395, and β'LEU399 hubs associated with the β'rudder domain (salmon). (H) Shows the β'ARG414 hub and its association with the β'SW2 domain (magenta). (I) Shows the close proximities of β'GLN435 and β'LEU436 hubs, and the β'GLY829 hub to the αII-NTD (teal) and the β'BH (green) domains respectively. (J) Displays the association of the β'ALA1213 hub with the β'jaw domain (dark purple).

Note that the β'BH domain plays a catalytic role wherein it switches between straight and kinked conformations to allow NTP binding during the nucleotide addition cycle,

and may also play a role in coordinating the conformational changes of surrounding domains (Tuske *et al.*, 2005; Weinzierl, 2010). Pertaining to the *BC* hubs that lost importance in the *mt* systems, we observed β residue LEU1036 along with β' residues ALA569 and GLN1001 for the unbound systems (**Figure 5.5B**). Among these residues, βLEU1036 is both located near the active site and follows absolutely conserved residue βHIS1035, whereas β'GLN1001 forms part of the catalytic β'TL domain. The β'TL domain is responsible for mediating active site entry of NTPs during the nucleotide addition cycle through a coordinated switch between open and closed conformations which is further modulated by interactions with the βFL2 and β'BH domains (Wang *et al.*, 2006). Regarding the RIF-bound systems, β residues PHE430, SER434, SER588, and LYS1040 along with β' residues ASP237, GLY393, GLY395, LEU399, ARG414, GLN435, LEU436, GLY829, and ALA1213 (**Figure 5.5C**) showed a loss of importance in all the *mt*s. Interestingly, the β1D residues βTHR433 and βSER434 indicated in **Figure 5.5D**, which also form part of the RIF-binding pocket residues, have been shown to increasingly fluctuate in the presence of some mutations and have therefore been implicated in RIF-resistance (Monama, Olotu and Tastan Bishop, 2023). βLYS1040 is however an absolutely conserved residue that directly contacts RNA and DNA (**Figure 5.5E**). As shown in **Figure 5.5F**, β'ASP237 appears to form contacts with the σ1.1 subdomain while β'GLY393, β'GLY395, and β'LEU399 form part of the β'rudder domain (**Figure 5.5G**). β'ARG414 is an absolutely conserved residue that forms part of the β'SW2 domain and makes essential contacts with the DNA template (**Figure 5.5H**). β'GLN435 and β'LEU436 form contacts with αII-NTD while β'GLY829 appears to contact the catalytic β'BH domain (**Figure 5.5I**). β'ALA1213 is part of the β'jaw domain (residues 1039–1218) (**Figure 5.5J**) which plays an important role in transcription pausing and the longevity of the open complex

(Drennan *et al.*, 2012). Finally, although it's worth noting that the RIF-bound I65T system observed the greatest differences in averaged *BC* deltas (*wt–mt* and *wt2–mt*) for both the β and β', it was not entirely surprising given that it exhibited the greatest RIF instabilities in our previous findings. More interesting, however, was the commonly occurring S456L system which also displayed large RIF instabilities and indicated the second largest difference in *BC* deltas for the β subunit. Overall, the observed changes to the *mt* system communication paths relative to the *wt* systems may be essential for inducing the dynamic residue-level structural effects associated with the RIF-resistant *Mtb*-RNAPs and may further indicate key drivers in establishing the compensatory conformations needed to shift from RIF's active orientation.

5.2.2. Averaged closeness centrality (CC) indicated residues that may play an important role in the catalytic efficiencies of wt and missense mutated Mtb-RNAPs in the presence and absence of RIF.

CC hubs are determined through a calculation that selects the residues most proximally located to all other residues within a protein over time (Goldbeck, 2013; Okeke *et al.*, 2021). It is for this reason that *CC* hubs tend to be core protein residues that can either uniquely and/or additionally be interface residues (Okeke *et al.*, 2021; Sheik Amamuddy *et al.*, 2021; Barozi, Edkins and Tastan Bishop, 2022). Consequently, this may also highlight residues that may be particularly efficient at residue-residue communication within the residue network (Goldbeck, 2013; Okeke *et al.*, 2021). Here, we observed that the mapped *CC* hubs were primarily distributed around the active site of *Mtb*-RNAP, as indicated in **Figure 5.6A-B**.

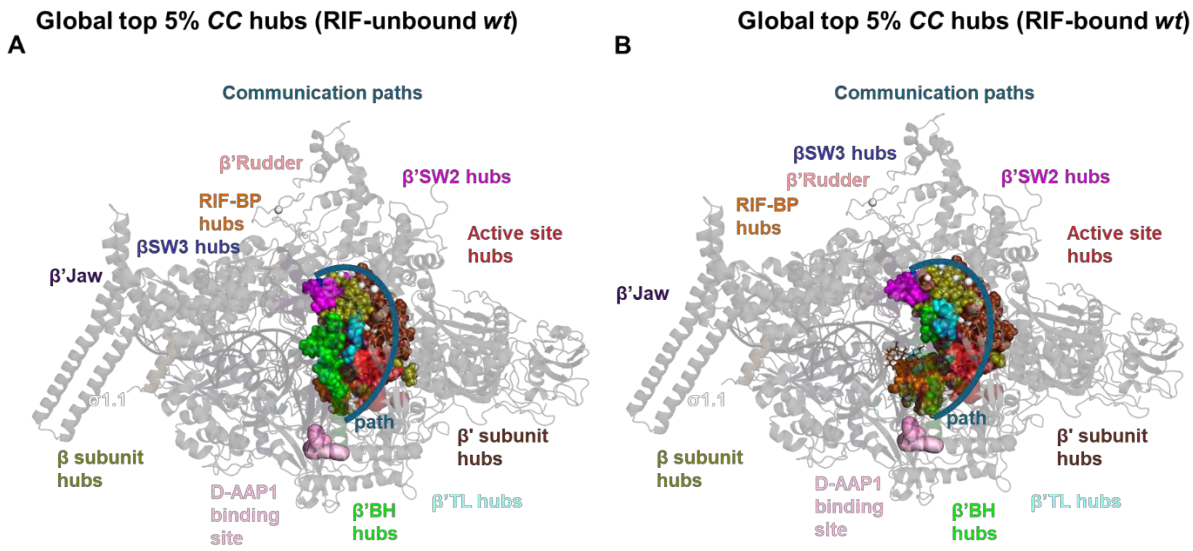


Figure 5.6. Distribution and paths of global 5% *CC* hubs and their structural associations for the (A) RIF-unbound and (B) bound *wt Mtb*-RNAPs.

Additionally, as similarly done in section **5.2.1** using the PyMOL interface script, we were able to confirm that several of the *CC* hubs were also interface residues (**Supplementary Figure S2.22**), further signifying their importance in protein function. As similarly portrayed through *BC* calculations, we also observed a communication path connecting the D-AAP1 allosteric site to the RIF-BP and the active site for the RIF-bound *wt* (**Figure 5.6B**). Peculiarly, this path was however disrupted in most of the *mt* systems (**Figure 5.7B**), which may suggest a reduced co-inhibitory effect upon D-AAP1 binding (Agoni, Ramharack and Soliman, 2018b).

Looking further at the distribution of *CC* hubs, we identified several of them as persistent (**Supplementary Table S2.1)**. Moreso, the persistent *CC* hub β residues ASP733–ILE735, ASP881–LEU885, LYS892–GLY896, LEU1033, VAL1037, LYS1040, ILE1041, GLN1066, PHE1068, GLY1069, and MET1071–ALA1075; and β' residues LYS420, VAL422, GLY426–VAL432, MET447, LEU498–ARG500, THR503, ARG506, GLY508–GLN510, LYS520–LEU524, CYS529, PHE532, and ALA534–VAL543, were observed in both RIF-unbound and bound systems (**Figures 5.7A-B**);

with some notable discrepancies in β'ASP535, β'ASP537, β'GLY538, and β'ASP539 for the unbound *wt* standards (**Supplementary Figure S2.12**). βASP733–βILE735 were found to form part of the structural core behind the active site. βASP733 in particular, however, is an absolutely conserved active site residue that forms critical interactions with the active site Mg^{2+}-chelating β'NADFDGD motif (residues 533–539) and the absolutely conserved βARG925, which itself further interacts with the active site Mg^{2+} and the γ-phosphate of an incoming nucleotide (Lane and Darst, 2010b). βASP881–βLEU885 and βLYS892–βGLY896 lie alongside one another while β'GLY426–β'VAL432 is oppositely positioned; however, all three regions, much like β'CYS529, lie behind the β'NADFDGD motif. Additionally, absolutely conserved βLYS884 makes further contacts with the nascent RNA transcript. βVAL1037 and βILE1041 are closely positioned near the absolutely conserved βLEU1033 and βLYS1040. βGLN1066 is adjacently located to absolutely conserved βGLY1065 which further lines the path of template DNA, whereas βPHE1068 and βGLY1069 form part of the βSW3 domain. βMET1071–βALA1075 appears as a helical structure with the DNA-interacting βMET1071. B'LYS420, with the inclusion of β'ASP423, β'GLU554, β'ASP580, and β'ARG857, showed the highest commonality among the identified hot spot residues (**Table 5.1**) through the SPOTON CASM approach (Moreira *et al.*, 2017), emphasizing functional importance (see **section 5.4**). Additionally, β'LYS420 along with β'VAL422 form part of the SW2 domain and lie adjacent to absolutely conserved β'ARG421 which makes critical contacts with the DNA template. β'ARG500 forms important interactions with the RNA transcript and the β'NADFDGD motif's ASN533, ALA534, and ASP539 residues. Between β'THR503 and β'ARG506 lie the residues β'LEU504 and β'PRO505 which form interactions with the absolutely conserved β'THR1080 β'TL residue and DNA respectively. β'GLY508–β'GLN510 appears as a

loop and forms contacts with the helical structure that leads to the β'NADFDGD motif. β'ARG535 forms critical interactions with the catalytic Mg^{2+} ion while the essential motif residues β'ASP535, β'ASP537, and β'ASP539 coordinate the active site Mg^{2+} ion.

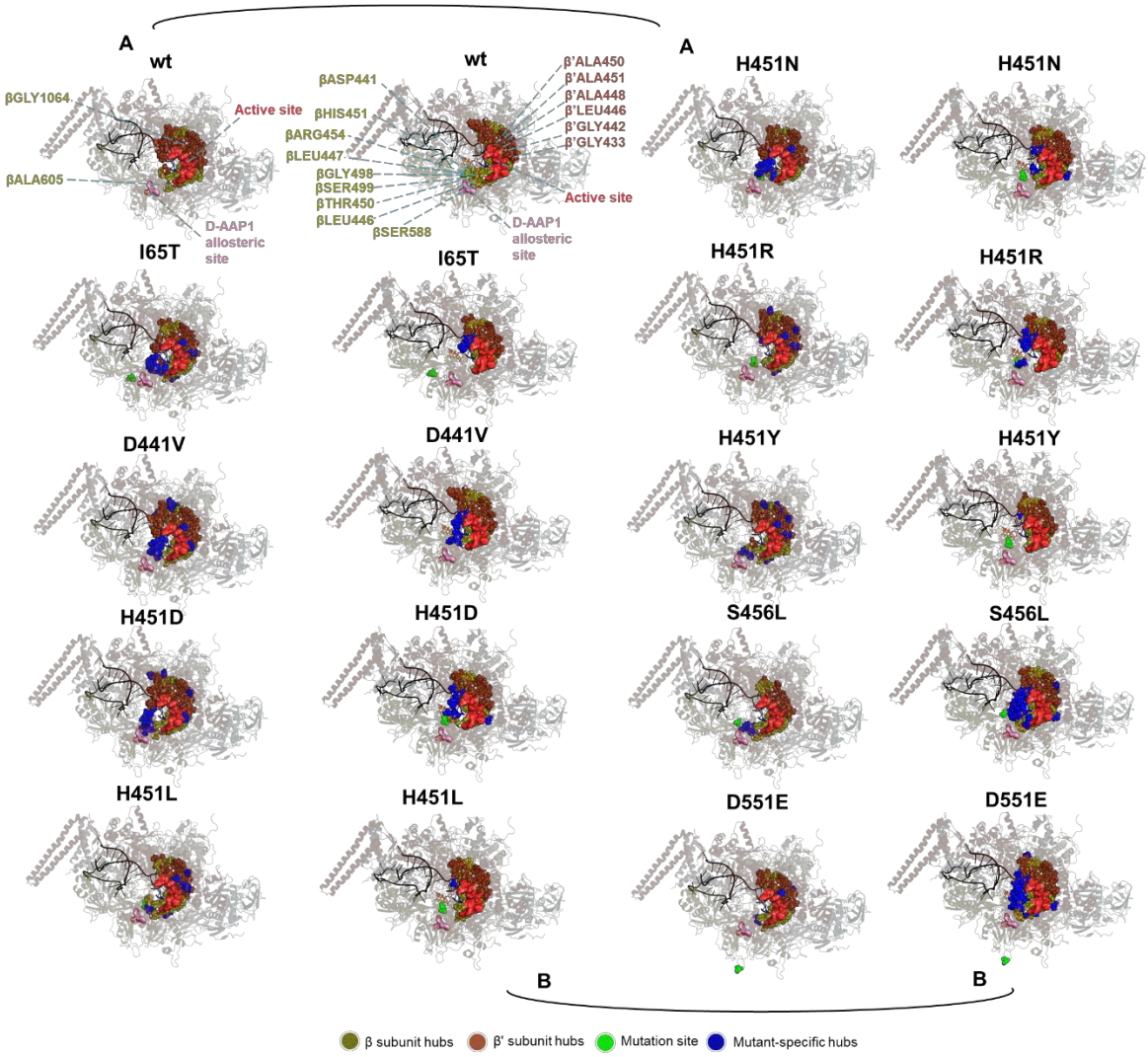


Figure 5.7. 3D *Mtb*-RNAP structures of the (A) RIF-unbound and (B) bound *wt* and representative *mt*s with mapped global top 5% *CC* hubs for the β and β' subunits. Hubs are shown in sphere representation and are colored olive and brown for the respective β and β' subunits. The blue- and purple-colored spheres represent the mutated site and *mt*-specific hubs respectively. Also, the active site (sphere representation), D-AAP1 allosteric site (surface representation), DNA (cartoon representation), RNA (cartoon representation), and RIF (stick representation) are each colored red, pink, black, dark blue, and orange. The residue annotations on the *wt* and *mt* systems represent hubs that are unique to the *wt* and all the *mt*s respectively for the bound and unbound states.

These altogether show that the identified persistent *CC* hubs are primarily involved in (either directly or through a structurally supportive role) *Mtb*-RNAP's catalysis, where transcriptional efficiency is crucial (Belogurov and Artsimovitch, 2019). To better understand the impact of mutation/ligand binding on *CC* hub distributions and potentially relay some details on the literature-reported decline in the catalytic activity of the mutated *Mtb*-RNAPs (Stefan, Ugur and Garcia, 2018), we looked at the introduced *CC* hubs for the RIF-unbound and bound *mt* systems and compared them to their corresponding *wt* standard. Firstly, it is worth noting that there were no consistent *CC* hubs introduced among the *mt* systems for the unbound and bound systems. In fact, the *mt* systems generally lost more *CC* hubs than they gained, which not only suggests a varying mutational effect but also shows agreement with the expected variable decline in catalytic function (Stefan, Ugur and Garcia, 2018). Furthermore, we found that βALA605 and βGLY1064 consistently lost importance in residue-residue communication, as shown by their relative loss of significance through the global top 5% analyisis among all the unbound *mt* systems. More notably however, β residues ASP441, LEU446, SER447, THR450, HIS451, ARG454, GLY498, SER499, and SER588; and β' residues GLY433, GLY442, LEU446, ALA448, GLU450, and LEU451 consistently lost importance in communication among the bound *mt* systems. Of particular interest was the consistent loss in centrality for the β residues, especially considering that βASP441, βHIS451, and βARG454 make up the RIF-BP while the rest of the β residues are located nearby. That said, the loss of specific *CC* hubs in the *mt* systems, particularly those of the active site and the RIF-BP, suggests that RIF binding decreases the efficiency of residue-residue communication in those regions for the missense mutated *Mtb*-RNAPs, the consequence of which may contribute to the reported transcriptional inefficiencies and reduced RIF effects. With that said,

when we observed the averaged *CC* deltas (*wt–mt* and *wt2–mt*), the H451D system showed the greatest difference overall for unbound β–β'complex, whereas the S456L system displayed the largest discrepancies in the bound form.

Complementary to the identification of hubs that may play a role in transcriptional efficiency, we assessed the motional dynamics of the active site through Cα-RMSD and Cα-RMSF calculations. We found that most of the *mt* systems indicated greater changes in stability relative to the *wt* standards, the purpose of which may be compensatory. This was particularly observed for the RIF-unbound D441V, H451N, and S456L systems and the RIF-bound I65T, H451L, and H451R systems as shown in **Figure 5.8A**.

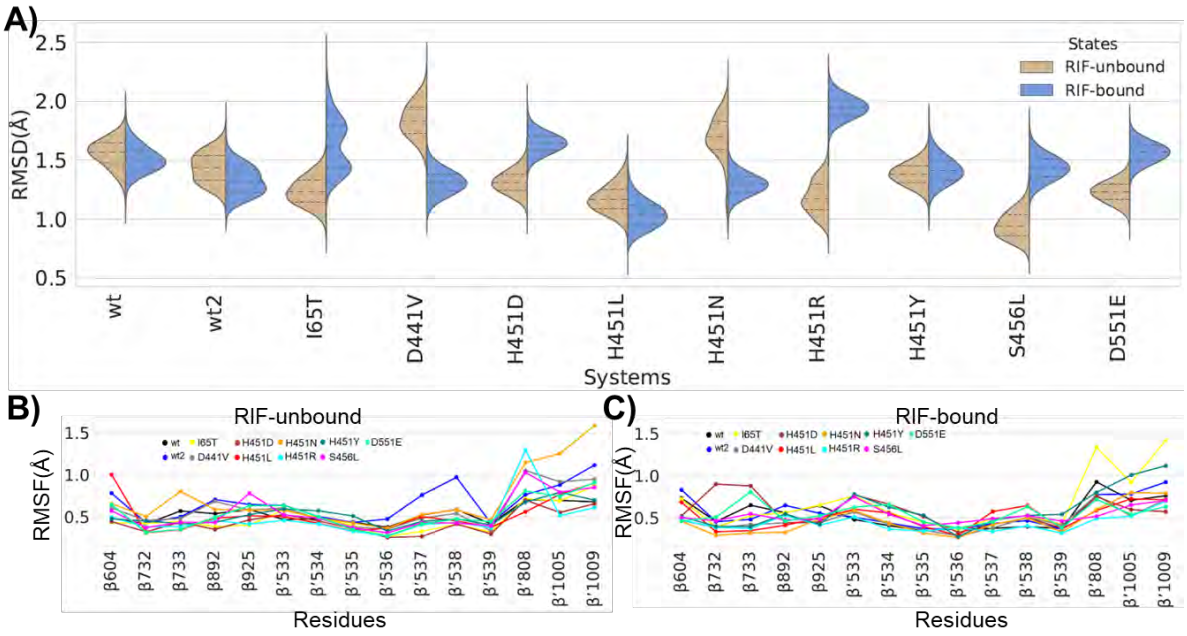


Figure 5.8. Graphical representation of the sampled *Mtb*-RNAP active site conformations and residue level fluctuations for the *wt*s and *mt* systems. A) Displays the Cα-RMSD violin plots for the RIF-unbound (brown) and bound active site (blue). B) Shows the Cα-RMSF line plot for the active site residues for the *wt*s and *mt*s in the RIF-unbound state. C) Shows the Cα-RMSF line plot for the active site residues for the *wt*s and *mt*s in the RIF-bound state.

We further observed that the differences in active site stabilities may be due at least in part to increases in residue-level fluctuations as indicated by changes in Cα-RMSF for residues β'THR808, β'GLU1005, and β'GLN1009 for the unbound H451N and RIF-bound I65T systems (**Figure 5.8B-C**). We also noted some discrepancies in the unbound *wt* standards residues β'ASP537 and β'GLY538, which likely contributed to the slight differences in active site stabilities.

5.2.3. Residues that may be particularly important in Mtb-RNAP catalytic function and RIF inhibition were captured through the averaged eigenvector centrality (EC) metric.

Eigenvector centrality (*EC*) measures the importance of a node by calculating its degree of influence (Ruhnau, 2000; Goldbeck, 2013; Sheik Amamuddy *et al.*, 2021). That is, the higher the connections a neighboring residue has, the higher the residue's centrality and vice versa. Looking at the distribution of mapped *EC* hubs, we were able to visually identify two highly clustered communication paths for the unbound and bound *wt* systems respectively (**Figure 5.9A-B**).

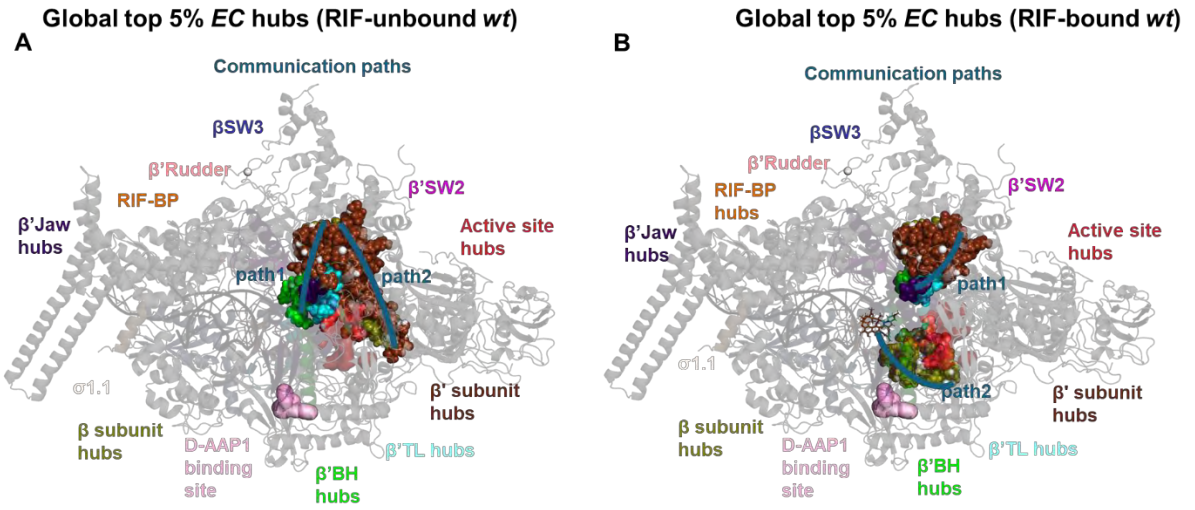


Figure 5.9. Distribution and paths of global 5% *EC* hubs and their structural associations for the (A) RIF-unbound and (B) bound *wt Mtb*-RNAPs.

More specifically, for the unbound *wt* system, the hub clusters were located closer to the active site than the RIF-binding site (**Figure 5.10A**), whereas for the RIF-bound *wt* system, the clusters were positioned both near RIF and the active site (**Figure 5.10B**). This showed that RIF binding may increase the importance of specific influential residues around RIF, an effect that is perhaps necessary to establish its inhibitory activities. In the unbound and RIF-bound *mt* systems, however, the distribution of *EC* hubs changed, resulting in a great loss/increase in the importance of hubs around RIF and/or the active site, thereby indicating a possible contribution toward the decline of RIF efficacy and *Mtb*-RNAP catalytic function. It's also worth noting that although there were no persistent hubs identified for the RIF-bound systems, most of the *EC* hubs highlighted in the unbound systems were also observed to show the greatest consistency in high centralities in the bound systems. As shown in **Supplementary Table S4.1**, only β residues GLY1083–LEU1085 and β' residues GLU550, ALA551, and ALA553–ALA555 were identified as persistent *EC* hubs for the unbound systems. βGLN1077 and βGLY1083–βLEU1085 appear to form contacts with the βclamp, β'SW5, and the β'BH domain which collectively play an important role in DNA interactions and catalysis.

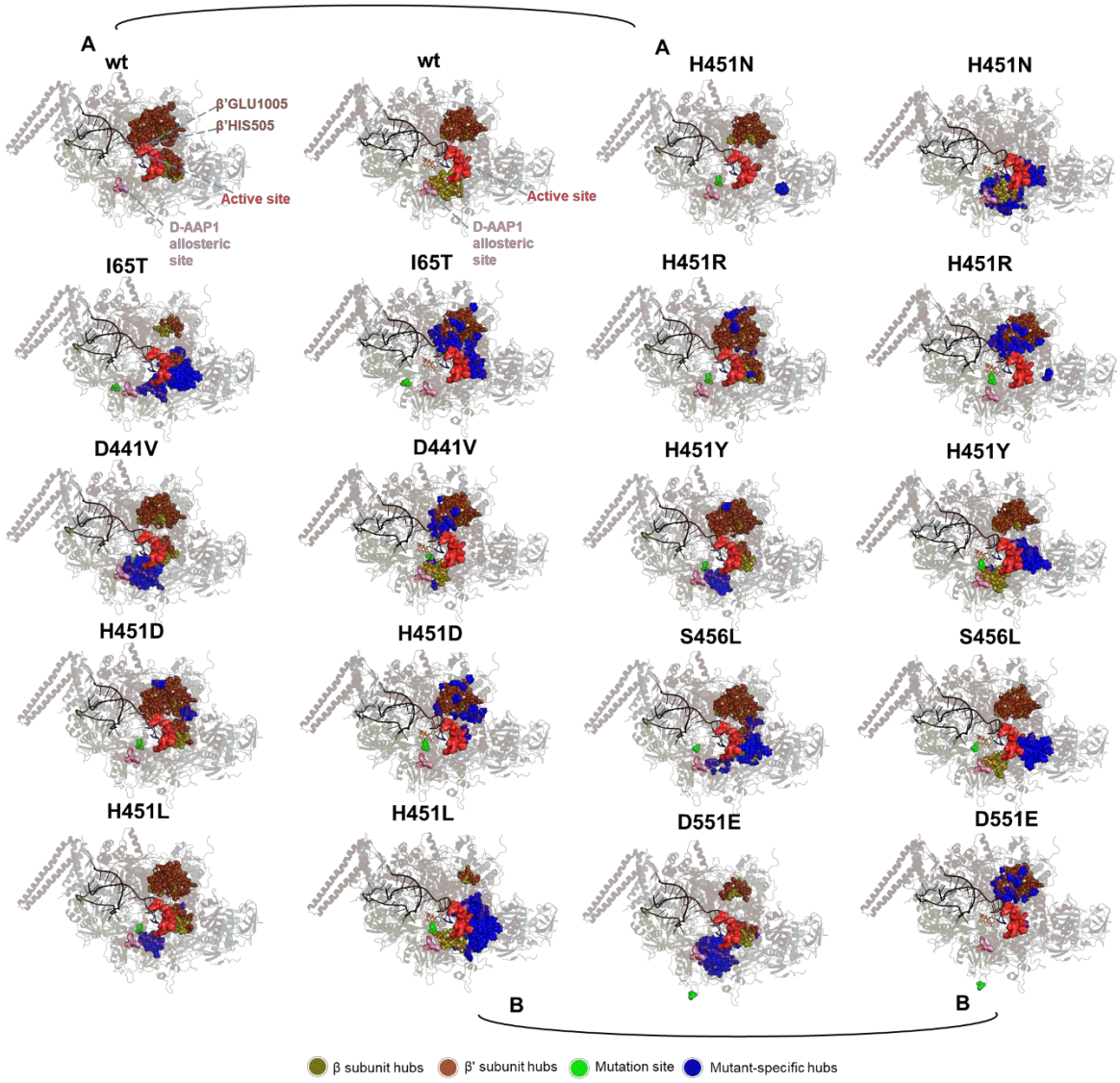


Figure 5.10. 3D *Mtb*-RNAP structures of the (A) RIF-unbound and (B) bound *wt* and representative *mt*s with mapped global top 5% *EC* hubs for the β and β' subunits. Hubs are shown in sphere representation and are colored olive and brown for the respective β and β' subunits. The blue- and green-colored spheres represent the *mt*-specific hubs and the mutated site respectively. Also, the active site (sphere representation), D-AAP1 allosteric site (surface representation), DNA (cartoon representation), RNA (cartoon representation), and RIF (stick representation) are each colored red, pink, black, dark blue, and orange. The residue annotations on the *wt* and *mt* systems represent hubs that are unique to the *wt* and all the *mt*s respectively for the bound and unbound states.

Next, we identified the mutational and/or RIF-binding effects on *EC* hub distribution by looking at those hubs that gained/lost importance in the *mt* systems for the RIF-unbound and bound systems. We found a consistent loss of the DNA-interacting β'HIS505 along with the active site residue β'GLU1005 among the unbound *mt*

systems, while no common loss of importance among the RIF-bound *mt* systems could be determined. Additionally, no commonality could be observed among the hubs that gained importance in the unbound and bound *mt* systems. Regarding the *mt* systems that showed the most change relative to the *wt* standards, we identified D551E and H451D for the β and β' respectively by quantifying the averaged *EC* deltas (*wt–mt* and *wt2–mt*). Additionally, the RIF-bound H451N and I65T systems, which were associated with the highest RIF instabilities in our previous study (Monama, Olotu and Tastan Bishop, 2023), showed the greatest differences in averaged *EC* deltas for the β and β' respectively. Briefly, our findings imply that the presence of the missense mutations and RIF binding result in varying local effects with great emphasis on the active site and RIF-binding site, particularly on residues considered to highly influence the dynamics of *Mtb*-RNAP in those regions. These local changes in centrality may therefore in part inform the observed differences in conformational sampling observed in our previous findings for the unbound and bound *mt* systems and additionally show a link towards both the reported loss of RIF's efficacy and *Mtb*-RNAP's catalytic efficiency (Stefan, Ugur and Garcia, 2018; Monama, Olotu and Tastan Bishop, 2023).

5.2.4. Averaged degree centrality (DC) indicated residues that may drive conformational changes in wt and mt Mtb-RNAPs.

Degree centrality (*DC*) is a simpler network metric compared to *EC* and associates the centrality of a node with the number of edges going through it, without considering the nature of the node's influence. That is, *DC* imparts that the higher the number of edges a particular node has, the higher the node's centrality and vice versa (Hanneman and Riddle, 2005; Zhan, Gurung and Parsa, 2017). Although we could not find a discrete communication path when we visually inspected the distribution of *wt Mtb*-RNAP *DC*-mapped hubs (**Figure 5.11A-B**), our inspection of the *DC* persistent hubs as outlined

in **Supplementary Table S2.1** revealed that β residues VAL236, LYS239, THR261, SER501, SER588, ILE650, ILE702, VAL803, VAL847, GLY937, CYS939, and ALA1082, as well as β' residues HIS94, PRO321, VAL422, VAL456, LEU460, ALA553, GLU554, TYR586, ALA631, GLY720, LEU925, VAL962, ALA994, ILE1167, ALA1201, ALA1260, and THR1262, were persistent for both the unbound and RIF-bound systems. βVAL236, βLYS239, and βTHR261 in particular form part of the β2 domain (residues 181–370) whereas βVAL803 and βVAL847 make up the βflap domain (residues 752–874) and βALA1082 appears to form contacts with the β' switch 5 (β'SW5) domain (residues 1257–1259). The β2 domain serves to guide non-template single-stranded DNA within the active center channel while the βflap domain plays roles in all phases of the transcription process including active site structural support and critical RNA interactions; and β'SW5 is important for clamp dynamics (Yang *et al.*, 2009; Lane and Darst, 2010b; Molodtsov *et al.*, 2015; Morichaud *et al.*, 2023). β'HIS94 makes up the β'ZBD (residues 24–98) which plays essential roles in multiple phases of the transcription process, including but not limited to transcriptional termination and antitermination (King *et al.*, 2004; Hu and Liu, 2022). Besides β'GLU554 being detected as a hot spot residue (**Table 5.1**), it has been reported to make structurally essential buried polar interactions with the other identified hot spot residue, β'ASP580, whereas β'TYR586 may form part of the NTP entry channel (Lane and Darst, 2010b). β'ALA994 is part of the catalytic β'TL domain while β'ALA1260 and β'THR1262 are positioned just beyond the β'SW5 domain. In summary, the scattered distribution of *DC* hubs not only displays the extensive network of *Mtb*-RNAP's local contacts but also shows the specific residues that may play an important role in contributing to *Mtb*-RNAP's conformational stability (Foutch, Pham and Shen, 2021).

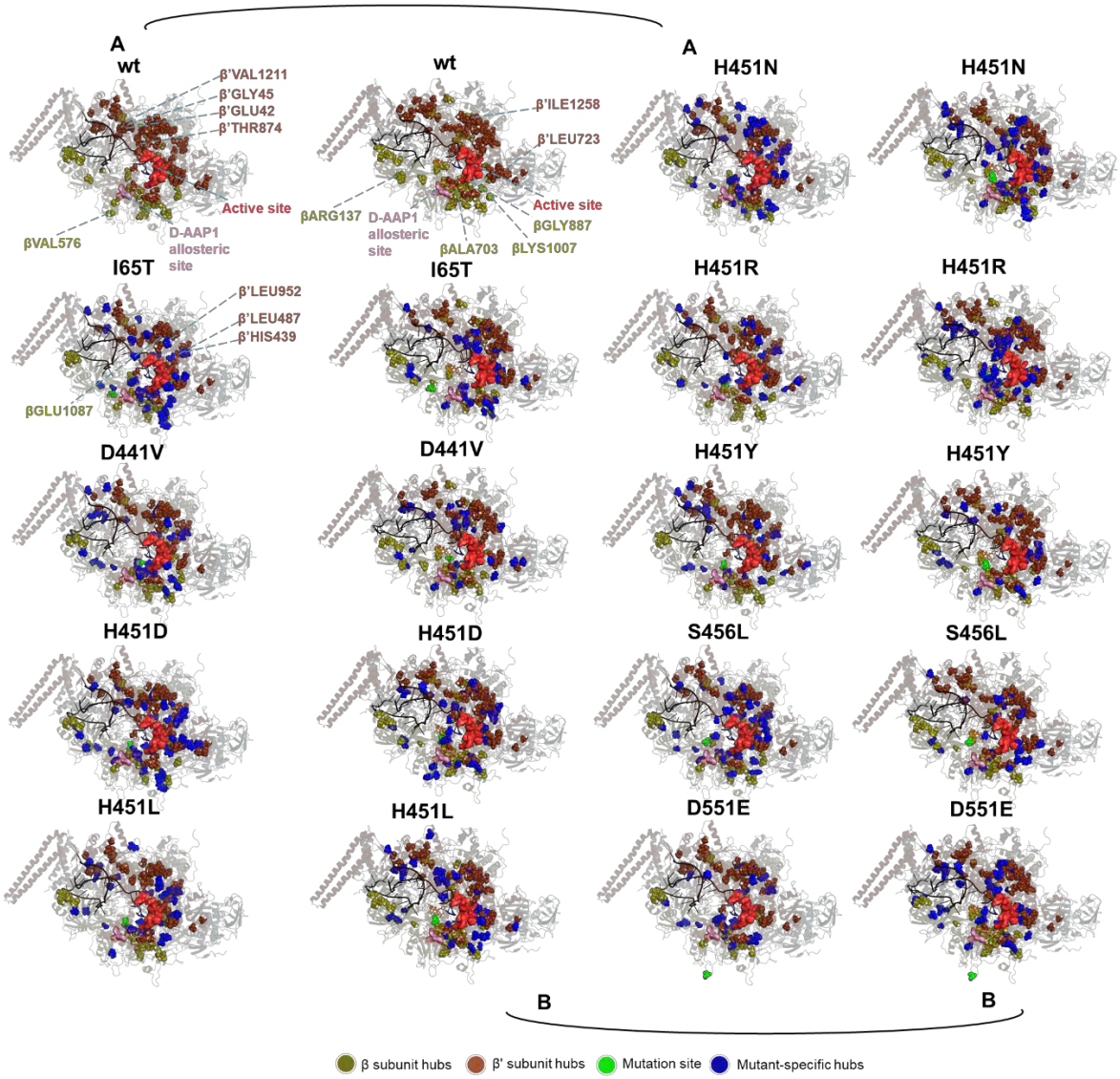


Figure 5.11. 3D *Mtb*-RNAP structures of the (A) RIF-unbound and (B) bound *wt* and representative *mt*s with mapped global top 5% *DC* hubs for the β and β' subunits. Hubs are shown in sphere representation and are colored olive and brown for the respective β and β' subunits. The blue- and purple-colored spheres represent the mutated site and *mt*-specific hubs respectively. Also, the active site (sphere representation), D-AAP1 allosteric site (surface representation), DNA (cartoon representation), RNA (cartoon representation), and RIF (stick representation) are each colored red, pink, black, dark blue, and orange. The residue annotations on the *wt* and *mt* systems represent hubs that are unique to the *wt* and all the *mt*s respectively for the bound and unbound states.

It's worth noting that several introduced *DC* hubs were observed both around and making up the active site, including those near the mutation site, for most of the RIF-unbound and bound *mt* systems. However, only the β residue GLU1087 which lies near absolutely conserved residue βLEU1089, which is itself further associated with

myxopyronin binding (Belogurov *et al.*, 2009), and β' residues HIS439, LEU487, and LEU952, were found to consistently gain importance in the unbound *mt* systems. Regarding those hubs that consistently lost importance, however, we identified β residue VAL576 and β' residues GLU42 of the promoter DNA-contacting β'zipper domain (residues 32–44) (Yuzenkova *et al.*, 2011), GLY45 and THR874 of the β'BH domain, and the β'jaw domain's VAL1121 for the unbound *mt* systems. Contrastingly, βARG137 of the β1D, βALA703, βGLY887, and βLYS1007, along with β'TRP723 which appears to contact the αII-NTD, and β'ILE1258 belonging to the β'SW5 domain were highlighted for the bound *mt* systems. Finally, we proceeded to analyse the quantified averaged *DC* deltas (*wt–mt* and *wt2–mt*). With that, we found that the unbound H451Y and H451N systems displayed the greatest changes for the β and β' respectively, while the H451N and S456L systems showed the largest changes in β and β' respectively in the RIF-bound states, signifying particularly extensive local effects.

5.2.5. Averaged Katz centrality (KC) highlighted catalytically important hubs including those that may serve in mediating RIF inhibition.

Katz centrality (*KC*) is a network metric that is similar to *EC* in highlighting influential nodes; however, instead of being limited to the immediately contacting high-degree nodes, it expands the range to include those high-degree nodes that are not immediately contacted (Koschützki and Schreiber, 2008; Zhan, Gurung and Parsa, 2017). Hence, it was not surprising that the distribution of *KC* hubs (**Figure 5.12A-B**) was similar to that of *EC*. We also noted path2 forming a communication path between the D-AAP1 binding site and the active site for both the RIF-unbound and RIF-bound *wt* unlike with *EC*.

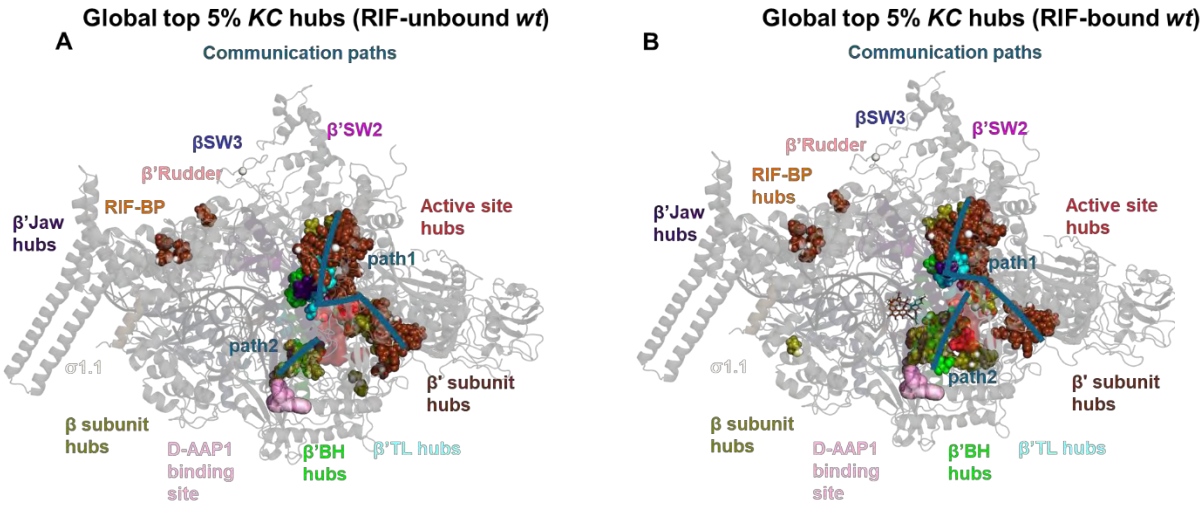


Figure 5.12. Distribution and paths of global 5% *KC* hubs and their structural associations for the (A) RIF-unbound and (B) bound *wt Mtb*-RNAPs.

This may again highlight the binding site's potential for introducing allosteric co-inhibitory effects. Furthermore, we noted a higher concentration of *KC* hubs near RIF in the RIF-bound *wt* as compared to its unbound counterpart, as similarly observed in the *EC* results (**Figure 5.13A-B**). βGLY1080–βLEU1085 and βclamp contacting βGLU1087; and β'VAL456, β'ALA549–β'ALA555, β'GLY584, the β'BH proximally positioned β'ALA994–β'ILE997, the β'SW5 β'ILE1258 residue and β'ALA1260–β'THR1262, were found to be persistent among the RIF-bound and unbound systems (**Supplementary Table S2.1)** with some similarities to the *EC* section. It's also worth noting that discrepancies were detected in βGLY896, βGLY1080, βALA1082, and β'GLY584, and β'ILE1258, β'ALA1260, and β'GLY1261 for the unbound and bound *wt* standards respectively (**Supplementary Figure S2.15**). Unlike *EC*, the concentration of *KC* hubs in the vicinity of RIF's binding site appeared to increase in a majority of the unbound *mt* systems, while a decrease in the concentration of *KC* hubs near RIF was observed for several of the bound *mt* systems, further implying the role of the mutations in reducing native binding site hub centrality.

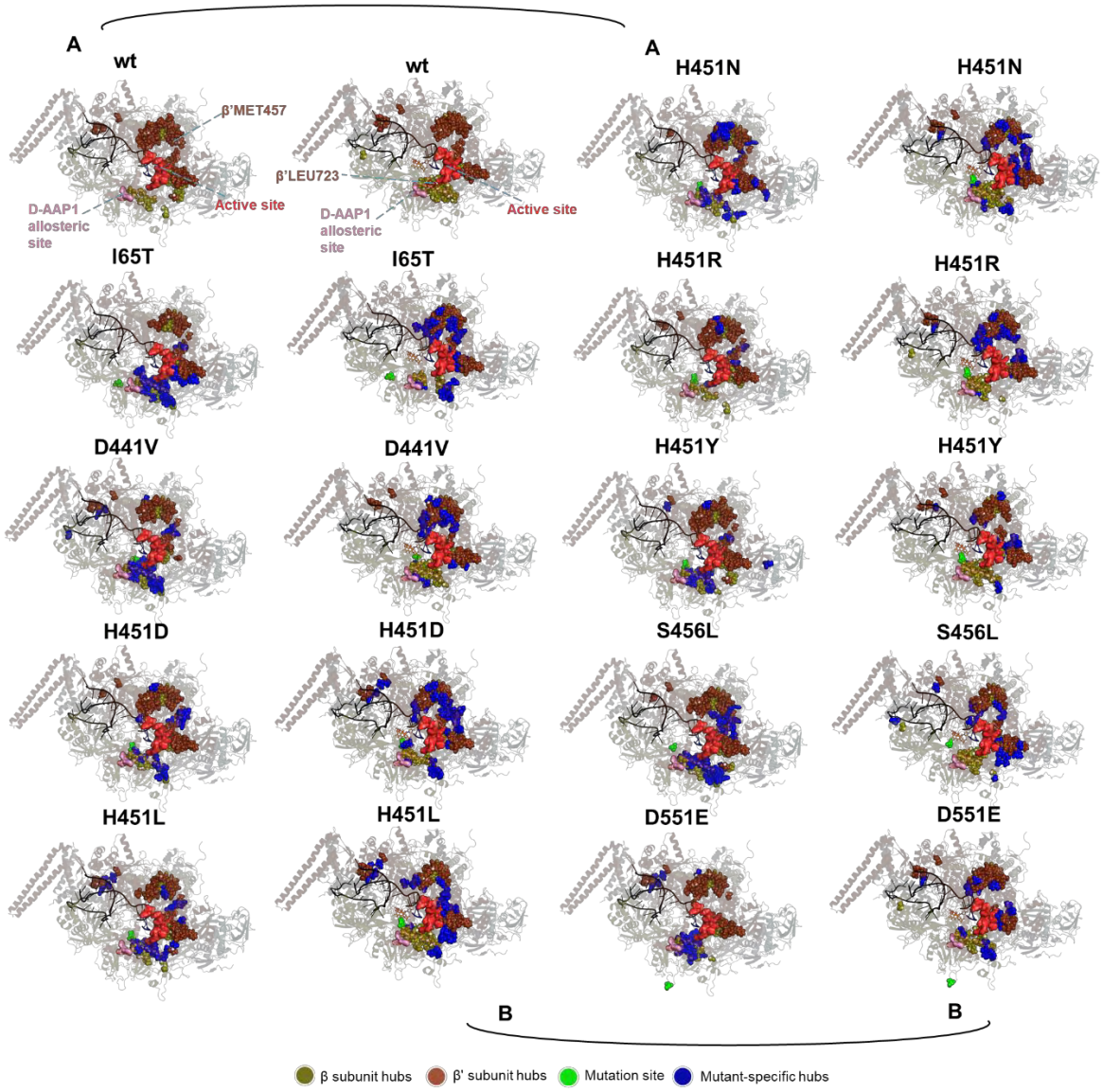


Figure 5.13. 3D *Mtb*-RNAP structures of the (A) RIF-unbound and (B) bound *wt* and representative *mt*s with mapped global top 5% *KC* hubs for the β and β' subunits. Hubs are shown in sphere representation and are colored olive and brown for the respective β and β' subunits. The blue- and purple-colored spheres represent the mutated site and *mt*-specific hubs respectively. Also, the active site (sphere representation), D-AAP1 allosteric site (surface representation), DNA (cartoon representation), RNA (cartoon representation), and RIF (stick representation) are each colored red, pink, black, dark blue, and orange. The residue annotations on the *wt* and *mt* systems represent hubs that are unique to the *wt* and all the *mt*s respectively for the bound and unbound states.

Additionally, βSER501, βSER588, the β'NADFDGD motif proximally located βGLY896, βGLY937, the β'SW5 proximally positioned βGLU1077, βALA1078,

Among the residues that lost importance, the σ3.2-contacting β'MET457 and β'ASN852 which form part of the catalytic β'BH domain were observed to consistently lose importance in the unbound and bound *mt* systems respectively. Interestingly, of the hubs that indicated a common gain in importance among the RIF-bound *mt* systems, we observed the catalytic β'TL domain β'GLN1001 to be the most consistent, which further collectively emphasized a local alteration affecting *Mtb*-RNAP catalytic activity. Finally, to identify the mutations displaying the largest overall influence on the *Mtb*-RNAP β–β' complex, we used the quantified averaged *KC* deltas (*wt*–*mt* and *wt*2–*mt*). Our analysis revealed that the unbound D441V and H451D systems displayed the largest discrepancies in the β and β' respectively. However, the RIF-bound H451N system indicated the greatest differences for the β subunit while the largest discrepancies for the β' were indicated by the H451D system.

5.3. Weighted contact map analysis revealed discrepancies in mutation site residue interactions in the RIF-unbound and bound mt systems.

To better decipher how the mutations/RIF binding induces local changes to *Mtb*-RNAP for the investigated *wt* and *mt* systems, we conducted a dynamic residue contact map analysis on the mutation sites to specifically determine their frequency of interactions with neighboring β residues in the presence and absence of RIF using the MDM-TASK-web weighted contact map script (Sheik Amamuddy *et al.*, 2021). The normalized frequency of contacts ranged between 0, which represented no contacts, and 1, which indicated sustained interactions throughout the simulation.

Based on **Figure 5.14A-B**, we were able to observe several noteworthy contact discrepancies between the *wt* and *mt* systems in both RIF-unbound and bound states. $_{mt}$THR65, for instance, showed noticeable differences in contact frequencies with the

β1D residues LEU86 and PHE112, relative to the $_{wt}$ILE65 system for both the unbound and RIF-bound forms. This may have had an impact on β1D dynamics in the *mt* system, given that the β1D RIF-BP residue THR433 was found to greatly fluctuate in the presence of RIF (Monama, Olotu and Tastan Bishop, 2023).

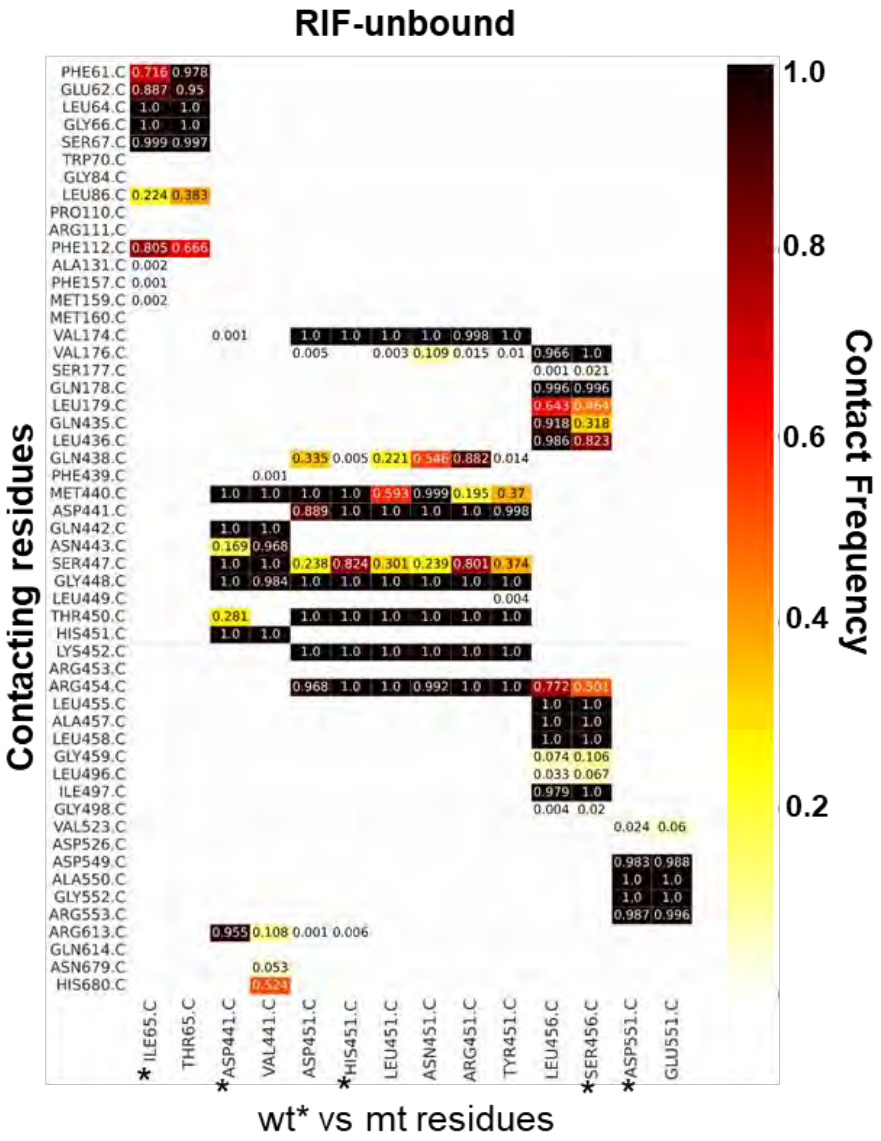


Figure 5.14. Heatmap representation of the weighted atomic contacts of the investigated *Mtb*-RNAP-β mutation sites for the RIF-unbound systems. The contact frequencies are denoted by colors ranging from yellow (non-existent to low-frequency interactions) to black (highly maintained interactions). The *wt* residues are marked with an asterisk (“*”).

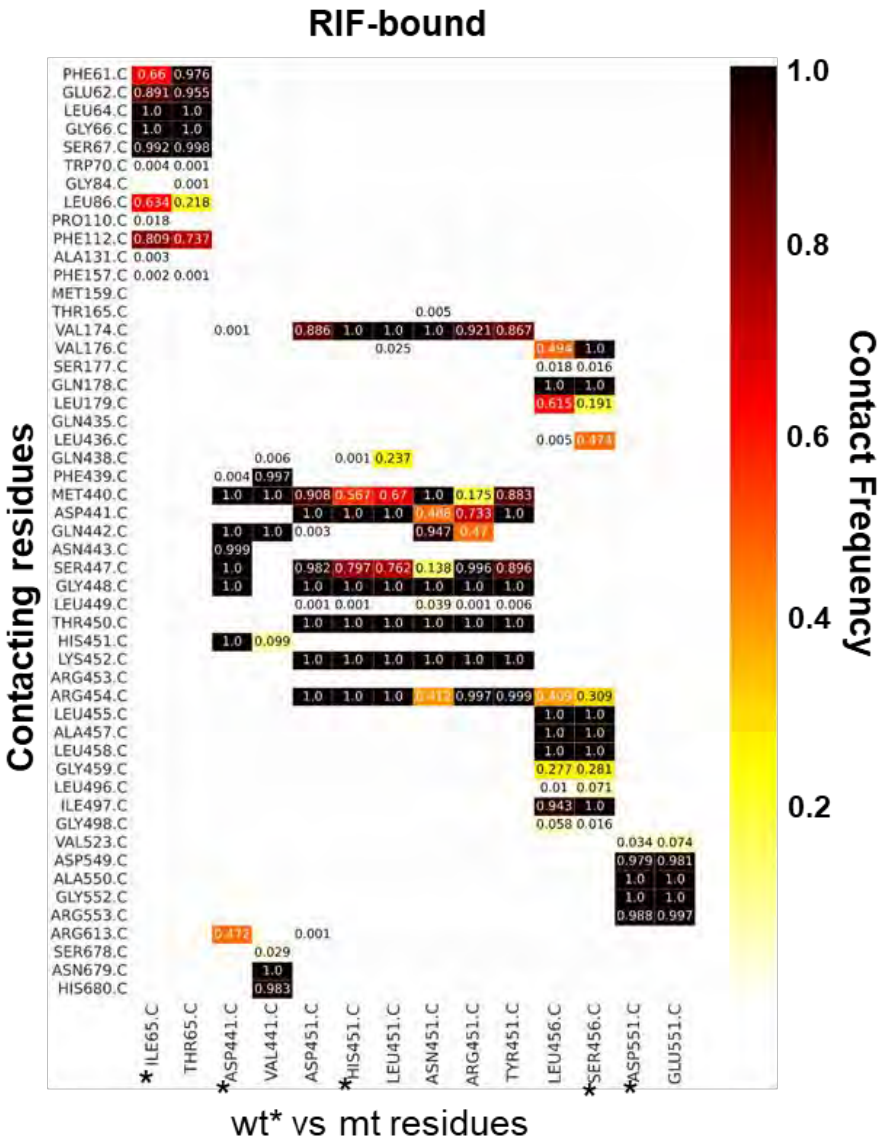


Figure 5.15. Heatmap representation of the weighted atomic contacts of the investigated *Mtb*-RNAP-β mutation sites for the RIF-bound systems. The contact frequencies are denoted by colors ranging from yellow (non-existent to low-frequency interactions) to black (highly maintained interactions). The *wt* residues are marked with an asterisk ("*").

RIF-BP $_{mt}$VAL441 indicated a great decrease in contact frequencies with ARG613 while displaying a drastic increase in interactions with ASN443 and HIS680 in the unbound form, relative to the $_{wt}$ASP441. In the RIF-bound form, however, $_{mt}$VAL441 appears to drastically increase interactions with PHE439, ASN679, and HIS680 while showing a great decrease in contacts with HIS451. Additionally, given the importance

of PHE439 and HIS451 in stabilizing RIF interactions, this shift from native contacts over time may likely greatly contribute to RIF's reorientation (Lin *et al.*, 2017; Monama, Olotu and Tastan Bishop, 2023). Looking at $_{mt}$ASP451, $_{mt}$LEU451, $_{mt}$ASN451, $_{mt}$ARG451, and $_{mt}$TYR451, we were able to determine a common trend among the contacting residues relative to the $_{wt}$HIS451 wherein we observed increases in interactions with GLN438 and/or decreases in interactions with MET440 and SER447 for the unbound systems. Interestingly, $_{mt}$ASN451, which displayed one of the most destabilized RIF orientations, indicated the most drastic differences in contacting frequencies among its neighboring residues, with ASP441, ASN443, and ARG454 in particular displaying some of the largest decreases in interactions for the bound systems (Monama, Olotu and Tastan Bishop, 2023). The most common RNAP missense mutation, $_{mt}$LEU456, which also clearly displayed a shifted RIF orientation in our previous work, displayed large discrepancies in VAL176, LEU179, and LEU436 for the bound form, which may suggest that they play a key role in its resistance mechanism. Finally, $_{mt}$GLU551 showed negligible differences concerning its frequency of interactions with neighboring residues relative to $_{wt}$ASP551, thereby suggesting a more subtle local effect.

5.4. Computational alanine scanning mutagenesis (CASM) analysis indicated a loss of hot spot statuses in residues that may be integral in forming a stable *Mtb*-RNAP β-β' complex.

Herein, we also investigated the changes in binding between the *wt* and *mt Mtb*-RNAP β-β' complexes in the RIF-bound and unbound forms by determining the binding energies of the interface residues through computational alanine scanning calculations. That is, individual interface residues were mutated to alanine to detect their energy contributions towards the stability of the protein-protein interaction

through the SpotOn server (Moreira *et al.*, 2017). Those interface residues that greatly contributed to β-β' complex binding (≥ 2 kcal/mol free energy binding difference) are referred to as hot spot residues, whereas those that only minimally contributed to complex binding (< 2 kcal/mol free energy binding difference) were regarded as cold spot residues.

Although the majority of residues were identified as cold spot residues, several residues were identified as hot spot residues (**Table 5.1**).

Table 5.1. Showing the identified hot spot residues (≥ 2 kcal/mol binding energy difference) for the β-β' complex of *Mtb*-RNAP *wt*s and *mt*s.

Systems	Hot spot residues (RIF-unbound)	Hot spot residues (RIF-bound)
wt	βGLU732, **β'LYS420**, **β'ASP423**, **β'GLU554**, **β'ASP580**, **β'ARG857**	β'LYS66, β'GLU90, **β'LYS420**, **β'ASP423**, β'LYS520, **β'ASP537**, **β'GLU554**, **β'ASP580**, **β'ARG857**
***wt*2**	β'ARG89, **β'LYS420**, **β'ASP423**, β'GLU450, β'ARG478, **β'GLU554**, **β'ASP580**, **β'ARG857**	**β'LYS420**, **β'ASP423**, **β'ASP537**, **β'GLU554**, **β'ARG857**
I65T	**β'LYS420**, **β'ASP423**, β'GLU450, **β'GLU554**, **β'ASP580**	βASP601, βGLU732, β'GLU90, **β'LYS420**, β'ARG421, **β'ASP423**, β'GLU450, **β'GLU554**, **β'ARG857**
D441V	βASP601, **β'LYS420**, **β'ASP423**, **β'GLU554**, **β'ASP580**, **β'ARG857**	**β'LYS420**, **β'ASP423**, β'GLU450, **β'GLU554**, **β'ARG857**
H451D	β'ARG89, **β'LYS420**, β'ASP537, **β'GLU554**, **β'ASP580**, **β'ARG857**	βASP474, **β'LYS420**, **β'ASP423**, β'GLU450, β'ASP537, **β'GLU554**, **β'ASP580**, **β'ARG857**
H451L	βASP474, β'ARG89, **β'LYS420**, **β'ASP423**, β'GLU450, β'ASP537, **β'GLU554**, **β'ASP580**, **β'ARG857**	βASP601, βGLU732, βASP881, **β'LYS420**, β'GLU450, **β'GLU554**, **β'ASP580**
H451N	βASP474, βASP601, βASP881, **β'LYS420**, **β'ASP423**, **β'GLU554**, **β'ASP580**, **β'ARG857**	βGLU732, **β'LYS420**, **β'GLU554**, **β'ARG857**
H451R	**β'LYS420**, **β'ASP423**, β'ASP537, **β'GLU554**, **β'ASP580**	βASP474, **β'LYS420**, **β'GLU554**, **β'ASP580**, **β'ARG857**
H451Y	βASP474, βASP601, β'GLU90, **β'LYS420**, **β'ASP423**, β'GLU450, **β'GLU554**, **β'ASP580**, **β'ARG857**	βASP474, β'ARG89, **β'LYS420**, **β'ASP423**, **β'GLU554**, **β'ARG857**
S456L	βGLU732, β'ARG89, **β'LYS420**, **β'ASP423**, **β'GLU554**, **β'ARG857**	βASP474, β'GLU90, **β'LYS420**, **β'ASP423**, β'GLU450, **β'GLU554**, **β'ASP580**, **β'ARG857**
D551E	βGLU732, β'GLU90, **β'LYS420**, **β'ASP423**, **β'GLU554**, **β'ASP580**, **β'ARG857**	βASP601, **β'LYS420**, **β'ASP423**, **β'GLU554**, **β'ASP580**, **β'ARG857**

Bolded- *residues that indicate commonality among the systems.*

Furthermore, the β'SW2 domain residues β'LYS420 and β'ASP423; and β'GLU554 and β'ASP580, which have been reported to form structurally essential polar contacts (Lane and Darst, 2010b), along with the β'BH domain residue β'ARG857, were identified as consistent among the RIF-bound *wt* standards. The same hot spot residues were also highlighted for the bound *wt*s with the addition of the catalytic

β'NADFDGD motif residue β'ASP537. Regarding the unbound *mt* systems, however, we observed that D441V, H451L, H451N, H451Y, and D551E retained the same residue hot spot statuses relative to the *wt*s. However, the overall hot spot profiles among the *mt*s are different from the *wt*s, as indicated by the observed losses and gains in hot spot statuses. Most interestingly, residue β'ASP537 which was also highlighted as a persistent *CC* hub, lost its hot spot status across all of the *mt*s, which may possibly signify that a local shift in β-β' interactions occurred that may play an active role in reducing RIF's catalytic inhibitory effects.

5.5. Conclusion

Our main findings are as follows: **I)** We were able to identify several residues that may be critical in establishing stable β-β' complex interactions which appeared to be altered in some of the *mt* proteins and might complement our previous results which indicated changes in β and β' functional domain stabilities (see **section 4.2**). **II)** We further identified a number of changes to residue contacts for the mutated residues relative to the *wt*. This further informed the suspected perturbation effects shown through changing H-bond interactions, binding affinities, and the shifts from the active orientation of RIF as indicated by deductions from the RMSD and CoM estimations along with time-based structural snapshots, (see **section 4.3-4.4**). **III)** Finally, apart from identifying residues that may play an important role in establishing the RIF-resistant state in the *mt* proteins, residues involved in the suspected mutation-induced compensatory conformational effects (see **section 4.1**), which has consequences in transcriptional processivity, may have been captured through identified persistent hubs for the investigated proteins.

Chapter 6
Conclusions and final remarks

TB continues to be one of the leading causes of mortality due to a unicellular pathogen. Although RIF is still a potent first-line antitubercular drug, the prevalence of RR-TB, MDR-TB and XDR-TB continues to soar. In other words, as a consequence of the increasing trend of drug resistance, we have observed a decline in favorable prognoses for patients, and increases in side effects due to combination therapies that may include second-line drugs like fluoroquinolones (Roberts, 2018; WHO, 2021). Additionally, global healthcare systems have experienced a serious financial burden, made worse by the recent COVID-19 pandemic (Roberts, 2018; WHO, 2021). It is therefore an existential imperative that presently used drugs such as RIF that are incapable of overcoming certain prevailing mutations are investigated to determine the mechanisms these mutations use in reducing drug efficacy. Hence, the purpose of this thesis was to elucidate the mechanism of RIF resistance at the molecular level using the *wt Mtb*-RNAP multisubunit protein and the I65T, D441V, H451D, H451L, H451R, H451N, H451Y, S456L, and D551E *Mtb*-RNAP proteins in the RIF-unbound and bound state.

Throughout this investigation, MD simulations were run for the RIF-bound and unbound *wt* and *mt Mtb*-RNAPs with bound DNA and RNA molecules through GROMACS 2016.1 following homology modelling through MODELLER. After post-MD preparations, a series of more traditional trajectory analytical approaches consisting of RMSD, RMSF, Rg, CoM, and H-bond occupancy analysis along with the newly developed comparative ED approach (Sheik Amamuddy *et al.*, 2021) were used to conduct a global analysis. This included binding free energy and per residue decomposition calculations between RIF and *Mtb*-RNAP β subunits through the

MM/GBSA approach (Valdés-Tresanco *et al.*, 2021). Moreover, we applied local analysis techniques such as CASM analysis, weighted contact map analysis, and DRN analysis techniques to decipher some of the more subtle changes that may be essential for conferring RIF resistance.

Global analysis revealed that the *mt* proteins may display increased conformational sampling and deviating structural dynamics of their functional domains which may suggest a compensatory mechanism needed to establish RIF resistance, even at the cost of transcriptional processivity (Stefan, Ugur and Garcia, 2018). We further complemented the aforementioned findings by indicating the change in RIF affinity to the RIF-BP through energetic calculations along with various other metrics. 3D structural snapshots indicating RIF's reorientation were also captured within the RIF-bound *mt Mtb*-RNAP RIF-BPs. This was especially interesting to see for the high incidence S456L mutated *Mtb*-RNAP. It was also shown that the non-RRDR β subunit missense mutations such as the I65T could still warrant some caution depending on their association with or effect on the RIF-BP. We also found some signs of perturbational effects on the catalytic nucleic acids, the impact of which may be important in the mechanism of resistance for some *mt Mtb*-RNAPs.

Finally, we supplemented our above-mentioned findings with some local analysis approaches which yielded some insights at the residue level. For instance, we identified residues that may play an important role in essential residue-residue communications needed to maintain transcriptional processivity, along with those needed to allow for effective compensatory motions necessary in reducing the functional efficacy of RIF. We also identified residues that may be particularly important for establishing the interactions needed to attain a stable β-β' complex, interactions which appeared to be altered in most of the *mt* proteins. This was further

informed by the changing interactions between mutated residues and the closely located contacting residues, further detailing some of the more direct effects of the introduced perturbations.

To overcome the limitations of the current study and improve the accuracy of similar future studies, parameterizing bound metal ions should be considered. Furthermore, the addition of multiple *wt* and *mt* MDs of improved *Mtb*-RNAP models should provide a more precise measure of the mutation and/or drug-induced perturbation effects.

This study can have real-world implications in future works, that is, these findings may have important ramifications for determining a means to overcome RIF resistance and potentially in the development of novel therapeutics by exploiting the identified *mt*-specific structural dynamics and conserved communication pathways. Overall, our RIF resistance study gives invaluable insights into the structural dynamical implications of the β subunit I65T, D441V, H451D, H451L, H451R, H451N, H451Y, S456L, and D551E mutated *Mtb*-RNAP proteins at both the global and local perspectives. Our novel study will therefore add to the growing knowledge surrounding RIF resistance and may lead to the discovery of novel drugs with a reduced inclination towards resistance through a similarly applied research protocol.

References

Abraham, M.J. *et al.* (2016) ‘GROMACS User Manual version 2016.1’.

Abrahams, K.A. and Besra, G.S. (2018) ‘Mycobacterial cell wall biosynthesis: A multifaceted antibiotic target’, *Parasitology* [Preprint]. Available at: https://doi.org/10.1017/S0031182016002377.

Acharya, B. *et al.* (2020) ‘Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis’, *Molecular Biology Reports* [Preprint]. Available at: https://doi.org/10.1007/s11033-020-05413-7.

Agnihotry, S. *et al.* (2021) *Protein structure prediction*, *Bioinformatics: Methods and Applications*. INC. Available at: https://doi.org/10.1016/B978-0-323-89775-4.00023-7.

Agoni, C., Ramharack, P. and Soliman, M.E.S. (2018a) ‘Co-inhibition as a strategic therapeutic approach to overcome rifampin resistance in tuberculosis therapy: Atomistic insights’, *Future Medicinal Chemistry* [Preprint]. Available at: https://doi.org/10.4155/fmc-2017-0197.

Agoni, C., Ramharack, P. and Soliman, M.E.S. (2018b) ‘Synergistic Interplay of The Co-administration of Rifampin And Newly Developed Anti-TB Drug: Could It Be a Promising New Line of TB Therapy?’, *Combinatorial Chemistry & High Throughput Screening* [Preprint]. Available at: https://doi.org/10.2174/1386207321666180716093617.

Amadei, A., Linssen, A. and Berendsen, H. (1993) ‘Essential dynamics of proteins’, *Proteins: Structure, Function, and Bioinformatics* [Preprint]. Available at: https://doi.org/10.1002/prot.340170408.

Amusengeri, A., Khan, A. and Tastan Bishop, Ö. (2022) 'The Structural Basis of Mycobacterium tuberculosis RpoB Drug-Resistant Clinical Mutations on Rifampicin Drug Binding', *Molecules* [Preprint]. Available at: https://doi.org/10.3390/molecules27030885.

Andersen, P. and Doherty, T.M. (2005) 'The success and failure of BCG - Implications for a novel tuberculosis vaccine', *Nature Reviews Microbiology* [Preprint]. Available at: https://doi.org/10.1038/nrmicro1211.

Andries, K. *et al.* (2005) 'A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis', *Science* [Preprint]. Available at: https://doi.org/10.1126/science.1106753.

Appelbaum, P.C. and Hunter, P.A. (2000) 'The fluoroquinolone antibacterials: Past, present and future perspectives', *International Journal of Antimicrobial Agents* [Preprint]. Available at: https://doi.org/10.1016/S0924-8579(00)00192-8.

Assefa, Y. *et al.* (2019) 'Screening tools to exclude active pulmonary TB in high TB burden countries: Systematic review and meta-analysis', *International Journal of Tuberculosis and Lung Disease* [Preprint]. Available at: https://doi.org/10.5588/ijtld.18.0547.

Atilgan, A.R., Akan, P. and Baysal, C. (2004) 'Small-World Communication of Residues and Significance for Protein Dynamics', *Biophysical Journal* [Preprint]. Available at: https://doi.org/10.1016/S0006-3495(04)74086-2.

Bahuguna, A. and Rawat, D.S. (2020) 'An overview of new antitubercular drugs, drug candidates, and their targets', *Medicinal Research Reviews* [Preprint]. Available at: https://doi.org/10.1002/med.21602.

Baker, J.J., Dechow, S.J. and Abramovitch, R.B. (2019) ‘Acid Fasting: Modulation of Mycobacterium tuberculosis Metabolism at Acidic pH’, *Trends in Microbiology* [Preprint]. Available at: https://doi.org/10.1016/j.tim.2019.06.005.

Barberis, I. *et al.* (2017) ‘The history of tuberculosis: From the first historical records to the isolation of Koch’s bacillus’, *Journal of Preventive Medicine and Hygiene* [Preprint].

Barozi, V. *et al.* (2022) ‘Deciphering Isoniazid Drug Resistance Mechanisms on Dimeric Mycobacterium tuberculosis KatG via Post-molecular Dynamics Analyses Including Combined Dynamic Residue Network Metrics’, *ACS Omega*, 7(15), pp. 13313–13332. Available at: https://doi.org/10.1021/acsomega.2c01036.

Barozi, V., Edkins, A.L. and Tastan Bishop, Ö. (2022) ‘Evolutionary progression of collective mutations in Omicron sub-lineages towards efficient RBD-hACE2: Allosteric communications between and within viral and human proteins’, *Computational and Structural Biotechnology Journal*, 20, pp. 4562–4578. Available at: https://doi.org/10.1016/j.csbj.2022.08.015.

Basu, R.S. *et al.* (2014) ‘Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme’, *Journal of Biological Chemistry* [Preprint]. Available at: https://doi.org/10.1074/jbc.M114.584037.

Bateman, A. *et al.* (2021) ‘UniProt: the universal protein knowledgebase in 2021’, *Nucleic Acids Research* [Preprint]. Available at: https://doi.org/10.1093/nar/gkaa1100.

Beckstein, S.S. and O. (2017) *AdKGromacsTutorial*. Available at: https://adkgromacstutorial.readthedocs.io/en/latest/analysis/rmsf.html.

Belogurov, G.A. *et al.* (2009) ‘Transcription inactivation through local refolding of the RNA polymerase structure’, *Nature* [Preprint]. Available at: https://doi.org/10.1038/nature07510.

Belogurov, G.A. and Artsimovitch, I. (2019) ‘The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase’, *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1016/j.jmb.2019.05.042.

Benkert, P., Biasini, M. and Schwede, T. (2011) ‘Toward the estimation of the absolute quality of individual protein structure models’, *Bioinformatics* [Preprint]. Available at: https://doi.org/10.1093/bioinformatics/btq662.

BIOVIA, D.S. (2012) *BIOVIA Discovery Studio Visualizer - Dassault Systèmes*, *Computational Molecular Bioscience*.

BIOVIA Discovery Studio Visualizer - Dassault Systèmes (2021). Available at: https://discover.3ds.com/discovery-studio-visualizer-download (Accessed: 22 October 2021).

Bishop, Ö.T. and Kroon, M. (2011) ‘Study of protein complexes via homology modeling, applied to cysteine proteases and their protein inhibitors’, *Journal of Molecular Modeling* [Preprint]. Available at: https://doi.org/10.1007/s00894-011-0990-y.

Bissantz, C., Kuhn, B. and Stahl, M. (2010) ‘A medicinal chemist’s guide to molecular interactions’, *Journal of Medicinal Chemistry* [Preprint]. Available at: https://doi.org/10.1021/jm100112j.

Bodnar, K.A., Serbina, N. V. and Flynn, J.L. (2001) ‘Fate of Mycobacterium tuberculosis within murine dendritic cells’, *Infection and Immunity* [Preprint].

Available at: https://doi.org/10.1128/IAI.69.2.800-809.2001.

Bolon, M.K. (2011) 'The Newer Fluoroquinolones', *Medical Clinics of North America* [Preprint]. Available at: https://doi.org/10.1016/j.mcna.2011.03.006.

Borukhov, S. and Nudler, E. (2008) 'RNA polymerase: the vehicle of transcription', *Trends in Microbiology* [Preprint]. Available at: https://doi.org/10.1016/j.tim.2007.12.006.

Borukhov, S. and Severinov, K. (2002) 'Role of the RNA polymerase sigma subunit in transcription initiation', *Research in Microbiology* [Preprint]. Available at: https://doi.org/10.1016/S0923-2508(02)01368-2.

Boyaci, H., Saecker, R.M. and Campbell, E.A. (2020) 'Transcription initiation in mycobacteria: a biophysical perspective', *Transcription* [Preprint]. Available at: https://doi.org/10.1080/21541264.2019.1707612.

Brickner, S.J. *et al.* (1996) 'Synthesis and antibacterial activity of U-100592 and U-100766, two oxazolidinone antibacterial agents for the potential treatment of multidrug-resistant gram-positive bacterial infections', *Journal of Medicinal Chemistry* [Preprint]. Available at: https://doi.org/10.1021/jm9509556.

Brown, D.K. *et al.* (2017) 'MD-TASK: A software suite for analyzing molecular dynamics trajectories', *Bioinformatics* [Preprint]. Available at: https://doi.org/10.1093/bioinformatics/btx349.

Brueckner, F., Ortiz, J. and Cramer, P. (2009) 'A movie of the RNA polymerase nucleotide addition cycle', *Current Opinion in Structural Biology* [Preprint]. Available at: https://doi.org/10.1016/j.sbi.2009.04.005.

Bruns, H. and Stenger, S. (2014) 'New insights into the interaction of Mycobacterium

tuberculosis and human macrophages', *Future Microbiology* [Preprint]. Available at: https://doi.org/10.2217/fmb.13.164.

Burley, S.K. *et al.* (2019) 'Protein Data Bank: The single global archive for 3D macromolecular structure data', *Nucleic Acids Research* [Preprint]. Available at: https://doi.org/10.1093/nar/gky949.

Bush, N.G. *et al.* (2020) 'Quinolones: Mechanism, lethality and their contributions to antibiotic resistance', *Molecules* [Preprint]. Available at: https://doi.org/10.3390/molecules25235662.

Butler, M.S., Blaskovich, M.A. and Cooper, M.A. (2013) 'Antibiotics in the clinical pipeline in 2013', *Journal of Antibiotics* [Preprint]. Available at: https://doi.org/10.1038/ja.2013.86.

Campbell, E.A. *et al.* (2001) 'Structural mechanism for rifampicin inhibition of bacterial RNA polymerase', *Cell* [Preprint]. Available at: https://doi.org/10.1016/S0092-8674(01)00286-0.

Campbell, E.A. *et al.* (2002) 'Structure of the bacterial RNA polymerase promoter specificity σ subunit', *Molecular Cell* [Preprint]. Available at: https://doi.org/10.1016/S1097-2765(02)00470-7.

Cavasotto, C.N. and Phatak, S.S. (2009) 'Homology modeling in drug discovery: current trends and applications', *Drug Discovery Today* [Preprint]. Available at: https://doi.org/10.1016/j.drudis.2009.04.006.

Chamberlin and J., M. (1976) 'RNA Polymerase—An Overview', *Cold Spring Harbor Monograph Archive* [Preprint].

Chandra, P., Grigsby, S.J. and Philips, J.A. (2022) 'Immune evasion and provocation

by Mycobacterium tuberculosis', *Nature Reviews Microbiology*, 0123456789. Available at: https://doi.org/10.1038/s41579-022-00763-4.

Chauhan, A. *et al.* (2021) 'Comprehensive review on mechanism of action , resistance and evolution of antimycobacterial drugs', *Life Sciences*, 274(November 2020), p. 119301. Available at: https://doi.org/10.1016/j.lfs.2021.119301.

Chen, J., Boyaci, H. and Campbell, E.A. (2021) 'Diverse and unified mechanisms of transcription initiation in bacteria', *Nature Reviews Microbiology* [Preprint]. Available at: https://doi.org/10.1038/s41579-020-00450-2.

Christ, C.D. *et al.* (2009) 'Basic Ingredients of Free Energy Calculations : A Review', *JOURNAL OF COMPUTATIONAL CHEMISTRY* [Preprint].

Ciciliato, I. *et al.* (2004) 'Antibiotics GE23077, novel inhibitors of bacterial RNA polymerase. I. Taxonomy, isolation and characterization', *Journal of Antibiotics* [Preprint]. Available at: https://doi.org/10.7164/antibiotics.57.210.

Cohen, S.B. *et al.* (2018) 'Alveolar Macrophages Provide an Early Mycobacterium tuberculosis Niche and Initiate Dissemination', *Cell Host and Microbe* [Preprint]. Available at: https://doi.org/10.1016/j.chom.2018.08.001.

Comas, I. *et al.* (2013) 'Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans', *Nature Genetics* [Preprint]. Available at: https://doi.org/10.1038/ng.2744.

Corleis, B. *et al.* (2012) 'Escape of Mycobacterium tuberculosis from oxidative killing by neutrophils', *Cellular Microbiology* [Preprint]. Available at: https://doi.org/10.1111/j.1462-5822.2012.01783.x.

Cramer, P. (2002) 'Multisubunit RNA polymerases', *Current Opinion in Structural*

Biology [Preprint]. Available at: https://doi.org/10.1016/S0959-440X(02)00294-4.

Darst, S.A. (2001) 'Bacterial RNA polymerase', *Current Opinion in Structural Biology* [Preprint]. Available at: https://doi.org/10.1016/S0959-440X(00)00185-8.

Dartois, V.A. and Rubin, E.J. (2022) 'Anti-tuberculosis treatment strategies and drug development: challenges and priorities', *Nature Reviews Microbiology 2022*, pp. 1–17. Available at: https://doi.org/10.1038/s41579-022-00731-y.

David, A. *et al.* (2022) 'The AlphaFold Database of Protein Structures: A Biologist's Guide', *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1016/j.jmb.2021.167336.

David, C.C. and Jacobs, D.J. (2014) *Chapter 11 the Essential Dynamics of Proteins*. Available at: https://doi.org/10.1007/978-1-62703-658-0.

Davidson, L.A. and Takayama, K. (1979) 'Isoniazid inhibition of the synthesis of monosaturated long-chain fatty acids in Mycobacterium tuberculosis H37Ra', *Antimicrobial Agents and Chemotherapy* [Preprint]. Available at: https://doi.org/10.1128/AAC.16.1.104.

Delano, W.L. (2002) 'The PyMOL Molecular Graphics System', *CCP4 Newsletter on protein crystallography* [Preprint].

DeLano, W.L. (2002) 'Pymol: An open-source molecular graphics tool', *CCP4 Newsletter on protein crystallography* [Preprint].

Dill, K.A. and MacCallum, J.L. (2012) 'The protein-folding problem, 50 years on', *Science* [Preprint]. Available at: https://doi.org/10.1126/science.1219021.

Dobson, A.P. and Robin Carper, E. (1996) 'Infectious diseases and human population history', *BioScience* [Preprint]. Available at:

https://doi.org/10.2307/1312814.

Donald, P.R. *et al.* (2018) ‘Droplets, dust and Guinea pigs: An historical review of tuberculosis transmission research, 1878-1940’, *International Journal of Tuberculosis and Lung Disease* [Preprint]. Available at: https://doi.org/10.5588/ijtld.18.0173.

Drennan, A. *et al.* (2012) ‘Key roles of the downstream mobile jaw of escherichia coli RNA polymerase in transcription initiation’, *Biochemistry* [Preprint]. Available at: https://doi.org/10.1021/bi301260u.

Drobniewski, F.A. and Wilson, S.M. (1998) ‘The rapid diagnosis of isoniazid and rifampicin resistance in Mycobacterium tuberculosis - A molecular story’, *Journal of Medical Microbiology* [Preprint]. Available at: https://doi.org/10.1099/00222615-47-3-189.

Duan, L., Liu, X. and Zhang, J.Z.H. (2016) ‘Interaction entropy: A new paradigm for highly efficient and reliable computation of protein-ligand binding free energy’, *Journal of the American Chemical Society* [Preprint]. Available at: https://doi.org/10.1021/jacs.6b02682.

Durrant, J.D. and McCammon, J.A. (2011) ‘Molecular dynamics simulations and drug discovery’, *BMC Biology* [Preprint]. Available at: https://doi.org/10.1186/1741-7007-9-71.

Ebright, R.H. (2000) ‘RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II’, *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1006/jmbi.2000.4309.

Eddy, S.R. (1998) ‘Profile hidden Markov models’, *Bioinformatics* [Preprint].

Available at: https://doi.org/10.1093/bioinformatics/14.9.755.

Ehrt, S. and Schnappinger, D. (2009) ‘Mycobacterial survival strategies in the phagosome: Defence against host stresses’, *Cellular Microbiology* [Preprint]. Available at: https://doi.org/10.1111/j.1462-5822.2009.01335.x.

Eisenhaber, F. *et al.* (1995) ‘The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies’, *Journal of Computational Chemistry* [Preprint]. Available at: https://doi.org/10.1002/jcc.540160303.

Emmerson, A.M. and Jones, A.M. (2003) ‘The quinolones: Decades of development and use’, *Journal of Antimicrobial Chemotherapy* [Preprint]. Available at: https://doi.org/10.1093/jac/dkg208.

Feig, M. *et al.* (2004) ‘Performance Comparison of Generalized Born and Poisson Methods in the Calculation of Electrostatic Solvation Energies for Protein Structures’, *Journal of Computational Chemistry* [Preprint]. Available at: https://doi.org/10.1002/jcc.10378.

Feklistov, A. *et al.* (2017) ‘RNA polymerase motions during promoter melting’, *Science* [Preprint]. Available at: https://doi.org/10.1126/science.aam7858.

Flandrois, J.P., Lina, G. and Dumitrescu, O. (2014) ‘MUBII-TB-DB: A database of mutations associated with antibiotic resistance in Mycobacterium tuberculosis’, *BMC Bioinformatics* [Preprint]. Available at: https://doi.org/10.1186/1471-2105-15-107.

Flentie, K., Garner, A.L. and Stallings, C.L. (2016) ‘Mycobacterium tuberculosis transcription machinery: Ready to respond to host attacks’, *Journal of Bacteriology* [Preprint]. Available at: https://doi.org/10.1128/JB.00935-15.

Flynn, J.A.L. (2006) ‘Lessons from experimental Mycobacterium tuberculosis infections’, *Microbes and Infection* [Preprint]. Available at: https://doi.org/10.1016/j.micinf.2005.10.033.

Forrest, G.N. and Tamura, K. (2010) ‘Rifampin combination therapy for nonmycobacterial infections’, *Clinical Microbiology Reviews* [Preprint]. Available at: https://doi.org/10.1128/CMR.00034-09.

Foster, I. (2011) ‘Globus online: Accelerating and democratizing science through cloud-based services’, *IEEE Internet Computing* [Preprint]. Available at: https://doi.org/10.1109/MIC.2011.64.

Foutch, D., Pham, B. and Shen, T. (2021) ‘Protein conformational switch discerned via network centrality properties’, *Computational and Structural Biotechnology Journal*, 19, pp. 3599–3608. Available at: https://doi.org/10.1016/j.csbj.2021.06.004.

Fox, W., Ellard, G.A. and Mitchison, D.A. (1999) ‘Studies on the treatment of tuberculosis undertaken by the British Medical Research Council Tuberculosis Units, 1946-1986, with relevant subsequent publications’, *International Journal of Tuberculosis and Lung Disease* [Preprint].

Genheden, S. *et al.* (2012) ‘The normal-mode entropy in the MM/GBSA method: Effect of system truncation, buffer region, and dielectric constant’, *Journal of Chemical Information and Modeling* [Preprint]. Available at: https://doi.org/10.1021/ci3001919.

Genheden, S. and Ryde, U. (2015) ‘The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities’, *Expert Opinion on Drug Discovery* [Preprint]. Available at: https://doi.org/10.1517/17460441.2015.1032936.

Ghanavi, J. *et al.* (2021) ‘The role of interferon-gamma and interferon-gamma receptor in tuberculosis and nontuberculous mycobacterial infections’, *International Journal of Mycobacteriology* [Preprint]. Available at: https://doi.org/10.4103/ijmy.ijmy_186_21.

Global TB report (2020) *Global tuberculosis report 2020*, *World Health Organization*.

Gnatt, A.L. *et al.* (2001) ‘Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution’, *Science* [Preprint]. Available at: https://doi.org/10.1126/science.1059495.

Goldbeck, J. (2013) *Jennifer Goldbeck: Analyzing the Social Web*.

Golden, M.P. and Vikram, H.R. (2005) ‘Extrapulmonary tuberculosis: An overview’, *American Family Physician* [Preprint]. Available at: https://doi.org/10.5772/intechopen.81322.

Goldstein, B.P. (2014) ‘Resistance to rifampicin: A review’, *Journal of Antibiotics* [Preprint]. Available at: https://doi.org/10.1038/ja.2014.107.

González, M.A. (2011) ‘Force fields and molecular dynamics simulations’, *École thématique de la Société Française de la Neutronique* [Preprint]. Available at: https://doi.org/10.1051/sfn/201112009.

Gostling, N.J. (2002) ‘From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design’, *Heredity* [Preprint]. Available at: https://doi.org/10.1038/sj.hdy.6800154.

Gradmann, C. (2001) ‘Robert Koch and the pressures of scientific research: tuberculosis and tuberculin.’, *Medical history* [Preprint]. Available at: https://doi.org/10.1017/s0025727300000028.

Grobbelaar, M. *et al.* (2019) ‘Evolution of rifampicin treatment for tuberculosis’, *Infection, Genetics and Evolution* [Preprint]. Available at: https://doi.org/10.1016/j.meegid.2019.103937.

Hanneman, R. a and Riddle, M. (2005) ‘Introduction to Social Network Methods’, *Riverside, CA: University of California, Riverside. On-line textbook* [Preprint]. Available at: https://doi.org/10.1016/j.socnet.2006.08.002.

Hardin, G. and Diamond, J. (1997) ‘Guns, Germs, and Steel: The Fates of Human Societies.’, *Population and Development Review* [Preprint]. Available at: https://doi.org/10.2307/2137390.

Hayman, J. (1984) ‘MYCOBACTERIUM ULCERANS: AN INFECTION FROM JURASSIC TIME?’, *The Lancet* [Preprint]. Available at: https://doi.org/10.1016/S0140-6736(84)91110-3.

Hayward, S. and De Groot, B.L. (2008) ‘Normal modes and essential dynamics’, *Methods in Molecular Biology* [Preprint]. Available at: https://doi.org/10.1007/978-1-59745-177-2_5.

Henry Boom, W., Schaible, U.E. and Achkar, J.M. (2021) ‘The knowns and unknowns of latent Mycobacterium tuberculosis infection’, *Journal of Clinical Investigation* [Preprint]. Available at: https://doi.org/10.1172/JCI136222.

Hoagland, D.T. *et al.* (2016) ‘New agents for the treatment of drug-resistant Mycobacterium tuberculosis’, *Advanced Drug Delivery Reviews* [Preprint]. Available at: https://doi.org/10.1016/j.addr.2016.04.026.

Hollingsworth, S.A. and Dror, R.O. (2018) ‘Molecular Dynamics Simulation for All’, *Neuron* [Preprint]. Available at: https://doi.org/10.1016/j.neuron.2018.08.011.

Hu, Y. and Liu, B. (2022) ‘Roles of zinc-binding domain of bacterial RNA polymerase in transcription’, *Trends in Biochemical Sciences*, 47(8), pp. 710–724. Available at: https://doi.org/10.1016/j.tibs.2022.03.007.

Huber, W. *et al.* (2007) ‘Graphs in molecular biology’, *BMC Bioinformatics* [Preprint]. Available at: https://doi.org/10.1186/1471-2105-8-S6-S8.

Humphrey, W., Dalke, A. and Schulten, K. (1996) ‘VMD: Visual molecular dynamics’, *Journal of Molecular Graphics* [Preprint]. Available at: https://doi.org/10.1016/0263-7855(96)00018-5.

Hunter, J.D. (2007) ‘Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering’, *Computing In Science & Engineering* [Preprint].

Hussell, T. and Bell, T.J. (2014) ‘Alveolar macrophages: Plasticity in a tissue-specific context’, *Nature Reviews Immunology* [Preprint]. Available at: https://doi.org/10.1038/nri3600.

Janin, J., Bahadur, R.P. and Chakrabarti, P. (2008) ‘Protein-protein interaction and quaternary structure’, *Quarterly Reviews of Biophysics* [Preprint]. Available at: https://doi.org/10.1017/S0033583508004708.

Jenkins, H.E., Zignol, M. and Cohen, T. (2011) ‘Quantifying the burden and trends of isoniazid resistant tuberculosis, 1994-2009’, *PLoS ONE* [Preprint]. Available at: https://doi.org/10.1371/journal.pone.0022927.

Johnsson, K. and Schultz, P.G. (1994) ‘Mechanistic Studies of the Oxidation of Isoniazid by the Catalase Peroxidase from Mycobacterium tuberculosis’, *Journal of the American Chemical Society* [Preprint]. Available at: https://doi.org/10.1021/ja00095a063.

Karabekmez, M.E. and Kirdar, B. (2016) ‘A novel topological centrality measure capturing biologically important proteins’, *Molecular BioSystems* [Preprint]. Available at: https://doi.org/10.1039/c5mb00732a.

Karplus, M. and McCammon, J.A. (2002) ‘Molecular dynamics simulations of biomolecules’, *Nature Structural Biology* [Preprint]. Available at: https://doi.org/10.1038/nsb0902-646.

Khanna, A., Saha, R. and Ahmad, N. (2023) ‘National TB elimination programme - What has changed’, *Indian Journal of Medical Microbiology* [Preprint]. Available at: https://doi.org/10.1016/j.ijmmb.2022.10.008.

King, R.A. *et al.* (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* [Preprint]. Available at: https://doi.org/10.1016/j.jmb.2004.07.072.

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 and Infections* [Preprint]. Available at: https://doi.org/10.1038/emi.2014.17.

Koschützki, D. and Schreiber, F. (2008) ‘Centrality analysis methods for biological networks and their application to gene regulatory networks’, *Gene Regulation and Systems Biology* [Preprint]. Available at: https://doi.org/10.4137/grsb.s702.

Koul, A. *et al.* (2007) ‘Diarylquinolines target subunit c of mycobacterial ATP synthase’, *Nature Chemical Biology* [Preprint]. Available at: https://doi.org/10.1038/nchembio884.

Koul, A. *et al.* (2014) ‘Delayed bactericidal response of Mycobacterium tuberculosis to bedaquiline involves remodelling of bacterial metabolism’, *Nature communications* [Preprint]. Available at: https://doi.org/10.1038/ncomms4369.

Krishnan, V. and Rupp, B. (2012) ‘ Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy ’, in *eLS*. Available at: https://doi.org/10.1002/9780470015902.a0002716.pub2.

Kumar, S. and Jena, L. (2014) ‘Understanding Rifampicin Resistance in Tuberculosis through a Computational Approach’, *Genomics & Informatics* [Preprint]. Available at: https://doi.org/10.5808/gi.2014.12.4.276.

Kurkela, J. *et al.* (2021) ‘Revealing secrets of the enigmatic omega subunit of bacterial RNA polymerase’, *Molecular Microbiology* [Preprint]. Available at: https://doi.org/10.1111/mmi.14603.

Kuznedelov, K. *et al.* (2002) ‘A role for interaction of the RNA polymerase flap domain with the σ subunit in promoter recognition’, *Science* [Preprint]. Available at: https://doi.org/10.1126/science.1066303.

Lane, W.J. and Darst, S.A. (2010a) ‘Molecular Evolution of Multisubunit RNA Polymerases: Sequence Analysis’, *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1016/j.jmb.2009.10.062.

Lane, W.J. and Darst, S.A. (2010b) ‘Molecular Evolution of Multisubunit RNA Polymerases: Structural Analysis’, *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1016/j.jmb.2009.10.063.

Larsson, P. *et al.* (2008) ‘Using multiple templates to improve quality of homology models in automated homology modeling’, *Protein Science* [Preprint]. Available at:

https://doi.org/10.1110/ps.073344908.

Laskowski, R.A. *et al.* (1993) 'PROCHECK: a program to check the stereochemical quality of protein structures', *Journal of Applied Crystallography* [Preprint]. Available at: https://doi.org/10.1107/s0021889892009944.

Lee, C., Su, B.-H. and Tseng, Y.J. (2022) 'Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors', *Briefings in Bioinformatics*, 23(5), pp. 1–7. Available at: https://doi.org/10.1093/bib/bbac308.

Lee, J. *et al.* (2020) 'CD11cHi monocyte-derived macrophages are a major cellular compartment infected by Mycobacterium tuberculosis', *PLoS Pathogens* [Preprint]. Available at: https://doi.org/10.1371/journal.ppat.1008621.

Lee, R.E. *et al.* (2003) 'Combinatorial lead optimization of [1,2]-diamines based on ethambutol as potential antituberculosis preclinical candidates', *Journal of Combinatorial Chemistry* [Preprint]. Available at: https://doi.org/10.1021/cc020071p.

Lesk, A.M. and Chothia, C.H. (1986) 'The response of protein structures to amino-acid sequence changes', *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* [Preprint]. Available at: https://doi.org/10.1098/rsta.1986.0044.

Levine, C., Hiasa, H. and Marians, K.J. (1998) 'DNA gyrase and topoisomerase IV: Biochemical activities, physiological roles during chromosome replication, and drug sensitivities', *Biochimica et Biophysica Acta - Gene Structure and Expression* [Preprint]. Available at: https://doi.org/10.1016/S0167-4781(98)00126-2.

Lilic, M. *et al.* (2020) 'The antibiotic sorangicin A inhibits promoter DNA unwinding in

a Mycobacterium tuberculosis rifampicin-resistant RNA polymerase', *Proceedings of the National Academy of Sciences of the United States of America* [Preprint]. Available at: https://doi.org/10.1073/pnas.2013706117.

Lin, W. *et al.* (2017) 'Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition', *Molecular Cell* [Preprint]. Available at: https://doi.org/10.1016/j.molcel.2017.03.001.

Lindorff-Larsen, K. *et al.* (2010) 'Improved side-chain torsion potentials for the Amber ff99SB protein force field', *Proteins: Structure, Function and Bioinformatics* [Preprint]. Available at: https://doi.org/10.1002/prot.22711.

Lindorff-Larsen, K. *et al.* (2012) 'Systematic validation of protein force fields against experimental data', *PLoS ONE* [Preprint]. Available at: https://doi.org/10.1371/journal.pone.0032131.

Liu, L. *et al.* (2012) 'Comparison of next-generation sequencing systems', *Journal of Biomedicine and Biotechnology* [Preprint]. Available at: https://doi.org/10.1155/2012/251364.

Liu, Y. *et al.* (2020) 'Active or latent tuberculosis increases susceptibility to COVID-19 and disease severity', *medRxiv* [Preprint].

Livermore, D.M. (2003) 'Linezolid in vitro: Mechanism and antibacterial spectrum', *Journal of Antimicrobial Chemotherapy* [Preprint]. Available at: https://doi.org/10.1093/jac/dkg249.

M Cristina, G. *et al.* (2005) 'Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis', *PLoS Pathogens* [Preprint]. Available at: https://doi.org/10.1371/journal.ppat.0010005.

Mao, C. *et al.* (2018) ‘Association of ω with the Cterminal region of the β’ subunit is essential for assembly of RNA polymerase in Mycobacterium tuberculosis’, *Journal of Bacteriology* [Preprint]. Available at: https://doi.org/10.1128/JB.00159-18.

Martin, C. *et al.* (2020) ‘Update on TB vaccine pipeline’, *Applied Sciences (Switzerland)* [Preprint]. Available at: https://doi.org/10.3390/app10072632.

Mazumder, A. *et al.* (2020) ‘Closing and opening of the RNA polymerase trigger loop’, *Proceedings of the National Academy of Sciences of the United States of America* [Preprint]. Available at: https://doi.org/10.1073/pnas.1920427117.

McKinney, W. (2011) ‘pandas: a Foundational Python Library for Data Analysis and Statistics’, *Python for High Performance and Scientific Computing* [Preprint].

Meji, Y.X., Nudler, E. and Bustamante, C. (2015) ‘Trigger loop folding determines transcription rate of escherichia coli ’s RNA polymerase’, *Proceedings of the National Academy of Sciences of the United States of America* [Preprint]. Available at: https://doi.org/10.1073/pnas.1421067112.

Meyer, P.A. *et al.* (2009) ‘Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data’, *Journal of Biological Chemistry* [Preprint]. Available at: https://doi.org/10.1074/jbc.M809199200.

Miotto, P. *et al.* (2017) ‘A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis’, *European Respiratory Journal* [Preprint]. Available at: https://doi.org/10.1183/13993003.01354-2017.

Mishanina, T. V. *et al.* (2017) ‘Trigger loop of RNA polymerase is a positional not acid-base, catalyst for both transcription and proofreading’, *Proceedings of the*

National Academy of Sciences of the United States of America [Preprint]. Available at: https://doi.org/10.1073/pnas.1702383114.

Mitusińska, K., Skalski, T. and Góra, A. (2020) 'Simple selection procedure to distinguish between static and flexible loops', *International Journal of Molecular Sciences* [Preprint]. Available at: https://doi.org/10.3390/ijms21072293.

Molodtsov, V. *et al.* (2015) 'X-ray Crystal Structures of Escherichia coli RNA Polymerase with Switch Region Binding Inhibitors Enable Rational Design of Squaramides with an Improved Fraction Unbound to Human Plasma Protein', *Journal of Medicinal Chemistry* [Preprint]. Available at: https://doi.org/10.1021/acs.jmedchem.5b00050.

Molodtsov, V. *et al.* (2017) 'Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in Mycobacterium tuberculosis', *Molecular Microbiology* [Preprint]. Available at: https://doi.org/10.1111/mmi.13606.

Monama, M.Z., Olotu, F. and Tastan Bishop, Ö. (2023) 'Investigation of Multi - Subunit Mycobacterium tuberculosis DNA - Directed RNA Polymerase and Its Rifampicin Resistant Mutants'.

Mooney, R.A., Darst, S.A. and Landick, R. (2005) 'Sigma and RNA polymerase: An on-again, off-again relationship?', *Molecular Cell* [Preprint]. Available at: https://doi.org/10.1016/j.molcel.2005.10.015.

Moreira, I.S. *et al.* (2017) 'SpotOn: High Accuracy Identification of Protein-Protein Interface Hot-Spots', *Scientific Reports* [Preprint]. Available at: https://doi.org/10.1038/s41598-017-08321-2.

Morichaud, Z. *et al.* (2023) ‘Structural basis of the mycobacterial stress-response RNA polymerase auto-inhibition via oligomerization’, *Nature Communications*, 14(1), pp. 1–13. Available at: https://doi.org/10.1038/s41467-023-36113-y.

Muschenheim, C. (1952) ‘Hydrazine derivatives of isonicotinic acid in the treatment of tuberculosis.’, *New York state journal of medicine* [Preprint].

Naji, S. *et al.* (2008) ‘Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement’, *Nucleic Acids Research* [Preprint]. Available at: https://doi.org/10.1093/nar/gkm1086.

Narayanan, A. *et al.* (2018) ‘Cryo-EM structure of Escherichia coli 70 RNA polymerase and promoter DNA complex revealed a role of non-conserved region during the open complex formation’, *Journal of Biological Chemistry*, 293(19), pp. 7367–7375. Available at: https://doi.org/10.1074/jbc.RA118.002161.

Naryshkina, T., Kuznedelov, K. and Severinov, K. (2006) ‘The Role of the Largest RNA Polymerase Subunit Lid Element in Preventing the Formation of Extended RNA-DNA Hybrid’, *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1016/j.jmb.2006.05.034.

Notredame, C. and Higgins, D.G. (1996) ‘SAGA: Sequence alignment by genetic algorithm’, *Nucleic Acids Research* [Preprint]. Available at: https://doi.org/10.1093/nar/24.8.1515.

Nusrath Unissa, A. *et al.* (2016) ‘Insights into RpoB clinical mutants in mediating rifampicin resistance in Mycobacterium tuberculosis’, *Journal of Molecular Graphics and Modelling* [Preprint]. Available at: https://doi.org/10.1016/j.jmgm.2016.04.005.

O’Garra, A. *et al.* (2013) ‘The immune response in tuberculosis’, *Annual Review of*

Immunology [Preprint]. Available at: https://doi.org/10.1146/annurev-immunol-032712-095939.

Okeke, C.J. *et al.* (2021) 'Allosteric pockets and dynamic residue network hubs of falcipain 2 in mutations including those linked to artemisinin resistance', *Computational and Structural Biotechnology Journal* [Preprint]. Available at: https://doi.org/10.1016/j.csbj.2021.10.011.

Pagán, A.J. and Ramakrishnan, L. (2018) 'The Formation and Function of Granulomas', *Annual Review of Immunology* [Preprint]. Available at: https://doi.org/10.1146/annurev-immunol-032712-100022.

Paget, M.S.B. and Helmann, J.D. (2003) 'The sigma70 family of sigma factors.', *Genome biology* [Preprint].

Paulson, T. (2013) 'Epidemiology: A mortal foe', *Nature* [Preprint]. Available at: https://doi.org/10.1038/502s2a.

Pavlopoulos, G.A. *et al.* (2011) 'Using graph theory to analyze biological networks', *BioData Mining* [Preprint]. Available at: https://doi.org/10.1186/1756-0381-4-10.

Penkler, D.L., Atilgan, C. and Tastan Bishop, Ö. (2018) 'Allosteric Modulation of Human Hsp90α Conformational Dynamics', *Journal of Chemical Information and Modeling* [Preprint]. Available at: https://doi.org/10.1021/acs.jcim.7b00630.

Piccaro, G. *et al.* (2013) 'Activities of drug combinations against mycobacterium tuberculosis grown in aerobic and hypoxic acidic conditions', *Antimicrobial Agents and Chemotherapy* [Preprint]. Available at: https://doi.org/10.1128/AAC.02154-12.

Place, V.A. and Thomas, J.P. (1963) 'Clinical pharmacology of ethambutol.', *The American review of respiratory disease* [Preprint].

Poce, G. *et al.* (2014) ‘SAR analysis of new anti-TB drugs currently in pre-clinical and clinical development’, *European Journal of Medicinal Chemistry* [Preprint]. Available at: https://doi.org/10.1016/j.ejmech.2014.08.066.

Roberts, J.R. (2018) ‘InFocus Fluoroquinolone Side Effects Just Got Scarier’, *Emergency Medicine News*, 40(10), pp. 26–27. Available at: http://journals.lww.com/00132981-201810000-00012.

Roberts, J.W. (2019) ‘Mechanisms of Bacterial Transcription Termination’, *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1016/j.jmb.2019.04.003.

Rodrigue, S. *et al.* (2006) ‘The σ factors of Mycobacterium tuberculosis’, *FEMS Microbiology Reviews* [Preprint]. Available at: https://doi.org/10.1111/j.1574-6976.2006.00040.x.

Roe, D.R. and Cheatham, T.E. (2013) ‘PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data’, *Journal of Chemical Theory and Computation* [Preprint]. Available at: https://doi.org/10.1021/ct400341p.

Rosignoli, S. and Paiardini, A. (2022) ‘Boosting the Full Potential of PyMOL with Structural Biology Plugins’, *Biomolecules* [Preprint]. Available at: https://doi.org/10.3390/biom12121764.

Rothchild, A.C. *et al.* (2019) ‘Alveolar macrophages generate a noncanonical NRF2-driven transcriptional response to Mycobacterium tuberculosis in vivo’, *Science Immunology* [Preprint]. Available at: https://doi.org/10.1126/sciimmunol.aaw6693.

Ruhnau, B. (2000) ‘Eigenvector-centrality - a node-centrality’, *Social Networks* [Preprint]. Available at: https://doi.org/10.1016/S0378-8733(00)00031-9.

Russell, D.G. (2007) ‘Who puts the tubercle in tuberculosis?’, *Nature Reviews*

Microbiology [Preprint]. Available at: https://doi.org/10.1038/nrmicro1538.

Sacchettini, J.C. and Blanchard, J.S. (1996) 'The structure and function of the isoniazid target in M. tuberculosis', in *Research in Microbiology*. Available at: https://doi.org/10.1016/0923-2508(96)80201-4.

Sachdeva, P. *et al.* (2010) 'The sigma factors of Mycobacterium tuberculosis: Regulation of the regulators', *FEBS Journal* [Preprint]. Available at: https://doi.org/10.1111/j.1742-4658.2009.07479.x.

Saecker, R.M., Record, M.T. and Dehaseth, P.L. (2011) 'Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis', *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1016/j.jmb.2011.01.018.

Sala, C. *et al.* (2010) 'Simple model for testing drugs against nonreplicating Mycobacterium tuberculosis', *Antimicrobial Agents and Chemotherapy* [Preprint]. Available at: https://doi.org/10.1128/AAC.00821-10.

Šali, A. and Blundell, T.L. (1993) 'Comparative protein modelling by satisfaction of spatial restraints', *Journal of Molecular Biology* [Preprint]. Available at: https://doi.org/10.1006/jmbi.1993.1626.

Sandgren, A. *et al.* (2009) 'Tuberculosis drug resistance mutation database', *PLoS Medicine* [Preprint]. Available at: https://doi.org/10.1371/journal.pmed.1000002.

Santucci, P. *et al.* (2021) 'Intracellular localisation of Mycobacterium tuberculosis affects efficacy of the antibiotic pyrazinamide', *Nature Communications* [Preprint]. Available at: https://doi.org/10.1038/s41467-021-24127-3.

Sanyanga, A.T., Nizami, B. and Tastan Bishop, Ö. (2019) 'Mechanism of action of

non-synonymous single nucleotide variations associated with α-carbonic anhydrase II deficiency', *Molecules* [Preprint]. Available at: https://doi.org/10.3390/molecules24213987.

Sanyanga, T.A. and Tastan Bishop, Ö. (2020) 'Structural characterization of carbonic anhydrase viii and effects of missense single nucleotide variations to protein structure and function', *International Journal of Molecular Sciences* [Preprint]. Available at: https://doi.org/10.3390/ijms21082764.

Sarathy, J.P., Gruber, G. and Dick, T. (2019) 'Re-understanding the mechanisms of action of the anti-mycobacterial drug bedaquiline', *Antibiotics* [Preprint]. Available at: https://doi.org/10.3390/antibiotics8040261.

Savojardo, C. *et al.* (2021) 'Solvent Accessibility of Residues Undergoing Pathogenic Variations in Humans: From Protein Structures to Protein Sequences', *Frontiers in Molecular Biosciences* [Preprint]. Available at: https://doi.org/10.3389/fmolb.2020.626363.

Scorpio, A. and Zhang, Y. (1996) 'Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus', *Nature Medicine* [Preprint]. Available at: https://doi.org/10.1038/nm0696-662.

Sensi, P. (1983) 'History of the development of rifampin', *Reviews of Infectious Diseases* [Preprint]. Available at: https://doi.org/10.1093/clinids/5.Supplement_3.S402.

Shea, J. *et al.* (2021) 'Low-Level Rifampin Resistance and rpoB Mutations in Mycobacterium tuberculosis: An Analysis of Whole-Genome Sequencing and Drug Susceptibility Test Data in New York', *Journal of Clinical Microbiology* [Preprint].

Available at: https://doi.org/10.1128/JCM.01885-20.

Sheik Amamuddy, O., Glenister, M., *et al.* (2021) 'MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics', *Computational and Structural Biotechnology Journal* [Preprint]. Available at: https://doi.org/10.1016/j.csbj.2021.08.043.

Sheik Amamuddy, O., Baoteng, R.A., *et al.* (2021) 'Novel dynamic residue network analysis approaches to study homodimeric allosteric modulation in SARS-CoV-2 Mpro and in its evolutionary mutations', *Computational and Structural Biotechnology Journal*, 19, pp. 6431–6455. Available at: https://chemrxiv.org/engage/chemrxiv/article-details/614acc296fc3a869b9a8e98e.

Sheik Amamuddy, O., Verkhivker, G. and Tastan Bishop, Ö. (2020) 'Impact of early pandemic stage mutations on molecular dynamics of SARS-CoV-2 MPro', *Journal of Chemical Information and Modeling* [Preprint]. Available at: https://doi.org/10.1021/acs.jcim.0c00634.

Shumway, M., Cochrane, G. and Sugawara, H. (2009) 'Archiving next generation sequencing data', *Nucleic Acids Research*, 38(SUPPL.1), pp. 2009–2010. Available at: https://doi.org/10.1093/nar/gkp1078.

Singh, A. *et al.* (2017) 'Mechanistic Principles Behind Molecular Mechanism of Rifampicin Resistance in Mutant RNA Polymerase Beta Subunit of Mycobacterium tuberculosis', *Journal of Cellular Biochemistry* [Preprint]. Available at: https://doi.org/10.1002/jcb.26124.

Siu, G.K.H. *et al.* (2011) 'Mutations outside the rifampicin resistance-determining region associated with rifampicin resistance in Mycobacterium tuberculosis', *Journal of Antimicrobial Chemotherapy* [Preprint]. Available at:

https://doi.org/10.1093/jac/dkq519.

Del Sol, A. *et al.* (2006) ‘Residues crucial for maintaining short paths in network communication mediate signaling in proteins’, *Molecular Systems Biology* [Preprint]. Available at: https://doi.org/10.1038/msb4100063.

Sousa Da Silva, A.W. and Vranken, W.F. (2012) ‘ACPYPE - AnteChamber PYthon Parser interfacE’, *BMC Research Notes* [Preprint]. Available at: https://doi.org/10.1186/1756-0500-5-367.

Van Der Spoel, D. *et al.* (2005) ‘GROMACS: Fast, flexible, and free’, *Journal of Computational Chemistry* [Preprint]. Available at: https://doi.org/10.1002/jcc.20291.

Srivastava, G. *et al.* (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* [Preprint]. Available at: https://doi.org/10.1016/j.ijbiomac.2018.06.184.

Stallings, C.L. and Glickman, M.S. (2010) ‘Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence’, *Microbes and Infection* [Preprint]. Available at: https://doi.org/10.1016/j.micinf.2010.07.014.

Stefan, M.A., Ugur, F.S. and Garcia, G.A. (2018) ‘Source of the fitness defect in rifamycin-resistant mycobacterium tuberculosis RNA polymerase and the mechanism of compensation by mutations in the = subunit’, *Antimicrobial Agents and Chemotherapy* [Preprint]. Available at: https://doi.org/10.1128/AAC.00164-18.

Steingart, K.R. *et al.* (2014) ‘Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults’, *Cochrane Database of Systematic Reviews*

[Preprint]. Available at: https://doi.org/10.1002/14651858.CD009593.pub3.

Stephanie, F., Tambunan, U.S.F. and Siahaan, T.J. (2022) ‘M. tuberculosis Transcription Machinery: A Review on the Mycobacterial RNA Polymerase and Drug Discovery Efforts’, *Life*, 12(11), p. 1774. Available at: https://doi.org/10.3390/life12111774.

Sun, H. *et al.* (2018) ‘Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches’, *Physical Chemistry Chemical Physics* [Preprint]. Available at: https://doi.org/10.1039/c7cp07623a.

Sun, Q. *et al.* (2020) ‘The molecular basis of pyrazinamide activity on Mycobacterium tuberculosis PanD’, *Nature Communications* [Preprint]. Available at: https://doi.org/10.1038/s41467-019-14238-3.

Surie, D. *et al.* (2019) ‘Policies, practices and barriers to implementing tuberculosis preventive treatment-35 countries, 2017’, *International Journal of Tuberculosis and Lung Disease* [Preprint]. Available at: https://doi.org/10.5588/ijtld.19.0018.

Sutherland, C. and Murakami, K.S. (2018) ‘ An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase ’, *EcoSal Plus* [Preprint]. Available at: https://doi.org/10.1128/ecosalplus.esp-0004-2018.

Takayama, K., Armstrong, E.L. and David, H.L. (1974) ‘Restoration of mycolate synthetase activity in Mycobacterium tuberculosis exposed to isoniazid’, *AMER.REV.RESP.DIS.* [Preprint].

Takayama, K., Wang, L. and David, H.L. (1972) ‘Effect of isoniazid on the in vivo mycolic acid synthesis, cell growth, and viability of Mycobacterium tuberculosis.’,

Antimicrobial agents and chemotherapy [Preprint]. Available at: https://doi.org/10.1128/AAC.2.1.29.

Tan, Y.Z. *et al.* (2020) 'Cryo-EM structure of arabinosyltransferase EmbB from Mycobacterium smegmatis', *Nature Communications* [Preprint]. Available at: https://doi.org/10.1038/s41467-020-17202-8.

Tastan Bishop, Ö., Musyoka, T.M. and Barozi, V. (2022) 'Allostery and Missense Mutations as Intermittently Linked Promising Aspects of Modern Computational Drug Discovery', *Journal of Molecular Biology*, 434(17), p. 167610. Available at: https://doi.org/10.1016/j.jmb.2022.167610.

Thompson, J.D. *et al.* (2011) 'A comprehensive benchmark study of multiple sequence alignment methods: Current challenges and future perspectives', *PLoS ONE* [Preprint]. Available at: https://doi.org/10.1371/journal.pone.0018093.

Toulokhonov, I. and Landick, R. (2003) 'The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination', *Molecular Cell* [Preprint]. Available at: https://doi.org/10.1016/S1097-2765(03)00439-8.

Travers, A.A. and Burgess, R.R. (1969) 'Cyclic Re-use of the RNA polymerase sigma factor', *Nature* [Preprint]. Available at: https://doi.org/10.1038/222537a0.

Tuske, S. *et al.* (2005) 'Inhibition of bacterial RNA polymerase by streptolydigin: Stabilization of a straight-bridge-helix active-center conformation', *Cell* [Preprint]. Available at: https://doi.org/10.1016/j.cell.2005.07.017.

Unissa, A.N. *et al.* (2016) 'Overview on mechanisms of isoniazid action and resistance in Mycobacterium tuberculosis', *Infection, Genetics and Evolution*

[Preprint]. Available at: https://doi.org/10.1016/j.meegid.2016.09.004.

Valdés-Tresanco, M.S. *et al.* (2021) 'Gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS', *Journal of Chemical Theory and Computation* [Preprint]. Available at: https://doi.org/10.1021/acs.jctc.1c00645.

Varadi, M. *et al.* (2022) '3D-Beacons: Decreasing the gap between protein sequences and structures through a federated network of protein structure data resources', *GigaScience* [Preprint]. Available at: https://doi.org/10.1093/gigascience/giac118.

Vassylyev, D.G. (2009) 'Elongation by RNA polymerase: a race through roadblocks', *Current Opinion in Structural Biology* [Preprint]. Available at: https://doi.org/10.1016/j.sbi.2009.10.004.

Vergne, I., Gilleron, M. and Nigou, J. (2014) 'Manipulation of the endocytic pathway and phagocyte functions by Mycobacterium tuberculosis lipoarabinomannan', *Frontiers in Cellular and Infection Microbiology* [Preprint]. Available at: https://doi.org/10.3389/fcimb.2014.00187.

Van Der Walt, S., Colbert, S.C. and Varoquaux, G. (2011) 'The NumPy array: A structure for efficient numerical computation', *Computing in Science and Engineering* [Preprint]. Available at: https://doi.org/10.1109/MCSE.2011.37.

Wang, D. *et al.* (2006) 'Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis', *Cell* [Preprint]. Available at: https://doi.org/10.1016/j.cell.2006.11.023.

Wang, E. *et al.* (2019) 'End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design', *Chemical Reviews*

[Preprint]. Available at: https://doi.org/10.1021/acs.chemrev.9b00055.

Wang, S. *et al.* (2022) 'MM/PB(GB)SA benchmarks on soluble proteins and membrane proteins', *Frontiers in Pharmacology* [Preprint]. Available at: https://doi.org/10.3389/fphar.2022.1018351.

Wang, T. *et al.* (2019) 'Improved fragment sampling for ab initio protein structure prediction using deep neural networks', *Nature Machine Intelligence* [Preprint]. Available at: https://doi.org/10.1038/s42256-019-0075-7.

Wang, Z. *et al.* (2019) 'Mode-of-action profiling reveals glutamine synthetase as a collateral metabolic vulnerability of M. tuberculosis to bedaquiline', *Proceedings of the National Academy of Sciences of the United States of America* [Preprint]. Available at: https://doi.org/10.1073/pnas.1907946116.

Waskom, M. (2021) 'seaborn: statistical data visualization', *Journal of Open Source Software* [Preprint]. Available at: https://doi.org/10.21105/joss.03021.

Watkins, R.R., Lemonovich, T.L. and File Jr, T.M. (2012) 'An evidence-based review of linezolid for the treatment of methicillin-resistant Staphylococcus aureus (MRSA): Place in therapy', *Core Evidence* [Preprint].

Weinzierl, R.O.J. (2010) 'The nucleotide addition cycle of RNA polymerase is controlled by two molecular hinges in the Bridge Helix domain', *BMC Biology* [Preprint]. Available at: https://doi.org/10.1186/1741-7007-8-134.

WHO (2021) 'Global tuberculosis report 2021'. Available at: https://www.who.int/publications/i/item/9789240037021.

WHO (2023) *Global Tuberculosis Report 2023*, *January*.

Winkelman, J.T. and Gourse, R.L. (2017) 'Open complex DNA scrunching: A key to

transcription start site selection and promoter escape', *BioEssays* [Preprint]. Available at: https://doi.org/10.1002/bies.201600193.

Wolfe, N.D., Dunavan, C.P. and Diamond, J. (2007) 'Origins of major human infectious diseases', *Nature* [Preprint]. Available at: https://doi.org/10.1038/nature05775.

World Health Organization (2018) *Global tuberculosis report 2018 - Geneva*, *WHO report*.

Wu, C.-W. and Oen, H. (1977) ' RNA Polymerase . R. Losick , M. Chamberlin ', *The Quarterly Review of Biology* [Preprint]. Available at: https://doi.org/10.1086/410153.

Xie, Z., Siddiqi, N. and Rubin, E.J. (2005) 'Differential antibiotic susceptibilities of starved Mycobacterium tuberculosis isolates', *Antimicrobial Agents and Chemotherapy* [Preprint]. Available at: https://doi.org/10.1128/AAC.49.11.4778-4780.2005.

Yam, W.C. *et al.* (2004) 'Direct detection of rifampin-resistant Mycobacterium tuberculosis in respiratory specimens by PCR-DNA sequencing', *Journal of Clinical Microbiology* [Preprint]. Available at: https://doi.org/10.1128/JCM.42.10.4438-4443.2004.

Yang, X. *et al.* (2009) 'The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA', *EMBO Reports* [Preprint]. Available at: https://doi.org/10.1038/embor.2009.155.

Yeager, R.L., Munroe, W.G. and Dessau, F.I. (1952) 'Pyrazinamide (aldinamide) in the treatment of pulmonary tuberculosis.', *American review of tuberculosis* [Preprint].

Yuzenkova, Y. *et al.* (2011) 'A new basal promoter element recognized by RNA

polymerase core enzyme', *EMBO Journal* [Preprint]. Available at: https://doi.org/10.1038/emboj.2011.252.

Zaw, M.T., Emran, N.A. and Lin, Z. (2018) 'Mutations inside rifampicin-resistance determining region of rpoB gene associated with rifampicin-resistance in Mycobacterium tuberculosis', *Journal of Infection and Public Health* [Preprint]. Available at: https://doi.org/10.1016/j.jiph.2018.04.005.

Zhan, J., Gurung, S. and Parsa, S.P.K. (2017) 'Identification of top-K nodes in large networks using Katz centrality', *Journal of Big Data* [Preprint]. Available at: https://doi.org/10.1186/s40537-017-0076-5.

Zhang, G. and Darst, S.A. (1998) 'Structure of the Escherichia coli RNA polymerase α subunit amino- terminal domain', *Science* [Preprint]. Available at: https://doi.org/10.1126/science.281.5374.262.

Zhang, L. *et al.* (2020) 'Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol', *Science* [Preprint]. Available at: https://doi.org/10.1126/science.aba9102.

Zhang, Q. *et al.* (2019) 'Uncovering the Resistance Mechanism of Mycobacterium tuberculosis to Rifampicin Due to RNA Polymerase H451D/Y/R Mutations From Computational Perspective', *Frontiers in Chemistry* [Preprint]. Available at: https://doi.org/10.3389/fchem.2019.00819.

Zhang, Y. *et al.* (1999) 'Role of acid pH and deficient efflux of pyrazinoic acid in unique susceptibility of Mycobacterium tuberculosis to pyrazinamide', *Journal of Bacteriology* [Preprint]. Available at: https://doi.org/10.1128/jb.181.7.2044-2049.1999.

Zhang, Y. *et al.* (2003) ‘Mode of action of pyrazinamide: Disruption of Mycobacterium tuberculosis membrane transport and energetics by pyrazinoic acid’, *Journal of Antimicrobial Chemotherapy* [Preprint]. Available at: https://doi.org/10.1093/jac/dkg446.

Zhang, Y. *et al.* (2012) ‘Structural basis of transcription initiation’, *Science* [Preprint]. Available at: https://doi.org/10.1126/science.1227786.

Zhang, Yu *et al.* (2014) ‘GE23077 binds to the RNA polymerase “i” and “i+1” sites and prevents the binding of initiating nucleotides’, *eLife* [Preprint]. Available at: https://doi.org/10.7554/eLife.02450.

Zhang, Ying *et al.* (2014) ‘Mechanisms of Pyrazinamide Action and Resistance’, *Microbiology Spectrum* [Preprint]. Available at: https://doi.org/10.1128/microbiolspec.mgm2-0023-2013.

Zhang, Y., Permar, S. and Sun, Z. (2002) ‘Conditions that may affect the results of susceptibility testing of Mycobacterium tuberculosis to pyrazinamide’, *Journal of Medical Microbiology* [Preprint]. Available at: https://doi.org/10.1099/0022-1317-51-1-42.

Zhou, P., Huang, J. and Tian, F. (2012) ‘Specific Noncovalent Interactions at Protein-Ligand Interface: Implications for Rational Drug Design’, *Current Medicinal Chemistry* [Preprint]. Available at: https://doi.org/10.2174/092986712803414150.

Supplementary materials: Section 1

Supplementary Text. 1.1. Showing the PIR file sample used for conducting homology modelling.

>P1;5uhc.pdb

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

ISQRPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVLHE

FTTVPGVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHNPG

MHIATLNDKGKLEVELVVERGRGYVPAVQNRASGAEIGRIPVDSIYSPVLKVTYKVDATR

VEQRTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELN/

RPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVLHEFTT

VPGVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHNPGMHI

ATLNDKGKLEVELVVERGRGYVPAVQNRA--AEIGRIPVDSIYSPVLKVTYKVDATRVEQ

RTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELNVEAEGI/

SNNSVPGAPNRVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVGGLE

EVLYELSPIEDFSGSMSLSFSDPRFDDVKAPVDECKDKDMTYAAPLFVTAEFINNNTGEI

KSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVYFDETIDKSTDKTLHSVKVIPS

RGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSEQIVERFGFSEIMRSTLEKDNT

VGTDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKRYDLARVGRYKVNKKLGLHVGEP

ITSSTLTEEDVVATIEYLVRLHEGQTTMTVPGGVEVPVETDDIDHFGNRRLRTVGELIQN

QIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPVVAAIKEFFGTSQLSQFMDQNNPLS

GLTHKRRLSALGPGGLSRERAGLEVRDVHPSHYGRMCPIETPEGPNIGLIGSLSVYARVN

PFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQANSPIDADGRFVEPRVLVRRKAGE

VEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHDDANRALMGANMQRQAVPLVRSEAPLVG

TGMELRAAIDAGDVVVAEESGVIEEVSADYITVMHDNGTRRTYRMRKFARSNHGTCANQC

PIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLVAIMPWEGHNYEDAIILSNRLVEEDVL

TSIHIEEHEIDARDTKLGAEEITRDIPNISDEVLADLDERGIVRIGAEVRDGDILVGKVT

PKGETELTPEERLLRAIFGEKAREVRDTSLKVPHGESGKVIGIRVFSREDEDELPAGVNE

LVRVYVAQKRKISDGDKLAGRHGNKGVIGKILPVEDMPFLADGTPVDIILNTHGVPRRMN

IGQILETHLGWCAHSGWKVDAAKGVPDWAARLPDELLEAQPNAIVSTPVFDGAQEAELQG

LLSCTLPNRDGDVLVDADGKAMLFDGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTG

PYSMITQQPLGGKAQFGGQRFGEMECWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIV

KGENIPEPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGEDE./

DVNFFDELRIGLATAEDIRQWSYGEVKKPETINYRTLKPEKDGLFCEKIFGPTRDWECYCGKYKRVRFKGIICERCGV
EVTRAKVRRERMGHIELAAPVTHIWYFKGVPSRLGYLLDLAPKDLEKIIYFAAYVITSVDEEMRHNELSTLEAEMAVE
RKAVEDQRDGELEARAQKLEADLAELEAEGAKADARRKVRDGGEREMRQIRDRAQRELDRLEDIWSTFTKLAPKQ
LIVDENLYRELVDRYGEYFTGAMGAESIQKLIENFDIDAEAESLRDVIRNGKGQKKLRALKRLKVVAAFQQSGNSPM
GMVLDAVPVIPPELRPMVQLDGGRFATSDLNDLYRRVINRNNRLKRLIDLGAPEIIVNNEKRMLQESVDALFDNGR
RGRPVTGPGNRPLKSLSDLLKGKQGRFRQNLLGKRVDYSGRSVIVVGPQLKLHQCGLPKLMALELFKPFVMKRLVD
LNHAQNIKSAKRMVERQRPQVWDVLEEVIAEHPVLLNRAPTLHRLGIQAFEPMLVEGKAIQLHPLVCEAFNADFD
GDQMAVHLPLSAEAQAEARILMLSSNNILSPASGRPLAMPRLDMVTGLYYLTTEVPGDTGEYQPASGDHPETGVY
SSPAEAIMAADRGVLSVRAKIKVRLTQLRPPVEIEAELFGHSGWQPGDAWMAETTLGRVMFNELLPLGYPFVNKQ
MHKKVQAAIINDLAERYPMIVVAQTVDKLKDAGFYWATRSGVTVSMADVLVPPRKKEILDHYEERADKVEKQFQ
RGALNHDERNEALVEIWKEATDEVGQALREHYPDDNPIITIVDSGATGNFTQTRTLAGMKGLVTNPKGEFIPRPVK
SSFREGLTVLEYFINTHGARKGLADTALRTADSGYLTRRLVDVSQDVIVREHDCQTERGIVVELAERAPDGTLIRDPYI
ETSAYARTLGTDAVDEAGNVIVERGQDLGDPEIDALLAAGITQVKVRSVLTCATSTGVCATCYGRSMATGKLVDIG
EAVGIVAAQSIGEPGTQLT----------
GGLPRVQELFEARVPRGKAPIADVTGRVRLEDGERFYKITIVPDDGGEEVVYDKISKRQRLRVFKHEDGSERVLSDG
DHVEVGQQLMEGSADPHEVLRVQGPREVQIHLVREVQEVYRAQGVSIHDKHIEVIVRQMLRRVTIIDSGSTEFLPG
SLIDRAEFEAENRRVVAEGGEPAAGRPVLMGITKASLATDSWLSAASFQETTRVLTDAAINCRSDKLNGLKENVIIG
KLIPAGTGINRYRNIAVQPTEEARAA.../

GYDTPLGITNPPIDELLDRVSSKYALVIYAAKRARQINDYYNQLGEGILEYVGPLVEPGL

QEKPLSIALREIHADLLEHTE/

DESEALRQARKDAELTASADSVRAYLKQIGKVALLNAEEEVELAKRIEAGLYATQLMTEL

SERGEKLPAAQRRDMMWICRDGDRAKNHLLEANLRLVVSLAKRYTGRGMAFLDLIQEGNL

GLIRAVEKFDYTKGYKFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLGRIQREL

LQDLGREPTPEELAKEMDITPEKVLEIQQYAREPISLDQTIGDEGDSQLGDFIEDSEAVV

AVDAVSFTLLQDQLQSVLDTLSEREAGVVRLRFGLTDGQPRTLDEIGQVYGVTRERIRQI

ESKTMSKLRHPSRSQVLRDYLD/

...../

...../

.../*

>P1;1zyr_TL_fit.pdb

structureX:1zyr_TL_fit.pdb:FIRST:@ :END:::::

---/

----------/

----------/

GEIQEVPVRSPLTCQTRYGVCQKCYGYDLSMARPVSIGEAVGIVAAQSIGEPGTQLTMRTFHTGGVAGAADITQGL
PRVIELFEARRPKAK----------
----------/

---/

---/*

>P1;5uhc

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

ISQRPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVLHE

FTTVPGVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHNPG

MHIATLNDKGKLEVELVVERGRGYVPAVQNRASGAEIGRIPVDSIYSPVLKVTYKVDATR

VEQRTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELN/

RPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVLHEFTT

VPGVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHNPGMHI

ATLNDKGKLEVELVVERGRGYVPAVQNRASGAEIGRIPVDSIYSPVLKVTYKVDATRVEQ

RTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELNVEAEGI/

SNNSVPGAPNRVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVGGLE

EVLYELSPIEDFSGSMSLSFSDPRFDDVKAPVDECKDKDMTYAAPLFVTAEFINNNTGEI

KSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVYFDETIDKSTDKTLHSVKVIPS

RGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSEQIVERFGFSEIMRSTLEKDNT

VGTDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKRYDLARVGRYKVNKKLGLHVGEP

ITSSTLTEEDVVATIEYLVRLHEGQTTMTVPGGVEVPVETDDIDHFGNRRLRTVGELIQN

QIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPVVAAIKEFFGTSQLSQFMDQNNPLS

GLTHKRRLSALGPGGLSRERAGLEVRDVHPSHYGRMCPIETPEGPNIGLIGSLSVYARVN

PFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQANSPIDADGRFVEPRVLVRRKAGE

VEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHDDANRALMGANMQRQAVPLVRSEAPLVG

TGMELRAAIDAGDVVVAEESGVIEEVSADYITVMHDNGTRRTYRMRKFARSNHGTCANQC

PIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLVAIMPWEGHNYEDAIILSNRLVEEDVL

TSIHIEEHEIDARDTKLGAEEITRDIPNISDEVLADLDERGIVRIGAEVRDGDILVGKVT

PKGETELTPEERLLRAIFGEKAREVRDTSLKVPHGESGKVIGIRVFSREDEDELPAGVNE

LVRVYVAQKRKISDGDKLAGRHGNKGVIGKILPVEDMPFLADGTPVDIILNTHGVPRRMN

IGQILETHLGWCAHSGWKVDAAKGVPDWAARLPDELLEAQPNAIVSTPVFDGAQEAELQG

LLSCTLPNRDGDVLVDADGKAMLFDGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTG

PYSMITQQPLGGKAQFGGQRFGEMECWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIV

KGENIPEPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGEDE./

DVNFFDELRIGLATAEDIRQWSYGEVKKPETINYRTLKPEKDGLFCEKIFGPTRDWECYCGKYKRVRFKGIICERCGV
EVTRAKVRRERMGHIELAAPVTHIWYFKGVPSRLGYLLDLAPKDLEKIIYFAAYVITSVDEEMRHNELSTLEAEMAVE
RKAVEDQRDGELEARAQKLEADLAELEAEGAKADARRKVRDGGEREMRQIRDRAQRELDRLEDIWSTFTKLAPKQ
LIVDENLYRELVDRYGEYFTGAMGAESIQKLIENFDIDAEAESLRDVIRNGKGQKKLRALKRLKVVAAFQQSGNSPM
GMVLDAVPVIPPELRPMVQLDGGRFATSDLNDLYRRVINRNNRLKRLIDLGAPEIIVNNEKRMLQESVDALFDNGR
RGRPVTGPGNRPLKSLSDLLKGKQGRFRQNLLGKRVDYSGRSVIVVGPQLKLHQCGLPKLMALELFKPFVMKRLVD
LNHAQNIKSAKRMVERQRPQVWDVLEEVIAEHPVLLNRAPTLHRLGIQAFEPMLVEGKAIQLHPLVCEAFNADFD
GDQMAVHLPLSAEAQAEARILMLSSNNILSPASGRPLAMPRLDMVTGLYYLTTEVPGDTGEYQPASGDHPETGVY
SSPAEAIMAADRGVLSVRAKIKVRLTQLRPPVEIEAELFGHSGWQPGDAWMAETTLGRVMFNELLPLGYPFVNKQ
MHKKVQAAIINDLAERYPMIVVAQTVDKLKDAGFYWATRSGVTVSMADVLVPPRKKEILDHYEERADKVEKQFQ
RGALNHDERNEALVEIWKEATDEVGQALREHYPDDNPIITIVDSGATGNFTQTRTLAGMKGLVTNPKGEFIPRPVK
SSFREGLTVLEYFINTHGARKGLADTALRTADSGYLTRRLVDVSQDVIVREHDCQTERGIVVELAERAPDGTLIRDPYI
ETSAYARTLGTDAVDEAGNVIVERGQDLGDPEIDALLAAGITQVKVRSVLTCATSTGVCATCYGRSMATGKLVDIG
EAVGIVAAQSIGEPGTQLTMRTFHQGGVG--
EDITGGLPRVQELFEARVPRGKAPIADVTGRVRLEDGERFYKITIVPDDGGEEVVYDKISKRQRLRVFKHEDGSERVL
SDGDHVEVGQQLMEGSADPHEVLRVQGPREVQIHLVREVQEVYRAQGVSIHDKHIEVIVRQMLRRVTIIDSGSTE
FLPGSLIDRAEFEAENRRVVAEGGEPAAGRPVLMGITKASLATDSWLSAASFQETTRVLTDAAINCRSDKLNGLKEN
VIIGKLIPAGTGINRYRNIAVQPTEEARAA.../

GYDTPLGITNPPIDELLDRVSSKYALVIYAAKRARQINDYYNQLGEGILEYVGPLVEPGL

QEKPLSIALREIHADLLEHTE/

DESEALRQARKDAELTASADSVRAYLKQIGKVALLNAEEEVELAKRIEAGLYATQLMTEL

SERGEKLPAAQRRDMMWICRDGDRAKNHLLEANLRLVVSLAKRYTGRGMAFLDLIQEGNL

GLIRAVEKFDYTKGYKFSTYATWWIRQAITRAMADQARTIRIPVHMVEVINKLGRIQREL

LQDLGREPTPEELAKEMDITPEKVLEIQQYAREPISLDQTIGDEGDSQLGDFIEDSEAVV

AVDAVSFTLLQDQLQSVLDTLSEREAGVVRLRFGLTDGQPRTLDEIGQVYGVTRERIRQI

ESKTMSKLRHPSRSQVLRDYLD/

...../

...../

.../*

Supplementary Text 1.2. Sample of the MODELLER script used to conduct homology modelling.

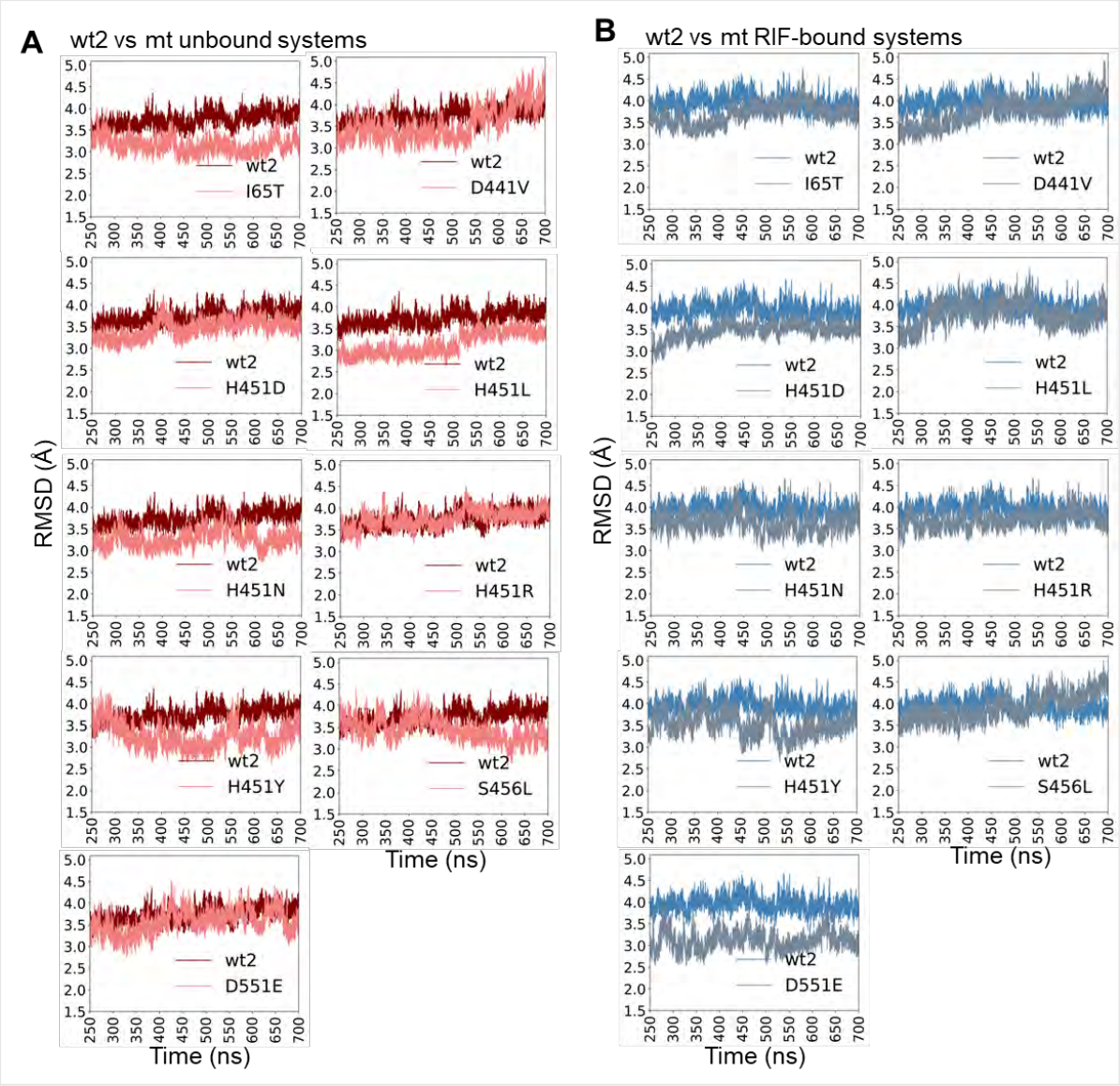
```
# Homology modelling by the automodel class

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 = ['/mounts/jabba/JMS/users/mokgerwa/modeling_job2/wt/Run/']
```

```
#read in HETATMS

env.io.hetatm = True

env.io.water = False




a = automodel(env,

        alnfile  = 'alignment2.pir', #'p9wgy9_rpoBprot.pir',                    # alignment
filename

        knowns   = ('5uhc.pdb','1zyr_TL_fit.pdb'),#(,'5uhc_C.pdb'),             # codes of
the templates

              sequence = '5uhc', assess_methods=(assess.DOPE, assess.GA341))  #'Pfalciparum',
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 = True                              # generate superimposed
templatesand 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])
```

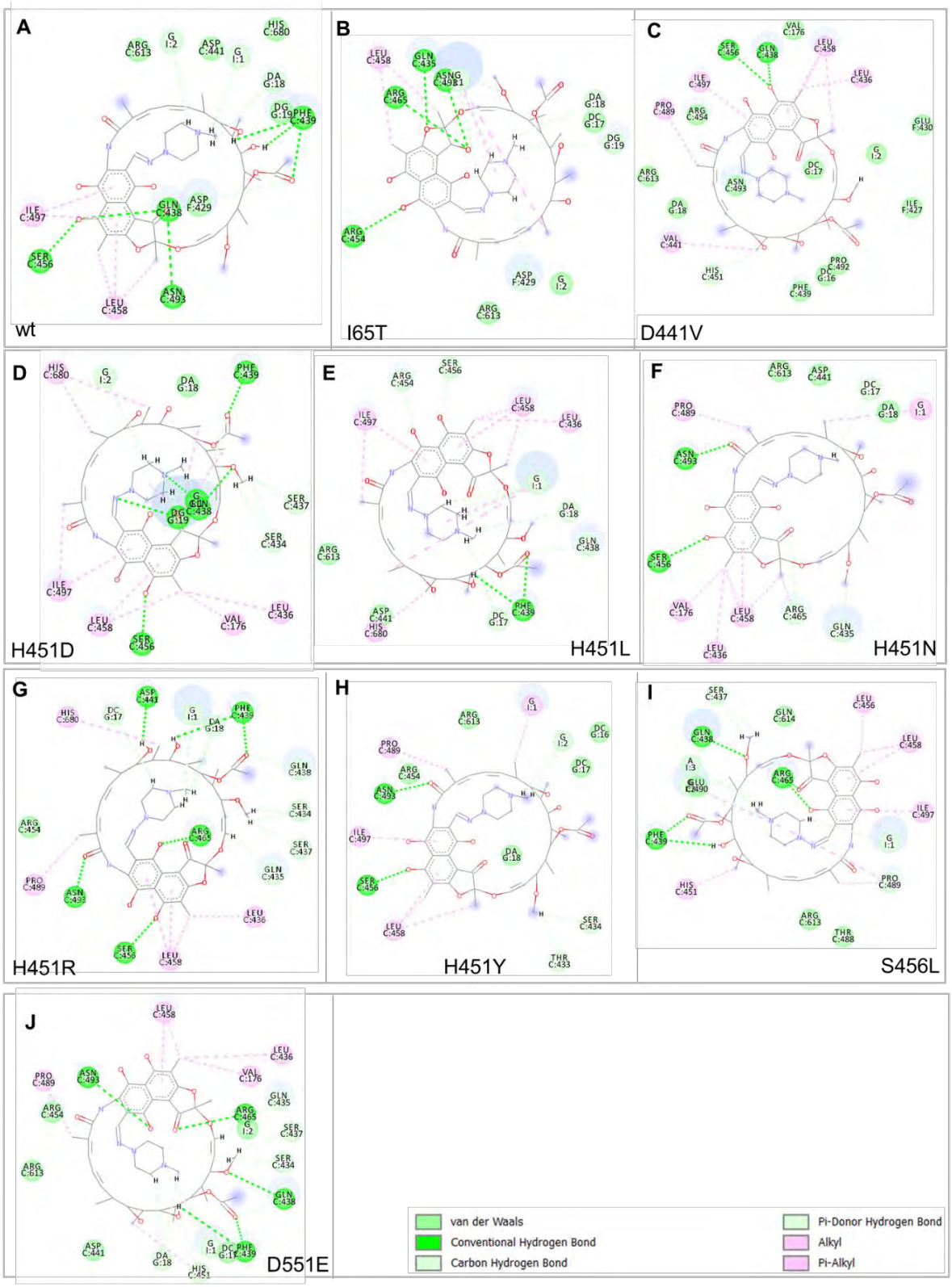
```
# Get top model
m = ok_models[0]

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 directory
```

Supplementary Table S1.1. Shows the homology modelling results through z-DOPE and RMSD along with the PROCHECK and QMEAN validation results.

Systems	z-DOPE score (MODELLER)	RMSD (Å) (5uhc *vs* model)	QMEAN6 (normalized score)	PROCHECK (Ramachandran plot)
wt	-0.52	0.21	0.71	Favoured, 94.1%; unfavoured, 0.1%
I65T	-0.52	0.20	0.72	Favoured, 94.1%; unfavoured, 0.1%
D441V	-0.52	0.20	0.71	Favoured, 94.0%; unfavoured, 0.1%
H451D	-0.52	0.21	0.71	Favoured, 93.9%; unfavoured, 0.1%
H451L	-0.52	0.20	0.71	Favoured, 94.0%; unfavoured, 0.1%
H451N	-0.52	0.21	0.71	Favoured, 93.8%; unfavoured, 0.1%
H451R	-0.52	0.22	0.71	Favoured, 93.8%; unfavoured, 0.1%
H451Y	-0.51	0.21	0.71	Favoured, 93.8%; unfavoured, 0.1%
S456L	-0.51	0.21	0.71	Favoured, 93.6%; unfavoured, 0.1%
D551E	-0.52	0.21	0.71	Favoured, 93.9%; unfavoured, 0.1%

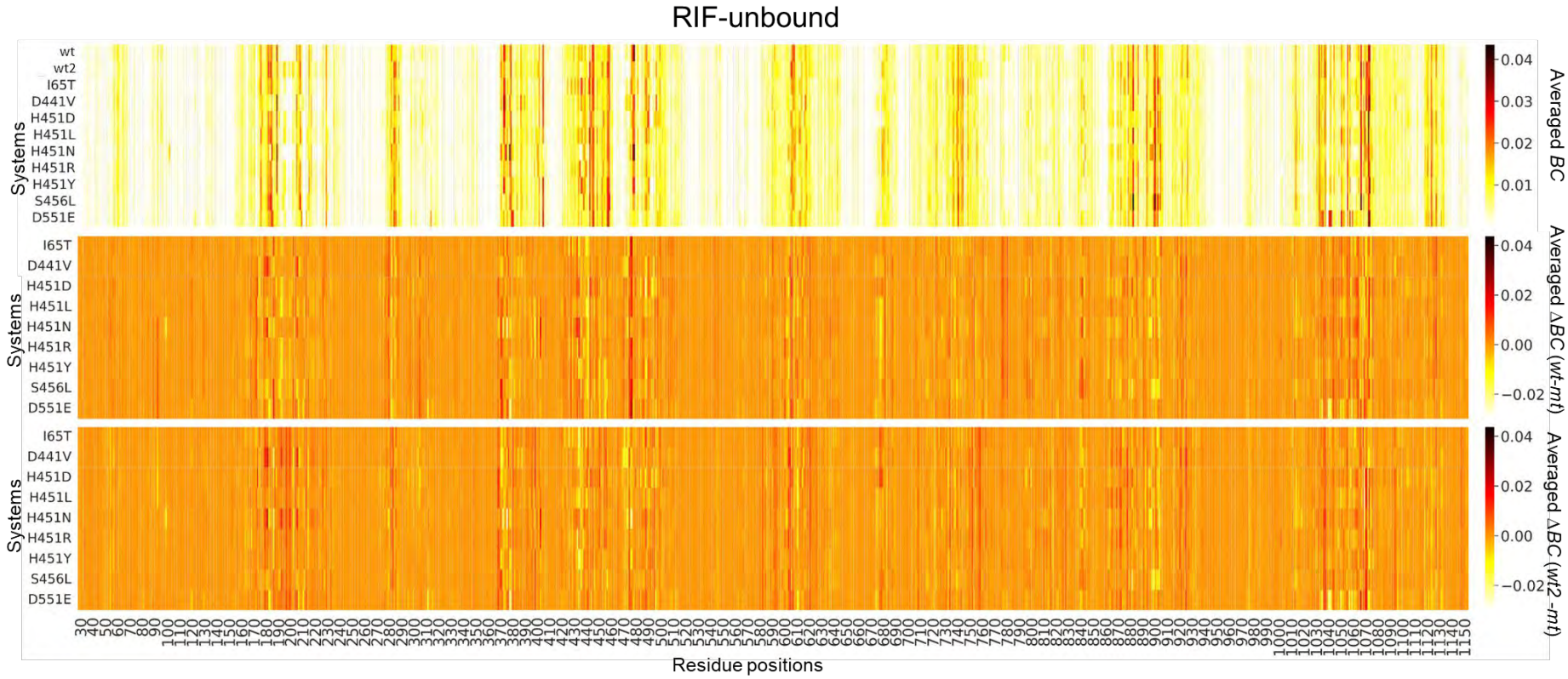


Supplementary Figure S1.1. Graphical analysis of the conformation stability of the *wt* and *mt Mtb*-RNAPs over time, as induced by the presence of the mutations and/or RIF. Cα-RMSD (Å) estimation line plots are shown for the unbound *wt* (colored maroon) and *mt* proteins as well as the RIF-bound *wt* (colored blue) and *mt* proteins (colored pink) with reference to their initial structures.

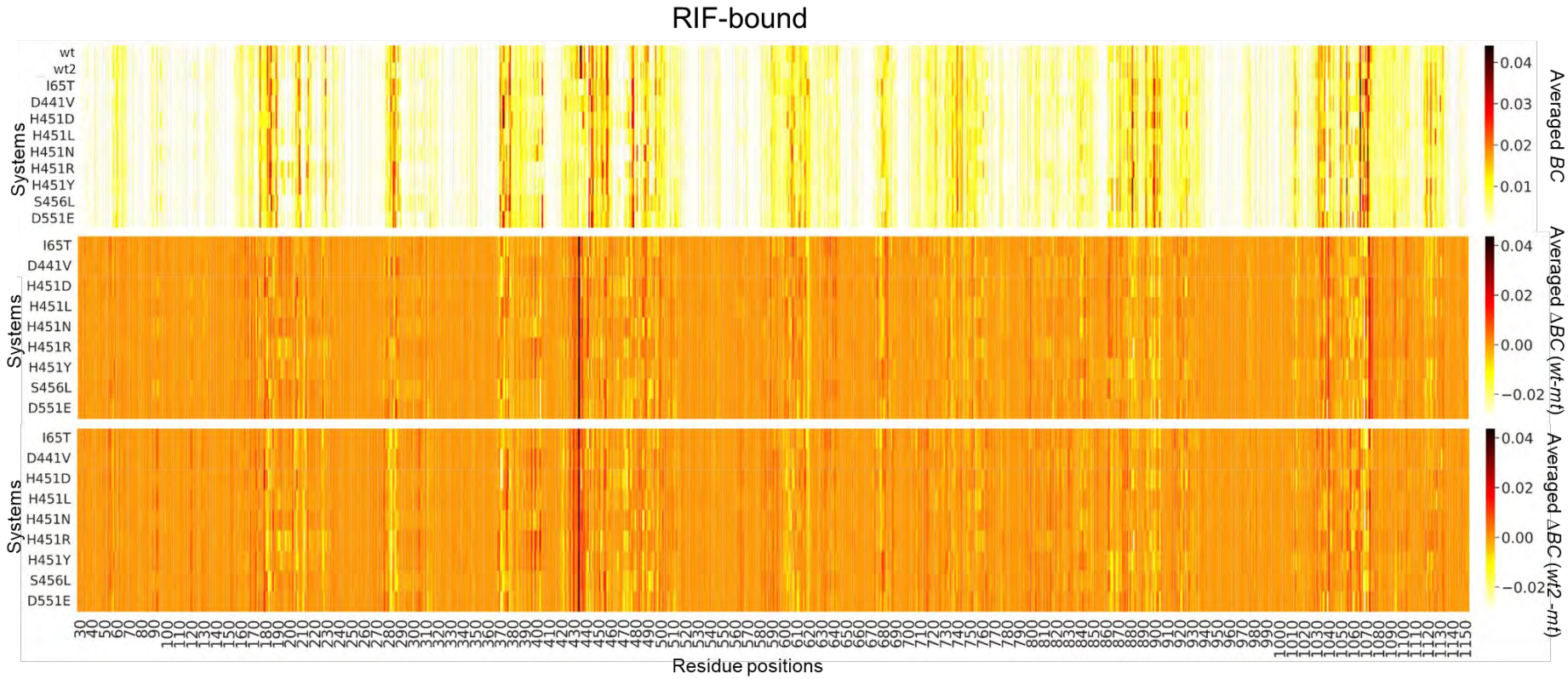


Supplementary Figure S1.2. (A-J) RIF interaction profiles as extracted from the low-energy minima RIF-BP structures of the *wt* and *mt* proteins**.**

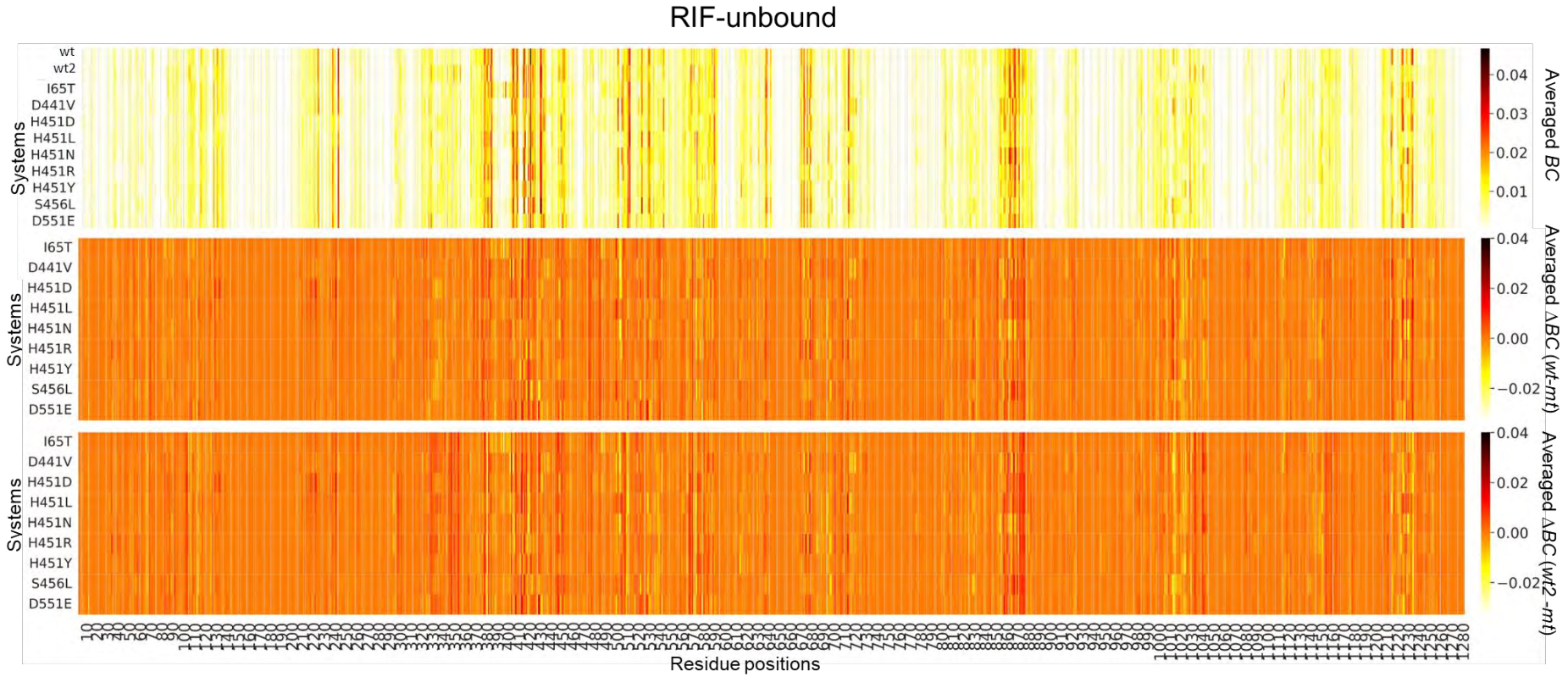
Supplementary materials: Section 2



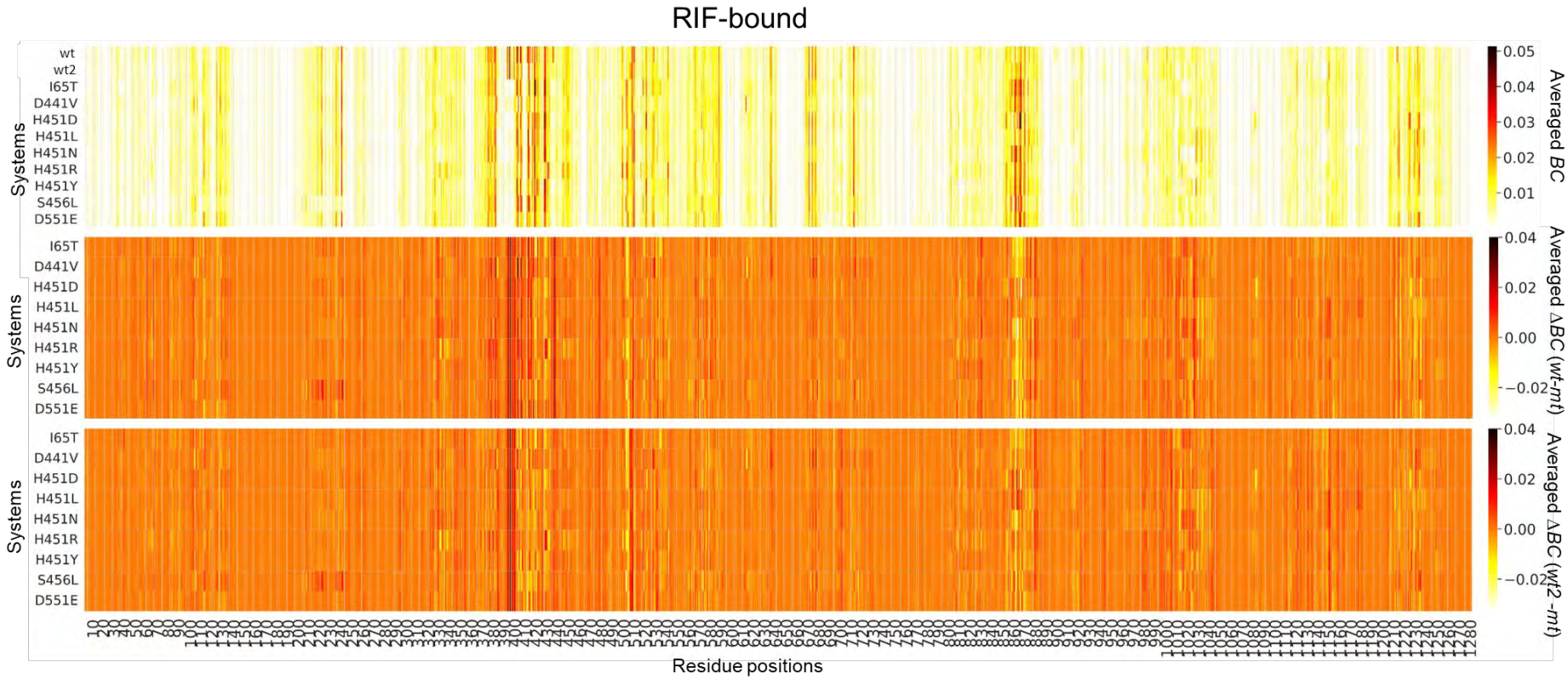
Supplementary Figure S2.1A. Heatmap representation of the raw and Δ*BC* values for the RIF-unbound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



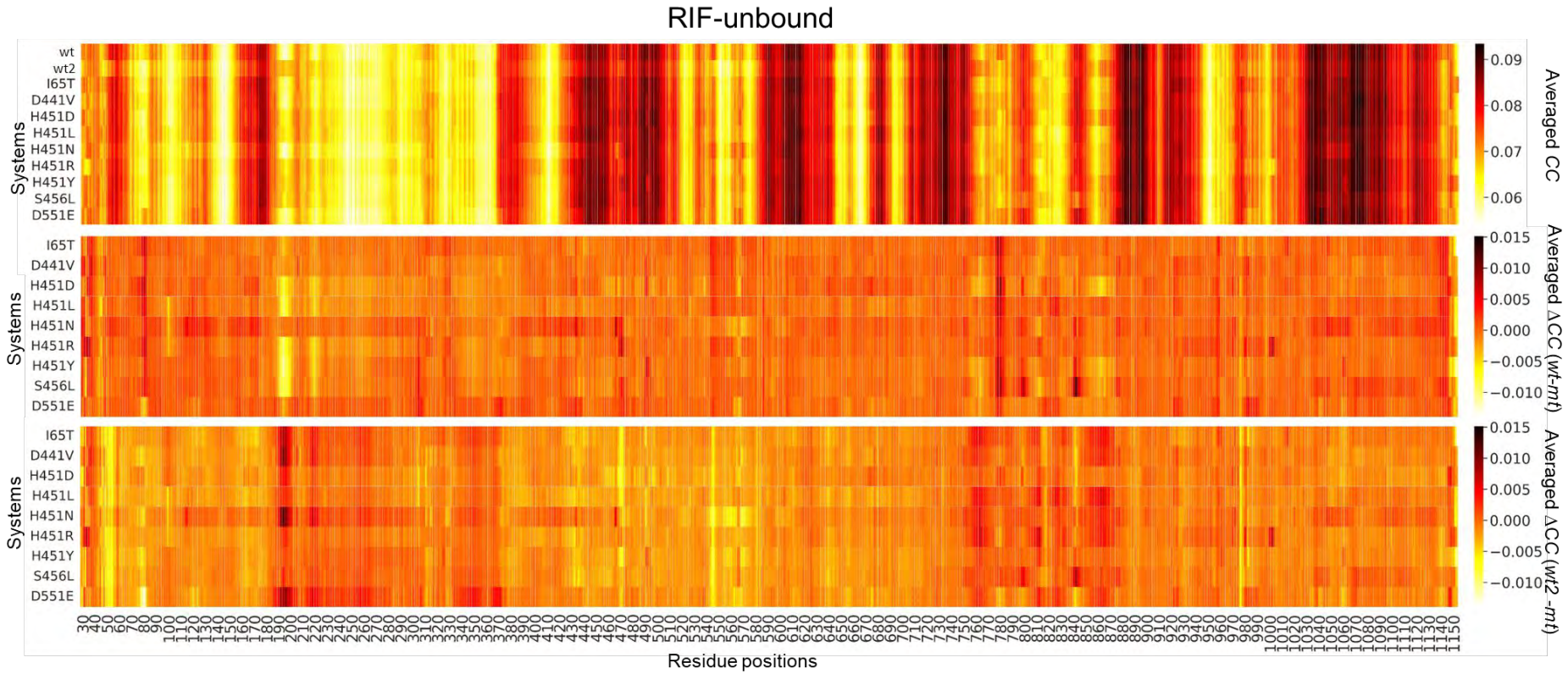
Supplementary Figure S2.1B. Heatmap representation of the raw and Δ*BC* values for the RIF-bound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



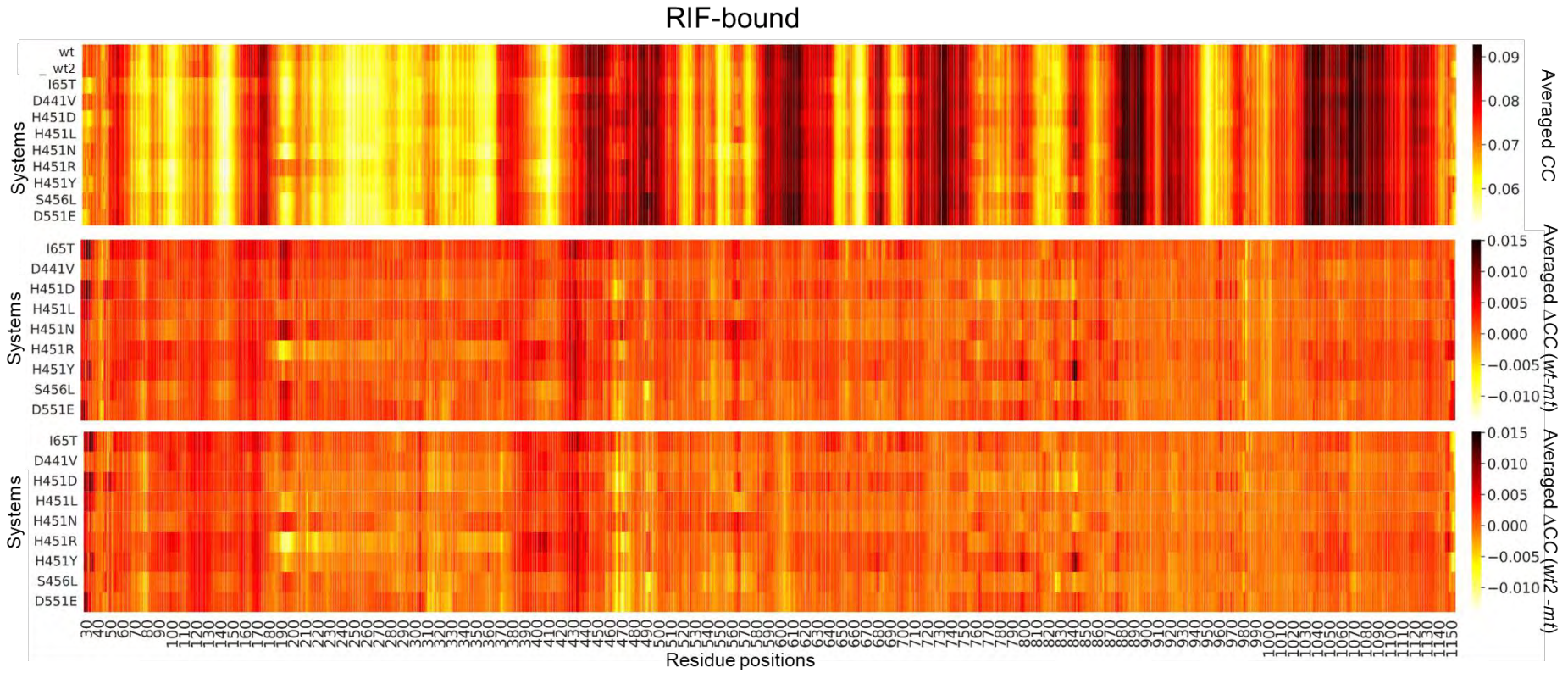
Supplementary Figure S2.1C. Heatmap representation of the raw and Δ*BC* values for the RIF-unbound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



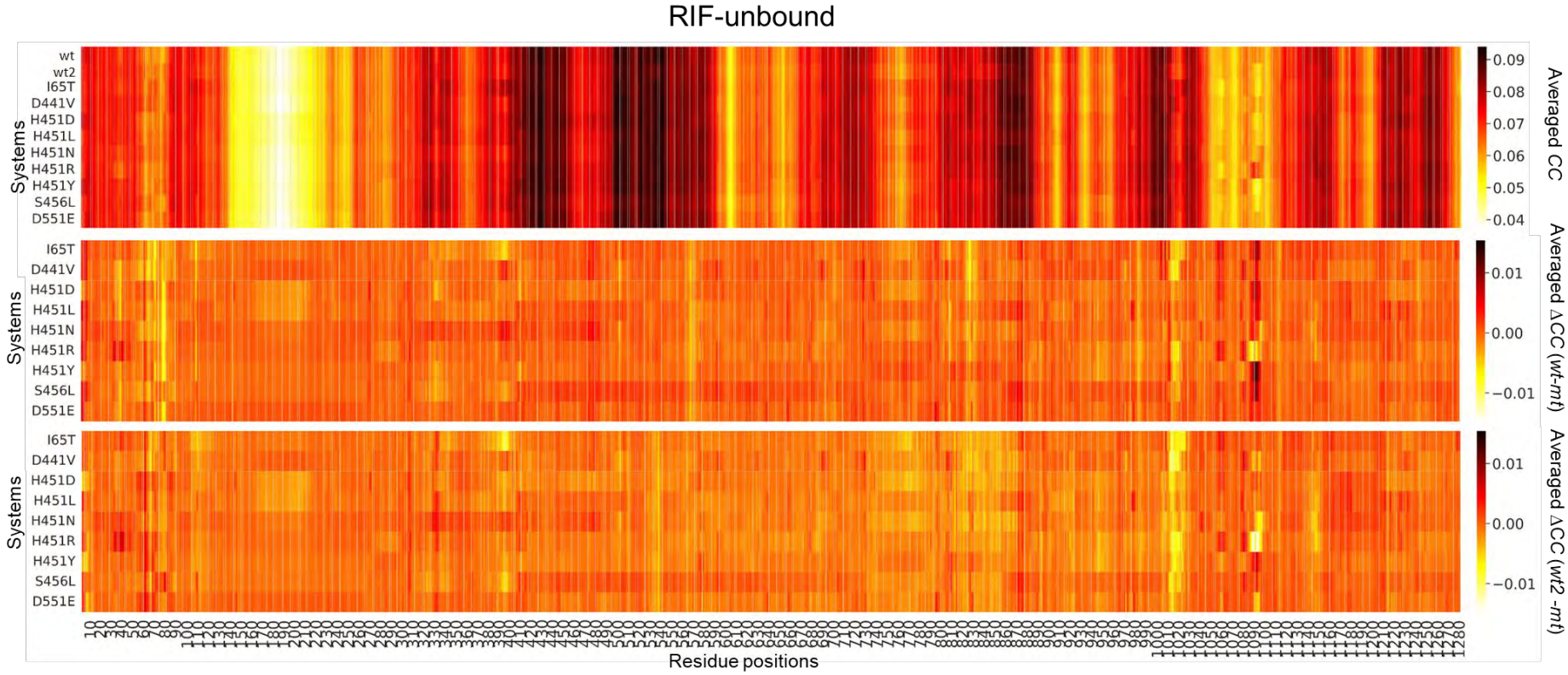
Supplementary Figure S2.1D. Heatmap representation of the raw and Δ*BC* values for the RIF-bound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



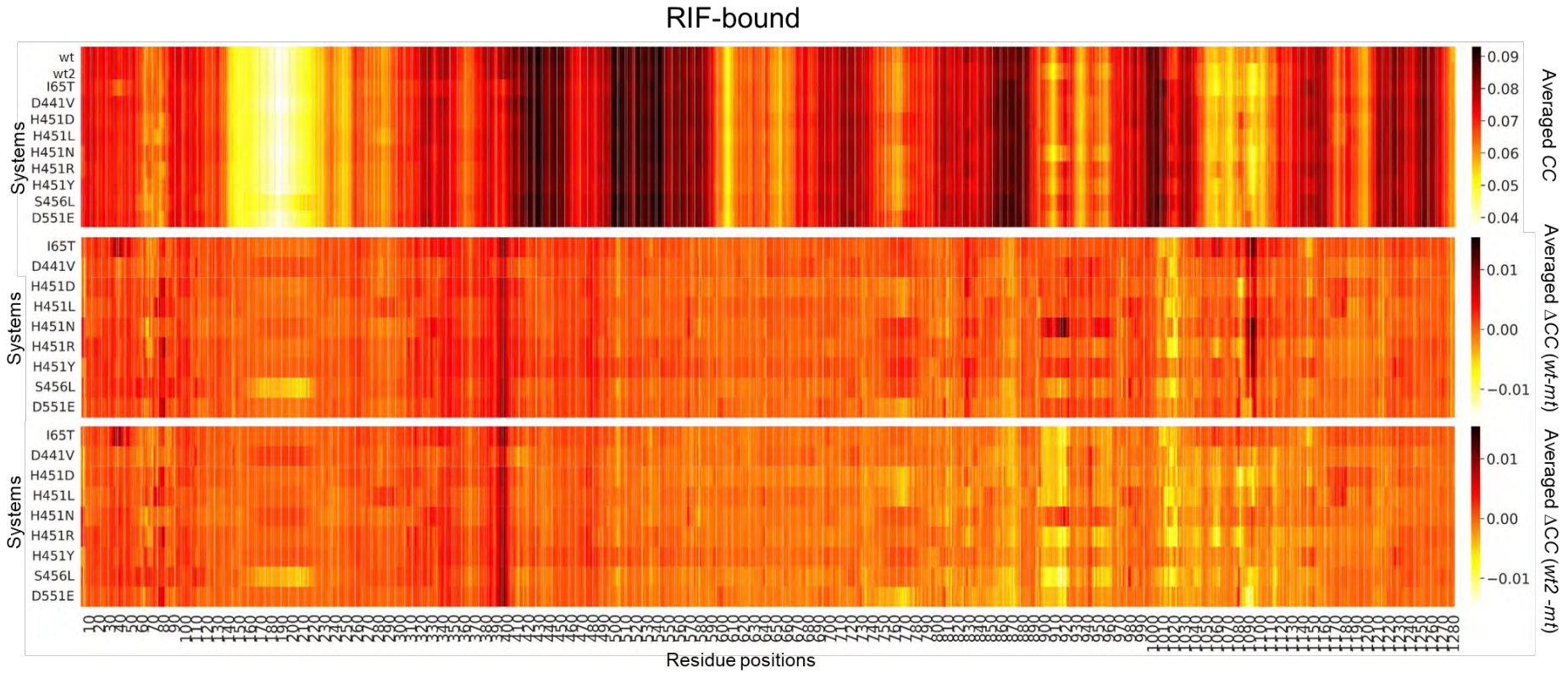
Supplementary Figure S2.2A. Heatmap representation of the raw and Δ*CC* values for the RIF-unbound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



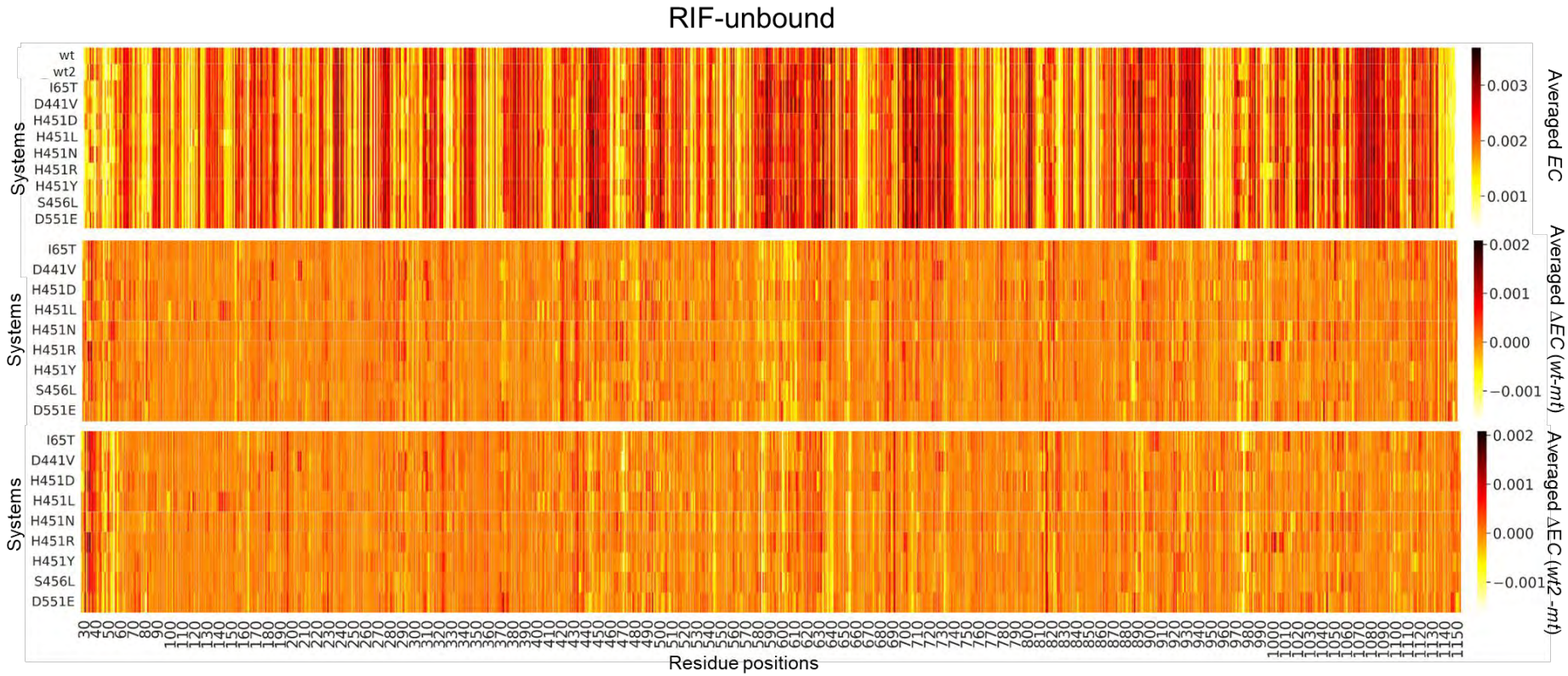
Supplementary Figure S2.2B. Heatmap representation of the raw and ΔCC values for the RIF-bound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



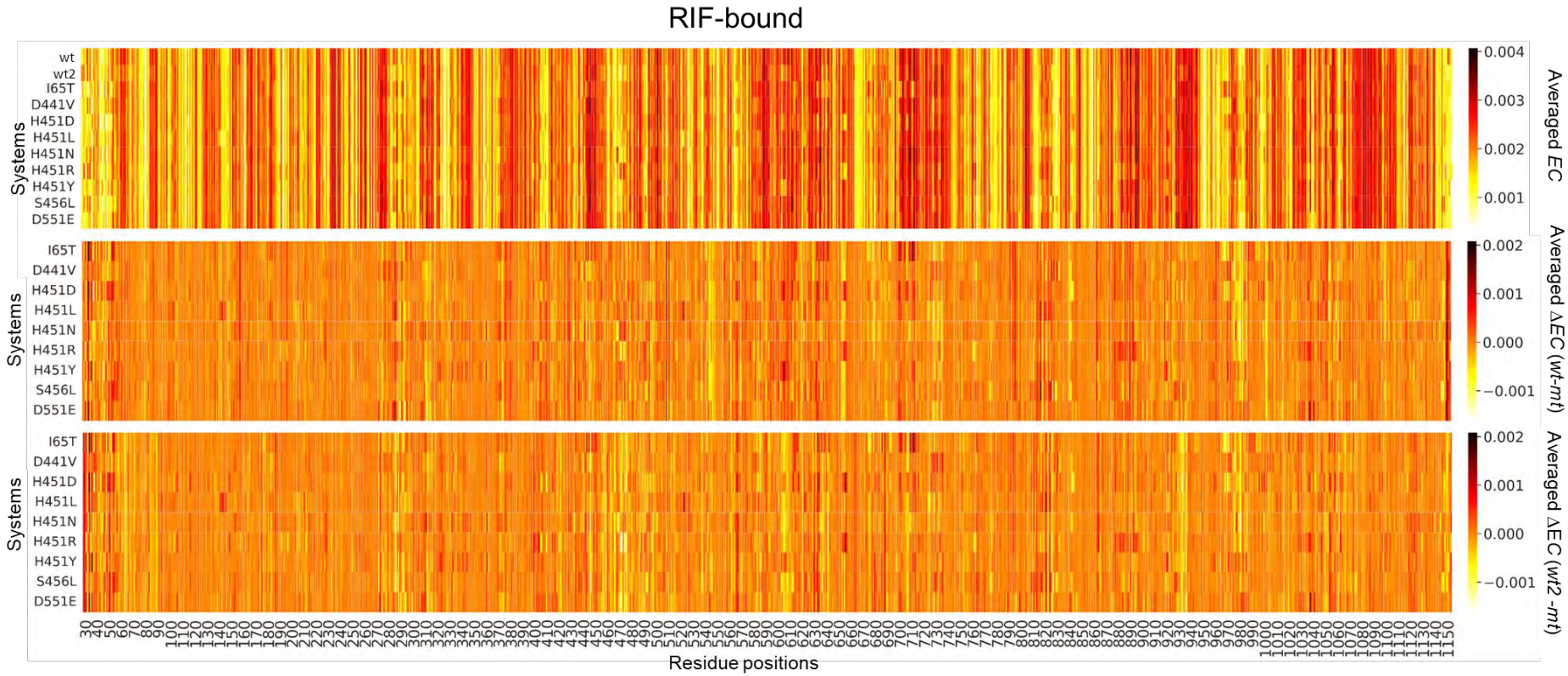
Supplementary Figure S2.2C. Heatmap representation of the raw and ΔCC values for the RIF-unbound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



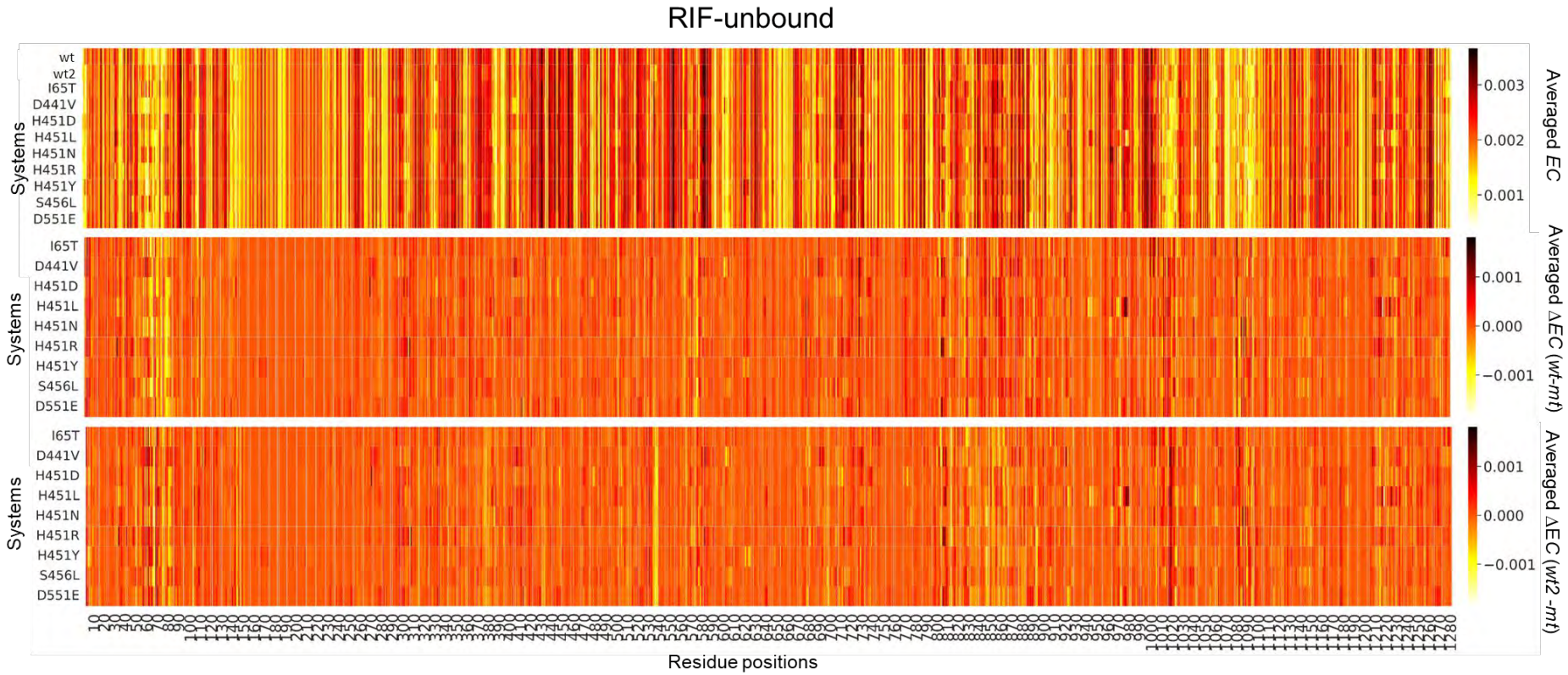
Supplementary Figure S2.2D. Heatmap representation of the raw and Δ*CC* values for the RIF-bound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



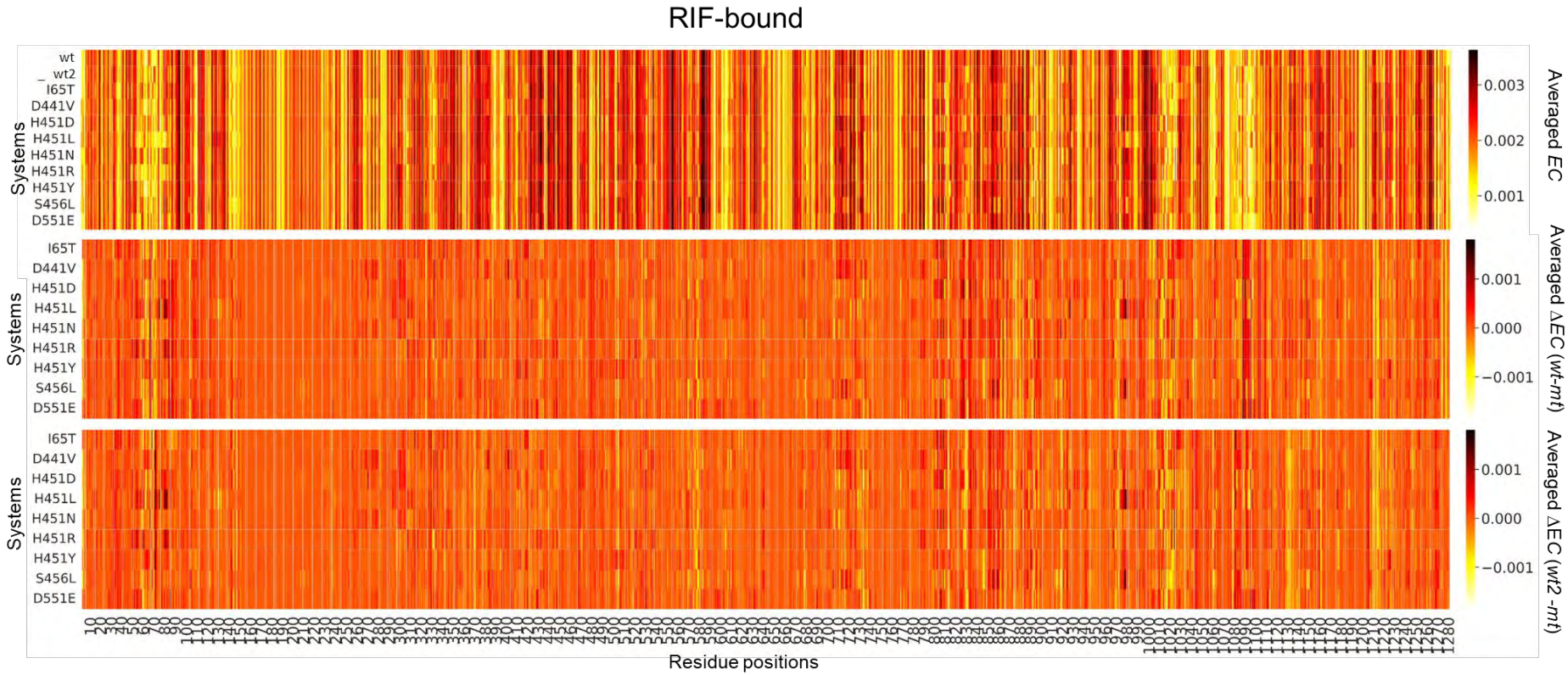
Supplementary Figure S2.3A. Heatmap representation of the raw and Δ*EC* values for the RIF-unbound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



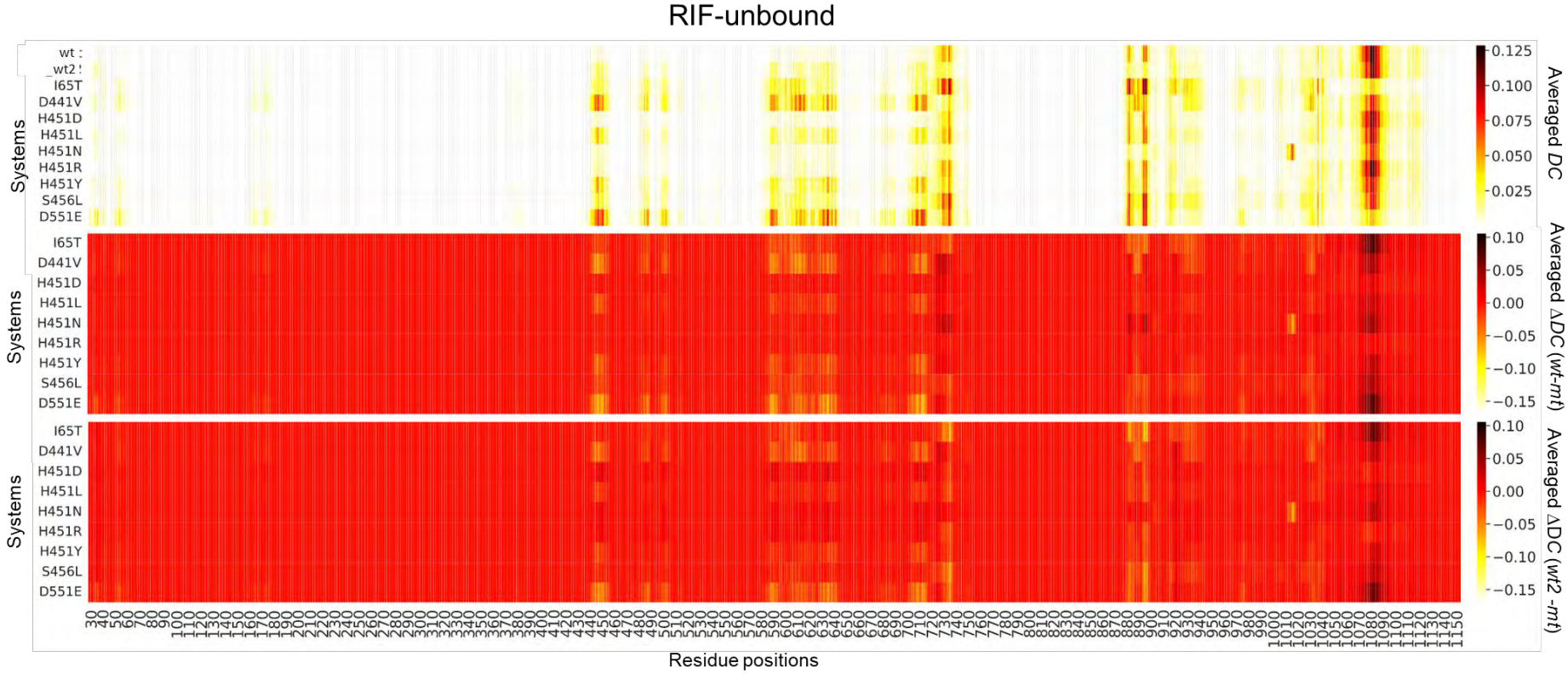
Supplementary Figure S2.3B. Heatmap representation of the raw and Δ*EC* values for the RIF-bound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



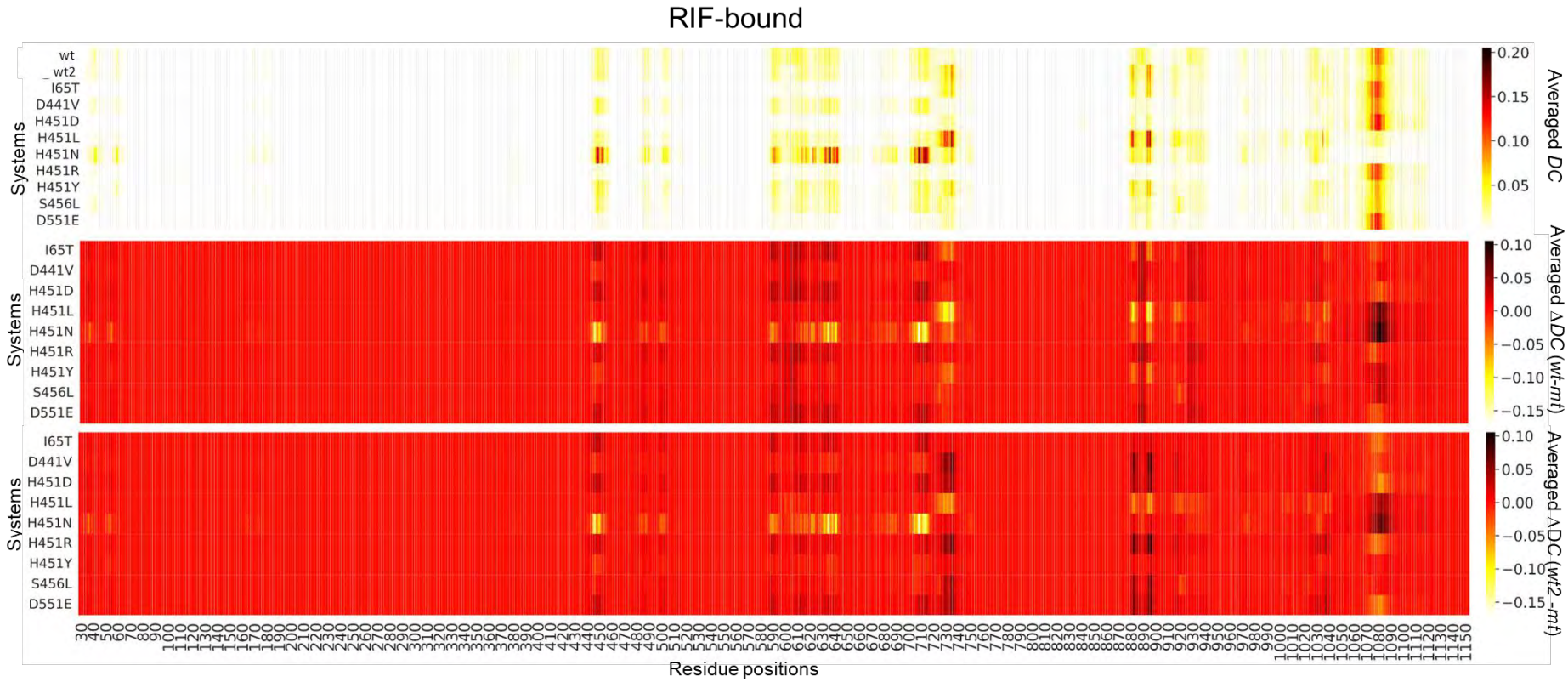
Supplementary Figure S2.3C. Heatmap representation of the raw and ΔEC values for the RIF-unbound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



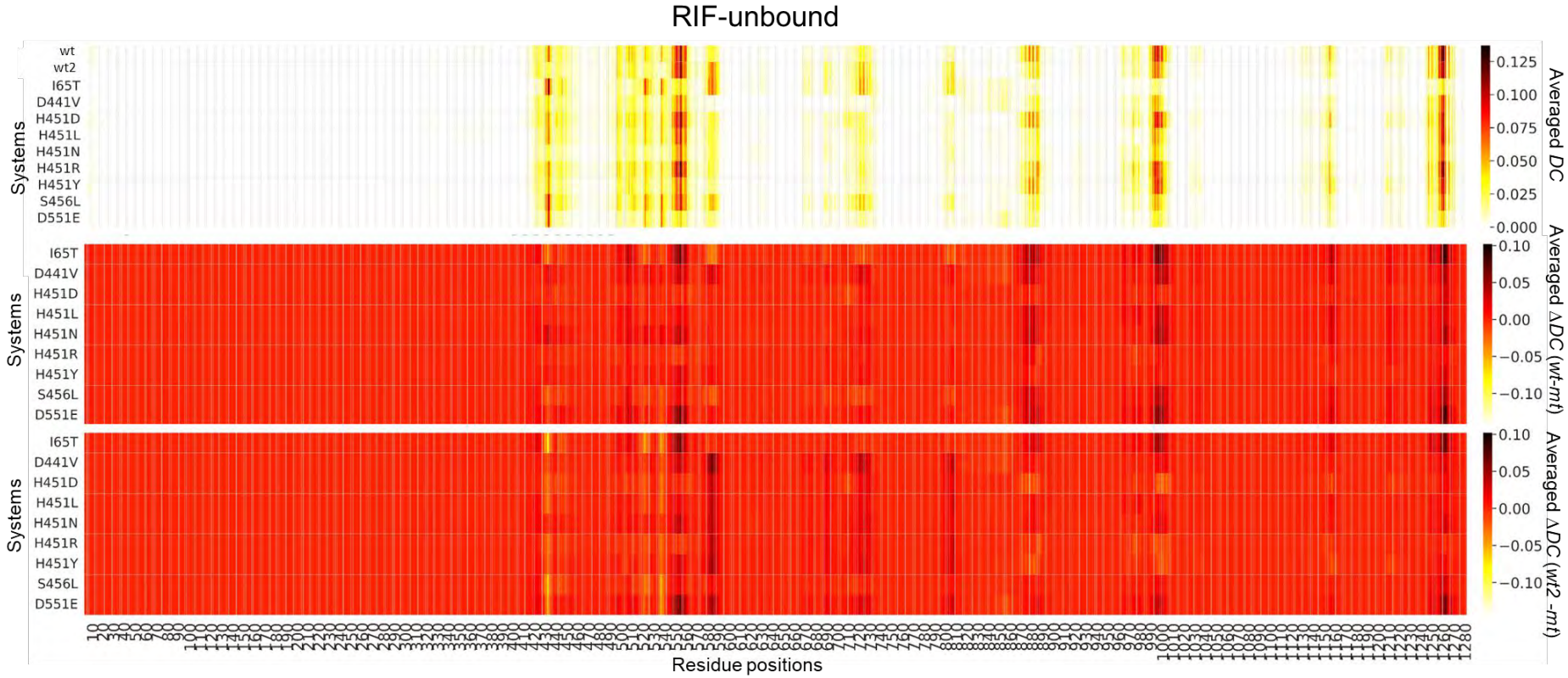
Supplementary Figure S2.3D. Heatmap representation of the raw and Δ*EC* values for the RIF-bound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



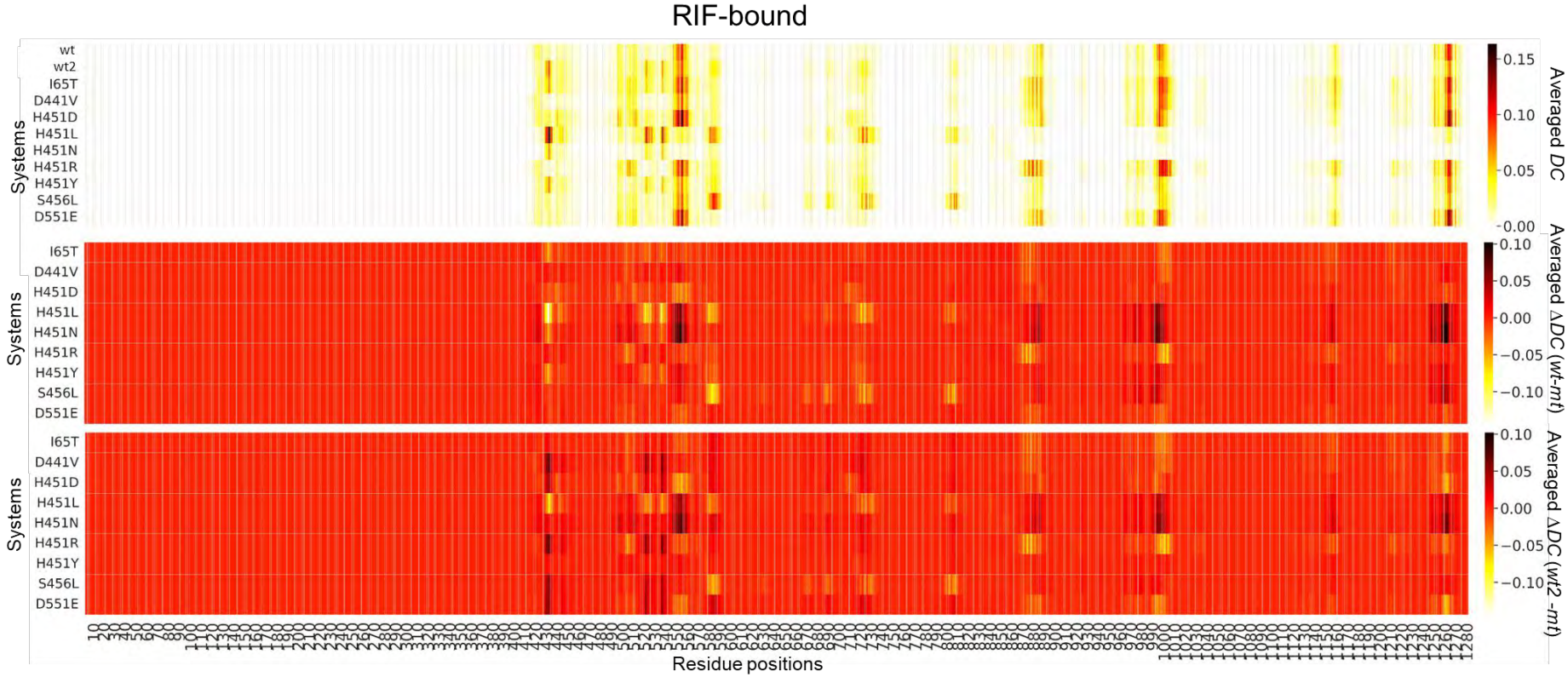
Supplementary Figure S2.4A. Heatmap representation of the raw and Δ*DC* values for the RIF-unbound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



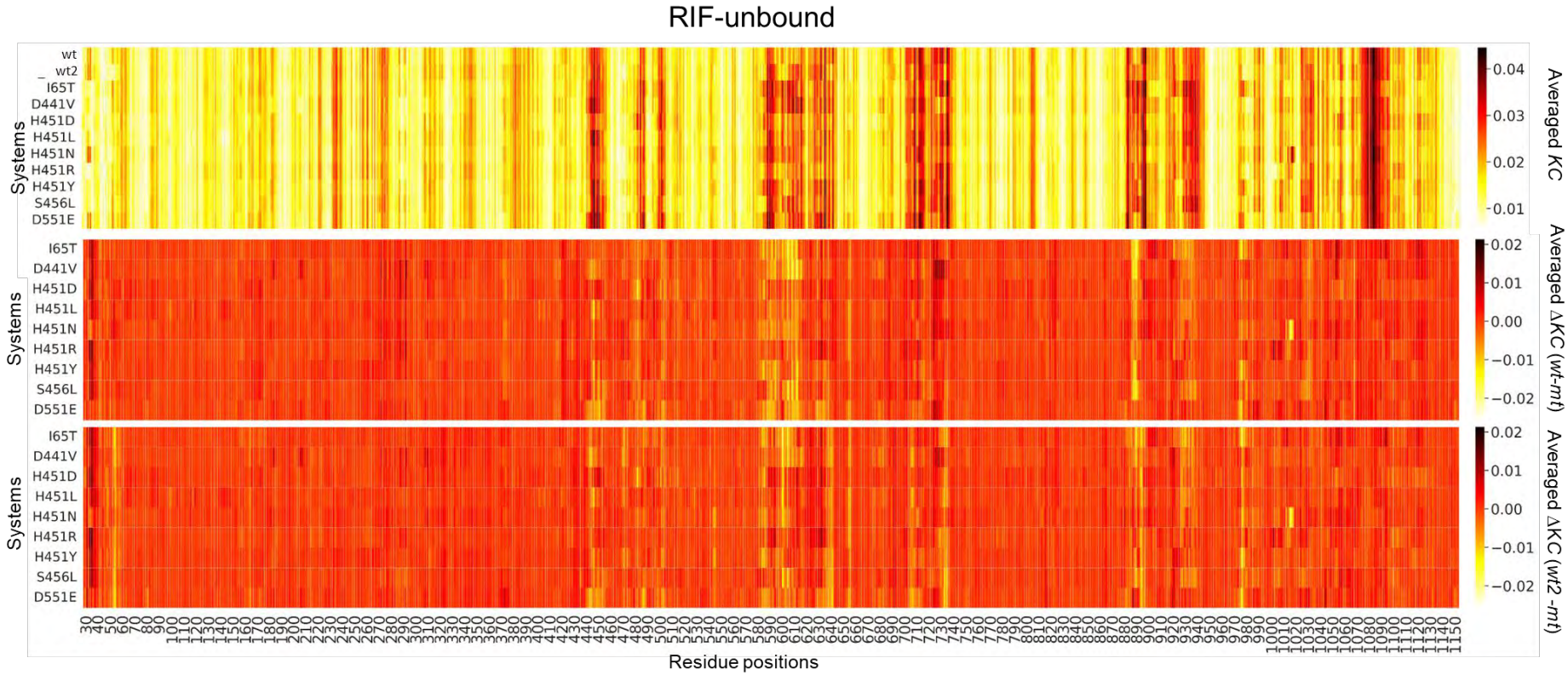
Supplementary Figure S2.4B. Heatmap representation of the raw and Δ*DC* values for the RIF-bound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



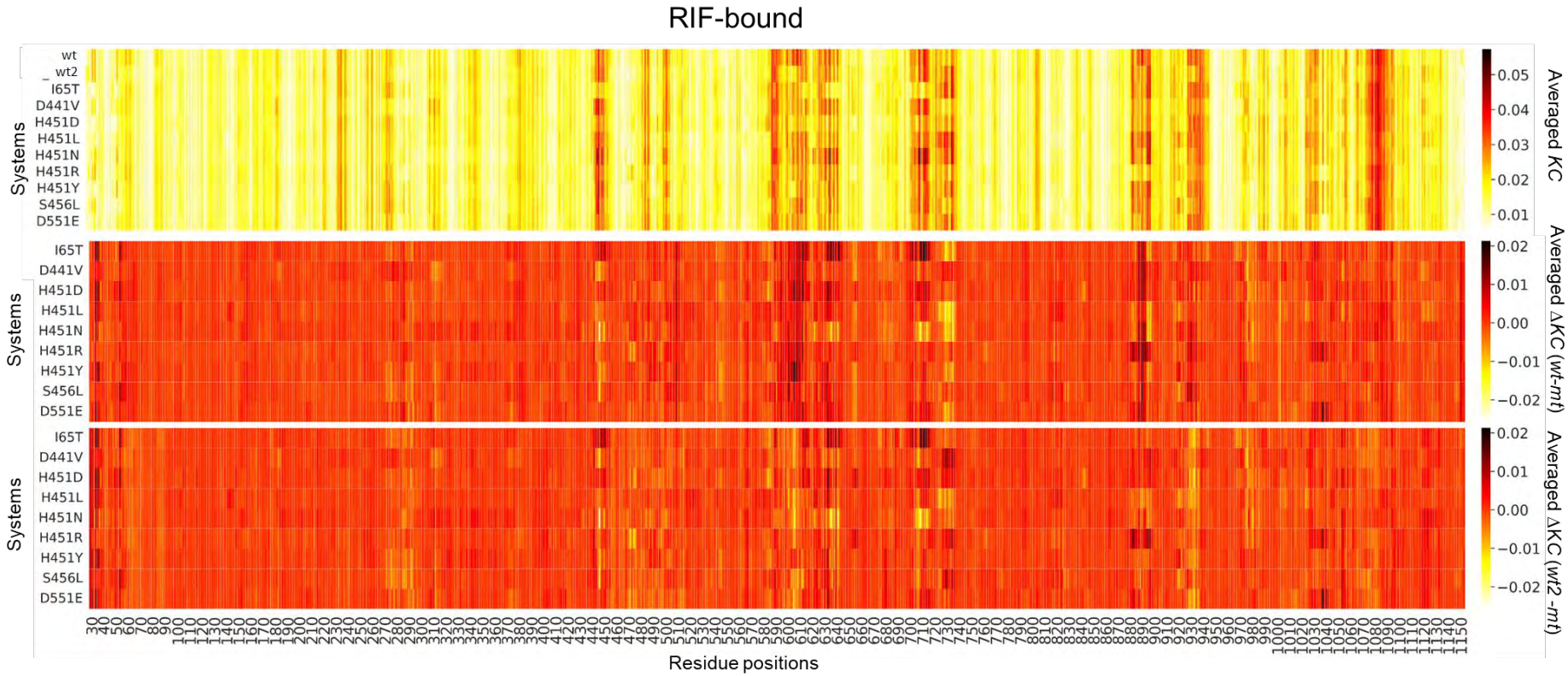
Supplementary Figure S2.4C. Heatmap representation of the raw and ΔDC values for the RIF-unbound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



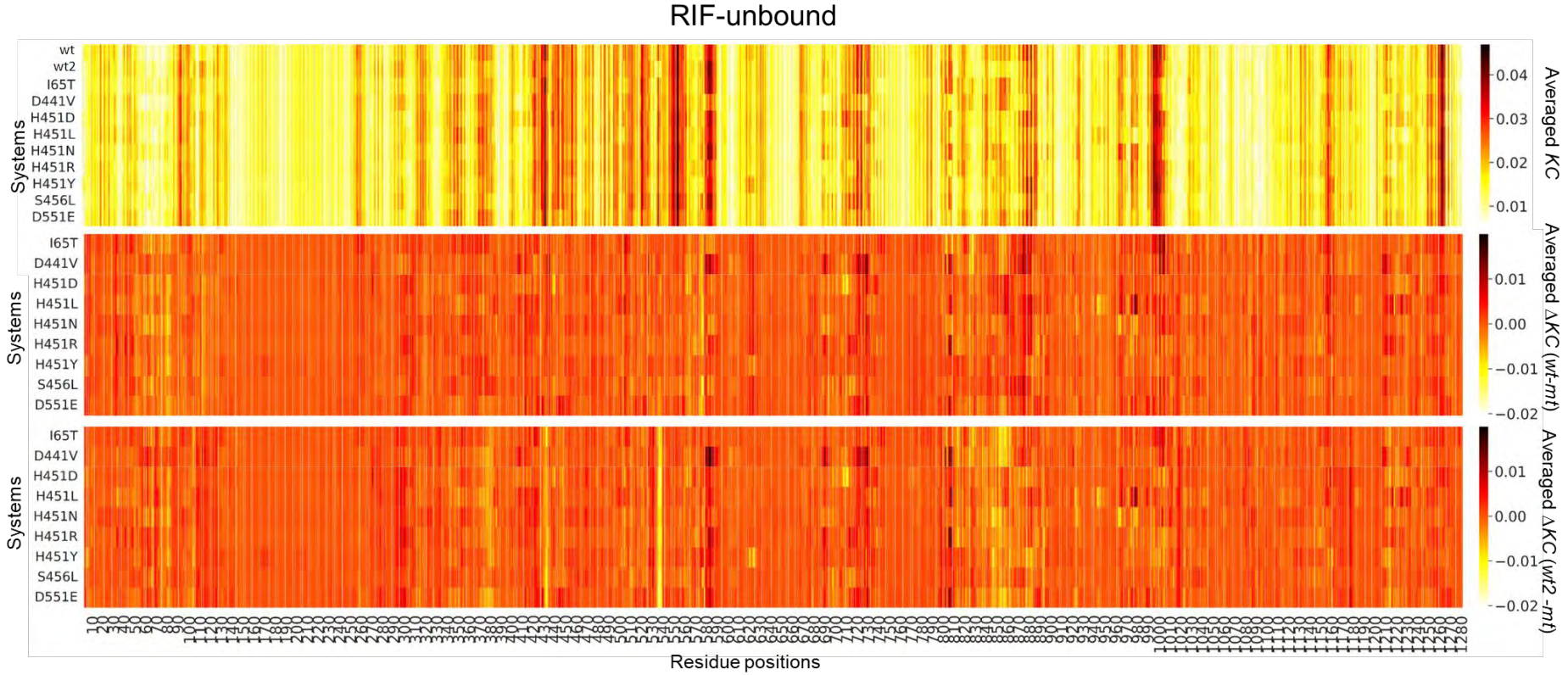
Supplementary Figure S2.4D. Heatmap representation of the raw and Δ*DC* values for the RIF-bound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



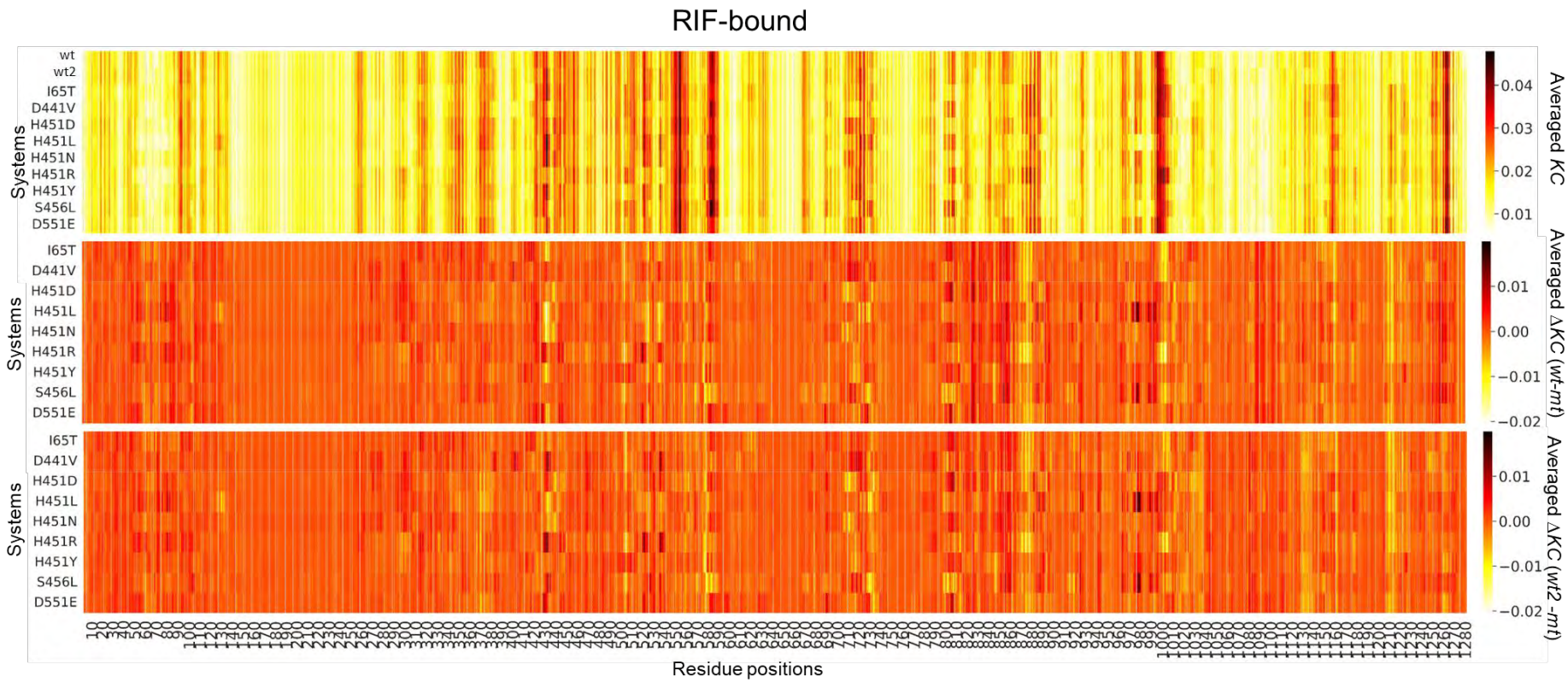
Supplementary Figure S2.5A. Heatmap representation of the raw and Δ*KC* values for the RIF-unbound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



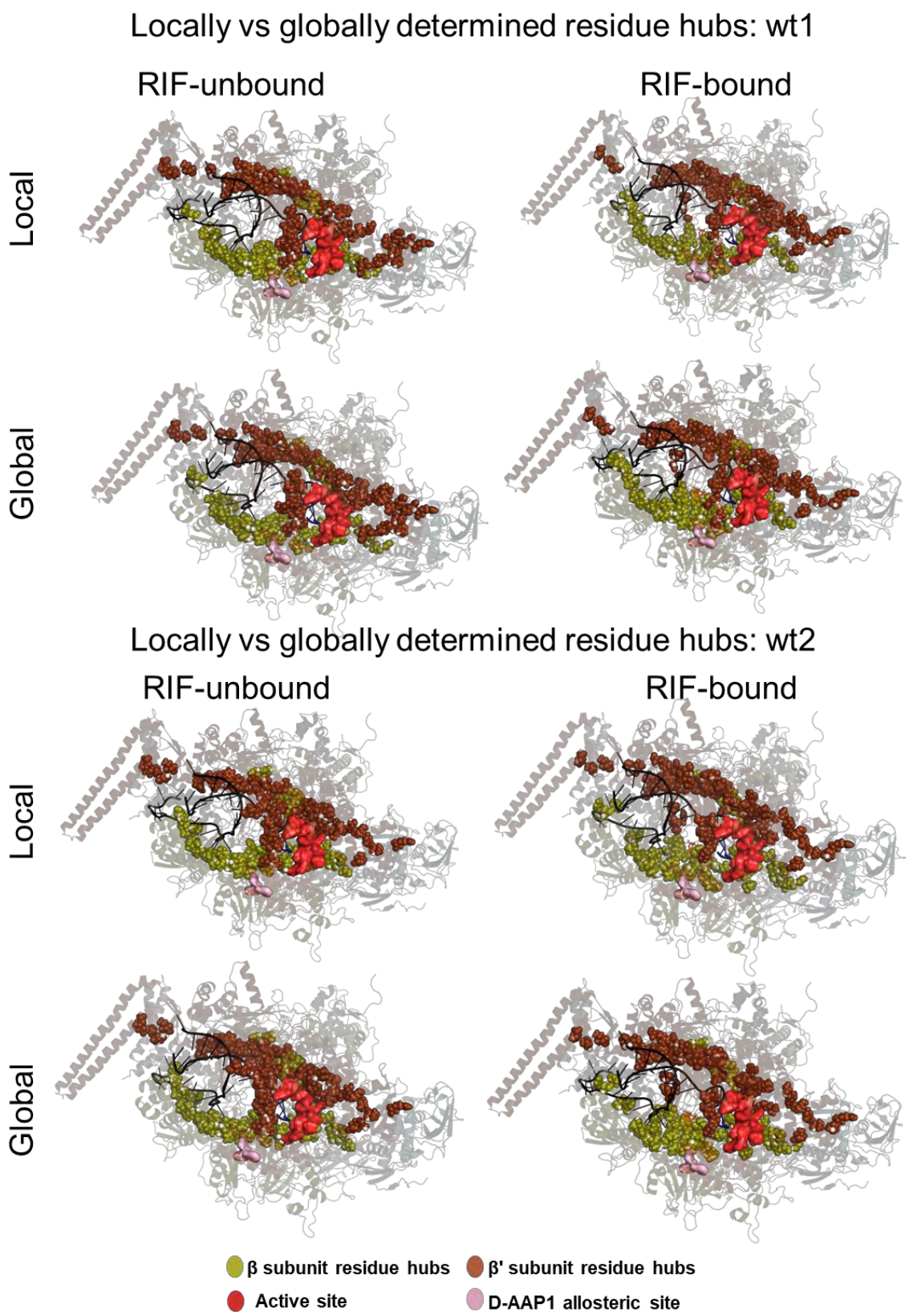
Supplementary Figure S2.5B. Heatmap representation of the raw and Δ*KC* values for the RIF-bound *Mtb*-RNAP β subunit *wt* standards and the respective *mt*s.



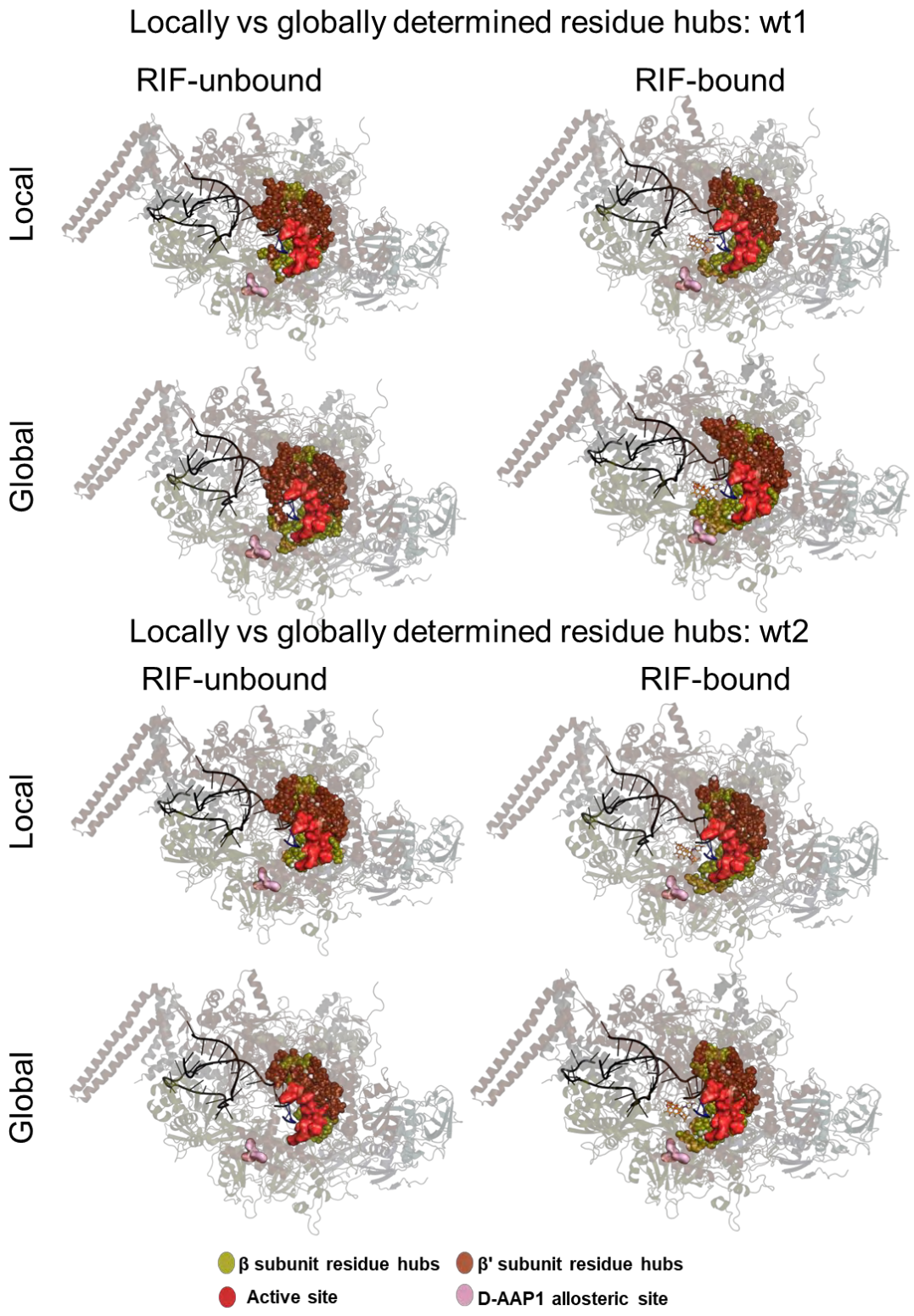
Supplementary Figure S2.5C. Heatmap representation of the raw and ΔKC values for the RIF-unbound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



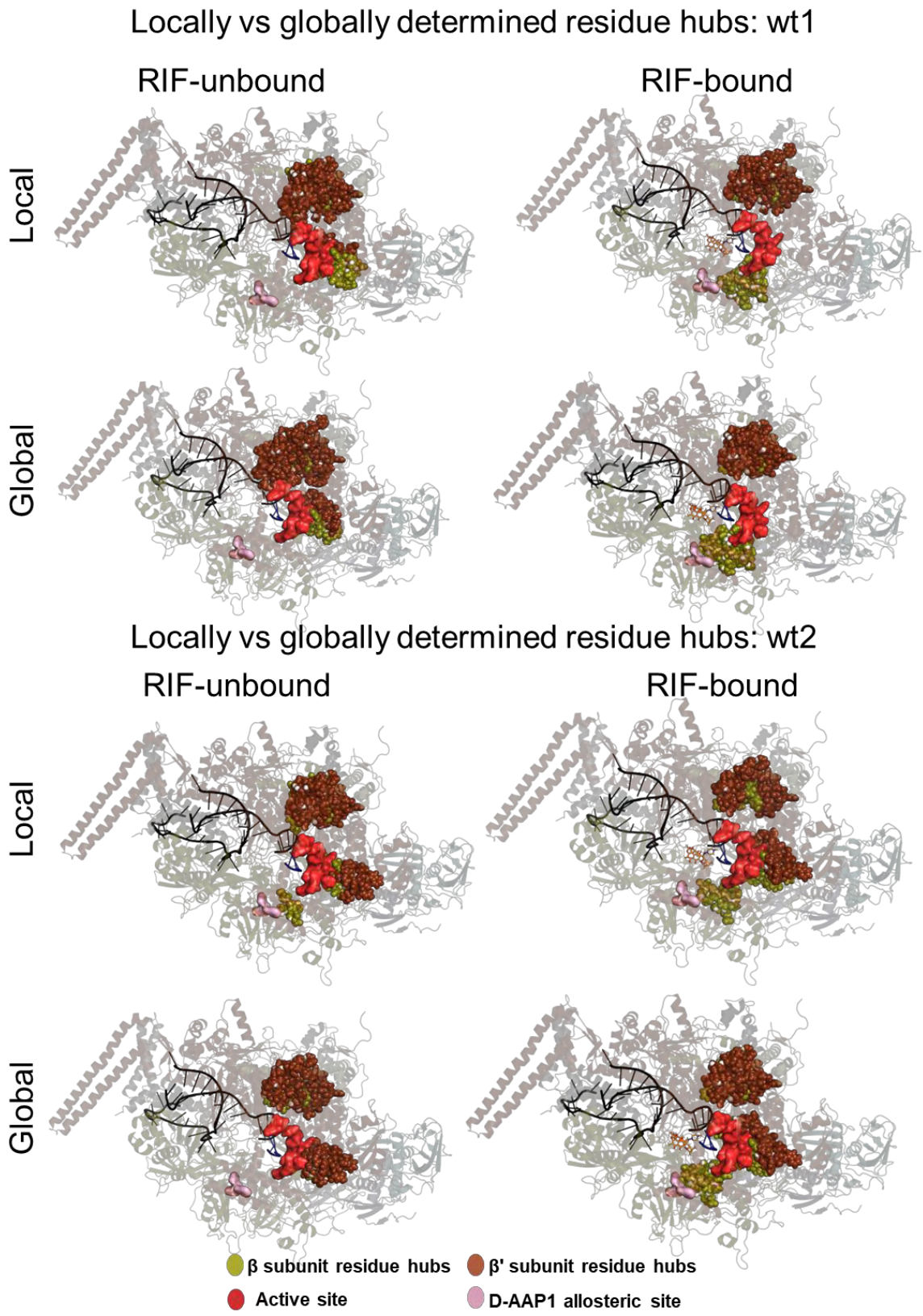
Supplementary Figure S2.5D. Heatmap representation of the raw and Δ*KC* values for the RIF-bound *Mtb*-RNAP β' subunit *wt* standards and the respective *mt*s.



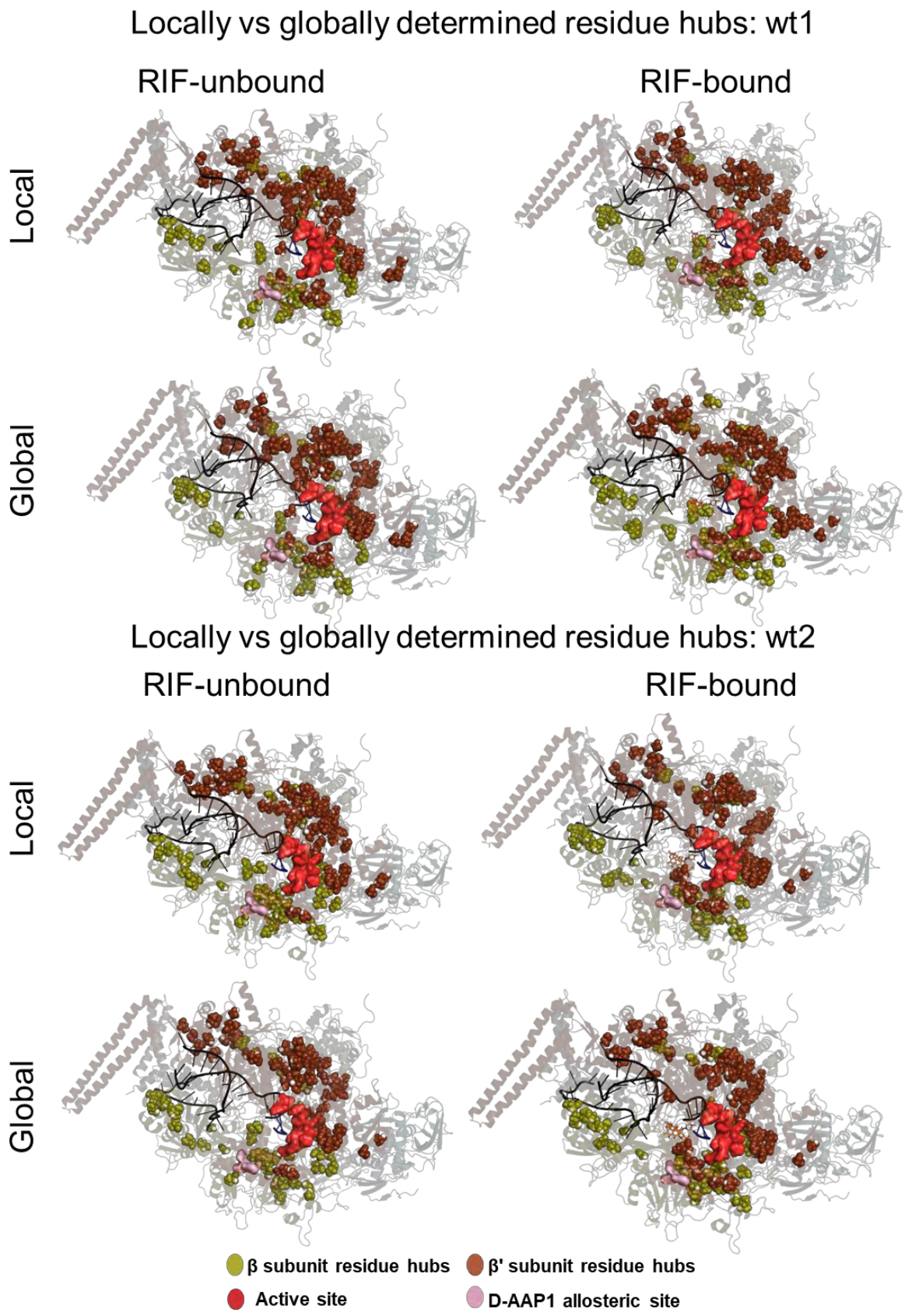
Supplementary Figure S2.6. Comparison of locally and globally determined *BC* hub residues mapped onto *Mtb*-RNAP.



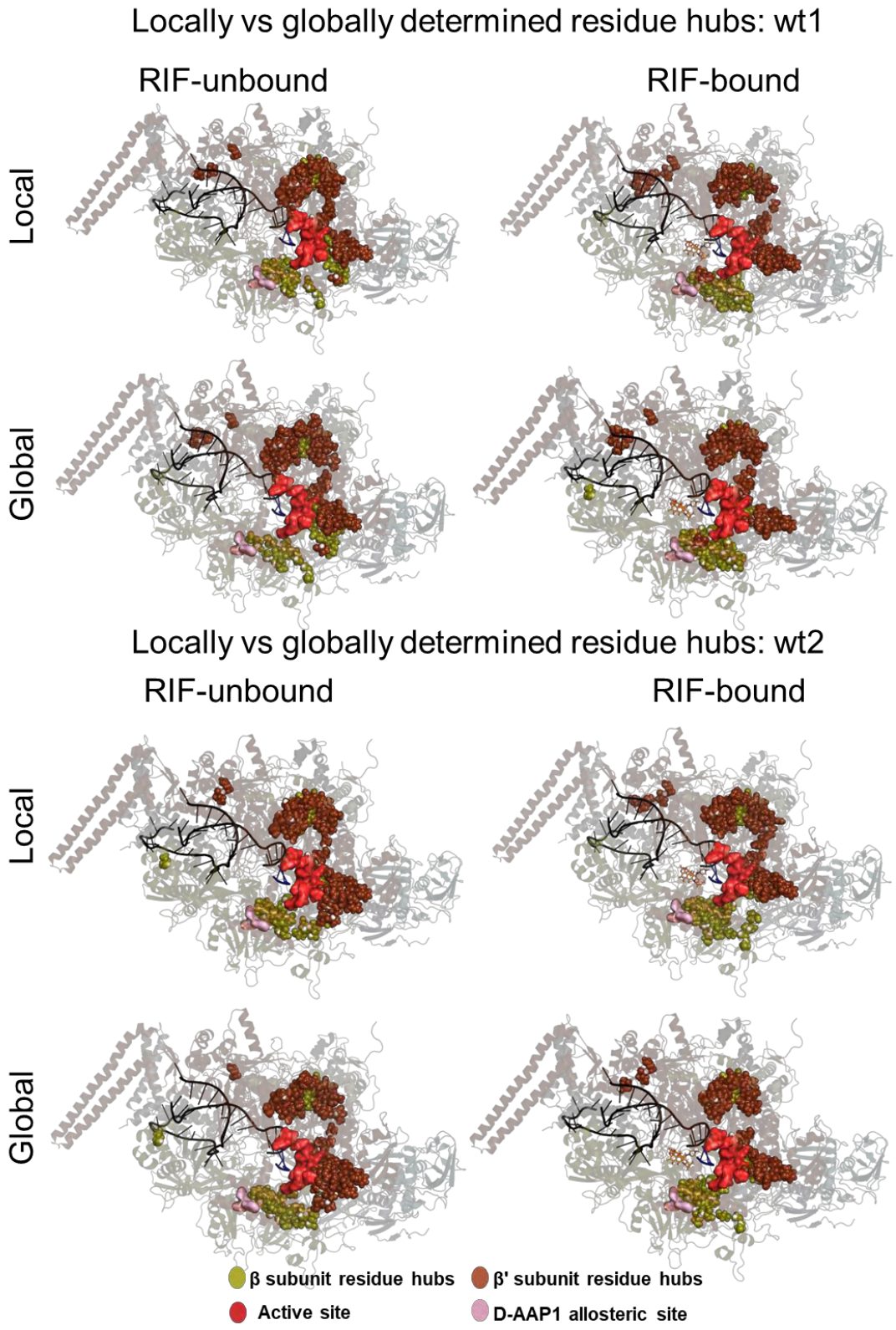
Supplementary Figure S2.7. Comparison of locally and globally determined *CC* hub residues mapped onto *Mtb*-RNAP.



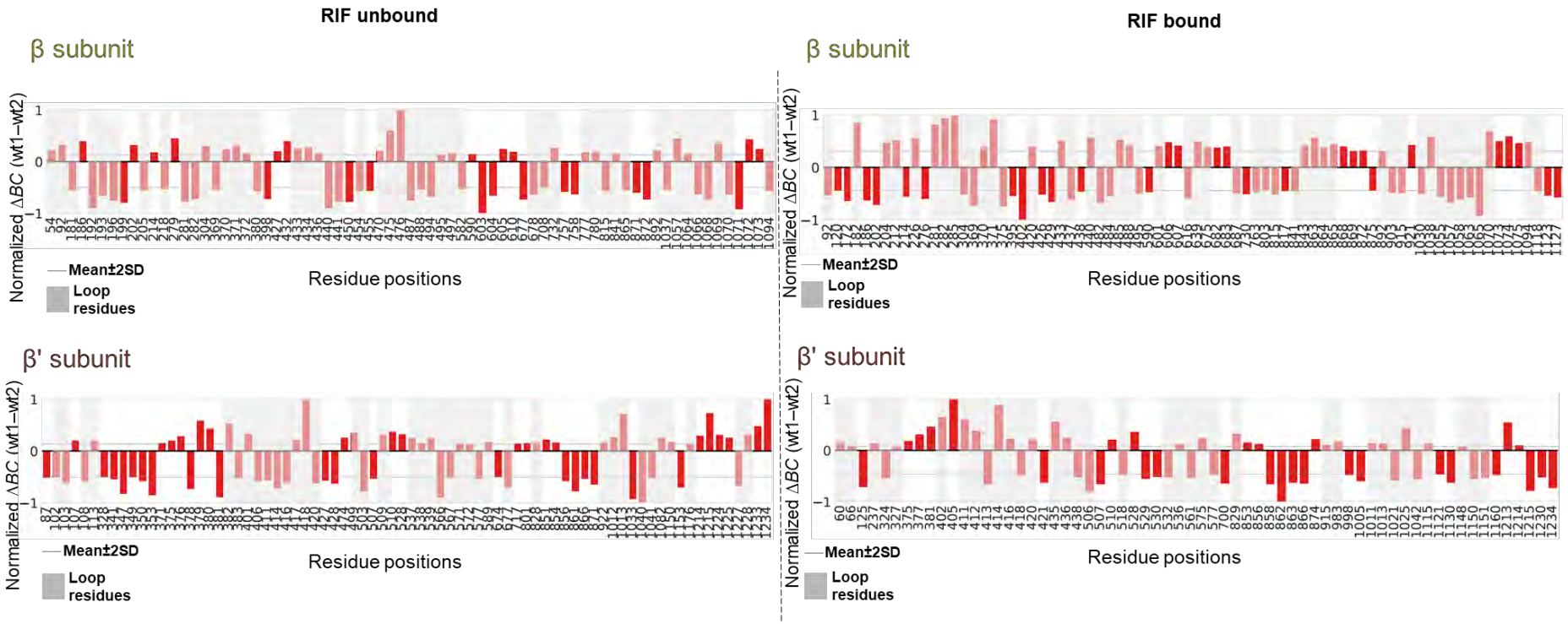
Supplementary Figure S2.8. Comparison of locally and globally determined *EC* hub residues mapped onto *Mtb*-RNAP.



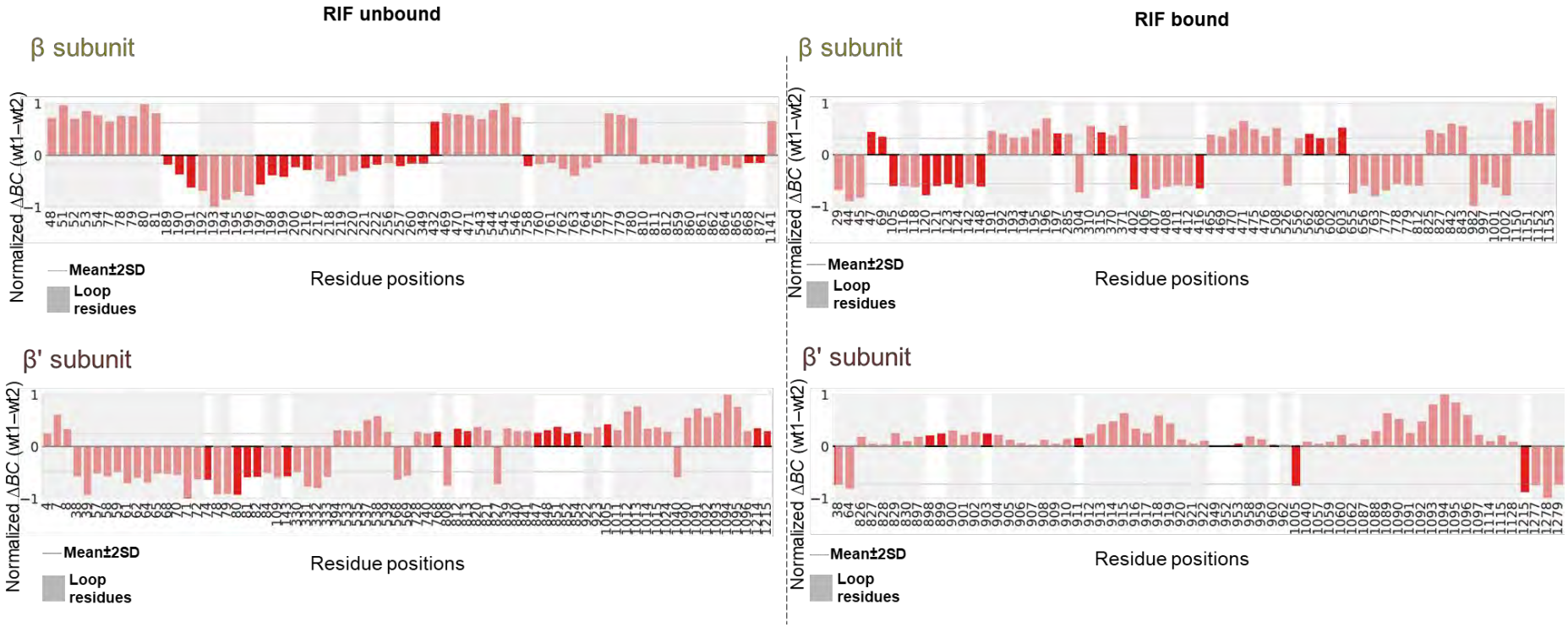
Supplementary Figure S2.9. Comparison of locally and globally determined *DC* hub residues mapped onto *Mtb*-RNAP.



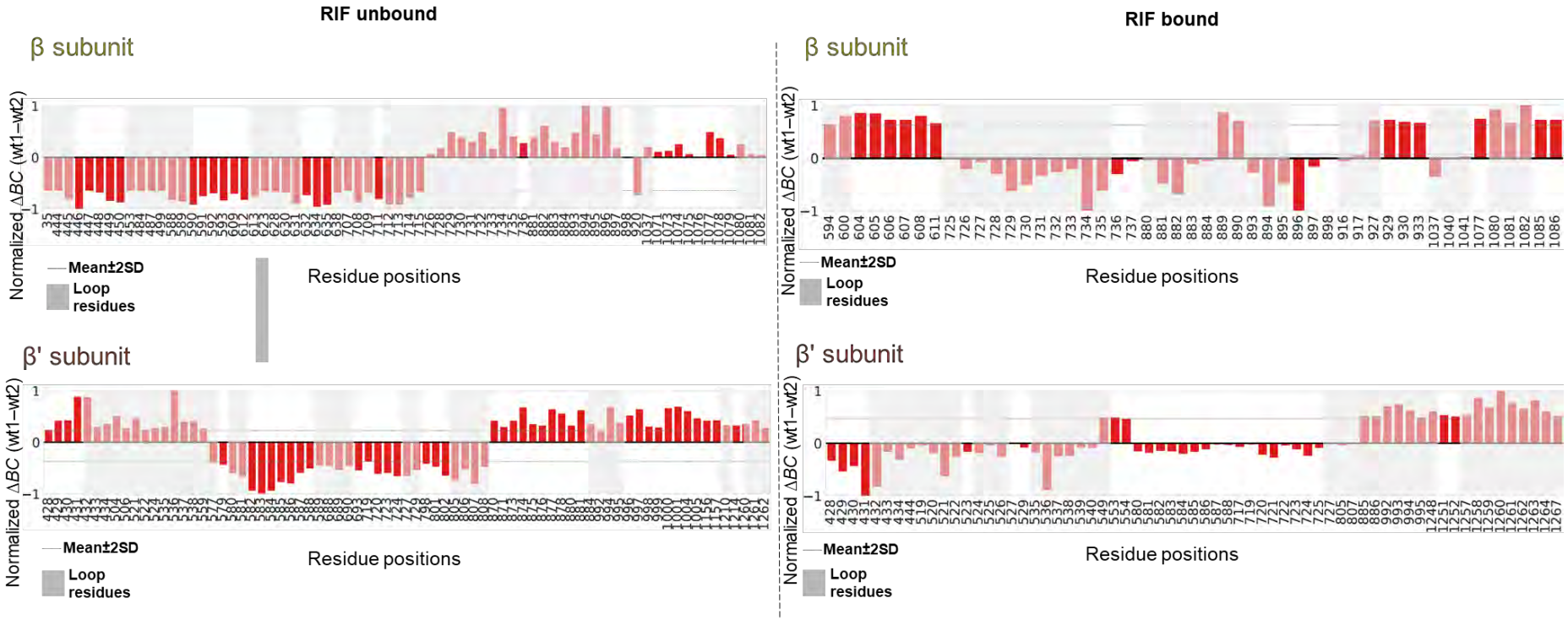
Supplementary Figure S2.10. Comparison of locally and globally determined *KC* hub residues mapped onto *Mtb*-RNAP.



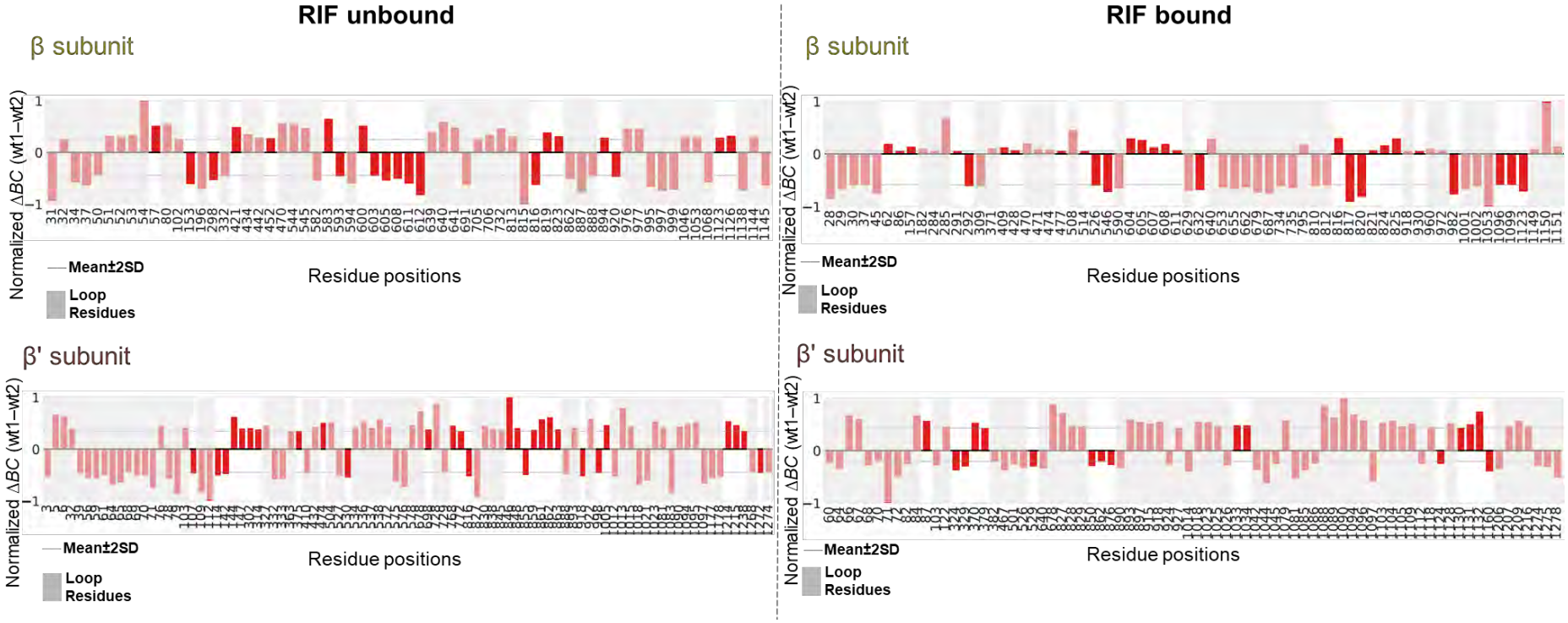
Supplementary Figure S2.11. Comparison of *Mtb*-RNAP β and β' subunits using normalized Δ*BC* (*wt*1–*wt*2) to highlight residues showing significant differences in centrality.



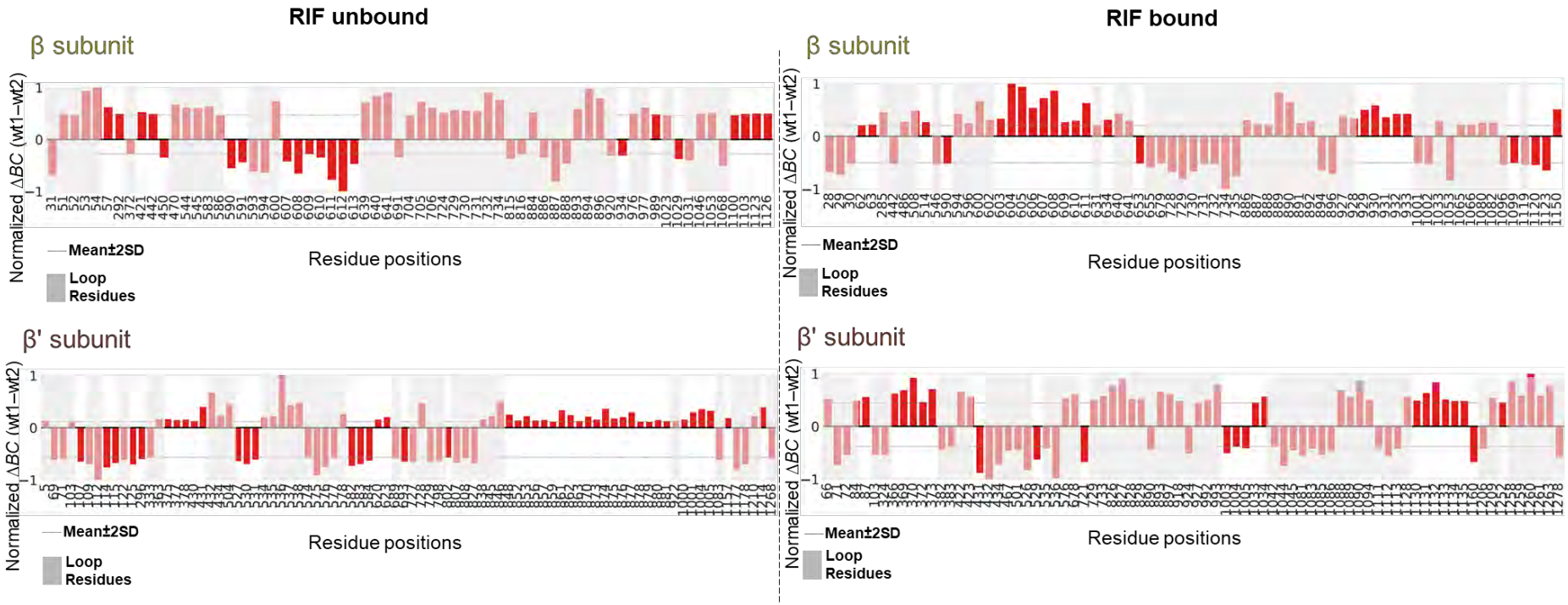
Supplementary Figure S2.12. Comparison of *Mtb*-RNAP β and β' subunits using normalized Δ*CC* (*wt*1–*wt*2) to highlight residues showing significant differences in centrality



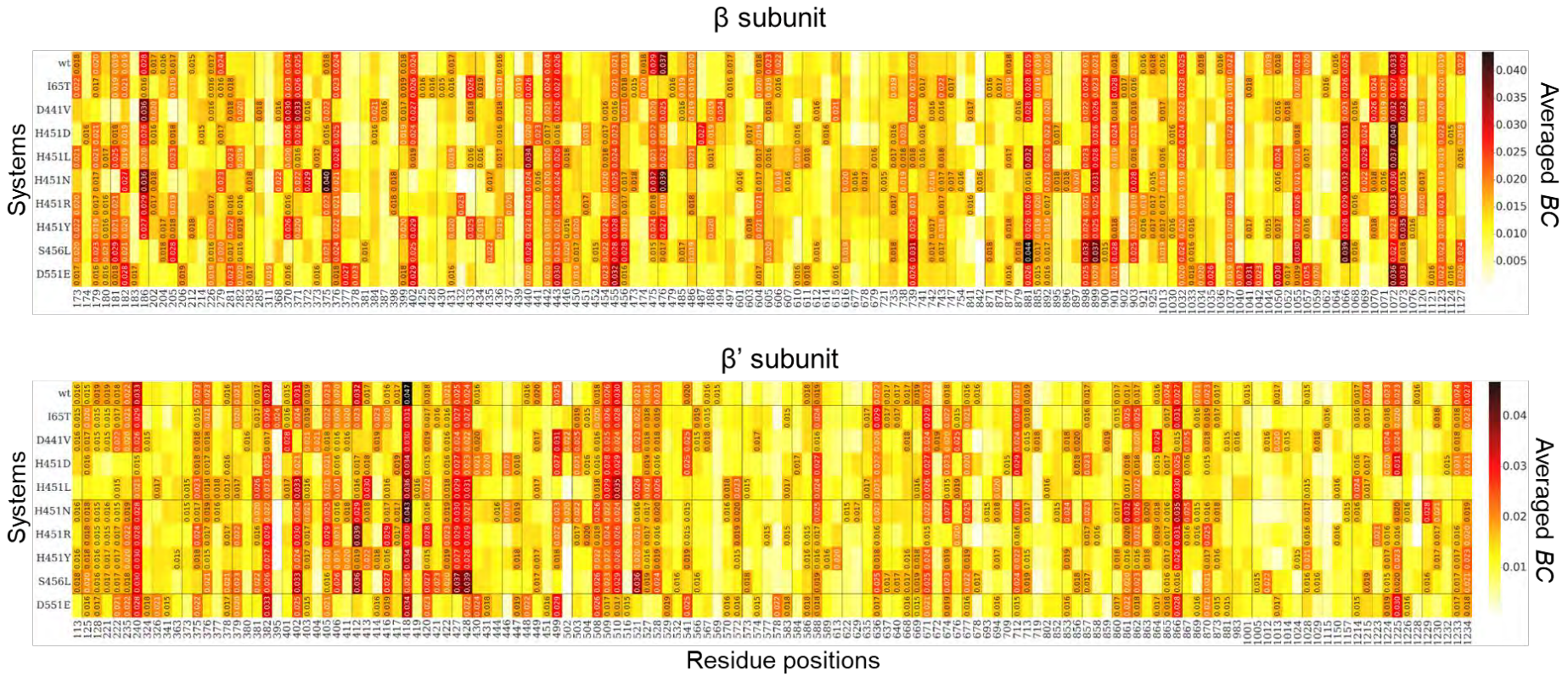
Supplementary Figure S2.13. Comparison of *Mtb*-RNAP β and β' subunits using normalized Δ*EC* (*wt*1–*wt*2) to highlight residues showing significant differences in centrality.



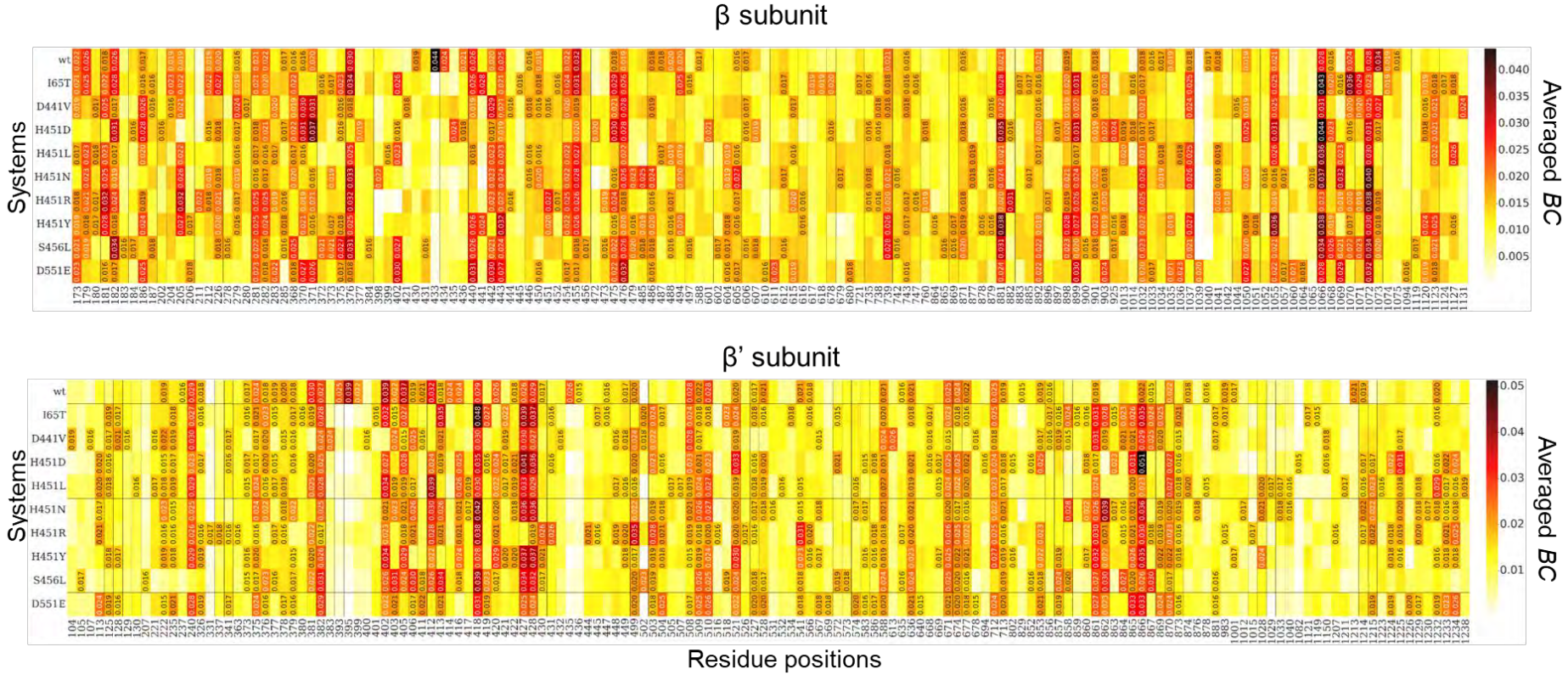
Supplementary Figure S2.14. Comparison of *Mtb*-RNAP β and β' subunits using normalized Δ*DC* (*wt*1–*wt*2) to highlight residues showing significant differences in centrality.



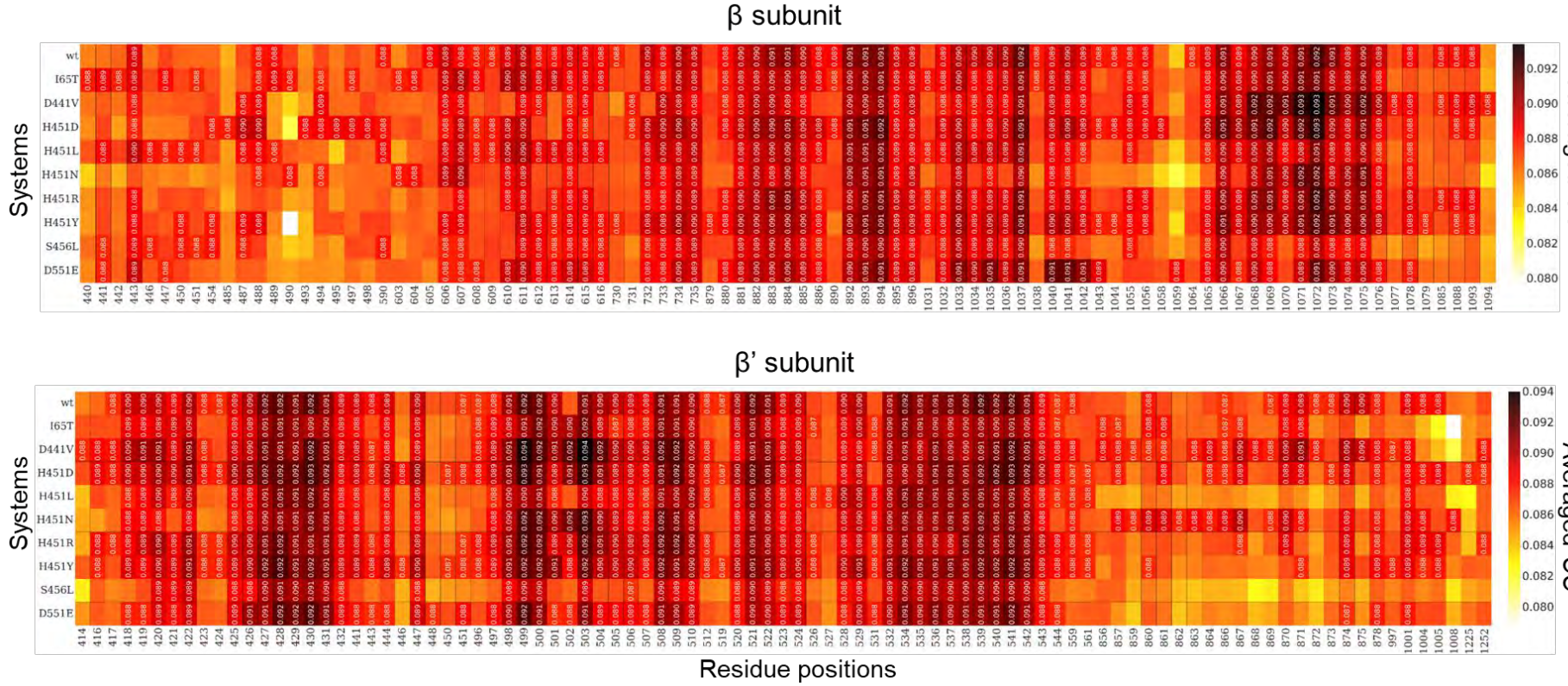
Supplementary Figure S2.15. Comparison of *Mtb*-RNAP β and β' subunits using normalized Δ*KC* (*wt*1–*wt*2) to highlight residues showing significant differences in centrality.



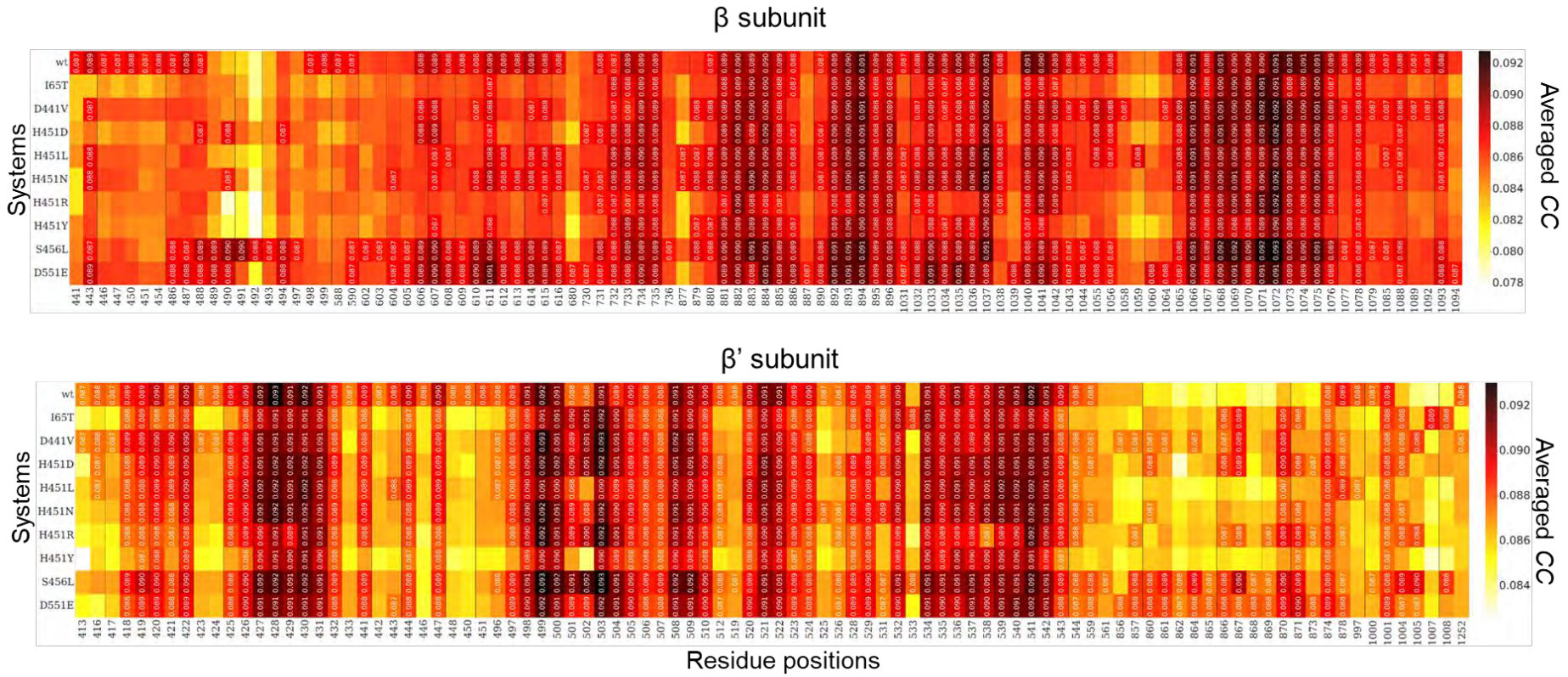
Supplementary Figure S2.16A. *Mtb*-RNAP β and β' global top 5% *BC* hub residues for the RIF-unbound *wt* and *mt* systems.



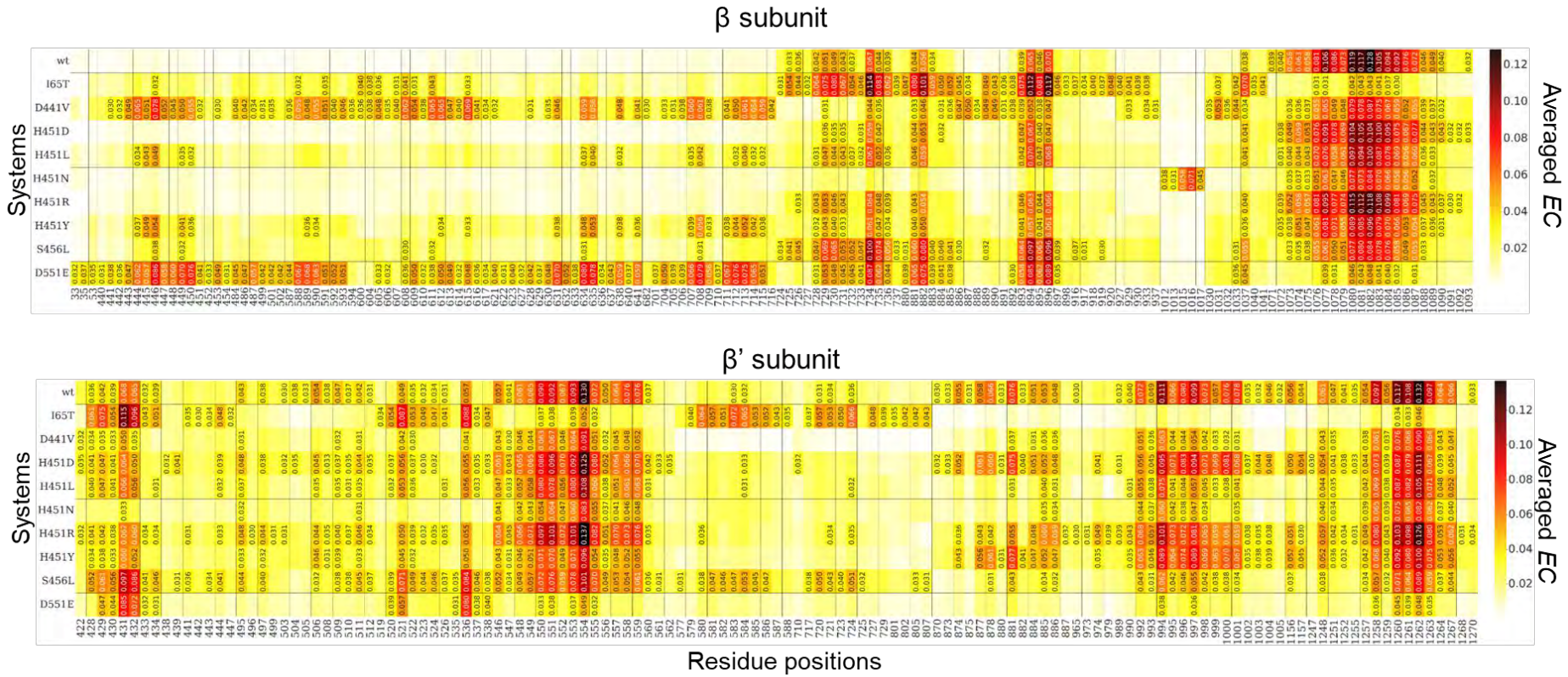
Supplementary Figure S2.16B. *Mtb*-RNAP β and β' global top 5% *BC* hub residues for the RIF-bound *wt* and *mt* systems.



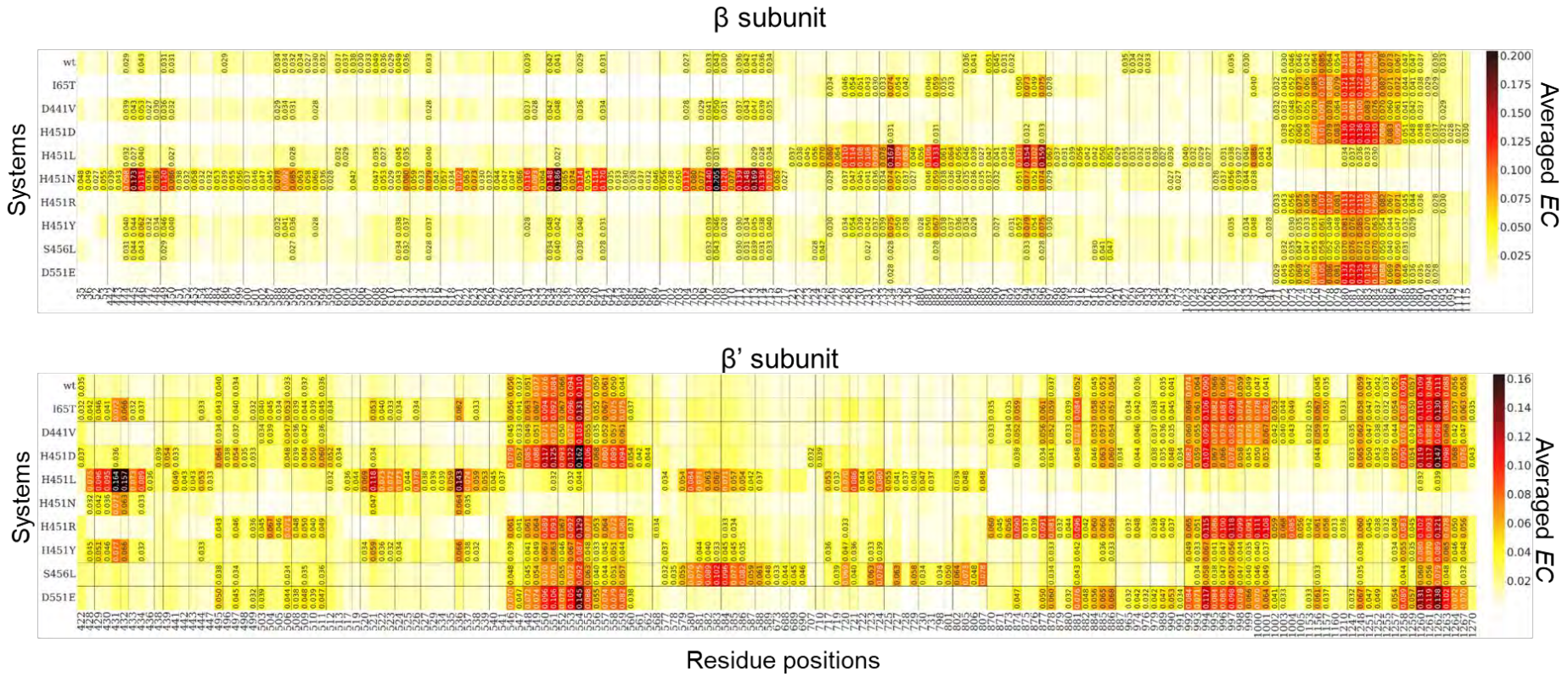
Supplementary Figure S2.17A. *Mtb*-RNAP β and β' global top 5% *CC* hub residues for the RIF-unbound *wt* and *mt* systems.



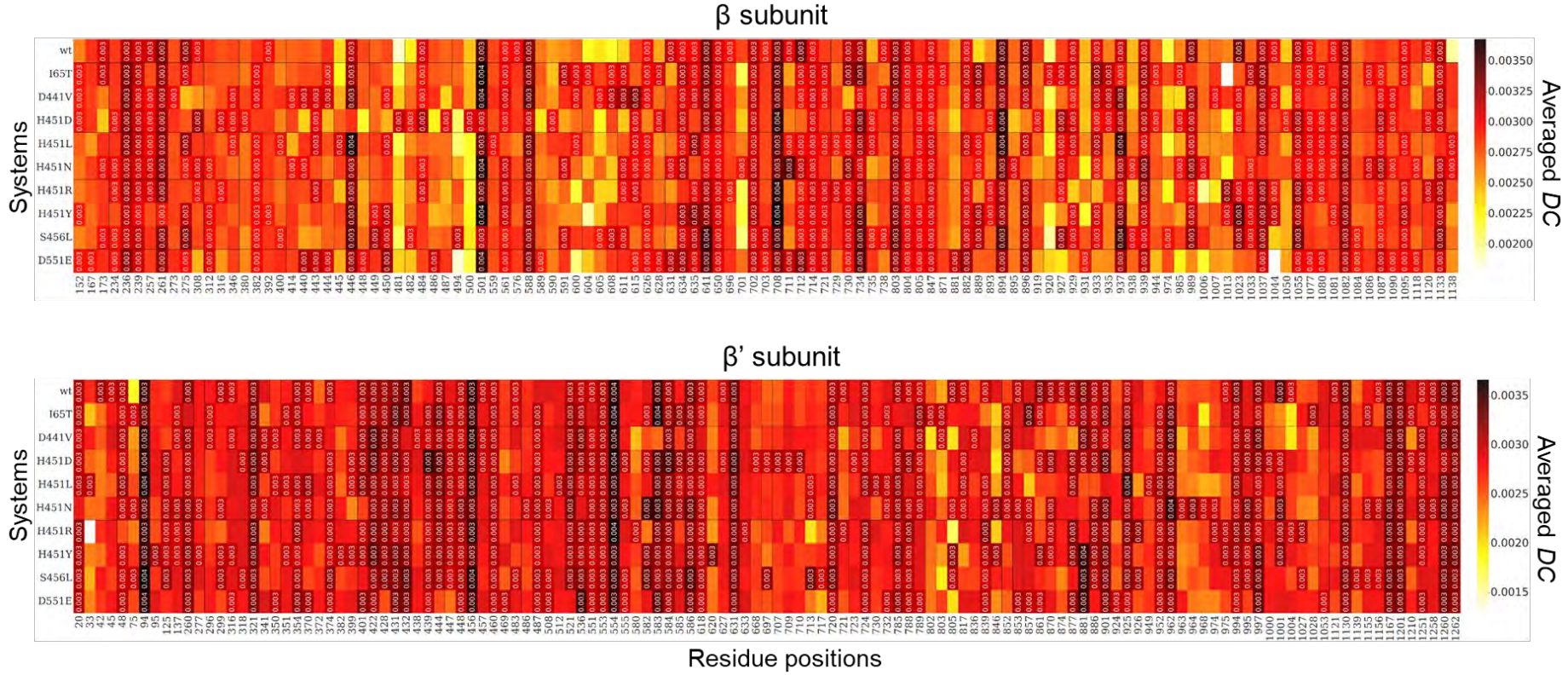
Supplementary Figure S2.17B. *Mtb*-RNAP β and β' global top 5% *CC* hub residues for the RIF-bound *wt* and *mt* systems.



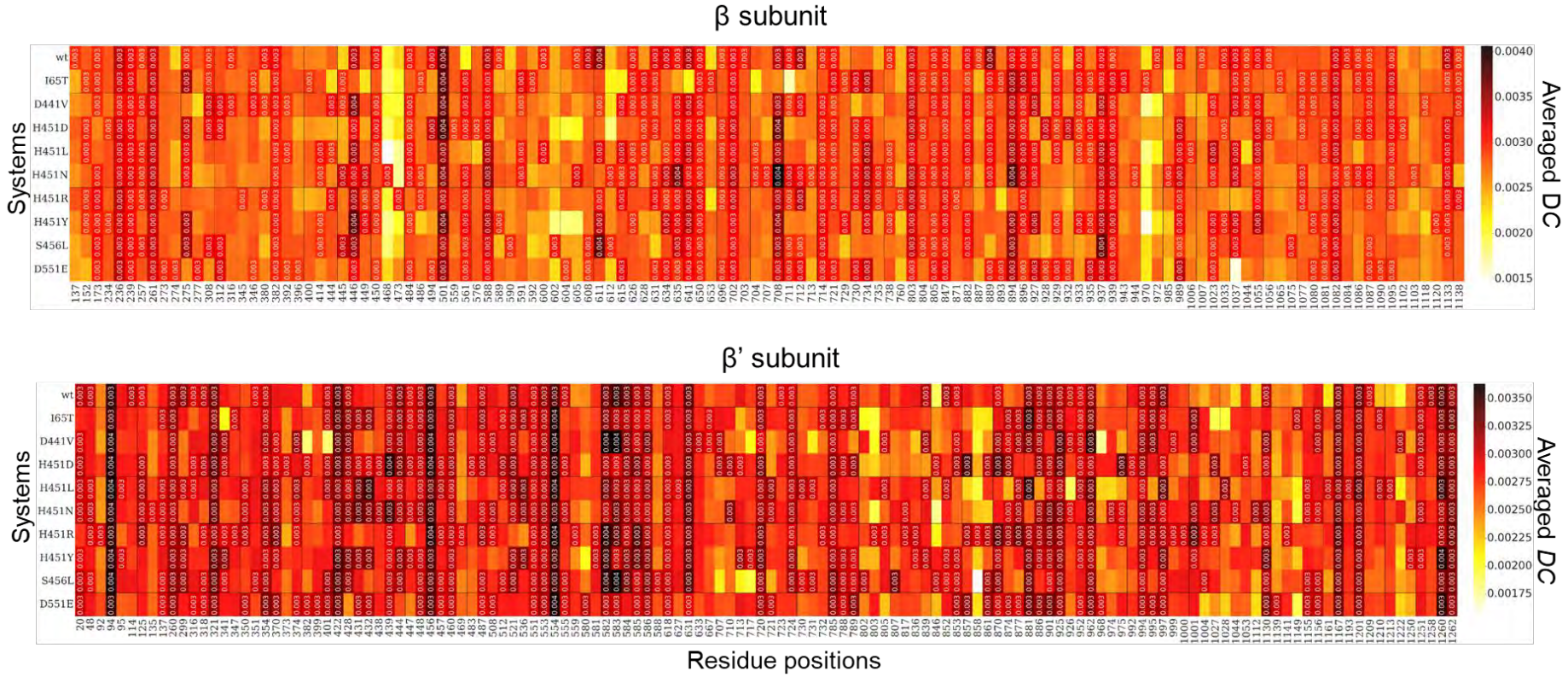
Supplementary Figure S2.18A. *Mtb*-RNAP β and β' global top 5% *EC* hub residues for the RIF-unbound *wt* and *mt* systems.



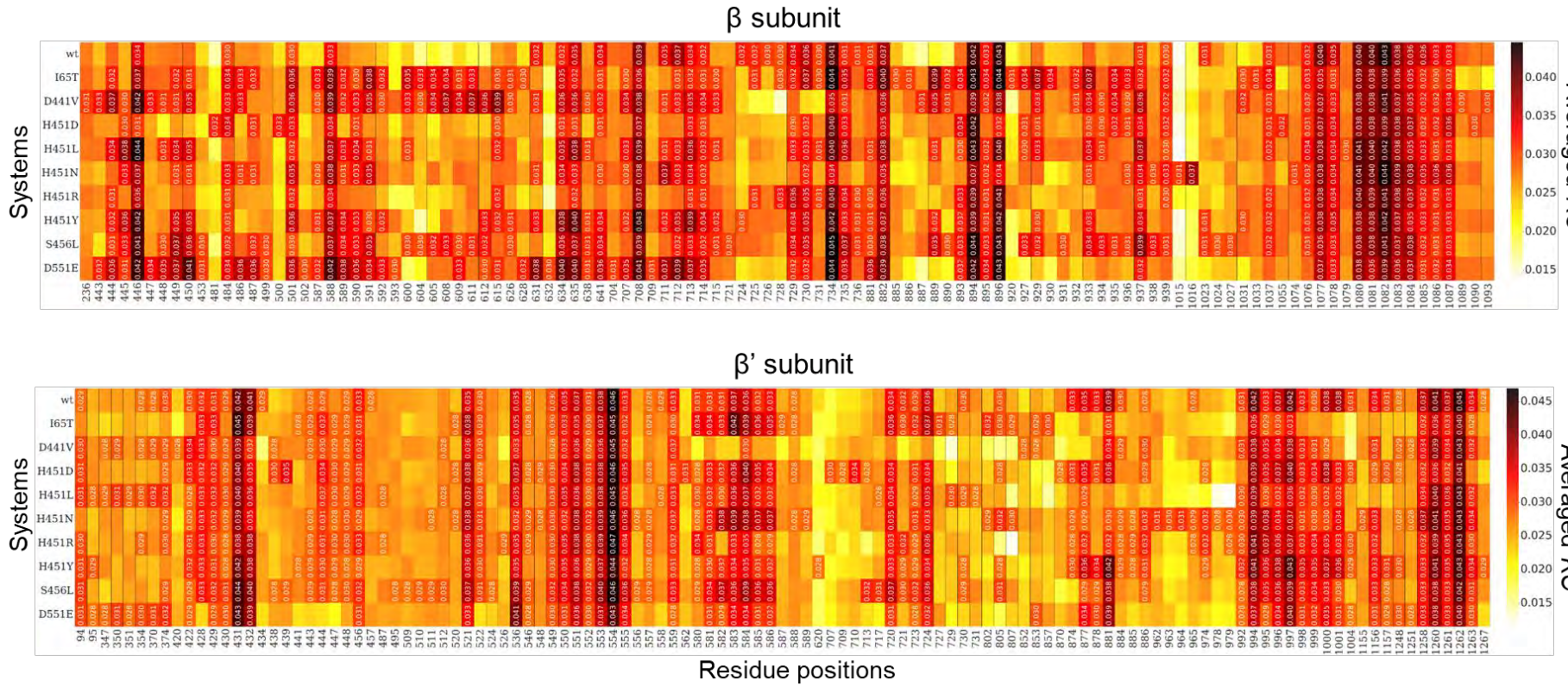
Supplementary Figure S2.18B. *Mtb*-RNAP β and β' global top 5% *EC* hub residues for the RIF-bound *wt* and *mt* systems.



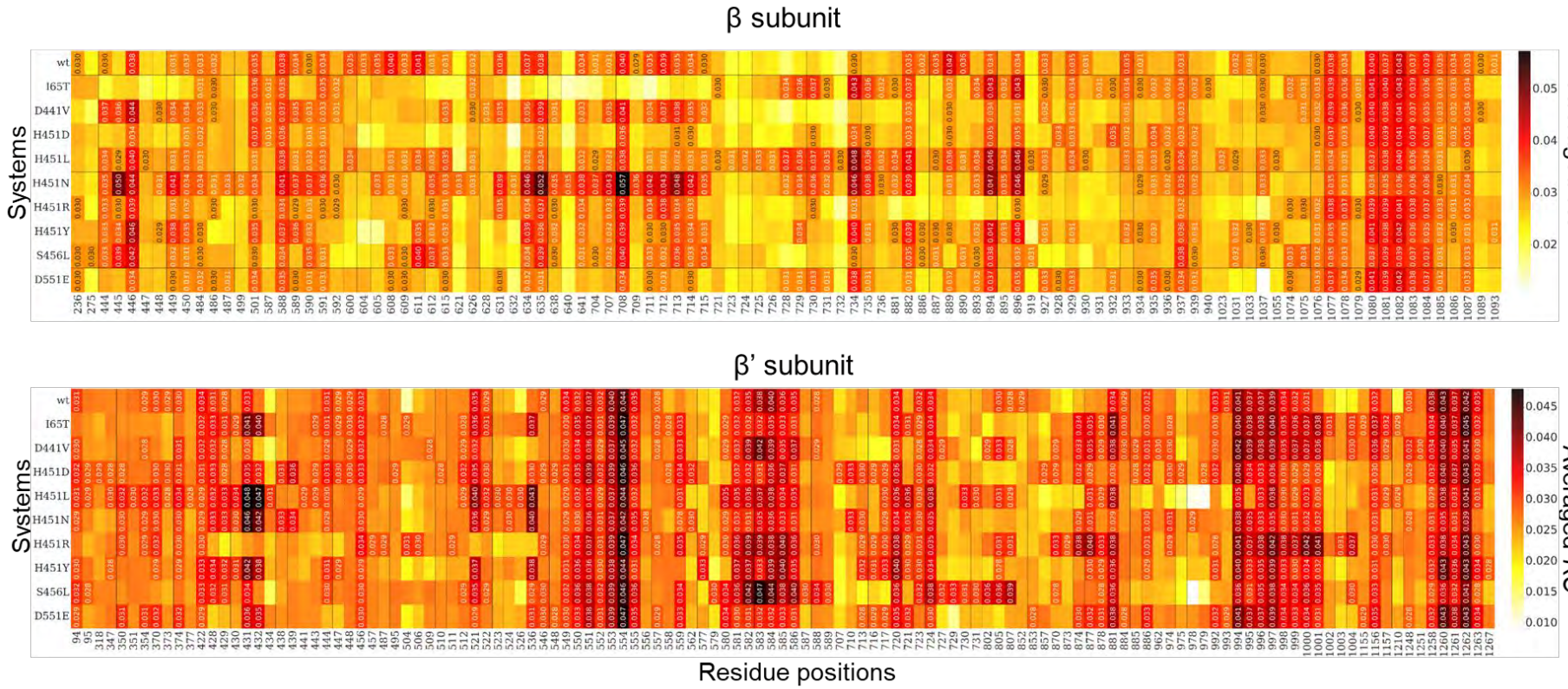
Supplementary Figure S2.19A. *Mtb*-RNAP β and β' global top 5% *DC* hub residues for the RIF-unbound *wt* and *mt* systems.



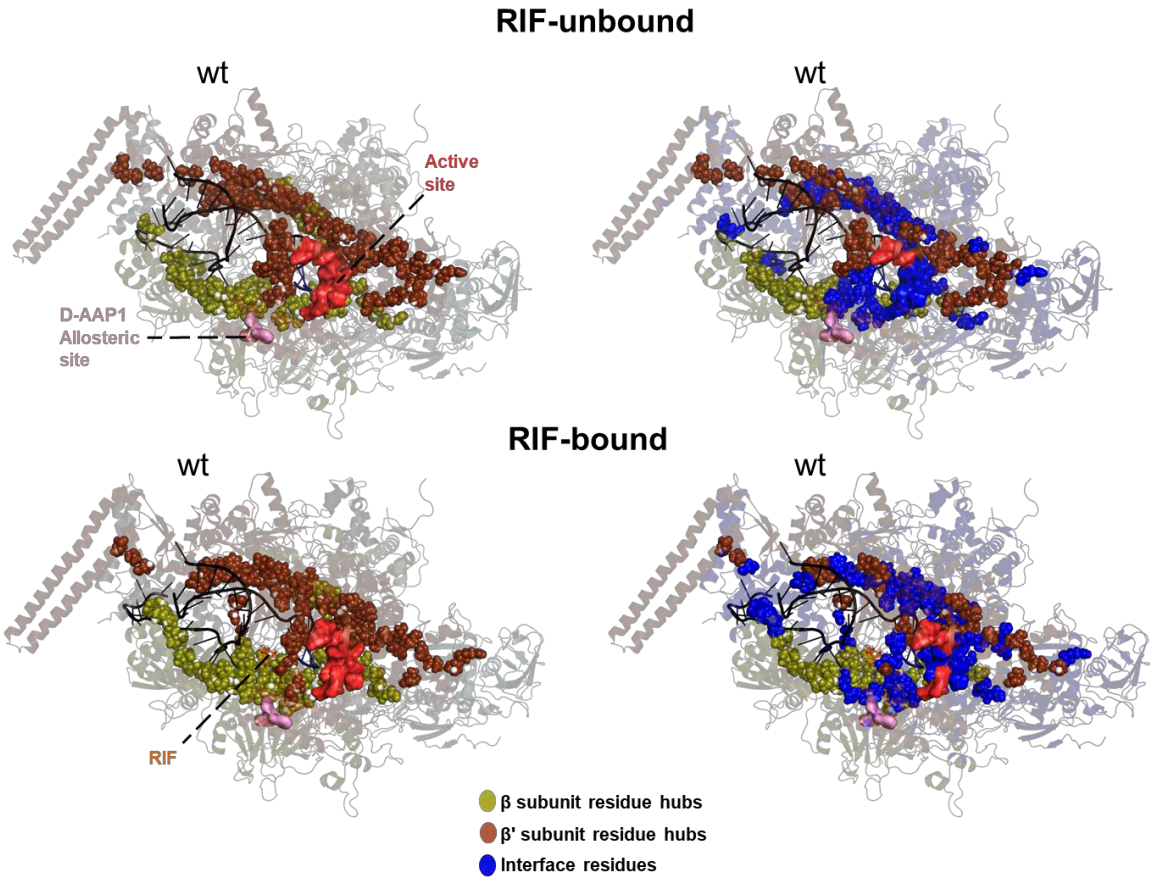
Supplementary Figure S2.19B. *Mtb*-RNAP β and β' global top 5% *DC* hub residues for the RIF-bound *wt* and *mt* systems.



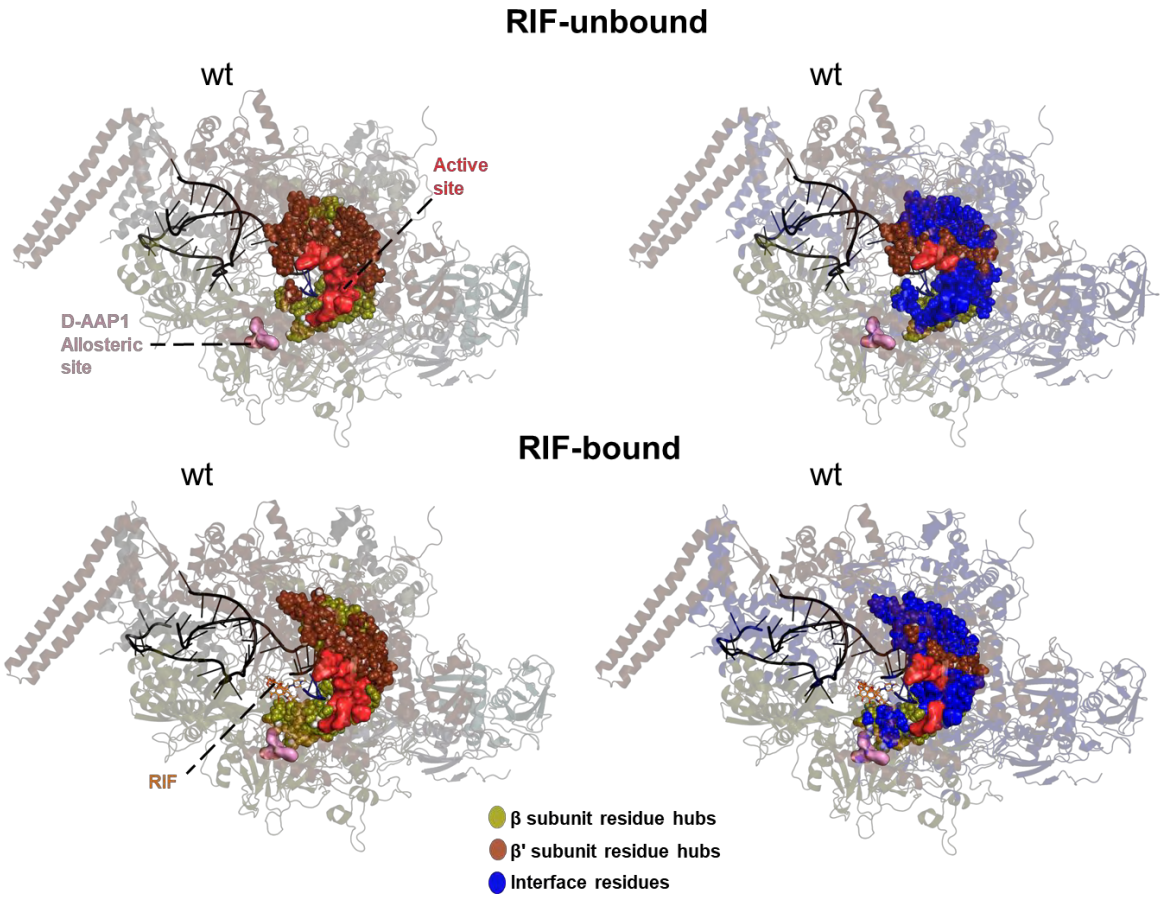
Supplementary Figure S2.20A. *Mtb*-RNAP β and β' global top 5% *KC* hub residues for the RIF-unbound *wt* and *mt* systems.



Supplementary Figure S2.20B. *Mtb*-RNAP β and β' global top 5% *KC* hub residues for the RIF-bound *wt* and *mt* systems.



Supplementary Figure S2.21. 3D structure of wt *Mtb*-RNAP with mapped *BC* hubs and highlighted interface residues as indicated through the PyMol interface script (4 Å cut-off).



Supplementary Figure S2.22. 3D structure of wt *Mtb*-RNAP with mapped *CC* hubs and highlighted interface residues as indicated through the PyMol interface script (4 Å cut-off).

Supplementary Table S2.1. Showing the list of persistent hubs identified through the global 5% calculation applied to the *Mtb*-RNAP β and β' subunits averaged DRN results respectively.

RNAP subunits	***BC* hubs**		***EC* hubs**		***CC* hubs**		***DC* hubs**		***KC* hubs**	
	RIF-unbound systems	RIF-bound systems	RIF-unbound systems	RIF-bound systems	RIF-unbound systems	RIF-bound systems	RIF-unbound systems	RIF-bound systems	RIF-unbound systems	RIF-bound systems
β	442, **443**, **455**, 899, **1032**, **1072**, 1123	182, **443**, **455**, 476, 881, **1032**, 1055, 1066, **1072**	1077, 1080, 1081, 1082, 1083, 1084, 1085		607, 611, 614, **733**, **734**, **735**, **881**, **882**, **883**, **884**, **885**, **892**, **893**, **894**, **895**, **896**, **1033**, 1035, **1037**, **1040**, **1041**, **1066**, **1068**, **1069**, 1071, **1072**, **1073**, **1074**, **1075**	732, **733**, **734**, **735**, **881**, **882**, **883**, **884**, **885**, **892**, **893**, **894**, **895**, **896**, **1033**, **1037**, **1040**, **1041**, **1066**, 1067, **1068**, **1069**, 1070, **1071**, **1072**, **1073**, **1074**, **1075**, 1078	**236**, **239**, **261**, **501**, **588**, 641, **650**, **702**, 708, 714, 734, **803**, **847**, 894, **937**, **939**, 1055, **1082**	**236**, **239**, **261**, 382, **501**, **588**, **650**, **702**, **803**, **847**, **937**, **939**, **1082**	446, **501**, **588**, 635, 708, 713, 714, 734, 882, 894, **896**, **937**, **1077**, **1078**, **1080**, **1081**, **1082**, **1083**, **1084**, **1085**, 1086, **1087**	**501**, **588**, **896**, **937**, **1077**, **1078**, **1080**, **1081**, **1082**, **1083**, **1084**, **1085**, **1087**
β'	240, **382**, 405, **418**,	375, **382**, **418**, **428**,	431, 550, 551, 553,		**420**, **422**, 425, **426**,	419, **420**, **422**, **426**, **427**, **428**,	**94**, **321**, **422**, 431,	**94**, 260, **321**, 354,	429, 431, 432, **456**,	422, **456**, **549**, **550**,

	420, **428**, 508, **509**, 510, **588**, 636, **671**, 862, **866**	**509**, 521, **588**, **671**, 861, **866**	554, 555, 1260, 1261, 1262		**427**, **428**, **429**, **430**, **431**, **432**, **447**, **498**, **499**, **500**, **503**, **506**, **508**, **509**, **510**, **520**, **521**, **522**, **523**, **524**, 528, **529**, **532**, **534**, **535**, **536**, **537**, **538**, **539**, **540**, **541**, **542**, **543**	**429**, **430**, **431**, **432**, 444, **447**, **498**, **499**, **500**, **503**, 504, 505, **506**, 507, **508**, **509**, **510**, **520**, **521**, **522**, **523**, **524**, **529**, **532**, **534**, **535**, **536**, **537**, **538**, **539**, **540**, **541**, **542**, **543**, 874, 1001	448, **456**, **460**, **553**, **554**, 584, **586**, **631**, **720**, **925**, **962**, **994**, 997, 1130, **1167**, **1201**, **1260**, **1262**	**422**, 439, **456**, **460**, 551, **553**, **554**, 582, 583, 585, **586**, **631**, **720**, 785, **925**, **962**, **994**, **1167**, **1201**, **1260**, **1262**	521, 536, **549**, **550**, **551**, **552**, **553**, **554**, **555**, 559, **584**, **994**, 995, **996**, **997**, **1258**, **1260**, **1261**, **1262**	**551**, **552**, **553**, **554**, **555**, 581, 582, 583, **584**, 585, 586, 720, 724, 881, **994**, **996**, **997**, 998, 999, 1000, 1156, **1258**, **1260**, **1261**, **1262**

* ***Bolded values*** *= hubs present in both RIF-unbound and bound system*