IN SILICO CHARACTERIZATION OF MISSENSE MUTATIONS IN INFECTIOUS DISEASES: CASE STUDIES OF TUBERCULOSIS AND COVID-19

by

Victor Barozi

Orcid: 0000-0002-1275-0891

A thesis submitted in conformity with the requirements for the degree of

DOCTOR OF PHILOSOPHY IN BIOINFORMATICS

at

RESEARCH UNIT OF BIOINFORMATICS (RUBi)

DEPARTMENT OF BIOCHEMISTRY AND MICROBIOLOGY

Faculty of Science

RHODES UNIVERSITY

SOUTH AFRICA

SUPERVISOR

Prof Özlem Tastan Bishop

Abstract

One of the greatest challenges facing modern medicine and the global public health today is antimicrobial drug resistance (AMR). This "silent pandemic," as coined by the world health organization (WHO), is steadily increasing with an estimated 4.95 million mortalities attributed to AMR in 2019, 1.27 million of which were directly linked to AMR. Some of the contributors to AMR include self-prescription, drug overuse, sub-optimal drug prescriptions by health workers, and inaccessibility to drugs, especially in remote areas, which leads to poor adherence. The situation is aggravated by the upsurge of new zoonotic infections like the coronavirus disease 2019, which present unique challenges and take the bulk of resources hence stunting the fight against AMR.

Quite alarming still is our current antimicrobial arsenal, which hasn't had any novel antimicrobial drug discovery/addition, of a new class, since the 1980s. This puts a burden on the existing broad-spectrum antimicrobial drugs which are already struggling against multi-drug resistant strains like multi-drug resistant tuberculosis (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB). Besides the search for new antimicrobial agents, the other avenue for addressing AMR is studying drug resistance mechanisms, especially single nucleotide polymorphisms (SNPs), that change drug target characteristics. With the advancement of computational power and data storage resources, computational approaches can be applied in mutational studies to provide insight into the drug resistance mechanisms with an aim to inform future drug design and development.

Therefore, in the first part of this thesis, we employ integrative *in silico* approaches, including 3D structure modeling, molecular dynamic (MD) simulations, comparative essential dynamics (ED), and protein network analysis approaches i.e., dynamic residue network (DRN) analysis to decipher drug resistance mechanisms in tuberculosis (TB). This involved an investigation of the drug resistance mutations in the catalase-peroxidase (KatG) and

pyrazinamidase (MtPncA) enzymes which are responsible for activation of TB first-line drugs; Isoniazid (INH) and Pyrazinamide (PZA), respectively. In the case of KatG, eleven high confidence (HC) KatG mutations associated with a high prevalence of phenotypic INH resistance were identified and their 3D structures modeled before subjecting them to MD simulations. Global analysis showed an unstable KatG structure and active site environment in the mutants compared to the wildtype. Active site dynamics in the mutants compromised cofactor (heme) interactions resulting in less bonds/interactions compared to the wildtype. Given the importance of the heme, reduced interactions affect enzyme function. Trajectory analysis also showed asymmetric protomer behavior both in the wildtype and mutant systems. DRN analysis identified the KatG dimerization domain and C-terminal domain as functionally important and influential in the enzyme function as per *betweenness centrality* and *eigenvector centrality* distribution.

In the case of the MtPncA enzyme, our main focus was on understanding the MtPncA binding ability of Nicotinamide (an analogue of PZA) in comparison to PZA, especially in the presence of 82 resistance conferring MtPncA mutations. Like in KatG, the mutant structures were modeled and subjected to MD simulations and analysis. Interestingly, more MtPncA mutants favored NAM interactions compared to PZA i.e., 34 MtPncA mutants steadily coordinated NAM compared to 21 in the case of PZA. Trajectory and ligand interaction analysis showed how increased active site lid loop dynamics affect the NAM binding, especially in the systems with the active site mutations i.e., H51Y, W68R, C72R, L82R, K96N, L159N, and L159R. This led to fewer protein-ligand interactions and eventually ligand ejection. Network analysis further identified the protein core, metal binding site (MBS), and substrate binding site as the most important regions of the enzyme. Furthermore, the *degree of centrality* analysis showed how specific MtPncA mutations i.e.,

C14H, F17D, and T412P, interrupt intra-protein communication from the MtPncA core to the MBS, affecting enzyme activity.

The analysis of KatG and MtPncA enzyme mutations not only identified the effects of mutations on enzyme behaviour and communication, but also established a framework of computational approaches that can be used for mutational studies in any protein.

Besides AMR, the continued encroachment of wildlife habitats due to population growth has exposed humans to wildlife pathogens leading to zoonotic diseases, a recent example being coronavirus disease 2019 (COVID-19). In the second part of the thesis, the established computational approaches in Part 1, were employed to investigate the changes in inter-protein interactions and communication patterns between the severe acute respiratory coronavirus 2 (SARS-CoV-2) with the human host receptor protein (ACE2: angiotensin-converting enzyme 2) consequent to mutations in the SARS-CoV-2 receptor binding domain (RBD). Here, the focus was on RBD mutations of the Omicron sub-lineages. We identified four Omicron-sub lineages with RBD mutations i.e., BA.1, BA.2, BA.3 and BA.4. Each sub-lineage mutations were modeled into RBD structure in complex with the hACE2. MD analysis of the RBD-hACE2 complex highlighted how the RBD mutations change the conformational flexibility of both the RBD and hACE2 compared to the wildtype (WT). Furthermore, DRN analysis identified novel allosteric paths composed of residues with high *betweenness* and *eigenvector centralities* linking the RBD to the hACE2 in both the wildtype and mutant systems. Interestingly, these paths were modified with the progression of Omicron sub-lineages, highlighting how the virus evolution affects protein interaction.

Lastly, the effect of mutations on S RBD and hACE2 interaction was investigated from the hACE2 perspective by focusing on mutations in the hACE2 protein. Here, naturally occurring hACE2 polymorphisms in African populations i.e., S19P, K26R, M82I, K341R, N546D, and

D597Q, were identified and their effects on RBD-hACE2 interactions investigated in presence of the Omicron BA.4/5 RBD mutations. The hACE2 polymorphisms subtly affected the complex dynamics; however, RBD-hACE2 interaction analysis showed that hACE2 mutations effect the complex formation and interaction. Here, the K26R mutation favored RBD-hACE2 interactions, whereas S19P resulted in fewer inter-protein interactions than the reference system. The M82I mutation resulted in a higher RBD-hACE2 binding energy compared to the wildtype meaning that the mutation might not favor RBD binding to the hACE2. On the other hand, K341R had the most RBD-hACE2 interactions suggesting that it probably favors RBD binding to the hACE2. N546D and D597Q had diminutive differences to the reference system. Interestingly, the network of high *betweenness centrality* residues linking the two proteins, as seen in the previous paragraph, were maintained/modified in presence of hACE2 mutations. HACE2 mutations also changed the enzyme network patterns resulting in a concentration of high *eigenvector centrality* residues around the zinc-binding and active site region, ultimately influencing the enzyme functionality.

Altogether, the thesis highlights fundamental structural and network changes consequent to mutations both in TB and COVID-19 proteins of interest using *in silico* approaches. These approaches not only provide a new context on impact of mutations in TB and COVID target proteins, but also presents a framework that be implemented in other protein mutation studies.

Declaration

I, Victor Barozi, declare that this is an original description of my research in the context of my degree of Doctor of Philosophy at Rhodes University and not submitted for any previous degree. Any collaborations and contributions have been clearly indicated and acknowledged.

Signature: …….

Acknowledgement

I would like to extend my sincere gratitude to my supervisor, Prof Özlem Tastan Bishop for always believing in me, her invaluable guidance and support offered throughout the course of my PhD Journey. I am forever grateful.

Further appreciation to the Center for High Performance Computing (CHPC), South Africa for providing the computational resources needed for this work.

I would also extend my sincerest gratitude to my family, Bridget Tusiime, Caroline Barozi and Ronald Barozi for the never-ending support and belief in me. Thank you.

Lastly, I am eternally grateful to my friends and all members of the Research Unit in Bioinformatics (RUBi), Rhodes University. The memories we made will forever stay with me.

The TB work was funded through the Grand Challenges Africa program [GCA/DD/rnd3/023]. Grand Challenges Africa is a program of the African Academy of Sciences (AAS) implemented through the Alliance for Accelerating Excellence in Science in Africa (AESA) platform, an initiative of the AAS and the African Union Development Agency (AUDA-NEPAD). GC Africa is supported by the Bill & Melinda Gates Foundation (BMGF), Swedish International Development Cooperation Agency (SIDA), German Federal Ministry of Education and Research (BMBF), Medicines for Malaria Venture (MMV), and Drug Discovery and Development Centre of University of Cape Town (H3D).

The COVID-19 research was supported by Funding for COVID-19 Research and Development Goals for Africa Programme (Grant No: SARSCov2-2-20-002) of the African Academy of Sciences (AAS). It is implemented through the Alliance for Accelerating Excellence in Science in Africa (AESA) platform, an initiative of the AAS and the African

Union Development Agency (AUDA-NEPAD). It was also supported by the South African Medical Research Council under a Self-Initiated Research Grant awarded to A.L.E.

Dedication

To Bridget, for the unwavering love, encouragement and support throughout this academic journey. The sacrifices and guidance have made this milestone possible. This thesis is dedicated to you.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Supplementary Figures

List of Supplementary Tables

List of Abbreviations

Abbreviation	Definition
3D	Three-dimensional
BC	*Betweenness centrality*
BCE	Before Common Era
CC	*Closeness centrality*
CHPC	Centre for high performance computing
COVID-19	Corona virus disease 2019
DC	*Degree of centrality*
DRN	Dynamic residue network
DRAGdb	Drugs resistance gene associated database
DOF	Degrees of freedom
DOPE	Discrete optimized protein energy
EC	*Eigenvector centrality*
ED	Essential dynamics
FDA	Food and drug authority
GISAID	Global initiative on sharing all influenza data
GROMACS	GROningen MAchine for Chemical Simulations
GMTV	Genome-wide Mycobacterium tuberculosis variation
gnomAD	Genome Aggregation Database
INH	Isoniazid
KatG	Catalase peroxidase
KC	*Katz centrality*
Mtb	*Mycobacterium tuberculosis*
MD	Molecular dynamics
MDR-TB	Multi drug resistant tuberculosis
NAD	Nicotinamide adenine dinucleotide
PANGOLIN	Phylogenetic assignment of named global outbreak lineages
PCA	Principal component analysis
PZA	Pyrazinamide
RR-TB	Rifampicin resistant tuberculosis
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
Rg	Radius of gyration
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
TB	Tuberculosis
WHO	World health organisation
WT	Wildtype
XDR-TB	Extensively drug resistant TB

List of Amino Acids

Single Letter Code	Three Letter Code	Amino Acid
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

List of Tools and Web servers used

Web server	URL
GISAID	https://gisaid.org/
gnomAD	https://gnomad.broadinstitute.org/
MD-TASK	https://md-task.readthedocs.io/en/latest/home.html
MDM-TASK	https://mdmtaskweb.rubi.ru.ac.za/
RSCB PDB	https://www.rcsb.org/
ProSA	https://prosa.services.came.sbg.ac.at/prosa.php
VERIFY3D	https://saves.mbi.ucla.edu/

Research Outputs

Primary Publications

1. **Barozi Victor**, Musyoka Thommas M, Sheik Amamuddy O, Tastan Bishop Ö. Deciphering Isoniazid Drug Resistance Mechanisms on Dimeric Mycobacterium tuberculosis KatG via Post-molecular Dynamics Analyses Including Combined Dynamic Residue Network Metrics. *ACS Omega* 2022 April 7 doi.org/10.1021/acsomega.2c01036.
2. **Barozi Victor**, Adrienne Edkins, Tastan Bishop Ö. 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* 2022 August 7 doi.org/10.1016/j.csbj.2022.08.015.
3. Tastan Bishop Ö, Musyoka Thommas M, **Barozi Victor**. Allostery and Missense Mutations as Intermittently Linked Promising Aspects of Modern Computational Drug Discovery. *Journal of Molecular Biology* 2022 April 22 doi.org/10.1016/j.jmb.2022.167610.

Secondary Publications

4. Okeke CJ, Musyoka TM, Sheik Amamuddy O, **Barozi Victor**, Tastan Bishop Ö Allosteric pockets and dynamic residue network hubs of falcipain 2 in mutations including those linked to artemisinin resistance. *Computational and Structural Biotechnology Journal* 2021 October 8 doi: 10.1016/j.csbj.2021.10.011.
5. Sheik Amamuddy O, Boateng RA, **Barozi Victor**, Nyamai DW, Tastan Bishop Ö. Novel dynamic residue network analysis approaches to study allosteric modulation:

SARS-CoV-2 M^{pro} and its evolutionary mutations as a case study. *Computational and Structural Biotechnology Journal* 2021 November 25 doi.org/10.1016/j.csbj.2021.11.016.

Conference Presentations

Barozi Victor, Adrienne Edkins, **Tastan Bishop** Ö. Evolutionary progression of collective mutations in Omicron sub-lineages towards efficient RBD-hACE2: Allosteric communications between and within viral and human proteins, Grand Challenges Annual Meeting 2022 Brussels, Belgium (Poster presentation)

Thesis Overview

The research presented in this thesis is divided into two main parts, I and II. Part I of the thesis encompasses chapters 2 and 3, while Part II is made up of chapters 4 and 5. **Chapter 1** is the introductory chapter and it includes an introduction to infectious diseases, the global prevalence and management of infectious diseases, the missense mutations in infectious diseases and current research on missense mutation studies in infectious diseases. The chapter also highlights the gaps in the research before addressing the aims and objectives of the thesis. Furthermore, chapter 1 explains in detail the *in silico* approaches used throughout the thesis.

Part I of the thesis presents the effects of TB drug resistance missense mutations in two separate chapters i.e., chapter 2 and chapter 3. The work in **chapter 2** is a continuation of my mini thesis for the one year MSc degree in Bioinformatics and Computational Biology [1]. The research in chapter 2 describes the effects of isoniazid drug resistant mutations on the dynamics and network patterns of the dimeric KatG protein. In this chapter, we employ integrative bioinformatics approaches including molecular dynamic (MD) simulations, comparative essential dynamics (ED) and dynamic residue network (DRN) analysis to characterize the effects of the high confidence mutations.

Chapter 3 employs both MD simulations and DRN analysis to study the effect of 82 Pyrazinamide (PZA) drug resistance mutations in the Pyrazinamidase (MtPncA) protein with respect to Nicotinamide (NAM). Furthermore, the MtPncA-NAM complex behaviour in the presence of mutations was compared to that of the MtPncA-PZA to identify differences in drug affinity and binding in a bid to inform future drug design.

Part II of the thesis focuses on the identification of the effects of naturally occurring mutations in the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) receptor

binding domain (RBD) (chapter 4) and human angiotensin converting enzyme 2 (hACE2) (chapter 5) on the RBD-hACE2 inter-protein interactions. **Chapter 4** characterizes the behaviour and implications of the progressive Omicron sub-lineage mutations in the SARS-CoV-2 spike (S) protein RBD on the RBD-hACE2 behaviour in comparison to the reference system (Wuhan sequence). Here, the Omicron RBD mutations are investigated in complex with the human receptor, hACE2, using MD simulations and trajectory analysis tools including DRN analysis to identify changes in communication patterns within the proteins. For each analysis, the mutant systems are compared to the reference structure.

Chapter 5 is complimentary to Chapter 4 and focuses on the BA.4/5 Omicron RBD and its interaction with the different naturally occurring polymorphisms of the hACE2. In this chapter we still employ MD simulations and DRN analysis to describe the effects of hACE2 polymorphisms on SARS-CoV-2 Spike protein binding and interaction.

Finally, the last sections present the wholesome summary of the research results, concluding remarks, limitations and the possible future work.

CHAPTER 1

Introduction

1.1 Infectious Diseases

Communicable/infectious diseases are an ensemble of infections with pathogenic aetiology ranging from viruses to bacteria, parasites and fungi [2, 3] that affect both human and animal hosts. These kinds of infections can be transmitted between humans, animals and from animals to humans, where they are classified as zoonotic diseases [4]. According to the world health organization (WHO), infectious diseases impose the most burden on population health and growth. One such example is tuberculosis (TB) which is responsible for approximately 10 million cases globally per year and at least 1.4 million deaths as of 2021 which is an increase from 1.3 million in 2020 [5]. The recent increase in the global TB burden is partly due to the outbreak of another infectious disease, coronavirus disease-2019 (COVID-19), in 2019. The infectious disease, COVID-19, is caused by severe acute respiratory coronavirus 2 (SARS-CoV-2) [6, 7] and it has recorded the highest prevalence and mortality annually, followed by TB, human immuno-deficiency virus (HIV) and malaria with approximately six million [8], 1.4 million [5], 680,000 [9] and 620,000 [10] mortalities as of 2021, respectively. The global pandemic, COVID-19, as was declared by the WHO [11], has also caused significant health, economic, financial and social burden since its outbreak in 2019 [12–14].

The aerosol transmission of some of these infectious diseases like TB presents a great challenge concerning both disease containment and management [15–18]. Besides the irrepressible spread, the development of drug resistant strains i.e., multidrug resistant TB (MDR-TB), extensively resistant TB (XDR-TB) [19–21] due to drug selective pressure [22–

24], especially to the most effective drugs presents another hurdle in infectious disease management i.e., TB.

In the case of COVID-19, the SARS-CoV-2 RNA (ribonucleic acid) virus is prone to myriad evolutionary mutations evidenced through the emergence of a number of SARS-CoV-2 variants including, Alpha, Delta, Beta, Gamma and Omicron variants. These variants have shown to either increase viral transmission and/or infectivity with the rise of host immunity [25–28].

1.2 Timeline of Major Infectious Disease Outbreaks

Over time, the human race has encountered a number of infectious disease outbreaks including the smallpox (as early as 10,000 BCE), the Spanish flue (1918), influenza (1957), polio (1955), HIV (1960s), hepatitis C (1967), severe acute respiratory syndrome; SARS (2002), H1N1 (2009), Middle East respiratory syndrome; MERS (2012), zika virus (2016), Ebola (2014 and 2018) and most recently, SARS-CoV-2 (2019) [29–35] (**Figure 1.1**). Smallpox on its own was responsible for an estimated 300 million mortalities in the 20th century [36–38] and as of 2021, HIV-AIDs is cumulatively responsible for over 40 million deaths and counting [39] while the death toll by COVID-19 pandemic has eclipsed 6 million as of January 2023 [40].

Notwithstanding the mortalities, the intermittent outbreaks of infectious diseases has coincided with novel advancement of medical infrastructure directed towards management and prevention of future infections. For example, the first successful vaccine produced was the smallpox vaccine, by Edward Jenner, in 1796 which led to the eradication of smallpox [41]. The eradication of smallpox entailed a collaborative effort involving the development and widespread distribution of the smallpox vaccine, along with a coordinated effort identify and isolate infected individuals. The last known natural case of smallpox occurred in Somalia

in 1977, and the WHO declared smallpox eradicated in 1980, which goes to show the power of vaccines and international collaboration in elimination of infectious diseases. Like smallpox, Poliomyelitis, a viral infection caused by an enterovirus which affects the nervous system [42] was previously eradicated through mass campaigns, the Global Polio Eradication Initiative (GPEI), and vaccine administration. As of May 2022, the wild polio virus (WPV) 2 and 3 have been globally eradicated with a few sporadic cases of WPV 1 in Afghanistan, Pakistan and Malawi [43] in 2020 to 2022. Quite recently however, we have seen an emergence of circulating vaccine-derived poliovirus (cVDPV) in the USA, UK, and Israel [44, 45]. The cVDPV causes poliomyelitis through prolonged circulation in under immunized populations allowing for reversion to virulent form [46]. The re-emergence of poliomyelitis is mainly attributed to logistical difficulties in reaching remote and conflict-affected areas, vaccine hesitancy, and emerging vaccine-derived polioviruses.

The advancement in technology has ushered in new diagnostic techniques like the polymerase chain reaction (PCR) cartridges to detect both susceptible and resistant TB strains using the GeneXpert technology [47] and advancement in the messenger ribonucleic acid (mRNA) vaccines development [48, 49] as seen in the case of COVID-19. However, as observed from the re-emergence of previously eradicated infections, mass campaigns and continued public sensitization about the importance of vaccinations and preventive therapy in general is equally important.

In spite of the advancement in medicines and technologies, mutations rendering infectious pathogens unsusceptible to antimicrobials (AMR) present the greatest challenges to public health. In addition, more novel infectious diseases are predicted in future based on the increased human interaction with wildlife through both the extensive encroachment of wildlife habitats and demand for game meat [50–53]. Either way, an urgent intervention is required.

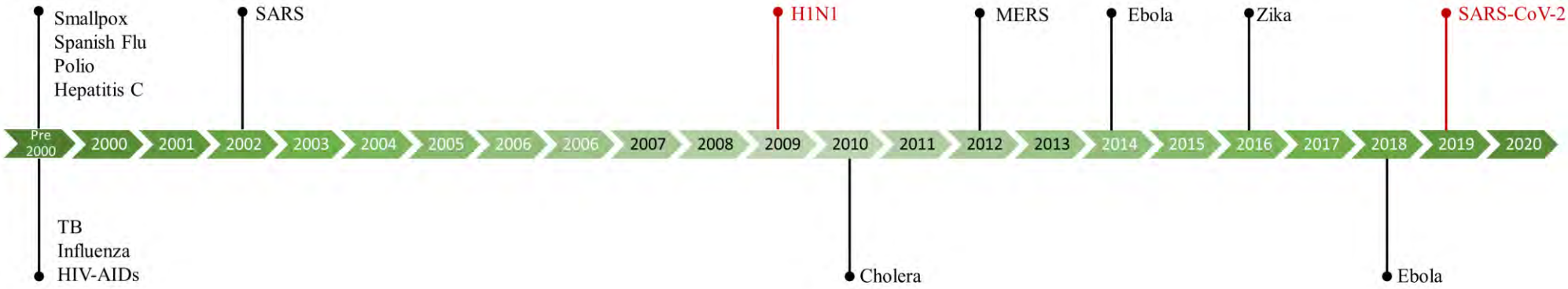

Figure 1. 1: Timeline of major infectious disease outbreaks [29–35]. Outbreaks that achieved pandemic status are shown in red.

1.3 Emergence of New Infectious Diseases

Besides the existent infectious diseases i.e., TB, HIV-AIDs, Hepatitis etc, the last two decades have witnessed an accelerated emergence and re-emergence of a plethora infectious pathogens, **Figure 1.1**, mainly of animal origin, which have had significant effect on the public health.

These outbreaks are mainly a circumstance of anthropogenic factors resulting from population growth, urbanization and agricultural expansion which have increased both the human encroachment on wildlife habitats and inter-species interaction [54]. Consequently, the wildlife pathogens have evolved to new hosts, domesticated animals and humans, through zoonosis. While anthropogenic factors have undoubtedly played a role in the increased incidence of infectious disease outbreaks, improved detection and surveillance mechanisms have also contributed to the higher number of reported cases unlike in the past where outbreaks might have gone undetected or misidentified.

Besides the anthropogenic influence on infectious disease outbreaks, climate change has been documented to aggravate human pathogenic infections [55, 56]. Researchers have identified a strong relationship between climate variables and the growth, mortality rate, reproduction, and distribution of disease vectors [57]. Climate change affects disease transmission through the impact on the vehicle or vector and the pathogen [58]. The changes in climate can increase the transmission cycles of certain vectors and result in the establishment of new diseases or increase prevalence of existent diseases [59]. The life cycles and transmission of most infectious agents are linked with climate, and the reported negative or uncertain responses of infectious diseases to climate change have been growing [60]. While the specific percentage of diseases aggravated by climate change is not mentioned, the evidence suggests

that climate change has a significant impact on the transmission and emergence of infectious diseases.

Furthermore, the advancement in global connectivity through fast and cheap travel has facilitated fast and global spread of the infectious agents from the spill over areas [54]. A recent example is the SARS-CoV-2 virus first identified in the Wuhan city of China in 2019, which has spread across the globe through air travel, attaining pandemic status [11, 61, 62].

Consequent to the unprecedented global spread, the human hosts have afforded SARS-CoV-2 ground for progressive evolution though a myriad of mutations in the various viral proteins, leading to the emergence of new variants. These SARS-CoV-2 variants i.e., Alpha, Beta, Delta, Gamma and Omicron, accommodate mutations (single nucleotide polymorphisms: SNPs, deletions and insertions) in key SARS-CoV-2 protein facilitating higher transmissibility, infectivity and antibody/vaccine escape properties compared to the Wuhan/reference strain. At present (August 2023), there isn't a well-established standard cure for COVID-19, however, there are treatments with antivirals (nirmatrelvir and ritonavir, remdesivir and molnupiravir) and neutralizing monoclonal antibodies (sotrovimab) that are used in the management of COVID-19 symptoms and prevention serious illness. Furthermore, significant advances have been made in vaccine production and characterization of viral mutations [63–68]

1.4 Mutations in Infectious Diseases

A mutation constitutes a change in the nucleotide sequence of the genome of an organism, whose effects can range from being benign to modifying the behaviour of an organism and in some cases deleterious [69]. This change in genetic constitution through mutations ensures the survival and progressive evolution of organisms in the context of natural selection. Notably, not all the mutations are beneficial as the process is sometimes a trade-off in a bid to

strike a balance between survival of the environmental conditions and maintaining efficient functionality [70]. In fact, in a constant environment, mutation rate would be minimal.

In the case of infectious diseases causing organisms like *Mtb*, mutations are known to occur in the proteins mostly targeted by or interacting with anti-microbial compounds like the *Mtb* catalase peroxidase (KatG) enzyme involved in the activation of the first-line anti-tubercular drug isoniazid and the DNA-dependent RNA polymerase β-subunit which is targeted by rifampicin [71, 72]. Evidently, some of these mutations (mostly SNPs) change the drug interaction profiles with the target proteins rendering them insensitive.

The RNA viruses like the SARS-CoV-2 are prone to mutations even with the presence of the proof reading mechanisms of the RNA-dependant RNA-polymerase [73]. The SARS-CoV-2 characteristically has a myriad of mutations in key proteins involved in viral replication and attachment to the host i.e., the Main protease (M^{pro}) [74–76] and Spike (S) protein [77–81].

The effects of such protein mutations are myriad, and they represent a primary area of investigation. Some mutational effects on protein behaviour include modifying the protein-substrate interaction profile, causing local structural perturbations, changing the protein secondary structure which in turn affects protein folding and stability and so much more [82–84]. Interestingly, a large number of drug-resistance mutations in the target proteins occur away from the active/drug binding site implying an allosteric mechanism of action [85–87]. In *Mtb* KatG, mutations occurring in C-terminal domain [88] away from the active site are associated with high level isoniazid resistance while in SARS-CoV-2, mutations in S protein domains other than the receptor binding domain (RBD) have been implicated in affecting viral binding to the host [89, 90].

1.4.1 Allosteric Effects of Mutations

Allostery is a fundamental concept in the life sciences that describes how biomolecules regulate each other in a remote action-at-a-distance mode. It involves the transmission of information between distinct sites of proteins, leading to conformational changes that affect protein function [91]. The allosteric effect of resistance mutations is multifaceted and involves different mechanism of action like; mutation of critical positions in the protein sequence which results in disruption of protein function [85, 86], shifting the conformational of the protein hence affecting its activity and drug binding [92], redistribution of the propagation pathways in the protein structure, or through the elimination or establishment of sites for allosteric posttranslational modification [93].

Given the complex signaling mechanisms involved in allosteric modulation, investigation of allosteric effects and mechanisms of action of mutations is challenging compared to orthosteric mutations [94]. Tastan Bishop and colleagues in [86] discuss deciphering mechanism of resistance mutations through combinatory investigation of allosteric and orthosteric mutations using Statistically Guided Network Analysis approach (SGNA) [95]. In [96], Tan and colleagues use a structure-based statistical mechanical model of allostery to discern the allosteric signaling path ways in the SARS-CoV-2 S protein. Additionally, various new tools have been developed to not only predict the effect of allosteric mutations but also the allosteric effects of ligand binding [97–99].

1.5 Drug Resistance in Infectious Diseases

Antimicrobial drug resistance is the significant decrease or complete loss of susceptibility of pathogens to a given drug or therapy [100, 101]. Various microorganisms have this inherent ability to adapt to an ever-evolving environment to survive, a process termed “natural selection” by Charles Darwin.

Globally, at least 1.2 million deaths directly resulted from drug resistance in 2019, with at least 4.5 million deaths associated with AMR [100]. In the case of TB, drug resistance can be classified into isoniazid (INH)-resistant TB, Rifampicin-resistant TB, MDR-TB (resistant to both INH and Rifampicin), pre-XDR-TB (resistant to INH, Rifampicin and second-line fluoroquinolones) and XDR-TB (resistant to rifampicin, fluoroquinolones and other priority drugs like bedaquiline and linezolid) [5, 102]. Approximately 132,222 and 25,681 diagnosed TB cases in 2021, were of MDR/Rifampicin resistant-TB and XDR-TB, respectively [5]. Furthermore, the number of TB case notifications and people enrolled onto MDR-TB treatment declined by 18% and 15%, respectively between 2019 and 2020 [5, 102]. These numbers are expected to decrease further from 2022 onwards due to the impact of the COVID-19 outbreak.

On the other hand, antiretroviral drug resistance in human immune-deficiency virus (HIV), a common TB comorbidity, has also increased, where at least 50% of infants diagnosed with HIV have the non-nucleoside reverse-transcriptase inhibitors (NNRTI) resistant strain of the virus in sub-Saharan Africa [101, 103].

AMR is a consequence of genetic changes in drug target proteins over time, through drug resistance mutations [100, 104]. Some of the key drivers of antimicrobial resistance include over-prescription of antibiotics, inaccessibility to quality drugs, and poor resistance surveillance systems, especially in low-income countries [103, 105]. Additionally, the misuse of antimicrobials in veterinary medicine and animal husbandry also contributes to the development of AMR especially in zoonotic pathogens, which can then be transmitted to humans. These multiple factors collectively contribute to the growing challenge of AMR, compromising our ability to treat infectious diseases effectively and posing a significant threat to global public health.

Low income and developing countries face the greatest impact of AMR on the account of insufficient resources [106–108]. Therefore, a concise understanding of the drug resistance mechanisms at a molecular level is essential to inform future drug design in infectious diseases.

1.6 Current Research on Infectious Diseases: TB and COVID-19

Extensive scientific research has focused on characterizing mutations in infectious pathogens in addition to deciphering the drug resistance consequent to mutations [109–115]. With the advancement of computational power and storage capabilities, a plethora of computational and bioinformatics approaches have been developed and employed in screening for potential drugs, drug-target structure prediction, *in vitro* simulation of drug-protein interaction, and prediction of drug resistance mechanisms [86, 116–118].

1.6.1 TB

Current *Mtb* research is focused on identifying genetic markers of drug resistance, understanding the mechanisms of drug resistance, and developing strategies to prevent the emergence of drug-resistant strains. The bulk of the *Mtb* mutation studies apply sequence analysis, molecular docking and molecular dynamics (MD) simulations to highlight the effect of mutations on active site stability coupled with the reduced first-line drug (INH and PZA) retention in the active site due to fewer ligand-protein bonds in the mutants compared to the wildtype [71, 119–131]. For example, Unissa et al. used MD simulations to show that KatG mutations at position 315 i.e., S315T, S315I, S315R, S315N and S315G experienced higher residue fluctuations for the INH binding residues compared to the wildtype (WT) [132]. Similarly, Singh and team analyzed the KatG G279D mutation and showed reduced INH binding to the active site in the mutant [133]. Khan and team focused on the L19R, R140H, and E144K mutations in MtPncA, where they observed a tendency of PZA to form less H-

bonds and a general reduction in the drug binding ability [134]. Similarly, Sheik Amamuddy et al., applied statistically guided network analysis to highlight conserved MtPncA motions associated with PZA unbinding in the mutant systems [135].

Another field of research has focused on building and applying Machine Learning (ML) approaches to genome wide sequencing (GWS) data to classify and predict genetic markers for drug resistance [136–142]. The high numbers of genome wide association studies (GWAS) to-date imply access to large datasets of sequenced data which can be used to train and build machine learning (ML) approaches with a higher efficiency for drug resistance prediction and classification. Recently, Deelder et al., presented a ML approach for predicting *Mtb* drug resistance mutations using genomic profiling [143]. Similarly, Green et al., used convolutional neural networks to predict *Mtb* resistance phenotypes to at least thirteen antibiotics including INH, PZA and ethambutol [144]. Other scientists are focusing on deep learning and artificial neutral networks to model more efficient ML approaches for phenotypic classification of *Mtb* resistance mutations [145–149]. Additionally, a number of online tools for prediction of phenotypic drug resistance of *Mtb* mutations have also been developed including GenTB [150] and resistance sniffer [151].

The existence of proteins as a network of residues means that graph theory principles can be applied to analyze protein behavior. Consequently, some researchers have focused on using network analysis and graph theory to describe protein networks [86, 152–156]. The principal behind this approach is that drug resistance mutations influence changes in the protein network patterns at the atomic level, which once studied could provide great insight in drug resistance mechanisms. Hitherto, this approach has not been used to study *Mtb* drug resistance especially in KatG and MtPncA proteins which presents a great knowledge gap. Information from *Mtb* mutant network analysis is a great source of insight into the drug resistance mechanisms.

1.6.2 COVID-19

The advent of COVID-19 prompted another upsurge of SARS-CoV-2 whole genome sequencing and data analysis, populating various databases like the Global Initiative on Sharing All Influenza Data (GISAID) [157], which consists of assorted viral sequences across different demographics. Initially, COVID-19 research focused on establishing the sequence and structure of different SARS-CoV-2 proteins [158–162]. However, with the emergence of SARS-CoV-2 variants of concern (VOC) and variants of interest (VOI) over time, vast research has gone into deciphering the epidemiological effect of SARS-CoV-2 mutations i.e., virus transmission and spread [163–170], mutational effect on clinical outcomes i.e., severity of the disease [163, 171–177], the evolutionary effect in terms of emergence of new variants [178–182], and molecular effect of mutations on RBD-hACE2 protein dynamics and interactions [183–189]. Of all the SARS-CoV-2 VOC, the Omicron variant has presented with the highest transmissibility [80, 190–193] afforded to it by the multiple mutations in the S protein, at least 30 SNPs [77–81]. Previous *in silico* and experimental studies have shown that the Omicron RBD mutations i.e., N439K, L452R, Y453F and N501Y are associated with increased RBD binding to the hACE2 [183, 194–197]. On the other hand, N440K, E484K and F486L mutations are linked to SARS-CoV-2 vaccine escape [198–201]. Furthermore, the Omicron variant has undergone further evolutionary mutations giving rise to the BA.1, BA.2, BA.3 and BA.4/5 sub-variants with unique mutations. For instance, the L452R and F486V RBD mutations unique to the BA.4/5 sub-variant are linked to enhanced RBD-hACE2 binding an vaccine escape, respectively [201].

In another study, Tan et al., 2022 [96] uses a structure-based statistical mechanical model of allostery (SBSMMA) to quantify the energetics of allosteric communication in the spike glycoprotein at the single-residue level. The results of which show a substantial involvement of allosteric mechanisms due to the high mutability of RNA viruses and the flexibility of viral

proteins. Tan and colleagues argue majority of the research is concentrated on S protein RBD interface, considering allosteric effects of mutations is as important in predicting and diagnosing new variants of concern [96].

The continuous evolution of SARS-CoV-2, let alone its most dominant variant, Omicron, beacons the need to characterize the variant behaviour and interaction with the hACE2. Previous computational studies have looked at how different variants interact with hACE2 [183, 202, 203] however, data on how the different Omicron sub-variants interact with the hACE2 and the patterns of communications between the two proteins is still wanting.

Furthermore, the existence of hACE2 polymorphisms in different populations creates another important variable to consider in the study of SARS-CoV-2 infectivity and transmission. Physiologically, the hACE2 is involved in blood pressure homeostasis under renin angiotensin systems (further explained in **chapter 5**). Given the complexity of RBD-hACE2 molecular interactions, changes in hACE2 genetic constitution, can have extensive effects on the affinity and type of molecular interactions. For instance, the rate of expression of hACE2 affects SARS-CoV-2 infectivity since the metalloenzyme mediates viral entry into the cell. The hACE2 is also highly polymorphic [204], where a number of these hACE2 mutations are linked to differential viral binding and infectivity in the host [205–208]. Besides, age, sex and race, the genetic variability of the hACE2 could be one of the contributing factors to the disparity in COVID-19 infectivity and case fatality rate among different populations. For instance, Paik et al., showed that hACE2 variants S43N and Q86R are associated with negative alchemical free energy, signifying stronger binding of the RBD to hACE2 compared to the wildtype [209]. Meanwhile, Ashoor et al., used the mCSM server to show that D355A, D355N, E35K, F40L, E35D, M82I, S19P and T27A hACE2 mutations have destabilizing effects of the RBD-hACE2 complex [210]. Similar studies have implicated the hACE2 S19P and K26R mutations as having protective and predisposing genetic factors to SARS-CoV-2

infection [207, 211–216]. Even with the available studies on hACE2 polymorphisms, understanding of complexity of RBD-hACE2 molecular interactions is still deficient, especially network analysis perspective. New insights from the mutation effect on the RBD-hACE2 inter-protein communication network patterns confer insights into drug design and vaccine discovery.

1.7 Knowledge Gap

The incidence of AMR has reached an unprecedented level, with an estimated 4.95 million deaths linked to AMR globally as of 2019. The highest death toll associated with AMR is in the sub-Saharan region, where at least 27.3 deaths per 100,000 are associated with AMR as of 2019 [100]. AMR in infectious diseases like TB has been observed in the second-line and third-line drugs, leading to multi-drug and extensively resistant strains of TB [5, 217]. Even with the identification of new antimicrobials for infectious disease management, only 6 out of 32 antibiotics identified by WHO for clinical development were categorized as innovative/new in 2019 [101]. This has created an urgent need to study and understand the drug resistance mechanisms in the phenotypically observed resistant mutations to better inform innovative drug design. Evidently, a lot of research as focused on deciphering the mechanisms of resistance in both the KatG and MtPncA proteins (**section 1.6**) however, analysis of enzyme communication patterns in the presence of resistance mutations is still lacking. Insight into the most important and influential residues in the mutant proteins is key in understanding drug resistance and drug design. The first part of the thesis focuses in describing the changes in protein dynamics and communication patterns in the TB first-line drug resistant proteins, KatG and MtPncA. Furthermore, a comparison of the PZA and nicotinamide (NAM) interaction profile and dynamics in the MtPncA mutants was completed to identify key differences in interaction to advise future drug design.

With the emergence of new infectious diseases, COVID-19, and the continued evolution of the etiologic agent, SARS-CoV-2, a great deal of research (section 1.6) focusing on understanding viral behaviour and characteristics is needed to better inform drug design and vaccine development. This is especially important given the fast paced rate of viral evolution towards a more efficient pathogen (section 1.6). The second part of the thesis capitalizes on the constantly evolving viral genome to characterize the effect of SARS-CoV-2 RBD mutations in the Omicron sub-variants on viral interaction with the hACE2. Using network analysis, this work highlights novel communication patterns between the two proteins. Furthermore, the thesis adds to the research on hACE2 polymorphisms, (section 1.6), by applying novel network analysis approaches to further inform on the viral and host interaction in the presence of both the Omicron BA.4/5 mutations and hACE2 mutations.

1.8 Research Aim

The work in this thesis was aimed using *in silico* approaches to 1) decipher and characterize drug resistance mechanisms of documented SNPs in TB with a main focus on the KatG and MtPncA enzymes and 2) characterize the interaction and network patterns between the SARS-CoV-2 RBD and the human ACE2 in the presence of both the Omicron sub-lineage mutations and naturally occurring hACE2 polymorphisms. To achieve this, the research is broken down into four sub-aims *viz*:

1. Using MD simulations and trajectory analysis to characterize the INH drug resistance mechanisms of eleven high confidence mutations in the dimeric KatG enzyme with the purpose of informing future drug design and development.
2. Deciphering the differences in ligand binding and interaction between the first-line drug, PZA, and its analogue, NAM, in the MtPncA enzyme and its mutants. Insights

from the which are meant to add on the research on *Mtb* drug resistance and advise future drug design.

3. Studying the effects of the progressive SARS-CoV-2 Omicron sub-lineage mutations in the SARS-CoV-2 RBD on the interaction with the human ACE2. This was motivated by the need to profile the changes in hACE2 interaction with the constantly evolving SARS-CoV-2, information that is vital for therapeutics development.
4. Analysing the effect of naturally occurring human ACE2 polymorphisms on the interaction and distribution of communication patterns between the host receptor and the BA.4/5 Omicron RBD of SARS-CoV-2. Given the disparity in SARS-CoV-2 infection among populations, insights from this analysis are meant to add on already exiting research on hACE2 and inform on the proclivity of hACE polymorphisms to bind SARS-CoV-2.

1.9 Research Objectives

1. Characterizing the effects of eleven high confidence KatG INH resistance mutations:
    - Identification of high confidence KatG mutations with phenotypic drug resistance from literature and TB databases for *in silico* analysis.
    - Modeling of the dimeric KatG mutant structures for MD simulation studies.
    - Access the changes in KatG dynamics of due to mutations through subjecting the modeled structures to MD simulations and trajectory analysis.
    - Perform dynamic residue network analysis to establish the changes in network patterns consequent to mutations.

2. Analysis of PZA and NAM interactions in wildtype and mutant MtPncA enzyme:

- Molecular docking of NAM to MtPncA to determine the ligand interaction profile.
- Modeling of 82 MtPncA mutant structures in complex with PZA and NAM for further *in silico* analysis.
- Molecular dynamic simulations of the modeled structures to access the mutation effects on PZA and NAM interaction in the MtPncA enzyme.
- Post MD trajectory analysis including dynamic residue network analysis to determine the effect of mutations on the protein networks.

3. Characterization of the effects of SARS-CoV-2 Omicron sub-lineage RBD mutations on the RBD-hACE2 dynamics and networks:

- Identification of SARS-CoV-2 Omicron sub-lineage RBD mutations for further analysis.
- Modeling of the RBD-hACE2 complex structures in the presence Omicron sub-lineage mutations for MD simulations.
- Molecular dynamic simulations of the wildtype and mutant structures to access the mutational effect on protein behavior.
- Dynamic residue network analysis of the RBD-hACE2 systems to determine the network patterns in the wildtype and mutant systems.

4. Characterization of the effects of hACE2 polymorphism on RBD-hACE2 interaction in the presence of BA.4/5 Omicron RBD mutations:

- Identification of the naturally occurring hACE2 polymorphisms in the African population for further analysis.

- Modeling of the RBD-hAC2 complex structures in the presence both RBD Omicron BA.4/5 mutations and hACE2 mutations for MD simulations.
- MD simulations of the modeled structures to access the mutational effect on protein behavior.
- Determination of hACE2 mutation effects on the RBD-hACE2 communication patterns through dynamic residue network analysis.

Given the *in silico* nature of the research objectives in the thesis, analogous computational approaches were applied including the group (RUBi: Research unit in Bioinformatics) developed technique of analyzing dynamic residue networks, to answer the research questions. The next sections delve into the principles of the various techniques.

1.10 Computational Approaches in Mutation Studies

The use of computational approaches to study protein behavior has grown staggeringly in the 21st century with the advancement in computational power and storage capabilities. This has created a perfect platform for development of and scaling of tools with more efficiency. For instance, technological and computational advances have led to more efficient genome-wide association studies (GWAS) [218–220] resulting in large volumes of sequencing data. Additionally, protein structure databases are i.e., Protein Data Bank (RSCB PDB) [221] are also growing tremendously due the large number of solved 3D protein structures (over one million computer structure models in PBD as of 01 February 2023). Public access to this data has encouraged the use and development of *in silico* tools/studies to investigate and characterize protein function and behaviour. Computational approaches like homology modeling and MD simulations can be used to model novel protein structures based on the sequence and mimic *in vivo* protein behaviour at different time scales, respectively. With the

ever increasing AMR and emergence of novel variations of infectious pathogens, computational approaches present a fast and efficient way to model variant strictures, study and characterize the effects of mutations on protein behaviour hence informing drug design and development. The use of computational approaches to inform on protein behavior and drug design is often referred to as Computer Aided Drug discovery (CADD).

In this section, we discuss in detail the different computational approaches used throughout the thesis to decipher and characterize the effects of both drug resistant mutations in *Mtb* and naturally occurring mutations in SARS-CoV-2 RBD and hACE2 proteins.

1.10.1 Sequence Alignment

Sequence alignment is one of the fundamental approaches in bioinformatics used for aligning evolutionary-related deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and protein sequences using an assortment of algorithms [222]. Through sequence alignment, phylogenetic relationships are established in addition to predicting of protein structure and function using homologous sequences with solved 3D structures [223, 224].

In mutation studies, sequence alignment allows scientists to compare sequences of two or more different proteins or nucleic acids to identify identical regions and potential mutations. Furthermore, the extent of agreement between any aligned sequences can be determined which informs on the degree of protein sequence identity, information which is important in 3D structure homology modeling. In this thesis, sequence alignment was used for template identification and homology modeling of 3D protein structures in the absence of solved mutant X-ray crystal structures.

Sequence alignments can be pairwise alignment, which involves alignment of two sequences, or multiple sequence alignment, where three or more sequences are aligned to each other. Furthermore, sequence alignment can be achieved through both local and global alignment.

In local alignment, similar regions of the sequences are aligned to each other creating homologous patterns, whereas in global alignment, the entire query sequence is aligned to the target sequence from end to end [225, 226].

In sequence alignment, the quality of the alignment or evolutionary relatedness is based on the alignment quality score as assigned by a scoring system/algorithm. The two main scoring methods are the dynamic programming; used primarily in pairwise alignments and the progressive iterative method used in multiple sequence alignment (MSA) [227]. Dynamic programming uses a substitution matrix to assign a score to residue pairings or gaps based on the likelihood of residue substitution [227]. Progressive alignment is a heuristic method where each possible pairwise alignment for the multiple sequences is done and scored using the global alignment method before a sequence phylogenetic/guide tree is generated based on a distant matrix from the scores. The guide tree then dictates the order of pairwise alignment where closely related sequences are aligned first to create a consensus sequence that is then aligned to the next sequence in line. The process is repeated until all the sequences are aligned [222, 227]. An improvement to the progressive method is the heuristic iterative method where the generated alignment is improved through numerous alignment iterations until the score cannot be improved further. Some of the established iterative alignment tools include ClustalΩ [228], MUSCLE [229] and MAFT [230]. This research employed the ClustalΩ tool.

1.10.1.1 ClustalΩ

The ClustalΩ tool allows for fast and accurate alignment of several sequences by generating the guide tree using a modified sequence embedding approach, mBed [231]. The modified mBed has a complexity of O(*NlogN*) where individual sequences are inserted in *n* dimensions corresponding to the log of *N* [228]. The sequences are then substituted for an *n* element

vector, before clustering using K-means. The alignment is then built using HAlign [232] that aligns two profile hidden Markov models [233].

1.10.2 Homology Modeling

Homology modeling is a combination of steps/techniques used to predict the 3D protein structure based on a homologous template [234–237]. The method relies on protein sequence conservation among an ensemble of related protein systems. The principle behind homology modeling is that similar protein sequences warrant similar 3D structures [235].

In *in silico* mutation studies, a comparison of mutant and wildtype protein behaviour requires the 3D structures of both proteins. In the absence of the mutant 3D structures in RSCB PDB [221], homology modeling is used to model the mutant protein structures using wildtype sequence as the template. In this thesis, we applied homology modeling to generate mutant 3D structures for structural and dynamics analysis.

The homology modeling process can be divided into four steps involving template identification, sequence alignment, model building, and model validation.

1.10.2.1 Template Identification

Since homology modeling is based on a template 3D structure homologous to the target sequence, template identification is one of the critical steps in modeling. Structural templates are identified based on percentage identity to the target sequence, which should be at least 30% [235, 238], and the crystallization quality of the template 3D structure in RSCB PDB. Crystallization quality is accessed through visualization, using the R-factor, R-value and resolution [239, 240]. The R-factor, R-value and resolution of the protein structures can be accessed in the RSCB PDB [221]. In addition to R-factor, R-value, the MolProbity tool [241] was used to further access the quality if the crystal structures by evaluating the percentage of

poor rotamers, favored rotamers, Ramachandran outliers, Ramachandran favored, Rama distribution score, bad bonds, bad angles carbon beta (C_β) deviations and carbon alpha (C_α) geometry outliers. The MolProbity webserver was used: http://molprobity.biochem.duke.edu.

1.10.2.2 Multiple Sequence Alignment

The next step in homology modeling is sequence alignment, where the target and template protein sequences are aligned to ensure that the template(s) account for all the protein regions. The number of alignment gaps should be minimal to ensure quality 3D structures. The process of MSA is explained in **section 1.10.1**.

1.10.2.3 Model Building

The template-target sequence alignment from sequence alignment is used to build 3D structure models using a several modeling tools like MODELLER [242], Protein Interactive Modeling (PRIMO) [243] and HHpred [244]. Here, MODELLER was used for mutant structure modeling. MODELLER uses the spatial restraint [245] model as building method where the restraints are placed on the target structure, ensuring that atoms are within a given distance of each other as in the template structure. MODELLER can also perform loop modeling and model refinement using fast, slow or very slow methods [242].

1.10.2.4 Model Validation

Since the homology modeling steps are interdependent, model validation is essential to ensure that a near-native structure is generated. A number of tools are available for model validation ,i.e., discrete optimized protein energy (z-DOPE) score [246], VERIFY3D [247] and Protein Structure Analysis (ProSA) [248]. Besides z-DOPE score, which uses atomic energy between atom pairs in the model to access quality as implemented by MODELLER, VERIFY3D and ProSA were used in this thesis for model validation.

1.10.2.4.1 VERIFY3D

This is a web-based tool that assesses the protein 3D quality by comparing the compatibility of the model structure to its amino acid sequence using the 3D profile. The 3D profiles are based on a precomputed database of 18 environmental profiles, and they consist of tables computed from the structure atomic coordinates. Each residue in the model is scored based on the environment characteristics like, area of the residue that is buried, fraction of side chains exposed to polar atoms and secondary structures. Good quality models have high profile scores [247].

1.10.2.4.2 ProSA

ProSA is a tool for protein structure validation and prediction which analyzes the model quality in comparison to all experimentally determined chains in PDB [221]. ProSA uses the protein C_α atoms to calculate the structural energy based on the distance-based pair potential [249] and the solvent exposure potential of the model residues [250]. From these computations, the model quality is displayed as a z-score which indicates the overall quality of the model in comparison to structures of similar characteristics [248].

1.10.3 Molecular Dynamic Simulation

In nature, protein molecules consist of atoms in constant motion and interaction with each other. Molecular dynamic (MD) simulations use Newton's law of motion to predict atomic and molecular motions. In this thesis, the generated mutant structures were subjected to MD simulations to study their behavior in relation to the wildtype. MD simulations mimic *in-vivo* atomic behavior by applying Newtonian physics in a form of force fields (FFs) to determine the atoms' spatial position as a function of time [251, 252]. There are several types of set FFs which include AMBER (Assisted Model Building with Energy Refinement) [253],

CHARMM (Chemistry at Harvard Macromolecular Mechanics) [254], GROMOS (GROningen MOlecular Simulation) [255] and OPLS (Optimized Potentials for Liquid Simulations) [256]. These FFs capture bonded and nonbonded interactions like the van der Waals and electrostatic interactions. With the advancement in computational power, near-native solvent MD simulations can now be performed up to a micro-second scale using SPC216 (single point charge 216) and TIP3P (transferable intermolecular potential 3P) water models [257]. Periodic boundary conditions (PBC) are used during MD simulations for approximating an infinite system using a unit cell [258] and in so doing, they reduce the computational cost of simulating a large system. Trajectories, which are conformational snap shots at different times, from MD simulations can be analyzed using several metrics and tools to determine system characteristics. The metrics used in this research include root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration (Rg), center of mass distance (COM) and principal component analysis (PCA). In the thesis, these metrics were used to compare the protein behaviour between the wildtype and mutant systems to inform on the effects of mutations on protein dynamics and interaction.

1.10.3.1 Root Mean Square Deviation

RMSD is used in MD trajectories analysis to determine the extent of structural variation over time from the initial/reference conformation [259]. RMSD is calculated from **Equation 1.1** where; N is the number of atoms and δ is the difference in distance between atom i in the initial and final structure.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \delta_i^2}$$

Equation 1.1: RMSD equation

1.10.3.2 Root Mean Square Fluctuation

RMSF is a measure of residue flexibility over an MD simulation. It is similar to RMSD, except here, the average individual residue flexibility/fluctuation is measured for the MD simulation [260]. RMSF is calculated from **Equation 1.2**, where; T is trajectory time, δ is the difference between residue position at time t and residue position at the reference time. The angle brackets mean that the average of the square distance is used for the atoms in a residue.

$$RMSF = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \langle \delta^2 \rangle}$$

Equation 1.2: RMSF Equation

1.10.3.3 Radius of Gyration

Rg measures the degree of atom spread/gyration from the center of mass as of the protein. Rg is calculated from **Equation 1.3**, where; N is the number of atoms and δ is the difference distance between residue i and the center of mass.

$$Rg = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \delta^2}$$

Equation 1.3: Rg Equation

1.10.3.4 Center of Mass Distance

COM distance measures the distance between a pair of selected residues over the simulation time. COM distance can also be measured between two groups of residue selections using the center of mass of each selection as the reference point.

1.10.3.5 Comparative Essential Dynamics

Principal component analysis (PCA) is a statistical technique used to reduce the dimensionality of a dataset [261, 262] , where the most dominant protein motions from an MD trajectory are extracted by first superimposing all the frames to a reference/initial structure before diagonalizing a variance-covariance matrix of the atomic displacements/cartesian coordinates [263]. The generated modes of displacement/ eigenvectors are associated with their respective magnitude of displacement (eigenvalues) based on the variance. However, the comparative ED approach goes a step further to align trajectories of systems to be compared via the C_α or C_β atoms prior to computation and decomposition of the covariance matrix. The result is a pair-wise comparison of systems' motions within the same eigen subspace [264]. In ED, the first two modes or principal components one and two (PC1 and PC2) explain the most variance in the dataset. ED analysis was used to differentiate prominent motions between the mutant and wildtype systems throughout the thesis.

1.10.3.6 Dynamic Cross-Correlation

Dynamic cross-correlation (DCC) is a technique that uses MD trajectories to measure the degree of correlation or anti-correlation between the *Cα* atoms in a protein system [265]. DCC calculates atomic correlations using **Equation 1.4**, where; Δr_i and $\langle\rangle$ represent the displacement of atom *i* from the average position and time average over the trajectory, respectively. DCC results range from 1 to -1 where 1 indicates a complete correlation, 0 shows no correlation, and -1 shows complete anti-correlation.

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

Equation 1.4: Calculation of dynamic cross-correlation

1.10.4 Dynamic Residue Network Analysis

DRN analysis is an MD trajectory analysis technique that applies graph theory to identify protein residue centralities based on distinct metrics. The DRN approach was developed by the RUBi group and is publicly available through the MDM-TASK web server [264, 266]. In graph theory, an approach borrowed from social sciences, properties of a network are identified by treating components/vertices of a network as nodes and the connection between them as edges [267]. In DRN analysis, protein residues are treated as nodes, and if a connection between them exists within a Euclidian distance of ≤ 6.7 Å [268], it is treated as an edge. With this representation, adjacency matrices of the residues interactions per MD trajectory frame can be created and analyzed. The centrality of residues in the network can be determined using different metrics, each describing unique properties in a network. The MDM-TASK web server [264] uses MD trajectories to calculate protein residue centrality using different DRN metrics.

In this thesis, five DRN metrics previously established in [76, 264, 266], i.e., averaged *betweenness centrality* (*BC*), averaged *closeness centrality* (*CC*), averaged *degree of centrality* (*DC*), averaged *eigenvector centrality* (*EC*), and averaged *Katz centrality* (*KC*) were used to 1) decipher the changes in the KatG and MtPncA residue network patterns as a consequence of drug resistant mutations. And 2) to identify the critical residues in the SARS-CoV-2 RBD and hACE2 involved in the inter-protein interaction and how the centrality of these residues' changes amidst both the SARS-CoV-2 and hACE2 naturally occurring mutations.

1.10.4.1 Averaged Betweenness Centrality

The *BC* metric assigns centrality based on the degree of involvement of a node/residue in the shortest path between all the residue pairs in a network (**Figure 1.2**). The more involved a

residue is, the higher the centrality values and hence its importance [269]. Working with the assumption that communication in molecular structures (proteins) is through the shortest paths, residues with high *BC* represent key residues involved in information dissemination, signal transduction and molecular interactions throughout the protein. Previous studies have also revealed that residues with high *BC* often overlap with highly functional sites of the protein [76, 152].

Averaged *BC* is calculated from **Equation 1.5** from the MDM-TASK web server [264, 266], where; V is the number of nodes, m the number of frames, $\sigma(s,t)$ is the number of shortest paths connecting nodes s and t, $\sigma(s,t \mid v)$ the paths passing another node v whereas i is the frame number. Through *BC* analysis, the changes in residue centrality between the wildtype and mutant systems was assessed.

$$\overline{BC}(v) = \frac{1}{m}\sum_{i=1}^{m}\sum_{s,t\in V}\frac{\delta(s_i,t_i|v_i)}{\delta(s_i,t_i)}$$

Equation 1.5: Calculation of averaged *BC*

1.10.4.2 Averaged Closeness Centrality

CC informs on the peripherality of a residue to all other residues in a network. Residues close to other residues, especially at the protein core or interface, usually have high *CC* values [115]. In context of protein structures, the *CC* metric measures the proximity of a residue to all other residues in a system and therefore access the residues' role in information dissemination within a protein network (**Figure 1.2**). High *CC* may represent residues that are spatially central and have a direct influence on nearby residues [270]. Biologically, these residues could be involved in local conformational changes and rapid communication within a protein's structure [76, 155]. Furthermore, residues at the protein interface that are involved

in intra-protein and inter-protein interaction and stability have been associated with high *CC* values [115, 271].

Equation 1.6 [264, 266] computes averaged *CC* where; $d(v, u)$ is the shortest-path distance between *v* and *u*, n is the number of nodes in the graph and *m* is the number of frames. *CC* analysis identified information disseminating residues in the wildtype and mutant systems.

$$\overline{CC}(v) = \frac{n-1}{m}\sum_{i=1}^{m}\sum_{u=1}^{n-1} d(v,u)$$

Equation 1.6: Computation of averaged

1.10.4.3 Averaged Degree of Centrality

DC is a measure of the connectedness of a node to the immediate neighbors (**Figure 1.2**). *DC* is based on unique immediate neighbors to a target node; hence, the higher the degree of connection to the immediate neighbors, the higher the *DC* values [268, 270]. In protein structures, *DC* is used to measure the exposure of residues in a network and hence highlights the residues with the highest number of immediate connections. *DC* can be used to quantify the effects of perturbations, ligand binding and/or mutations on the local residue interactions in a protein structure.

DC is computed from **Equation 1.7** [264, 266], where; *n* is the number of nodes; A_{ijk} is the jk^{th} adjacency for the i^{th} frame and *m* is the number of frames.

$$\overline{DC}(k) = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1, j\neq i}^{n} A_{ijk}$$

Equation 1.7: Averaged *DC* calculation

1.10.4.4 Averaged Eigenvector centrality

Averaged *EC* measures the degree of node influence in a network based centrality in a network and that of its neighboring nodes, both high and low scoring [270]. *EC* and *DC* demonstrate a correlated association, as the initial centrality assessment in *EC* is based on *DC*. In protein networks, the *EC* metric is useful in discerning how central a given residue is in the entire protein and its influence on information flow [86, 264].

EC is calculated from **Equation 1.8** [264, 266], where, for the decomposed adjacency matrix A, *EC* is the eigenvector, and λ is the eigenvalue. Averaged *EC* is for each i^{th} residue by averaging the vector for each MD frame, **Equation 1.9**, where *m* is the number of frames.

$$A \cdot \overrightarrow{EC} = \lambda \cdot \overrightarrow{EC}$$

Equation 1.8: *EC* calculation

$$\overline{EC}(i) = \frac{1}{m}\sum_{k=1}^{m} EC_{ik}$$

Equation 1.9: Averaged *EC* calculation

1.10.4.5 Averaged Katz Centrality

In *KC*, the degree of node influence in a network is assigned based on the centrality of its neighbors and the neighbor's neighbors [272]. In protein networks, *KC* can be used to determine how much influence a specific given residue has in relation to the other connected residues within a network [76].

Averaged *KC* is computed from **Equations 1.10** and **1.11** from the MDM-TASK web server [264, 266], where; *A* is the adjacency matrix for residue *i*, λ is the eigenvalue, α is the attenuation factor, β the weight assigned of the connected neighbors of residue *i* and *m* is the number of frames.

$$KC(i) = \alpha \sum_{j=1}^{n} A_{ij}\, KC_j + \beta$$

Equation 1.10: *KC* calculation

$$\overline{KC}(i) = \frac{1}{m} \sum_{k=1}^{m} KC_{ik}$$

Equation 1.11: Averaged *KC* calculation

The application of graph theory to the analysis of protein structures (DRN analysis) offers a powerful framework to unravel complex interactions within biomolecular systems. By representing proteins as nodes and their interactions as edges, network models provide a visual and analytical tool to comprehend the intricate web of relationships governing protein function. This approach allows researchers to identify key residues or domains crucial for ligand and substrate binding through *BC* analysis as previously show in [76, 115, 152, 155, 156]. Network analysis can also provide insight into the residue level changes in protein interactions consequent to perturbations from mutations and ligand binding through *DC*, *EC* and contact map analysis [115, 155, 266, 271]. Furthermore, *CC* analysis in chapter 2 and chapter 4 shows a correlation between *CC* values at the interface and the intra-protein and inter-protein interaction distances making it informative on changes in inter-molecular interaction. Overall, the application of network analysis in the study of protein structures provides invaluable insight into the global and local changes in protein networks which can be applied to drug discovery, mutational studies and prediction of functional regions in novel protein structures.

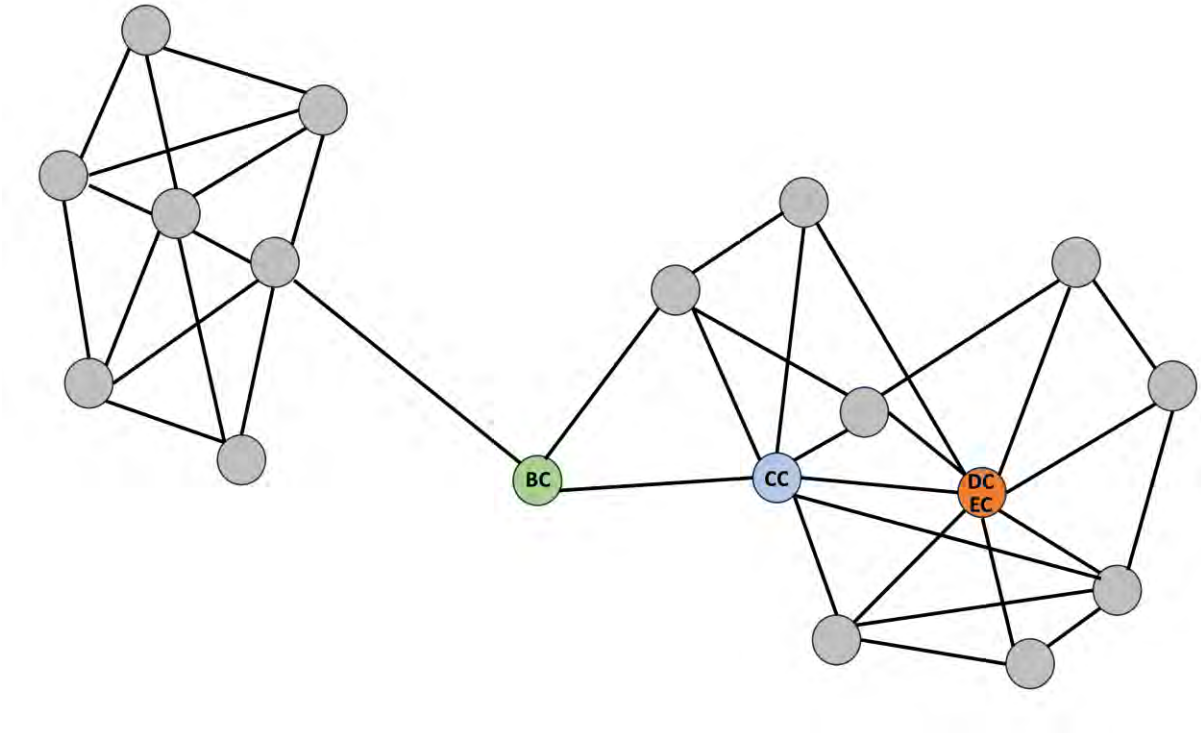

Figure 1. 2: Schematic representation of the DRN metric properties. The nodes represent the residues, and the edges represent the non-directed connection between residues in a protein. *BC* is represented by the green node, *CC* by the blue node making it the node, whereas *EC* and *DC* by the orange node.

1.10.5 Identification of the Global Top "x%" High Centrality Residues

Ensemble/global top high centrality residues were determined for each DRN metric using an in-house Python script written by Olivier Sheik Amamuddy. The script identified high centrality residues based on a set percentage cut-off (x%). This was motivated by the fact that residues with the highest centrality values across similar protein systems are observed to have functional importance [268, 273]. The script identified high centrality residues, "hubs" [76, 86, 155], by vectorizing all the residue centrality values before ordering them in descending order to determine the threshold as per the set percentage (x%). A data frame of the metric-specific centrality values for all systems is then made before the established threshold is used to make a binary matrix where one (1) corresponds to residues with centrality ≥ the threshold

and 0 for values < threshold. Rows summing to zero are dropped from the matrix and the idea here is to have homologous residues from other systems for any system residue with high centrality values. Finally, the script generates a heat map of the high centrality residues for the protein ensemble where, the residue hubs are annotated with their corresponding centrality values. The residues with hub status across all systems are referred to as *persistent* hubs [76, 86, 155].

This RUBi pioneered approach was employed to identify the residues with high centralities across both the wildtype and mutant systems in addition to highlighting the changes in centralities consequent to the mutations.

1.10.6 Contact Map Analysis

Protein structures exist as a network of residues whose properties and interactions facilitate protein folding and unfolding. A change residue property through mutations/SNPs affects the network interaction profile and in turn the protein behaviour and characteristics.

Contact maps were used here to assess the effect of mutations on the protein local residue interaction profile at the points of mutations and interface.

Contact map analysis, as executed using the *contact_map.py* tool from the MDM-TASK [264, 265] web server, enables profiling of target residue contacts within an Euclidian distance of 6.7 Å using a course grained trajectory (C_α and C_β atoms). Additionally, the *contact_heatmap.py* tool enables presentation of contact map analysis results as a heat map.

PART I: CHARACTERIZATION OF DRUG RESISTANCE MISSENSE MUTATIONS IN *Mycobacterium tuberculosis*

CHAPTERS 2 AND 3

CHAPTER 2

2. Understanding Isoniazid Drug Resistance Mechanisms in the Dimeric *Mycobacterium Tuberculosis* Catalase-Peroxidase Mutants

This chapter reports on the conformational changes in the catalase-peroxidase enzyme as a consequence of eleven high confidence mutations associated with phenotypic drug resistance to the first line drug, Isoniazid. Additionally, the changes in heme-cofactor coordination and protein communication networks consequent to mutations are discussed.

> **Contributions**: Mutation identification, homology modeling, MD simulations, trajectory analysis and all data analysis including script writing was done by Victor Barozi. The Python scripts used for the global x% DRN analysis and heatmap were generated by Olivier Sheik Amamuddy. Hub visualization scripts in PyMOL were generated by Olivier Sheik Amamuddy, Victor Barozi and Özlem Tastan Bishop.

The work in chapter 2 is reproduced in part from the following publication:

- **Barozi V**, Musyoka TM, Sheik Amamuddy O, Tastan Bishop Ö (2022). *Deciphering Isoniazid Drug Resistance Mechanisms on Dimeric Mycobacterium tuberculosis KatG via Post-molecular Dynamics Analyses Including Combined Dynamic Residue Network Metrics*. ACS Omega. https://doi.org/10.1021/acsomega.2c01036.

Authors Contributions: Ö.T.B. conceived the project. V.B. performed the calculations and data analysis under the guidance of T.M.M., O.S.A. and Ö.T.B.

2.1 Introduction

Up until 2020, TB, an airborne infection caused by the gram-positive *Mycobacterium tuberculosis (Mtb)* has been the leading cause of death from a single infection [274]. With the emergence of the COVID-19 global pandemic [11], TB is estimated to rank second as the leading cause of mortality from a single infection [5] in 2021. Globally, an estimated 9.9 million cases of TB were recorded in 2020 with the most burdened regions being South-East Asia, Africa and Western pacific, accounting for 43%, 25% and 18% of the TB incidence, respectively [5]. The infectious disease spreads through inhalation of *Mtb*-contaminated aerosols from an infected host and is characterized by fever, night sweats, weight loss and cough [275, 276]. The active form of TB is manifested through clinical symptoms whereas in latent TB infection (LTBI), the host immunological response is enough to keep the infection at bay with no clinical symptoms or transmission [277]. Active infection is further classified into pulmonary TB (PTB), which affects the lungs, and extrapulmonary TB (EPTB) which is known to infect other organs including, lymphatic TB, pleural TB, brain TB, central nervous system TB, abdominal TB, bone and joint TB, Genito-urinary TB and miliary TB [276, 278].

2.1.1 Diagnosis of TB

There are several TB diagnosis methods with varying sensitivity and turnaround time including the Ziehl-Neelsen (ZN) method, where stained sputum samples are examined under a microscope [279] and the X-ray technique which uses lung radiographic images to identify TB cavities. The other is the mycobacteria culture method which is also considered the gold standard and it involves growing the sampled bacteria in culture media at specific conditions in an incubator. This method however is expensive and has a long turnaround time [276]. More recently, polymerase chain reaction (PCR) techniques like GeneXpert which amplify the mycobacterium DNA for detection of both susceptible and resistant TB are being used

[280, 281]. Early identification of resistant TB using these methods is particularly useful especially in informing the choice of drug regimen.

2.1.2 Management of TB

There are different classes of drugs used for TB management depending on the strain of *Mtb*. First-line drugs are used for susceptible TB and consist of a combination of isoniazid, rifampicin, pyrazinamide and ethambutol. The ensuing classes like the second-line, third-line and fourth-line drugs, consisting of Fluoroquinolones and injectables are used in the management of the drug-resistant strains of *Mtb* [102, 282, 283]. The first-line of drugs are considered the most effective and with the least adverse effects [5].

2.1.3 Isoniazid as a TB Pro-drug

Isoniazid (INH: isonicotinic acid hydrazide) is a TB first-line pro-drug whose activity against susceptible *Mtb* depends on activation by the *Mtb* catalase peroxidase (KatG) enzyme [284]. Chemically, INH is made of a pyridine ring and hydrazide group (**Figure 2.1**). INH activation involves the cleavage of the hydrazide group by KatG to form an isonicotinoyl radical which then reacts with nicotinamide adenine dinucleotide (NAD^+) to form a INH-NAD adduct [285]. The formed INH-NAD adduct tightly binds to the enoyl-acyl carrier protein reductase, InhA, which is involved in mycolic acid biosynthesis, hence interrupting the cell wall biosynthesis process. In susceptible *Mtb* management, INH is administered *per os* with the other first-line drugs i.e., rifampicin, pyrazinamide and ethambutol for a duration of two months before switching to just INH and rifampicin for four and a half months [284]. Due to the drug dependency on KatG for activation, INH resistant *Mtb* strains are characterized by mutations in the KatG enzyme [286–290].

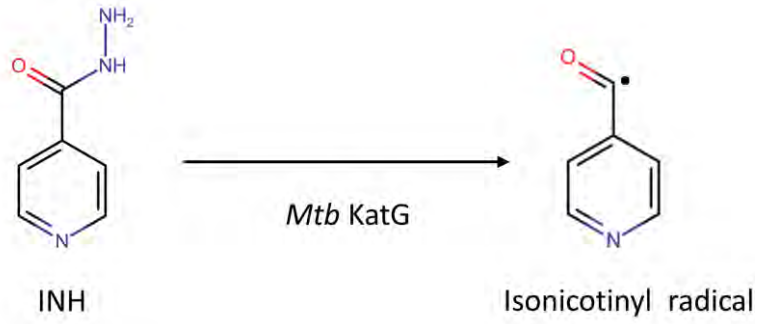

Figure 2. 1: Simplified illustration of the reaction between INH and the *Mtb* KatG leading to the formation of isonicotinoyl radical.

2.1.4 Structural Architecture of KatG

The *Mtb* KatG is a bi-functional enzyme responsible for both the catalase and peroxidase activity in the acid-fast bacilli [291]. The dual-functional enzyme belongs to the class I of the plant, fungal and bacterial super family of peroxidases [292]. In solution, KatG exists as a homodimer with each protomer/monomer consisting of a heme containing N-terminal domain and a cofactor deficient C-terminal domain. The enzyme active pocket is delimited by residues Ser315 and Asp137 at the access channel and the pocket accommodates a heme group at the lateral end [293]. The heme cofactor is bordered by active site residues His270, Trp321 and Asp381 on the proximal side and Arg104, Trp107 and His108 on the distal end (**Figure 2.2**). All catalase peroxidases i.e., cytochrome c peroxidase, ascorbate peroxidase and KatG have a conserved three residue adduct consisting of Trp107, Tyr229, and Met255 (MYW) at the distal end of the heme, which are important in the catalase activity of the enzymes [71, 294]. Chaplin et al., 2021 identified the INH binding site as next to the δ-meso edge of the heme [71].

The C-terminal domain consists of residues 436 to 740 [291] and has no documented activity in the *Mtb* KatG however, prior studies have indicated that it offers architectural support to the active site. Interestingly, truncation of the C-terminal domain in *Mtb* KatG by Baker and group curtailed the catalytic activity of the enzyme [295]. Furthermore, site directed mutation

of the C-terminal I'-helix region resulted in an unstable N-terminal and active site, hinting on the remote role the C-terminal domain on active site stability [296]. The KatG enzyme also has a dimerization domain comprising of the first 100 residues in each protomer, which facilitate dimerization through inter-protomer interactions with the opposite domains [297–299].

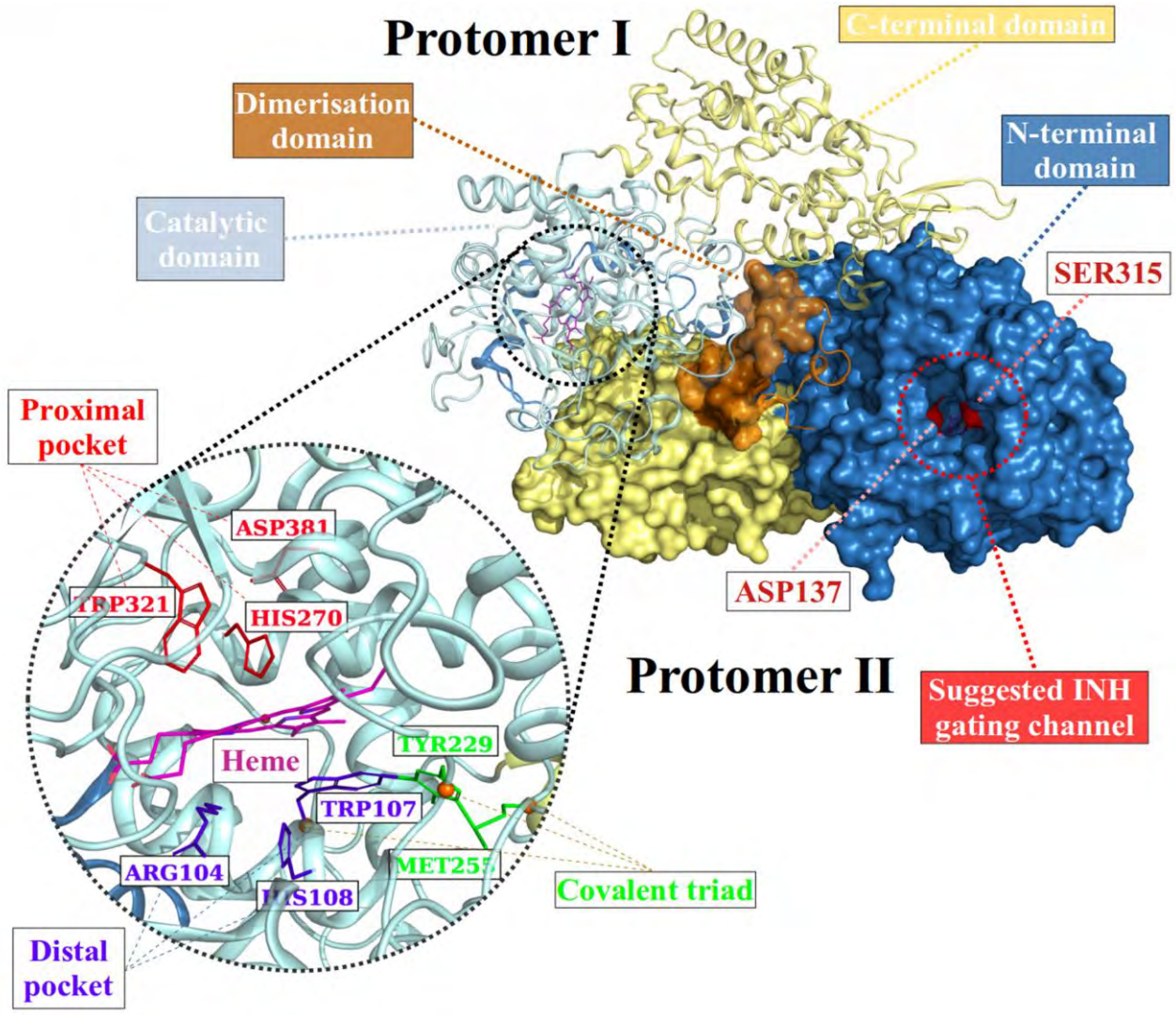

Figure 2. 2: Homodimeric structure of the *Mt*b KatG consisting of protomer A and B in cartoon and surface representation, respectively. The detailed view of the active site shows heme as magenta sticks and the active site residues proximal and distal to it as red and blue sticks, respectively. The KatG conserved triad (MYW) is shown as green (Tyr229, Met255) and blue (Trp107) sticks. The figure was generated by Olivier Sheik Amamuddy and adapted from Barozi et al., **[115]**.

2.1.5 Drug Resistance in TB

Due to drug and environment selective pressure, *Mtb* has over time developed mutations in key drug targets/activation proteins leading phenotypic drug-resistance. In the case of the most effective first-line drugs like INH, rifampicin and pyrazinamide, mutations in the drug-activating protein genes i.e., *katG*, *pncA* and *rpoB*, respectively result in drug-insensitive proteins [300–304]. 60 to 90% of the isoniazid phenotypic resistance is attributed to mutations in the *katG* gene [287, 305] with the S315T SNP documented as the most prevalent KatG mutation associated with INH resistance [306–311].

The marked increase in mycobacterial drug resistance especially to the most effective first-line drugs coupled with the scarcity of new drug interventions has created an urgency to decipher the behavior and mechanisms of action of the resistance-conferring mutations, to better inform drug design.

2.2 Chapter Aims and Objectives

This chapter is aimed at decoding the changes in conformational behavior and residue network patterns of the high confidence resistance conferring KatG mutations using computational approaches. To this end, the study objectives included i) identification of high confidence KatG mutations, ii) modeling of the mutant 3D structures, iii) molecular dynamic simulation of the wildtype and mutant structures and iv) post-MD trajectory analysis including DRN analysis to decipher communication patterns.

2.3 Materials and Methods

2.3.1 Identification and Retrieval of KatG Wildtype and Mutant Sequences

The *Mtb* KatG reference (WT) sequence and structure were retrieved from the Universal Protein Resource (UniProt) [312] and RSCB Protein Data Bank [313], respectively using the UniProt ID: P9WIE5 and PBD ID: 2CCA. KatG specific SNPs associated with phenotypic resistance to INH were retrieved from the TB Drug Resistance database (TBDReaMDB) [314] which classifies mutations based on prevalence, as identified through INH minimum inhibitory concentration (MIC) tests and genomic sequencing. Eleven highly prevalent mutations were retrieved under the high confidence TBDReaMDB classification.

2.3.2 Homology Modeling of Mutant KatG Structures

In the absence of mutant KatG structures in the PDB, models were generated using the heme containing WT (PBD ID: 2CCA) as a template in MODELLER [242]. Here, SNPs were individually introduced into the dimeric reference sequence using an in-house Python script before the mutant and WT sequences were aligned using ClustalΩ [228]. The alignment files were edited to fit the MODELLER *pir* format, and 100 models were generated for each mutant system using auto model and slow refinement MODELLER methods. The best models were selected based on the lowest z-DOPE score [246]. Model quality was further assessed using VERIFY3D [247] and ProSA [248].

2.3.3 All-atom Molecular Dynamic Simulations and Trajectory Analysis

The WT and mutant structures were subjected to 300 ns all-atom MD simulations in GROMACS [315], version 2019.4. Prior to the generation of GROMACS topology files, the protein structures were protonated at a pH of 7.0, based on the KatG functional kinetics [127], using the PROPKA tool from PDB2QR (version 2.1.1) [316]. The GROMOS54a7

force fields [317] were chosen as the FFs of choice based on their inclusivity of heme group parameters. To mimic *in-vivo* conditions, the protein structures were solvated with SPC216 water model in a cubic box of at least 1.0 nm solute-box edge clearance distance. After, the system charge was neutralized using 0.15M NaCl ions prior to energy minimization along the steepest descent minimization algorithm. Minimization was performed with a step size of 0.01 and without constraints until a threshold of 1000 kJ/mol/nm. Temperature equilibration under the NPT ensemble; constant number of particles, volume and temperature ensued using the Berendsen temperature coupling at 300 K for 100 ps followed by pressure equilibration under the NPT ensemble; constant number of particles, pressure and temperature using the Parinello-Rahman barostat [318] at 1 atm and 300 K for 100 ps. Subsequently, MD production runs for 300 ns were performed for the eleven mutant systems and the WT. All bonds were constrained during the equilibration and production runs under the LINCS algorithm [319]. Furthermore, for the long-range electrostatics, Particle mesh Ewald electrostatics [320] were used with Fourier spacing of 0.16 nm. For the short-range Coulomb and van der Waals interactions, a cut-off distance of 1.4 nm was used. Prior to MD trajectory analysis, the periodic boundary conditions were removed, and the trajectories fitted to the reference structure using *gmx trjconv* tool. Post-MD analysis involved calculation of RMSD, RMSF, Rg clustering and COM distance using the GROMACS in built tools, *gmx rms*, *gmx rmsf*, *gmx gyrate*, *gmx cluster* and *gmx distance*, respectively. The GROMACS generated results were analyzed and visualized using the following Python libraries; Seaborn [321], Pandas [322], pytraj [323], matplotlib [324], Numpy [325], and NGLview [326]. We also calculated pairwise RMSD calculations where a Python script was used to concatenate the all the system trajectories via the C_α atoms prior to calculating the RMSD for each frame in comparison to itself and all the other frames in the trajectory. System dynamics were visualized using the Visual Molecular Dynamics (VMD) tool [327].

2.3.4 Comparative Essential Dynamics

Comparative ED was used to identify the most dominant mutant system motions in relation to the WT using the *compare_essential_dynamics.py* script (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) from MDM-TASK web server [264]. Here, each mutant system trajectory was aligned to the WT trajectory via the C_α atoms before decomposition of the covariance variance matrix. Furthermore, the script allowed for exclusion of particularly highly flexible regions from the trajectory to reduce the noise from highly flexible loop regions. Here, the three C-terminal residues were excluded due to their pronounced flexibility. Results from comparative ED were presented as time-based scatter plots describing mutant motions in relation to the WT as explained by principal components one and two (PC1 and PC2). The scatter plots also indicated the time stamp in pico-seconds of the lowest energy conformation throughout the simulation.

2.3.5 Analysis of the Heme-Protein Interactions and COM Distance

The proclivity of heme to form H-bond interactions in the WT and mutant KatG systems was investigated using the *gmx hbond* tool from GROMACS and presented as line plots.

Besides the H-bonds, cofactors/ligands form other kinds of short-range interactions with the protein systems including van der Waals interaction, within a given interaction space 6.7 Å. Herein, the heme short-range interactions were determined using an ad hoc Python script. The script employed MDTraj [328], NetworkX [329], Numpy, Pandas and matplotlib Python libraries to determine the frequency of heme atomistic contacts within a Euclidian distance of 6.7 Å while using the last 50 ns of the trajectory and topology files as input. The obtained heme contacts were normalized across the different systems and presented as a heat map. Asymmetric protomer behaviour was assessed through determining the difference in heme

contacts between the protomer A and B (protomer A contact frequency - protomer B contact frequency).

Heme positional stability was also investigated through heme COM distance measurement from the respective protomer active pocket residues COM using the *gmx distance* tool. The last 50 ns of the trajectory were used.

2.3.6 Dynamic Residue Network Analysis

Residue level changes in network and communication patterns imposed by the mutations were investigated using DRN analysis. Herein, course-grained trajectories consisting of C_α and C_β atoms for the equilibrated last 50 ns of the trajectories were used with a default step size of 1. Five DRN metrics including *BC*, *CC*, *DC*, *EC* and *KC* were calculated using the *calc_network.py* Python script (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) from the MDM-TASK web server [264]. Here, residues with centralities ≥ the global top 4% were taken as hubs. The global top 4% DRN analysis results were presented as heat maps. Furthermore, residue hubs were mapped on the katG structures.

2.3.7 Contact Map Analysis

The changes in residue interactions at the mutation sites and *persistent hubs* within 6.7 Å were investigated using the *contact_map.py* and *contact_heatmap.py* Python scripts from the MDM-TASK web server [264]. The last 50 ns of each protein trajectory were used.

2.3.8 Analysis of Inter-protomer Binding Energy through Alanine Scanning

Dimerization of the KatG enzyme depends on the cross interaction between the dimerization domains of protomer A and B. The effects of the HC mutations on the dimerization behavior of KatG were investigated through inter-protomer binding energy calculation through Alanine scanning. Here, the interface face residues were individually substituted to an alanine

residue and the binding energy contribution at that position measured. Alanine scanning enabled the identification of stabilizing, neutral and destabilizing residues based in their binding energy contribution. The ROBETTA web server [330] was used for alanine scanning using the low energy basin structures from ED. The server identified interface residues as those within 4 Å of the opposite protomer/protein. Residues with binding energy ≥ 1 kcal/mol are regarded as destabilizing, < -0.8 kcal/mol as stabilizing and those with binding energy between -0.8 and 0.99 kcal/mol as neutral [331].

2.4 Results and Discussion

The results in this chapter are divided into two parts; Part one is focused on the global structural and conformational changes in enzyme due to mutations whereas part two delves further into the residue level effects of mutations while including the changes in cofactor interaction and residue networks.

2.4.1 Part One: Global Analysis

2.4.1.1 Distribution and Characteristics of KatG High Confidence Mutations

The eleven HC mutations with documented phenotypic INH-resistance as per the MIC and genomic studies [308, 332–335] *viz*: S140N, S140R, G279D, G285D, S315I, S315N, S315R, S315T, G316D, S457I and G593D were retrieved from the TBDReaMDB [314]. Of these mutations, nine are in the N-terminal domain (S140N, S140R, G279D, G285D, S315I, S315N, S315R, S315T and G316D), whereas the other two (S457I and G593D) are in the C-terminal domain (**Figure 2.3**). The most dominant KatG mutation, S315T [308, 310, 311, 335–338], is located at the access channel to the enzyme active site where the substitution of the a serine for a bigger threonine is believed to cause steric hinderance due to the methyl group [128, 132]. Besides the high prevalence, position 315 also had the most SNPs studied here i.e., S315I, S315N, S315R and S315T.

The mutations involved similar phytochemical residue substitutions for S140N, S140R, S315R, S315N and S315T which had polar for polar residue substitutions, whereas G279D, G285D, G316D and G593D involved substitution of a hydrophobic residue for a polar one. In S315I, the small and polar serine was replaced with the hydrophobic and aliphatic isoleucine.

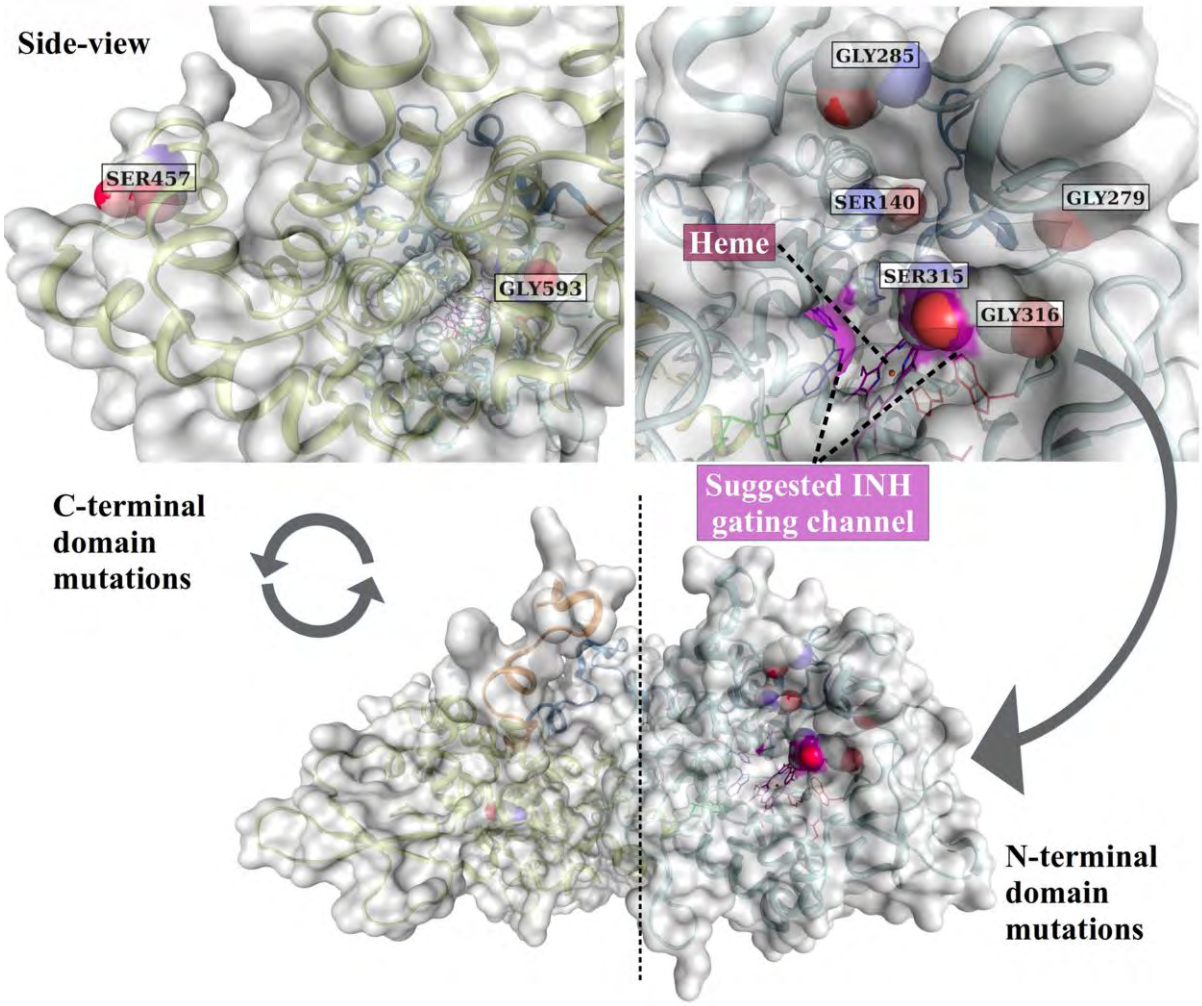

Figure 2. 3: Different orientations of the KatG structure showing the location of the 11 HC mutations. Mutation positions are labelled. The figure was generated by Olivier Sheik Amamuddy and adapted from Barozi et al., [115].

2.4.1.2 Conformational Diversity of the KatG Mutants

The global impact of mutations on KatG dynamics was assessed through assorted RMSD calculations. Firstly, the squared mean deviation line plots highlighted how both WT and mutant systems equilibrated early on in the simulation at approximately the 50 ns. Additionally, the line plots (**Figure 2.4A**) also showed a higher RMSD for G279D, G285D, S315N, S457I, and G593D compared to the WT. Of these systems, S457I was the most conformationally diverse. Further characterization of protomer stability using split violin plots showed that the WT and majority of the mutant systems *viz*: S140R, G279D, G285D, S315I, S315N, S315R, G316D, S457I and G593D experienced a multi-model RMSD

distribution. Additionally, the split violin plots showed asymmetric RMSD distribution between system protomers indicating contrasting protomer behaviour in solution as previously documented [76, 339]. The multimodal RMSD distribution suggested that on average, the systems sampled more than one conformation. Of all the systems, only S140N and S315T displayed a unimodal symmetric behaviour between protomers (**Figure 2.4B**).

The conventional RMSD calculations compare all the trajectory frames to the initial frame and as a result it informs on the gradual conformational evolution in relation to the first frame. In all versus all RMSD calculation, one is able to identify the time scales with similar conformational ensembles based on the comparison of each frame to itself and all other frames in the trajectory. Consequently, all versus all RMSD showed that the WT sampled two main conformations, the first being between 0 and 200 ns and the other from 200 ns onwards (**Figure 2.4C**). Interestingly, this was also true for most of the mutant systems but at differing time stamps. Like in the RMSD line plots, all versus all RMSD also identified S457I as having most structural diversity compared to other systems (> 4 Å RMSD). Furthermore, a statistical comparison of the WT and each mutant system mean RMSD using the Mann-Whitney U-test (given that the system RMSDs were not uniformly distributed) indicated no significant differences between the WT and each of the mutant systems with p-values > 0.05. These results however were not representative as only one data point (mean RMSD) was used for sample comparison. A more robust statistical comparison would require an n >1 for each sample.

System behavior was further investigated through Rg calculations which reflected nominal degree of gyration across all the protein system based on the scale, 2.775–2.975 nm (**Figure 2.4B**). The earlier observed asymmetric protomer behavior from RMSD was also evident in Rg where protomer A displayed a higher degree of gyration than B in most systems. Like in RMSD, statistical analysis using the Mann-Whitney U-test shows no significant difference

between the WT and mutant systems Rg will and intra-system protomer comparison with p-values > 0.05.

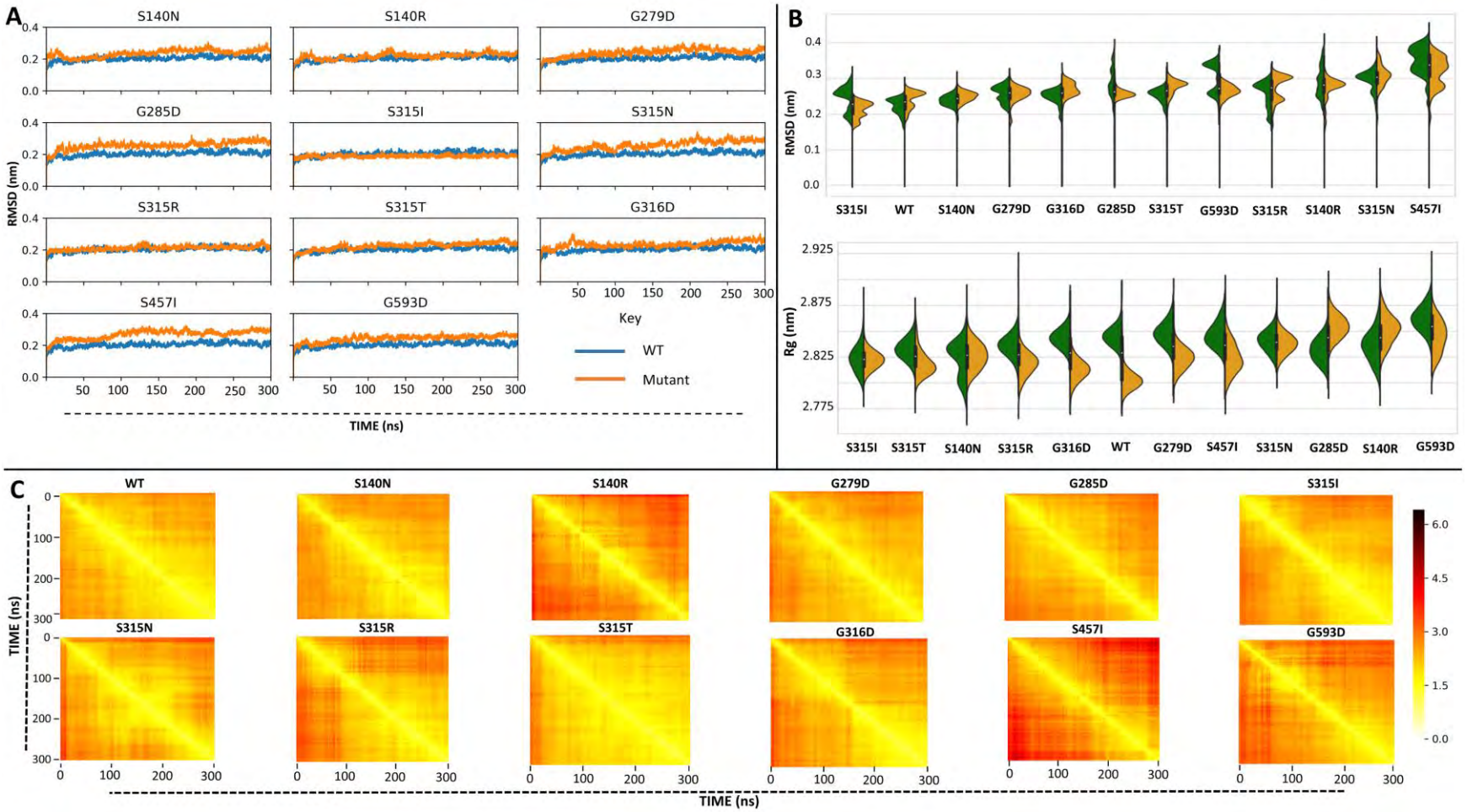

Figure 2. 4: A shows superimposed RMSD line plots of the WT in blue and mutant systems in orange. The x and y-axes indicate the time (ns) and RMSD (nm), respectively. B shows split violin plot showing protomer A (green) and protomer B (orange) RMSD and Rg distribution. C presents all versus all RMSD heat maps for each protein system. The color scale from white to dark red in C highlights the degree of conformational deviation. Adapted from Barozi et al., **[115]**.

2.4.1.3 Geometric Clustering Identified Two Main Conformational Ensembles

The GROMACS geometric clustering algorithm identifies, and groups MD trajectory frames based on their geometric similarity. Here, the mean conformational structures were identified and the MD frames assigned to these clusters based on an established standard deviation while ensuring that the frames in each group were unique [340]. Clustering identified two main conformations in majority of the systems with one of the clusters (C1) dominating the others in terms of number of frames (**Table 2.1**). The differences in the conformational ensembles were attributed to KatG flexible loop regions (**Figure S1**). Like in RMSD, S457I had the most conformational clusters (12) together with G593D. Evidently, the presence of these mutations in the KatG influence protein folding and dynamics. To gain a bigger picture into the most prominent protein motions, essential dynamics were determined for each mutant system in comparison to the WT in the next section.

Table 2.1: WT and mutant conformational frequency for each geometric cluster (C)

System												
	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**
WT	2306	561	56	38	34	6	-	-	-	-	-	-
S140N	2097	578	196	62	61	4	2	1	-	-	-	-
S140R	1144	863	482	361	44	41	33	20	10	2	1	-
G279D	2005	608	211	119	40	9	6	1	1	1	-	-
G285D	1523	1016	300	125	18	9	7	2	1	-	-	-
S315I	2112	813	43	15	15	1	1	1	-	-	-	-
S315N	2130	497	231	109	20	9	5	-	-	-	-	-
S315R	2003	793	105	55	41	2	1	1	-	-	-	-
S315T	2695	230	75	1	-	-	-	-	-	-	-	-
G316D	1785	772	212	145	40	20	16	8	3	-	-	-
S457I	1005	842	571	326	142	50	34	16	10	3	1	1
G593D	1311	880	380	227	102	52	23	16	7	1	1	1

2.4.1.4 Comparative Essential Dynamics

The application of the comparative ED approach from MDM-TASK web server [264] enabled the pairwise comparison of the most dominant motions between the WT and mutant systems within the same eigenspace. Here, comparative ED analysis was performed for the whole the protein, each protomer and active site to exhaustively explain the mutation effects on protein dynamics.

2.4.1.4.1 Mutant Systems Explored a more Diverse Conformational Space to the WT

The dynamics of each whole mutant system were individually compared to the WT along PC1 and PC2 which accounted for most variance [262]. From **Figure 2.5**, majority of the mutant systems explored a more diverse conformational area along PC1 compared to the WT. In S140N, PC1 and PC2 accounted for ~46 % and ~7% of the variance, respectively. Similarly, S140R, G279D, G285D, S315I, S315R, G316D, S457I and G593D had the most variation explained by PC1, where the total variance in these systems was ~ 40, 40, 46, 42, 49, 41, 58 and 57%, respectively (**Figure 2.5**). Comparative ED is preferred to conventional PCA because it enables a more representative comparison of systems dynamics. The comparative ED tool also identified simulation time stamps corresponding to the lowest energy structures, as shown in the scatter plots (**Figure 2.5**). Consequently, structural deviation of the low-energy mutant systems from the WT was investigated through the extraction of the mutant conformations at these time stamps and alignment to the WT. S315N had the highest RMSD of 3.396 Å (**Table S1**). The larger eigenspace in the mutant systems coupled with the low RMSD between their low energy structures and that of the WT implies that the mutant systems sampled a higher conformational space prior to equilibration.

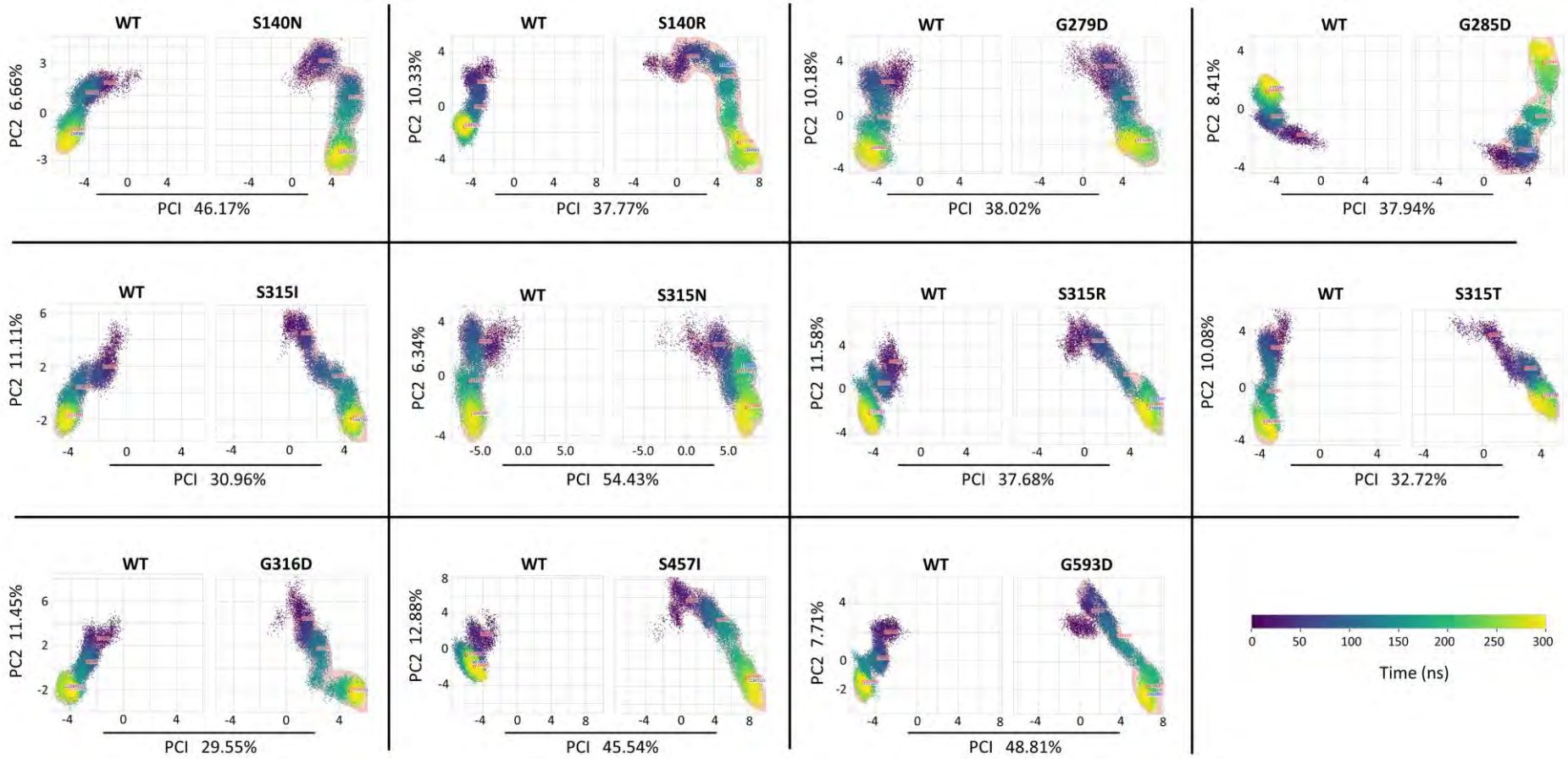

Figure 2. 5: Scatter plots of the comparative ED between each mutant system and the WT. the x and y-axes represent the variance as per principal components 1 and 2, respectively. The color scale from dark blue to yellow illustrates the progression of the simulation to 300 ns. Adapted from Barozi et al., **[115]**.

2.4.1.4.2 The Mutant Systems had Asymmetric Protomer Dynamics

Based on the asymmetry in the protomer RMSD results, **Figure 2.4B**, protomer behavior was further assessed through comparative ED relative to the WT. Here, each mutant protomer dynamics, A and B, were compared to the respective WT protomer. From **Figure 2.6**, each subplot consists of WT protomers on the top and mutant protomers at the bottom with protomer A on the left and B on the right. Focusing on the WT, it is apparent that in majority of the WT-mutant eigen subspaces, WT protomer A and B behave alike *viz*: S140N, S140R, G279D, S315N, S315R, S457I and G593D plots. Conversely, mutant dynamics shows asymmetric protomer behavior across all systems denoting that the mutations contribute to pronounced dimer asymmetry. To further explain the changes in protomer dynamics, ED were accompanied with RMSF calculations. For S140N, protomer B sampled more conformational space along PC1 (22.8%) compared to S140N protomer A and both WT protomers. RMSF results reflected a highly flexible S140N loop region (residues 359–375) in protomer B (**Figure S2**) which could be contributing protomer dynamics. Protomers A and B in S140R had a more spread-out distribution along PC1 and PC2 axes, respectively compared to the WT. Like in S140N, S140R also had a disordered loop region, 359–375 in addition to distinct flexibility for region 245–350 specifically in protomer A (**Figure S2**). Protomer A of G279D had a more span out distribution along PC2 (15.48%) compared to G279D protomer B and both WT protomers. Similarly, the G285D protomer A explored a more diverse conformational space than protomer B and both WT protomers, along PC1 (19.85%). This was also evident from the RMSD violin plots which showed multimodal RMSD distribution of protomer A (**Figure 2.4A**). The RMSF calculations also indicated a highly flexible 405-415 region in G285D protomer A which explains the dynamics. In S315N and S315R, the WT protomers had a more compact conformational distribution compared to the mutant protomers. Here, the mutant protomer dynamics were in different directions further

highlighting the asymmetry in these systems. In S457I and G593D, both mutant protomers had a more spread-out distribution compared to the compact WT.

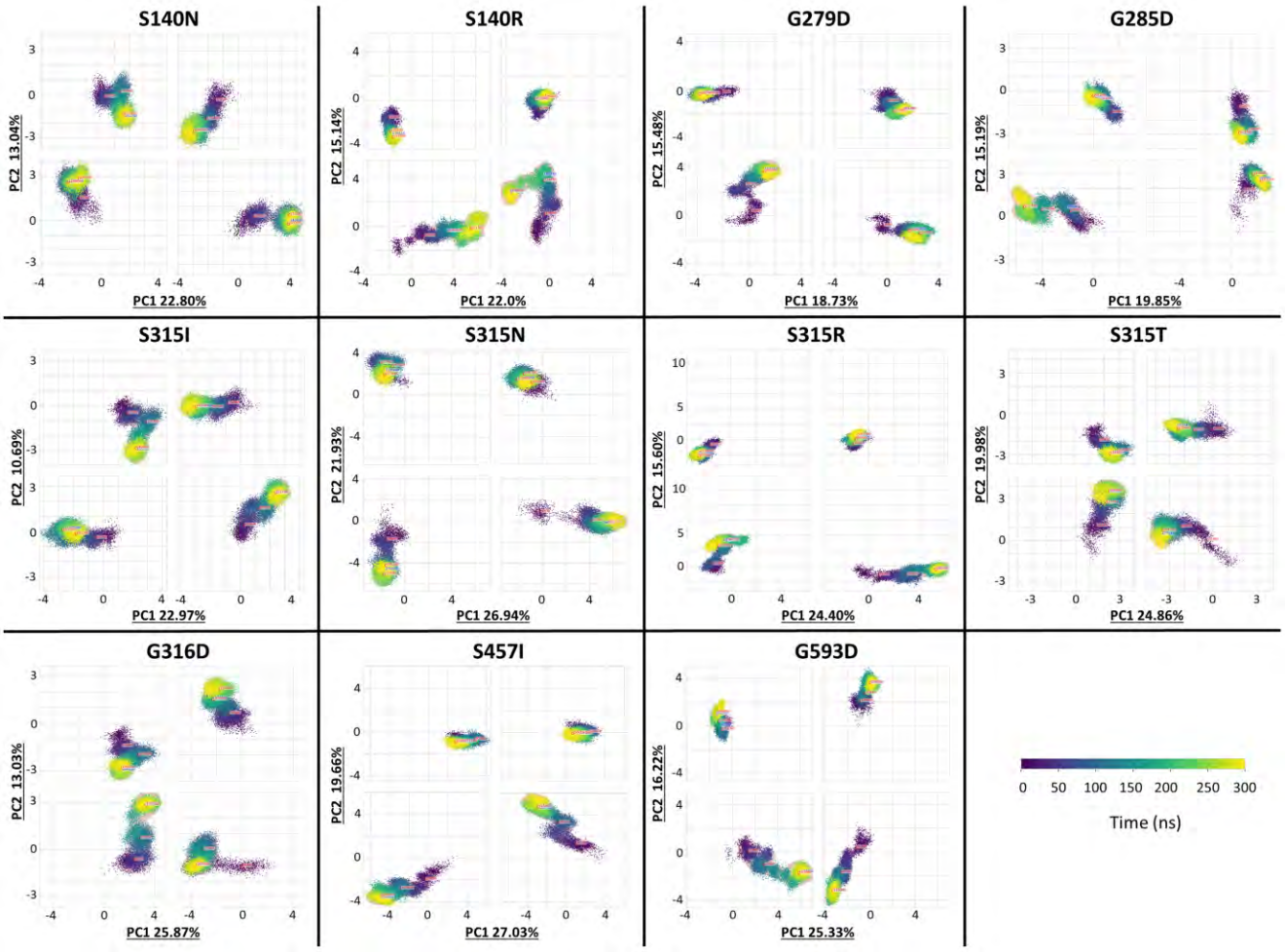

Figure 2. 6: Scatter plots of the protomer specific ED in comparison to the WT. For each subplot, the mutant and WT protomers are analyzed in the same eigen subspace. The WT protomers occupy the upper half and mutant protomers the lower half of each subplot whereas protomer A is to the left and B to the right. The x and y-axes show the percentage variance as per PC1 and PC2, respectively. Adapted from Barozi et al., **[115]**.

Interestingly a swap in protomer behaviour was noted in S315I, where protomer A of S315I behaved like WT protomer B while S315I protomer B like WT protomer A. This behaviour is not uncommon in homodimeric proteins especially in presence of mutations as recently observed in the SARS-CoV-2 M^{pro} enzyme [76]. Both protomers of S315T had more conformational space available to them than in the WT where PC1 and PC2 accounted for

24.86% and 19.98% of the variance. Similar behaviour was noted in G316D with where the variance was 25.87% and 13.03% for PC1 and PC2, respectively.

It is evident from both protein and protomer ED that mutations affect the conformational evolution resulting in a more unstable KatG with asymmetric protomer behaviour. Of all the mutants here, the S457I SNP showed the greatest effect on protein dynamics from RMSD, RMSF and ED calculations. These conformational changes appear to be driven mostly by marked N-terminal residue fluctuations while leaving the C-terminal domain steadily coordinated. Subsequently, the mutation effects on the predicted INH binding cavity were investigated next through binding pocket ED.

2.4.1.4.3 The Mutant Binding Pocket Experienced Assorted Conformational Changes

Chaplin et al., [71] recently illustrated how the KatG mutations, W107R and T275P significantly affected the binding pocket environment resulting in an unstable heme coordination with some mutant protomers failing to retain heme upon inspection of the Cryo-EM KatG structures. Here, ED analysis of the heme encompassing active pocket residues i.e., 91–95, 97, 98, 100, 101, 103, 104, 107, 108, 136, 137, 139, 140, 205, 224, 227–233, 248, 252, 265, 266, 269, 270, 272–276, 281, 309, 312–315, 317, 321, 326, 350, 378, 380, 381, 408, 412, and 415 somewhat mimicked the respective protomer behavior. The high conformational diversity observed at the protomer level was translated to the binding pocket as well (**Figure 2.7A**). And like in the protomer, asymmetry in the binding pocket dynamics was observed between mutant protomers of the following systems, S140R, G285D, S315I, S315R, G316D, and G593D. G285D protomer A binding pocket residues explored the most conformational space in comparison the WT where the total variance per PC1 and PC2 was ~49%. Furthermore, RMSD and Rg analysis showed that the mutant binding pocket residues had a higher degree of deviation and gyration (except S315N) than the WT (**Figure 2.7B**).

From the global perspective, the RMSD, RMSF, Rg and ED results inform on the diverse conformational sampling and instability of KatG mutant systems. The protomer and binding pocket-focused analysis further confirm that the observed global changes affect the binding pocket environment and dynamics. The residue level mutation effects are discussed in part two of the discussion.

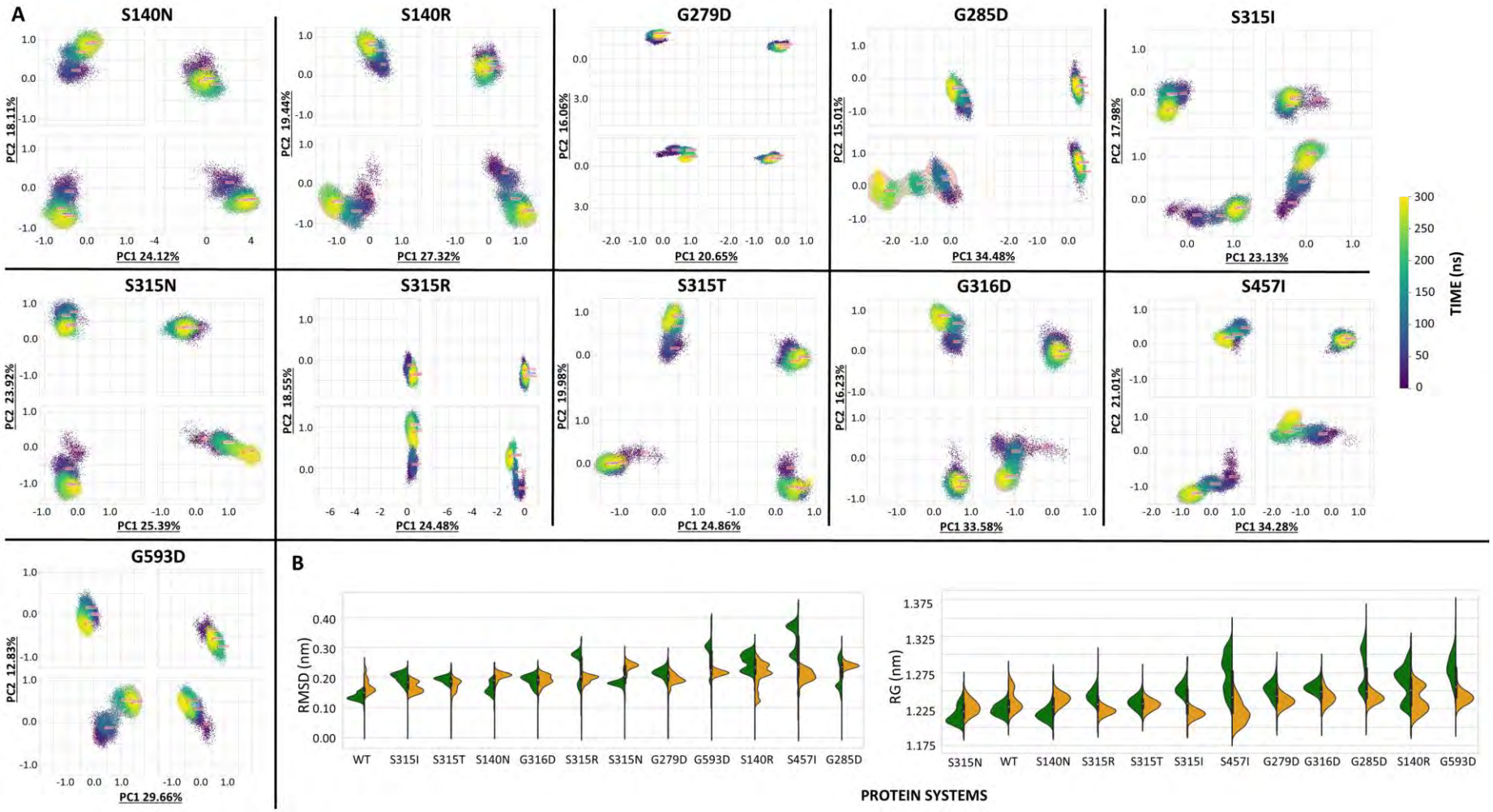

Figure 2. 7: A shows scatter plots of the mutant binding pocket ED in comparison to the WT. In each subplot, the WT protomers occupy the upper half and mutant protomers the lower half of where protomer A is to the left and B to the right. B shows binding pocket RMSD and Rg for protomers A (green) and B (orange). Adapted from Barozi et al., **[115]**.

2.4.2 Part Two: Local Analysis

This section focuses on the analysis heme and residue level mutation effects using interaction analysis approaches. Furthermore, the mutation induced changes on residue importance and influence in the protein network were investigated using DRN analysis.

2.4.2.1 Some Mutants had Variations in Heme Interaction Profile Compared to the WT

The marked changes in protein dynamics at the protomer and binding pocket level prompted further investigation into heme behaviour and coordination in the binding pocket. As recently shown [71], KatG mutations, W107R and T275P create a disorder of the heme environment affecting heme coordination. Heme forms H-bonds, van der Waals and covalent interactions with proximal and distal resides in the binding pocket. Therefore, the change/increase in binding pocket dynamics is bound to affect the ability of heme to form bonds. The propensity of heme to form H-bonds in the mutant systems was examined through H-bond measurement over the last 50 ns of the trajectory. From **Figure 2.8A**, heme group formed more hydrogen bonds in protomer A in majority of the mutant systems compared to protomer B. Interestingly, the reverse was true for the WT. As earlier observed from **section 2.4.1.4.2** and [76], mutations can have reverse effects on protomer behavior in homodimers. Generally, heme formed more H-bonds in the WT than in majority of the mutant systems especially in protomer A. G279D, G285D and G316D formed the least number of H-bonds in the mutant systems.

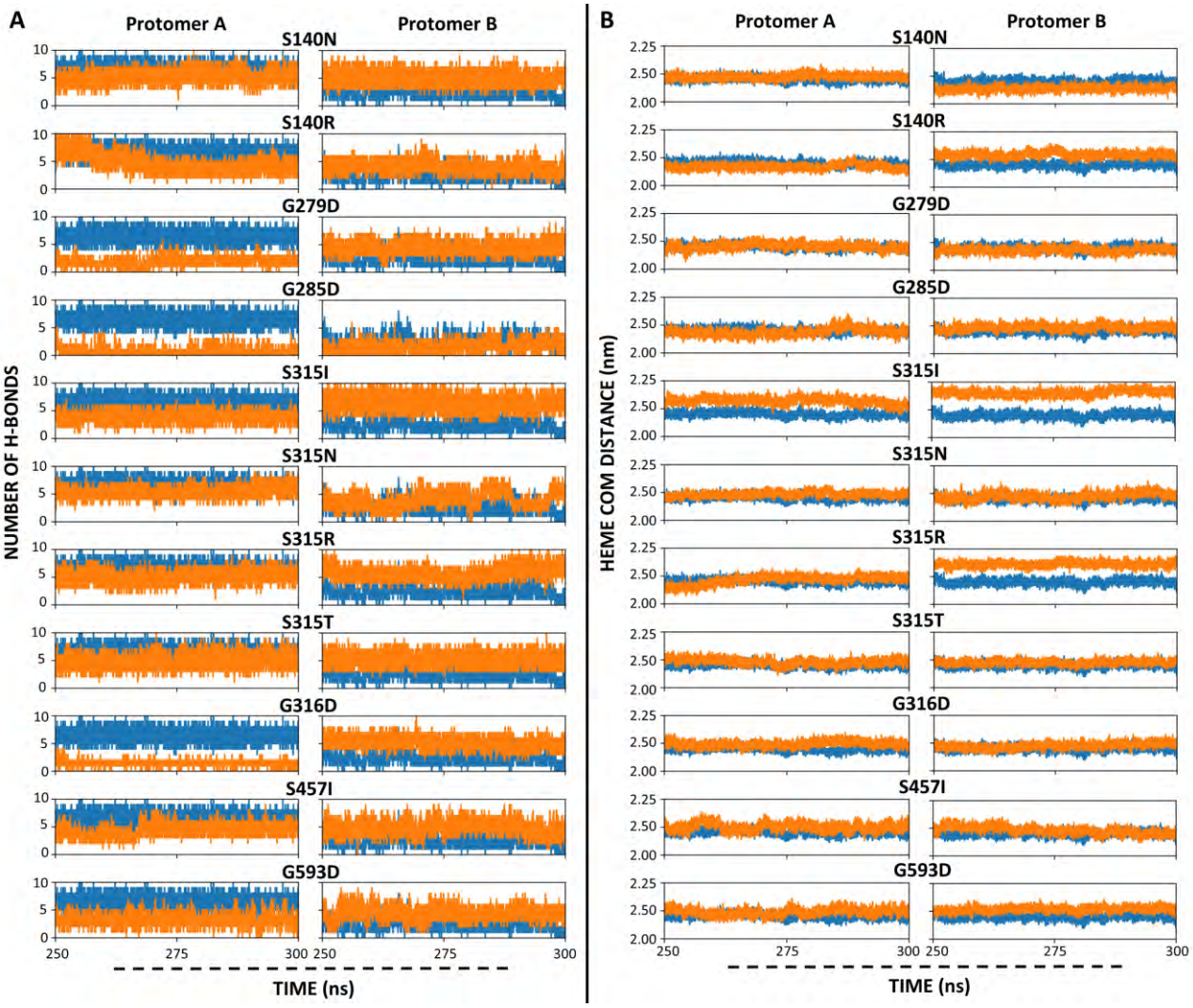

Figure 2. 8: A shows superimposed WT (blue) and mutant (orange) line plots of the heme H-bond progression over the last 50 ns. B shows comparative line plots of the heme COM distance to the active site COM between the WT (blue) and mutants (orange). Adapted from Barozi et al., [115].

The effect of reduced H-bonds on the heme positional stability was appraised through COM distance measurement between heme and binding pocket residues as determined by the CASTP web server [341]. Here, majority of the mutant systems showed nominal differences to the WT except for both protomers of S315I and protomer B of S315R which showed increased COM distance between the heme group and pocket residues (**Figure 2.8B**). The notable reduction in H-bonds especially in the mutant protomer A suggests decreased heme

stability in the binding pocket. Consequently, all other heme interaction and contact frequencies over the equilibrated trajectory were analysed in the in the next section.

2.4.2.2 Characterization of the Mutant-Heme Atomic Contacts Relative to the WT

The WT and mutant heme interaction frequency in each protomer was quantified using a tweaked version of *contact_map.py* (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) tool from the MDM-TASK web server [264]. The modification enabled summation of the residue specific atomic contact frequencies before normalising them across all the systems. Normalization allowed for ranking of contacts from 0 to 1, with 0 signifying no contact and 1 meaning full contact throughout the simulation period. Results from heme contact frequencies were presented as heat maps per protomer.

Congruent to the previous results, we observed slight asymmetric behavior in the WT heme contacts between the protomers. In WT protomer A, heme maintained five main residue contacts including, Pro100, Arg104, Trp107, His270, His276, Thr314 and Ser315. Of these, Arg104 and Trp107 are distal heme residues, His270 is proximal whereas Ser315 gates the access channel to the heme-containing active site pocket. In protomer B, less heme contacts were noted compared to protomer A with heme exhibiting a 6-fold reduction in Arg104 contact and a complete loss of contact with His276.

Compared to the WT, considerable differences in heme-protein interactions were observed in the mutant systems. In the WT, heme forms hydrophobic interactions with Ile103, Arg104, Trp107, Pro232, Ile248, Phe252, Lue265, Ile266, Thr275, Trp321, Leu378, Thr380, Phe408 and Trp412; H-bonds with Lys274, His276, Thr315 and Ser315; Pi-Pi stacking interactions with Trp107; salt bridges with Arg104 and His276; metal complexes with His270, and water bridges with Arg104 and His270. A number of these interactions were significantly reduced

and even lost in some of the mutant systems. Additionally, the mutation effect on heme interaction was specific in each protomer.

In protomer A, heme of S140N exhibited change contact frequencies with a 1.6-fold reduction in contact with Arg104, 4.3-fold reduction with His270, 2-fold reduction with His276 and 3.5-fold reduction with Ser315 (**Figure 2.9A**). Similarly, S140R had a 1.9-fold reduction in hydrophobic interaction with Arg104, 2.05-fold reduction in H-bond with Ser315 and a complete loss of metal complexes and H-bond with His270, His276 and Thr314, respectively. Complete loss of heme interaction for His270 was also observed in G285D and S315I. Likewise, the heme group in G279D, G285D and G316D had a total loss of interaction with His276. Interestingly, the most loss of heme interaction in protomer A in the mutants compared to the WT was at residue Thr314 where all the mutant systems lost interaction except for S140N, S315N and S315T. At Ser315, mutants G316D and S457I had a 2.4-fold and 4.4-fold reduction in heme interaction, respectively compared to the WT, whereas G279D, G285D, S315I, S315N and G593D completely lost contact with heme at this position.

The WT heme interactions in protomer B were comparable to protomer A except for the Arg104 and His276 heme contacts which were reduced. More asymmetry in heme interaction was observed at residue His270 where S140R exhibited some heme interaction in protomer B compared to the complete loss of interaction in protomer A. Similarly, heme interactions at His270 in protomer A of S315N and S315R were completely lost in protomer B of the same systems. It is important to note that His270 is a key residue in the KatG activity, where it forms electrostatic interactions with iron in heme ensuring a stable heme coordination [71, 291]. Another key residue in heme interaction is His276 which anchors heme in the active pocket [122]. Compared to protomer A, heme in protomer B gained interaction with His276 in S141R, G279D and G316D.

The asymmetry between protomers across all systems was clearly illustrated in **Figure 2.9B**, where the delta in heme-protein contacts (protomer A minus protomer B) was calculated and presented as a heat map. The WT showed minimal differences between protomers whereas pronounced inter-protomer differences were observed in S140N, S140R, S315N, S315R, G316D and G593D as earlier discussed.

Based on the importance of heme in the catalytic activity of KatG [342], the observed reduction, and in some cases, complete loss of heme interaction with key functional residues like His270, His276 and Ser315 informs on the destabilizing effect of mutations on the heme environment as a resistance mechanism. This is in agreement with finding from Chaplin et al., [71] who showed a destabilized heme environment as a result of W107R and T275P KatG mutations.

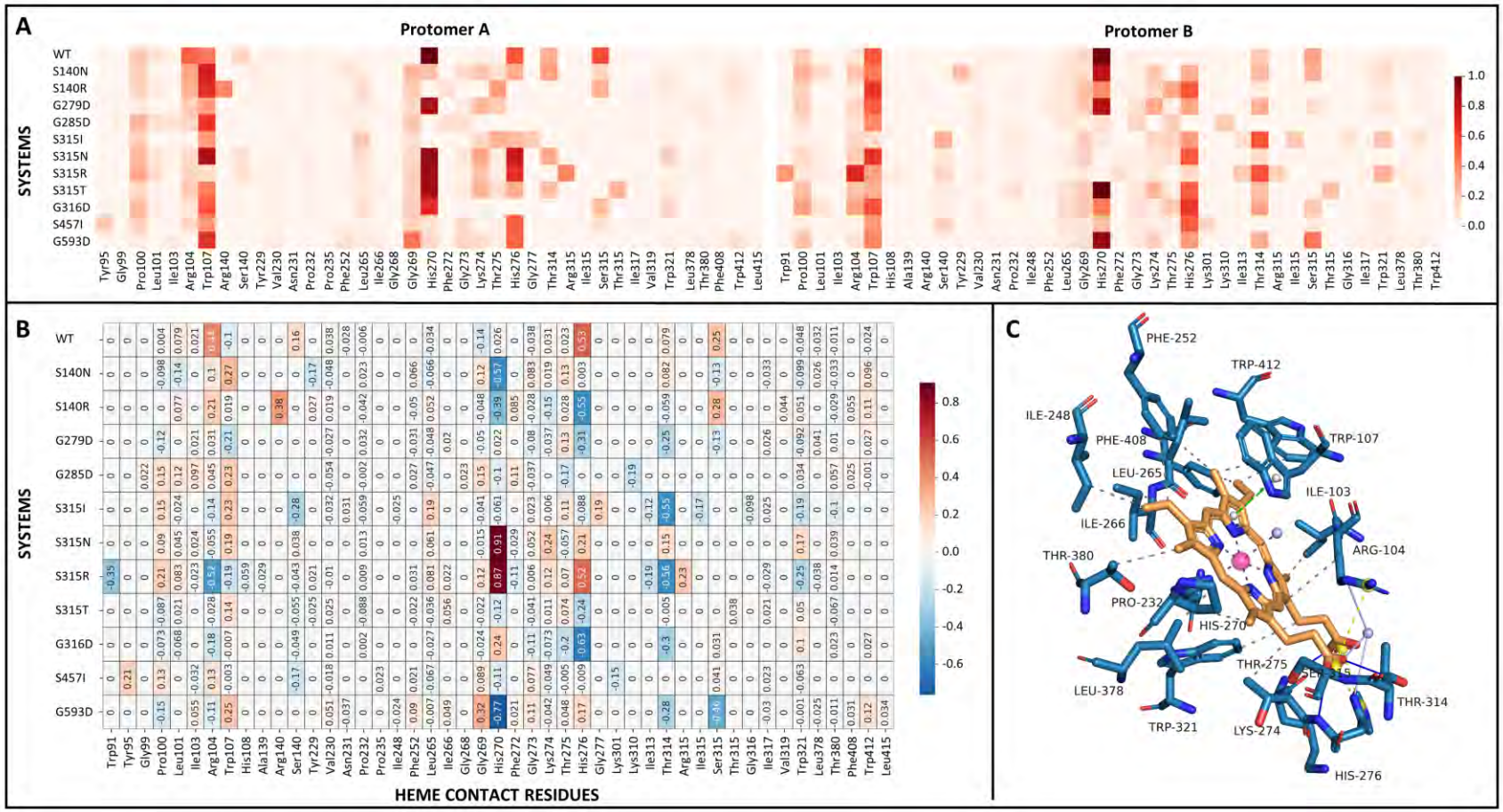

Figure 2. 9: A is a heat map of the heme contacts in protomer A and B. The color scale from brown to dark red shows the frequency of interaction. B shows the delta heme contacts between protomer A and B (A minus B), where blue indicates increased contact frequency in Protomer B and red increased frequency in A. C is a spatial representation of the heme (orange) contacts highlighting the hydrogen bond (blue), hydrophobic (dotted line), Pi-Pi stacking interactions (green), and water-bridged (grey) interactions in the reference structure (PDB: 2CCA) as viewed in PLIP **[343]**. The water molecules are shown as light blue spheres. Adapted from Barozi et al., **[115]**.

2.4.2.3 Dynamic Residue Network Analysis

Based on the KatG protein size, we defined the global high centrality threshold as the top 4% as opposed to 5% or 3% as these thresholds resulted in noisy and insufficient data, respectively.

2.4.2.3.1 Averaged Betweenness Centrality

From **Figure 2.10**, the high centrality residues in each protomer per system are annotated with the centrality values while homologous residues from other system are not annotated, but rather included for comparability. Asymmetric distribution of hubs was noted between systems and protomers. No *persistent hubs* were identified in either system protomers however, protomer A residues, Pro29, Leu43, Leu48, His49, Glu195, Gln190 and Leu616; and protomer B residues Leu43, Asn44, Leu48, His49, Glu195, Tyr197, Phe483 and Leu616 had hub status in at least six of the twelve systems. Most importantly, residues Asn35, Tyr197, and Phe483 were hubs exclusively in the mutant systems *viz*: Asn35 in A140N, G279D, G285D, G316D, and S457I; Tyr197 in G285D, S315I, S315N, S315T, S457I, and G593D; and Phe483 in S140R, G279D, G285D, S315N, S315R, G316D, and G593D. Of these residues, Tyr197 is implicated in forming hydrophobic interactions with Tyr28 from the opposite protomer, facilitating protein dimerization. Because of its low oxidation potential, Tyr197 is also believed to be the site for the porphyrin π-cation radical formation upon the reaction of KatG with hydrogen peroxide [344]. Furthermore, majority of the hubs present in at least six systems are part of the first 100 residues in each protomer that make up the dimerization domain. In KatG, the dimerization domains from opposite protomers form a hook-like interaction around each other, enabling homodimeric structure formation in KatG [291]. The high centrality of the domain especially in the mutant systems highlights its

importance and signals to a compensatory mechanism to sustain protein function amidst mutations.

Previous application of DRN analysis have identified distinct allosteric paths connecting one end of the protein to the other [76, 155]. Here, mapping of the hubs onto the protein structures showed clustering around the protomer interface and dimerization domain (**Figure S3**). The protein interface region is important in facilitating protein-protein interaction which is key in molecular interaction and overall functioning of organisms [345].Furthermore, protein interfaces are important in maintaining structure integrity as in the case of KatG [346].

Interestingly, several hubs were lost in the mutant systems compared to the WT however, compensatory hub gains in the mutants were centered around the binding cavity and the dimerization domain (**Figure S3**). Of the mutant exclusive hubs (Ala109, Trp135, Ala221, Thr251, Phe252, Met255, Thr275 and Thr314), residues Thr275 and Thr314 also showed increased interaction frequency with heme (section 2.4.2.2) in the mutant systems compared to the WT *viz*: protomer A for S140N and protomer B for S315R and S315T (**Figure 2.11**).

Another interesting observation from *BC* analysis was the allosteric effect of the C-terminal mutations, S457I and G593D, on the distribution *BC* hubs around binding pocket illustrated through the clustering of new *BC* hubs around the cavity. S457I is located approximately 60 Å from the heme group in the active pocket however, the mutation resulted in clusters of new hubs around the active pocket in the N-terminal domain compared to the WT (**Figure 2.12**). However, some of the key active site residues, Met255 Asn138 and Ala139 lost the hub status in S457I. Similarly, the active site access channel residues Asp137 and Ser315 had high *BC* values in the WT (A: D137-0.036, B: D137-0.012, A: S315-0.006, B: S315-0.191) compared to S457I (A: D137-0.011, B: D137-0.009, A: S315-0.004, B: S315-0.007). A keen

look at residues Asp137 and Ser135 behaviour through contact map analysis and RMSF showed that the residues had increased flexibility in S457I which resulted in less interaction and low *BC* values (**Figure S4**). The *BC* results not only highlight the importance of the dimerization but also illustrate the allosteric effect of mutations on protein communication patterns.

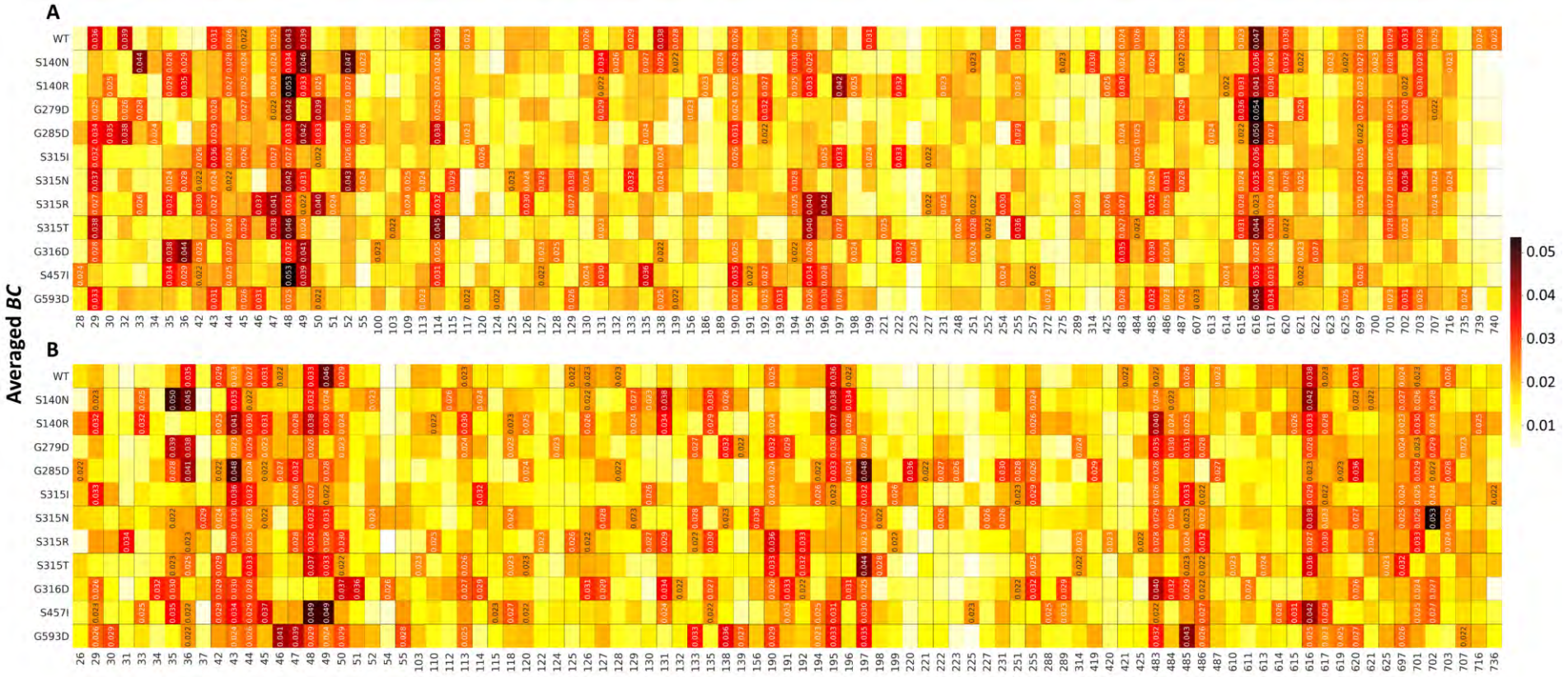

Figure 2. 10: A and B are heat maps representations of the *BC* hubs for protomers A and B, respectively as determined by the global 4% cut-off. Hubs in each system are annotated with centrality values whereas the homologous residues from the other systems are not. The color scale from yellow to dark red shows the degree of variation of *BC* values across the residues. Adapted from Barozi et al., **[115]**.

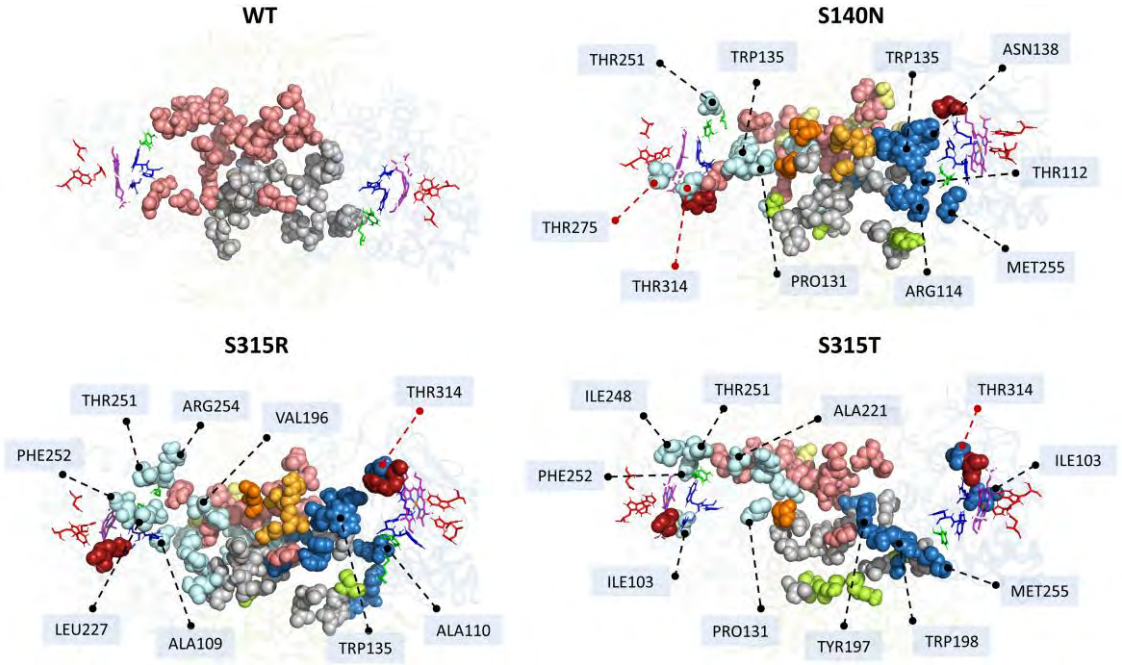

Figure 2. 11: WT and mutant structures showing the distribution of *BC* hubs. Hubs unique to mutant protomer A and B are shown as cyan and sky-blue spheres, respectively whereas the WT hubs as salmon and grey spheres. Sticks show the heme (magenta) proximal (red) and distal (blue) residues whereas mutation positons are shown as firebrick spheres. Mutant unique hubs around the active site are labelled. Adapted from Barozi et al., **[115]**.

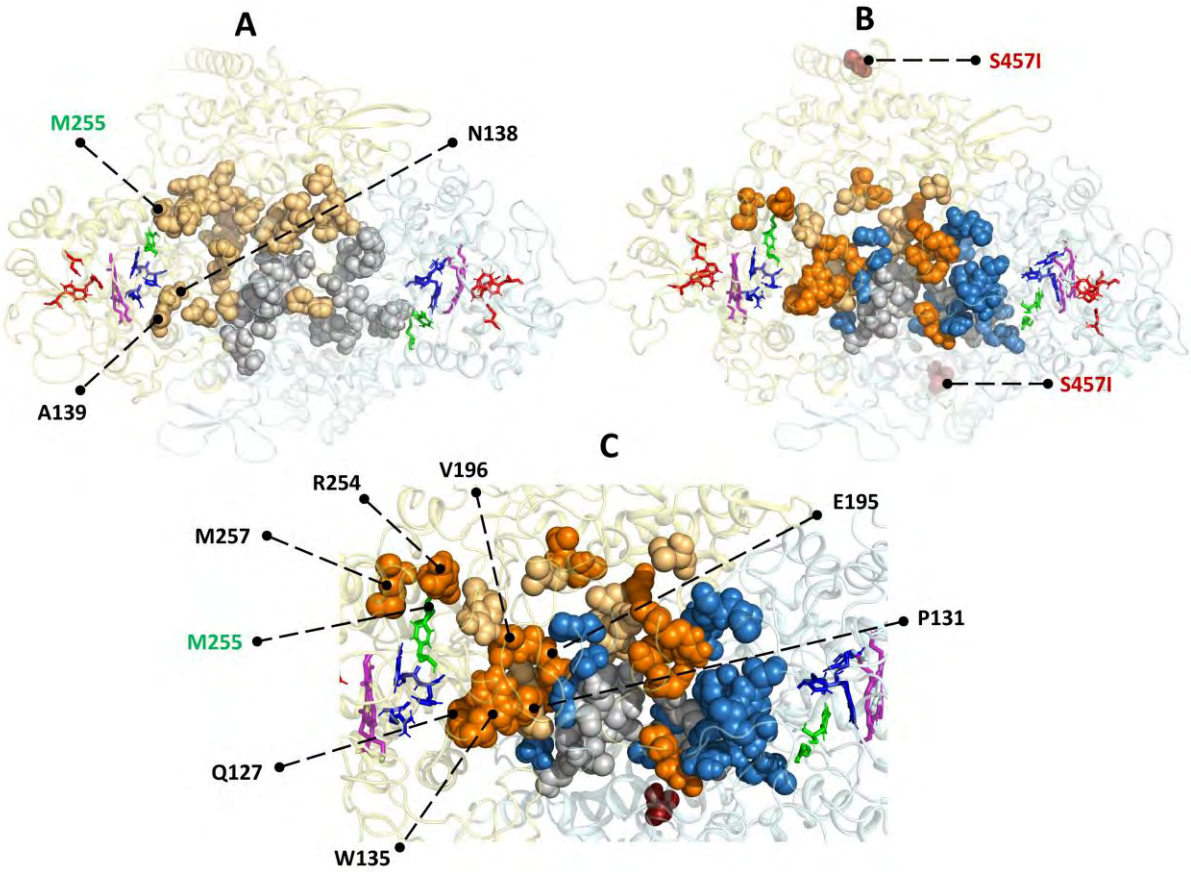

Figure 2. 12: KatG structures showing the distribution of *BC* hubs in the WT (A) and S457I (B and C). Functional residues with high *BC* in the WT and not the mutants (S457I) are labelled in A. Additionally, hubs unique to the S457I are labelled in C (black), whereas triad residue Met255 with low BC values in S457I is shown in green. The position of the S457I mutation is shown as firebrick spheres. Adapted from Tastan Bishop et al., [86].

2.4.2.3.2 Averaged Closeness Centrality

Analysis of the global top 4% *CC* hubs showed that the interface and dimerization domain had with highest concentration of *CC* hubs (**Figure S5 and 2.13**). Here, *persistent hubs* were identified as Asn44, Leu45, Val47, Leu48, His49, Ala621, and Glu703. Majority of the *CC persistent hubs* (Asn44, Leu45, Val47, Leu48, His49) form part of the dimerization domain whose importance was earlier highlighted under *BC* analysis. Based on the importance of the interface residues in protein-protein interaction, the relationship between *CC* hub distribution and the closeness of interaction between protein protomers was investigated through COM distance measurement.

Minimal differences were observed between the WT and mutant inter-protomer distances except for S315N and S457I where an increase of approximately 0.3 nm compared to the WT was noted between the mutant protomers (**Figure S6**). Interestingly, these systems also showed fewer number of *CC* hubs compared to the WT and other mutant systems, **Figure 2.13** (G, K).

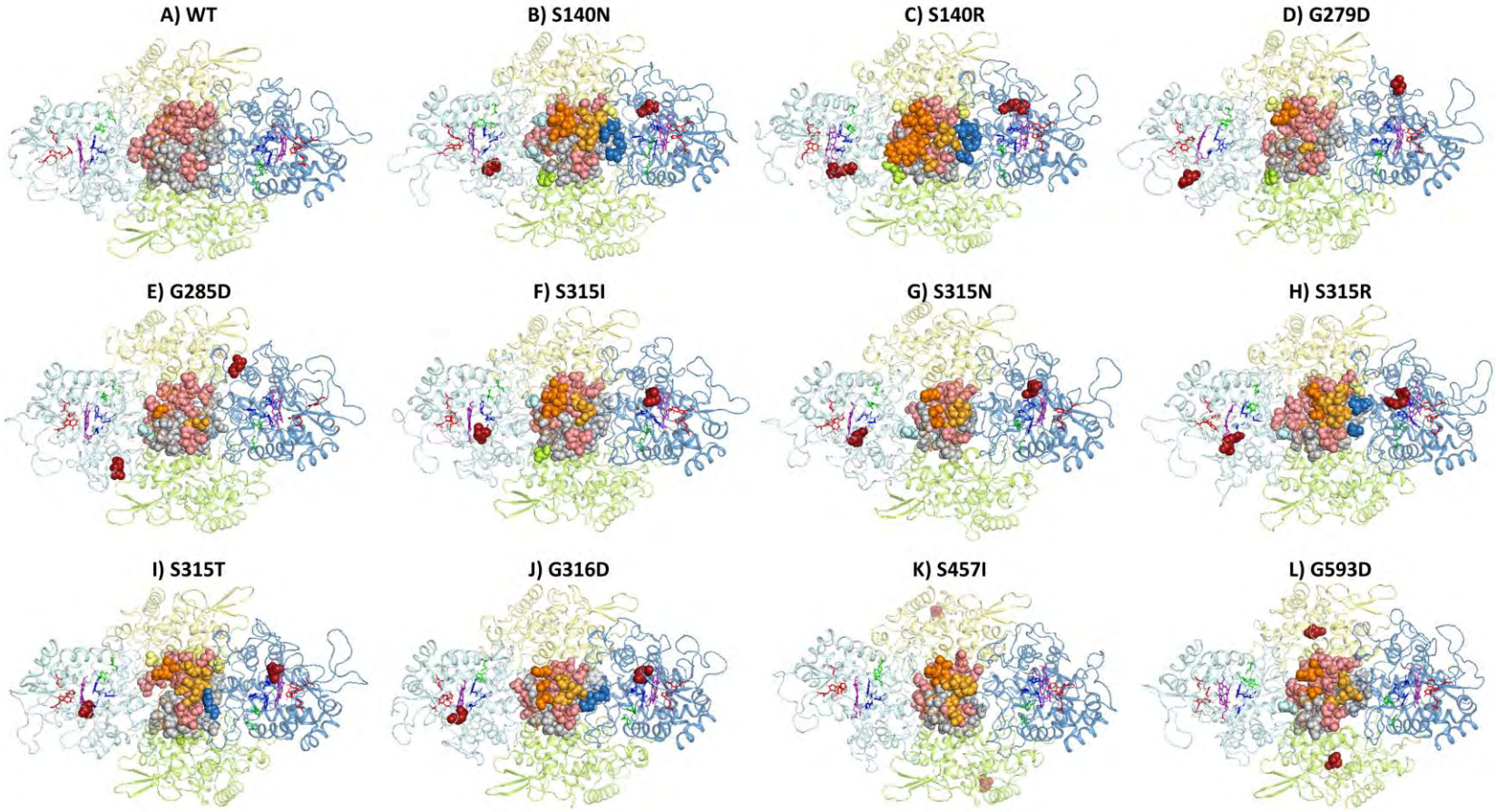

Figure 2. 13: Cartoon presentation of the WT (A) and mutant structures (B-L) showing the distribution of the global top 4% *CC* hubs. N and C-terminal domains of protomer A are shown as pale cyan and yellow, respectively whereas those of protomer B are sky-blue and lime. Dimerization domain is orange and bright yellow for protomer A and B, respectively. WT and mutant common hubs are shown are grey spheres (protomer A) and salmon spheres (protomer B). Mutant unique hubs are colored as the structure domains. Firebrick spheres are mutation positions. Adapted from Barozi et al., **[115]**.

2.4.2.3.3 Averaged Degree of Centrality

Averaged *DC* informed on the local connectivity of a residue in the protein network based on the immediate neighbors within 6.7 Å. From the global top 4% high centrality calculation, **Figure S7**, a close range of *DC* values implied a uniform degree of connectivity across all the systems. Furthermore, mapping of the *DC* hubs showed a uniform distribution of hubs in the N and C-terminal domains across the mutants however, more *DC* hubs were noted in the N-terminal domain of the WT compared to the mutants especially around the binding pocket (**Figure S8**). This was attributed to the pronounced flexibility of the binding pocket and heme environment area in the mutants as earlier discussed in **section 2.4.1.4.3**.

2.4.2.3.4 Averaged Eigenvector centrality

Residue influence in the protein systems was evaluated through averaged *EC* calculations where the assignment of centrality is based on both the residue connectivity and that of its neighbors [267]. The *EC* results further highlighted the asymmetry between system protomers, and between the WT and mutants (**Figure 2.14**). For the WT, only five residues, Leu472, Val473, Ala476, Gly547, and Gly548 were identified as hubs in protomer A compared to at least seventeen in the mutant systems. In protomer A, majority of the mutant systems had their hubs concentrated around the C-terminal region, residues 468-479 and 545-557. Additionally, only S140R, G279D, S315I and S315T had hubs in the N-terminal domain region 105-189 viz: **S140R**: 105, 106, 109–127, 162, 165, 166, 187, 190–194, 418, 419; **G279D**: 102–106, 109, 110, 121, 122, 165, 166, 169; **S315I**: 102–113, 118, 120–189, 256, 415, 416, 418, 419, and **S315T**: 106, 109–126, 166, 164, 196, 418, 419.

Interestingly in protomer B, S140N, G285D, S315I and G593D didn't have any hubs under the global top 4% criteria. Furthermore, the WT had the most hubs in protomer B contrary to protomer A. Mapping of the *EC* hubs (**Figure 2.15**) accentuated the influence of the C-

terminal domain for KatG functionality, as majority of the *EC* hubs were concentrated here. The C-terminal domain has previous been thought to be important KatG activity by maintaining the structural integrity of the active site through the cross talk with the N-terminal domain [295, 347].

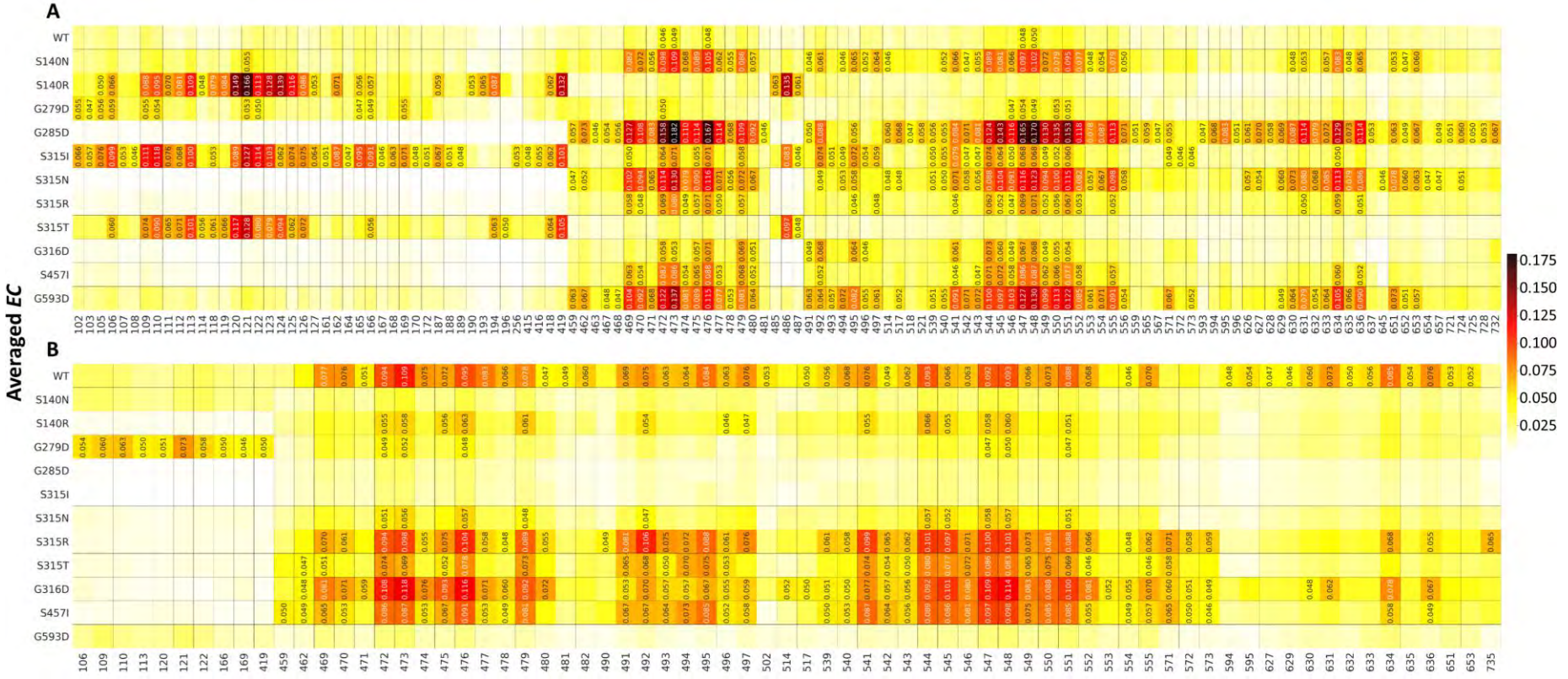

Figure 2. 14: A and B are heat maps showing *EC* hubs in protomers A and B, respectively as determined by the global 4% cut-off. The hubs per system are annotated with centrality values, whereas the homologous residues from the other systems are not. The color scale from yellow to dark red shows the degree of variation of *EC* values across the residues. Adapted from Barozi et al., **[115]**.

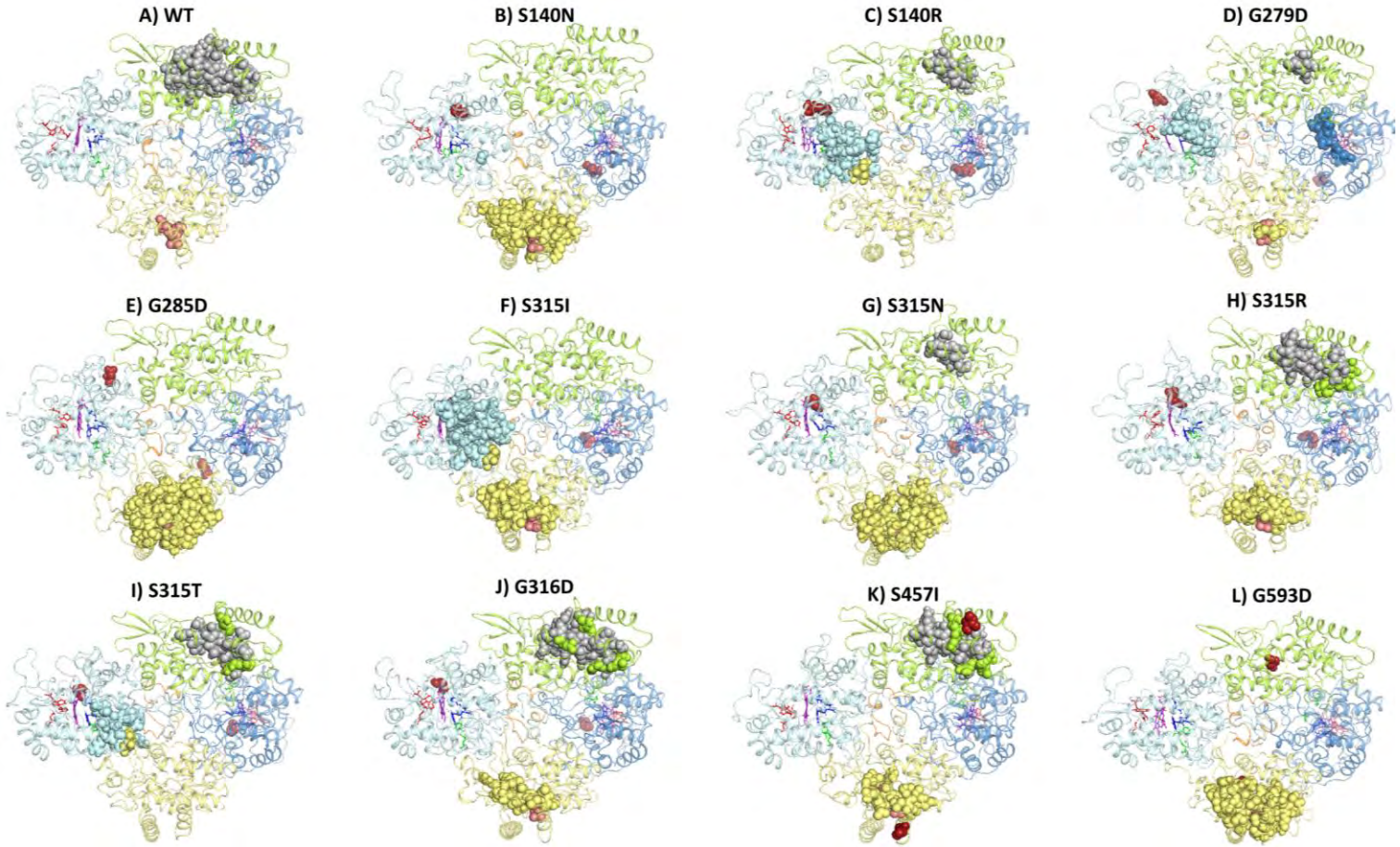

Figure 2. 15: Cartoon presentation of the WT (A) and mutant structures (B-L) showing the distribution of the global top 4% *EC* hubs. N and C-terminal domains of protomer A are shown as pale cyan and yellow, respectively, whereas those of protomer B are sky-blue and lime. The dimerization domain is orange and bright yellow for protomer A and B, respectively. Hubs common to the WT and mutant are shown as grey spheres (protomer A) and salmon spheres (protomer B). Mutant unique hubs are colored as the structure domains. Firebrick spheres are mutation positions. Adapted from Barozi et al., **[115]**.

2.4.2.3.5 Averaged Katz Centrality

Here, a more uniform distribution of *KC* hubs was observed between protomers A and B (**Figure S9)** with the C-terminal domain accommodating majority of the hubs (**Figure S10**). Furthermore, the *persistent KC hubs* i.e., Leu472, Val473, Gly547, Gly548 and Ala551 were all in the C-terminal domain further highlighting its importance.

Mapping of all the DRN *persistent hubs* (**Figure 2.16**) emphasized the importance of the dimerization domain at the interface in addition to the C-terminal domain which harbour all the *persistent hubs* from the all the five DRN metrics.

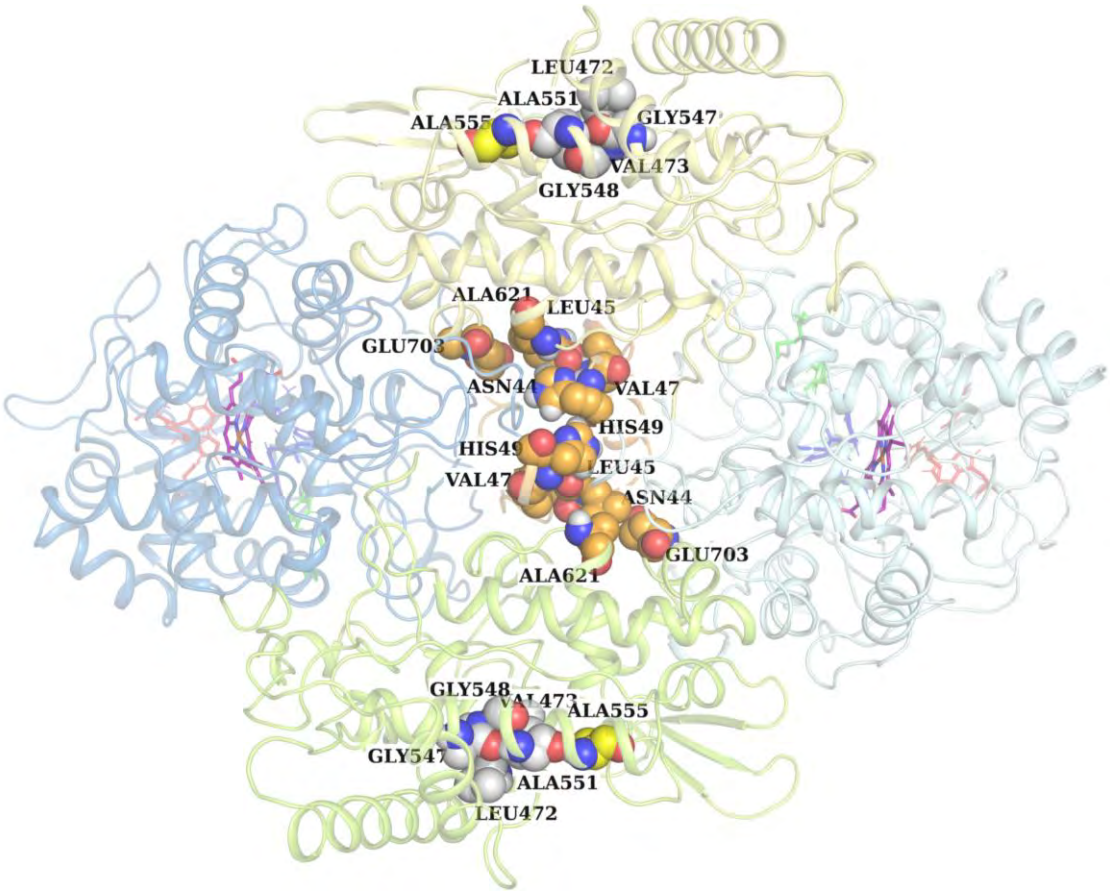

Figure 2. 16: Cartoon representation of KatG protein showing the concatenated distribution of *persistent hubs* across five DRN metrics, *BC*, *CC*, *DC*, *EC* and *KC*. The *persistent hubs* were localized at the dimerization and C-terminal domains. Adapted from Barozi et al., **[115]**.

2.4.2.4 Contact Map Analysis of Mutation Spots and Persistent Hubs

The local changes in atomistic interactions at the mutation positions and *persistent hubs* were investigated through contact map analysis [264]. From **Figure 2.17**, each mutated residue was compared to the WT and the color scale from white to dark red indicates degree of contact frequency. The change of residue side chains and size as a result of the SNPs caused both losses and gains in residue interactions at the different mutation positions. Furthermore, due to the asymmetry of the protomer dynamics, differences in positional contacts were noted between protomers.

In protomer A, the S140R mutation resulted in the reduction in contact frequency between Ser140 and His276 in contrast to the WT (**Figure 2.17**). Other mutation positions with reduced interaction included **G279D** with Leu283, Tyr304 and Ile313; **G285D** with Ser303 and **G316D** with Asp316, Pro232, His276, Ile313 and Ala348. In G593D, there was a complete loss of interaction with Arg595, Asn596, Ala606, Glu607, and Val628 coupled with reduced interaction with Val694 compared to the WT. Gains in interaction at these positions were also observed i.e., at position 140, arginine gained H-bond interactions with Lys143, Trp300, and Ser315 whereas asparagine gained H-bonds with Lys143, Trp300, and Ala131. Similarly, in G285D, there were slight compensatory gains with Ala291 and Gln295 which resulted in less residue flexibility in the mutant compared to the WT. G316D showed compensatory gains with Gly279 and Pro280 whereas G593D gained contact with Glu588 and Asn602. Residue position 315 harbours the most prevalent KatG mutations and here S315I and S315N exhibited loss of H-bond interaction with Thr275 and His276 whereas S315R and S315T lost interaction with Ala139 compared to the WT

In protomer B, similar trends of loss and gain in residue interactions were observed across all the mutant systems (**Figure 2.17)** highlighting the mutational effect on residue interactions.

Of most interest were the gains in interactions at position 315 i.e., Thr275, His276, Gly277, Gly349, and Ala350 for the S315T mutation. Previous studies have indicated that this mutation narrows the active site access channel preventing accessibility by INH [122, 348, 349]. Congruent to these observations, increased interaction at this site could imply a more compact access channel which compromises access to the binding pocket.

Similarly, the *persistent hubs* displayed an assortment of gains in residue interactions with the neighboring interface residues at the dimerization domain, explaining the observed centrality as per DRN calculations (**Table S2 and S3**). Taken together, contact map analysis supplements the previous observations from DRN analysis that showed changes in the protein interaction networks consequent to mutations.

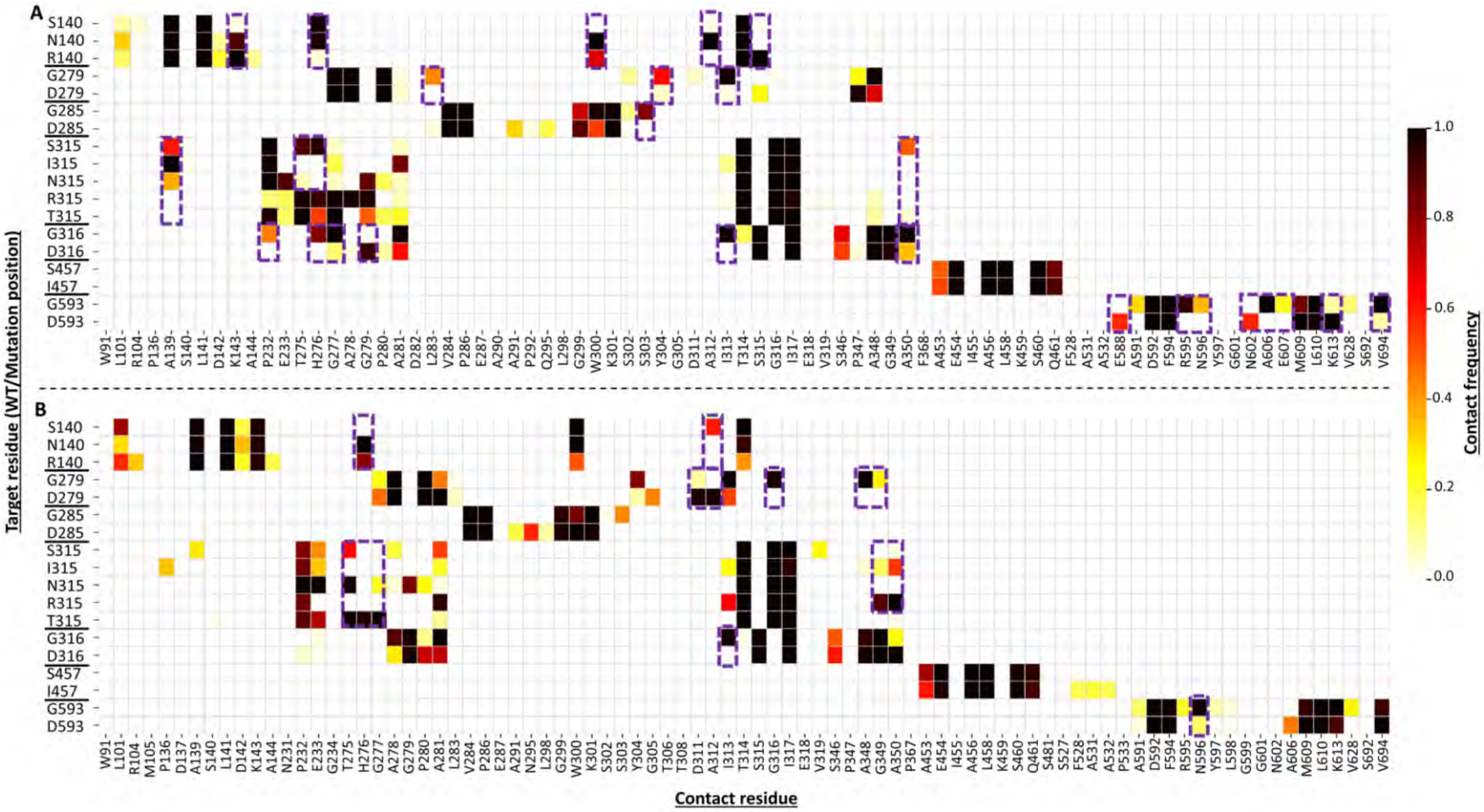

Figure 2. 17: Heat map showing the changes in the residue interaction frequencies at the mutation position for protomer A (A) and B (B). The color scale from yellow to dark red indicates the degree of interaction between the residues. The purple boxes indicate areas of gain/loss or reduction residue contact in the mutants compared the WT. The x and y-axes show the contact residues and mutated residues, respectively. Adapted from Barozi et al., **[115]**.

2.4.2.5 Alanine Scanning and System Binding energy

In protein interactions, binding energy is one of the key measures of optimal protein-protein interactions (PPI) [350]. Using alanine scanning, we determined the energy contributions of the key residues in the dimerization domain. Here, each interface residue was individually mutated to alanine and the change in binding energy calculated to quantify its stabilizing or destabilizing ability [351]. Using the ROBETTA web server [352], and the low energy structures from ED, we identified the destabilizing and stabilizing residues, **Table S4**. A cut-off of ≥ 1 kcal/mol was used to identify the destabilizing residues, whereas residues with binding energy < -0.08 kcal/mol were defined as stabilizing [331]. Interestingly, dimerization domain residues, Val30, Asn35, Gln36, Asn44, Lys46, His49, Glu192, and Glu195, previously identified as *BC* and *CC* hub (**section 2.4.2.3**) were identified as destabilizing residues here. Most importantly, no mutations were listed at positions Leu43, Asn44, Leu45 and Lys46 in the TBDReaMDB [314], drugs resistance gene associated database (DRAGdb), [288] and genome-wide Mycobacterium tuberculosis variation (GMTV) [353] databases. Such protein regions, invulnerable to mutations are termed as 'mutation cold spots' [354] and they portray highly functional regions of the protein. Evidently, the dimerization domain presents as one of the most important regions in the KatG not only for facilitating dimer formation but also as a region of rapid information diffusion in the protein network. The dimerization region therefore is an area of research interest especially in regard to new drug design and protease (peroxidase) functionality.

2.5 Chapter Conclusion

This chapter explores the mechanisms of resistance as well as the global and local structural changes associated with the HC KatG mutations. Here, the HC mutations were retrieved from literature and TBDReaMDB [314] before modeling the mutant protein structures and subjecting them to 300 ns MD simulations.

Trajectory analysis through RMSD, RMSF, Rg and comparative ED revealed increased protein dynamics for the mutant systems compared to the WT especially for the INH binding pocket residues. Additionally, asymmetric protomer behaviour was observed particularly in the mutant systems compared to the WT. These findings were in agreement with the research by Chaplin and colleagues [71] who demonstrated a disordered heme biding environment consequent to KatG mutations. Characterization of heme behaviour presented 1) asymmetric heme interactions in the mutant protomers and 2) a significant reduction and loss of interaction between the heme group and mutant residues especially for residues His270 and His276. His270 are essential for heme coordination and stabilization. The compromise of heme interactions in the mutants is indicative of a disordered binding pocket environment consequent to HC mutations.

DRN analysis established the protease dimerization domain as important and influential based on the residue *BC* and *CC* distribution. Furthermore, *BC* analysis highlighted the changes in network pattern consequent to mutations. The *BC* changes in the mutants included loss of hubs in contrast to the WT accompanied by compensatory gains of mutant unique hubs especially around the binding pocket area. *CC* calculations established both in the interface and dimerization domain as key information dissemination regions in the protein. Furthermore, we established a correlation between the distribution of *CC* hubs at the interface and the inter-protomer distance. Systems S315N and S457I with the least number of *CC* hubs

at the interface were associated with slightly increased inter-protomer distance compared to the WT. The *EC* metric 1) emphasized the inter-protomer asymmetry of network patterns of both the WT and mutant systems and 2) highlighted the importance of the C-terminal domain in the protease (catalase and peroxidase) function. The highly influential residues as per the global top 4% cut-off were concentrated in the C-terminal domain which has been implicated in offering architectural support to the active site in the N-terminal domain [295, 297].

Taken together, the findings of this research bring to light both the global and local changes in KatG behavior in the presence of resistance conferring mutations and go further identify the key regions in both the WT and mutant protein function through network analysis.

CHAPTER 3

3. *In silico* Characterization of the Effects of *Mycobacterium tuberculosis* Pyrazinamidase Mutations on Nicotinamide and Pyrazinamide Interactions and Network Patterns

This chapter focuses on characterizing drug resistance mutations in the *Mtb* pyrazinamidase enzyme (MtPncA) through analysis of mutation effects on the MtPncA-Nicotinamide complex in comparison to the MtPncA-Pyrazinamide complex. The comparison between MtPncA-Nicotinamide and the previously published work on MtPncA-Pyrazinamide [135] is aimed at identifying key drug interaction differences that can be honed in future drug design.

Contribution: Mutation identification, molecular docking and generation of 3D structure of wildtype and mutant was done by Thommas M. Musyoka and Rita Afriyie Boateng. Homology modeling, MD simulations were done by Rita Afriyie Boateng. Trajectory analysis and all data analysis including script writing was done by Victor Barozi. The Python scripts used for DRN to x% calculation and heat maps were generated by Olivier Sheik Amamuddy. Hub visualization scripts in PyMOL were generated by Olivier Sheik Amamuddy, Victor Barozi and Özlem Tastan Bishop

3.1 Introduction

Subsequent to the identification and isolation of *Mtb*, the TB etiologic agent, by Robert Koch in 1882 [355], various effective anti-tubercular compounds have since been identified. Among the most effective and corner stone therapies for TB management is pyrazinamide (PZA) [356–358]. PZA is an analogue to nicotinamide (NAM) which is natural substrate of the *Mtb* pyrazinamidase/nicotinamidase enzyme (MtPncA).

3.1.2 Analogous Nicotinamide and Pyrazinamide

NAM, which is sometimes known as niacin, is a substrate for MtPncA enzyme and precursor for nicotinamide adenine dinucleotide (NAD+) in *Mtb* [359]. NAD+ and its derivatives including NADH, NADP, and NADPH are important hydride acceptors and donors involved in a number of essential processes like calcium regulation [360], numerous cell redox reactions [361] and also act as substrates for various process in the cell like DNA repair, adenine diphosphate (ADP) acetylation and protein deacetylation [361]. Undoubtably, the efficient regulation of NAD+ in these organisms is paramount. In *Mtb*, NAD+ can be synthesized either under the *de novo* or the *salvage* pathways. Under the Preiss-Handler *salvage* pathway [359, 361], *Mtb* synthesizes NAD+ from NAM salvaged from the host or as biproduct of NAD+ biosynthesis. Here, NAM is converted to nicotinic acid (NA), under normal cellular MtPncA activity, and further acted upon by nicotinate phosphoribosyl transferases, PncB1 and PncB2, to yield NAD+ [361, 362]. Interestingly, NAM also has anti-tubercular activity and has previous been used as monotherapy for *Mtb* before the discovery and use of the a more effective structural analogue, PZA [363].

The analogous NAM and PZA only have one structural difference involving and extra nitrogen at the PZA aromatic ring. Previous studies have shown that difference in the aromatic ring influences the PZA activity and binding of to MtPncA compared to NAM,

where the change in the amide moiety and electrostatic potential due to the extra nitrogen makes PZA binding less energetically favorable compared to NAM [364]. This was further emphasized by Sun & Zhang [365] who showed that the MtPncA enzyme has an approximately 94-fold higher activity for NAM than PZA.

Structurally, PZA consists of a pyrimidine ring connected to an amide group. In its natural form, PZA is inactive against *Mtb* and requires activation by the MtPncA enzyme to pyrazinoic acid (POA^-). The entry of PZA into the bacterial cells is thought to be both through passive diffusion and active transport. Once in the *Mtb* cell, PZA is activated to POA^- which is then transported out of the cell through an inefficacious efflux pump. The extracellular POA^- is protonated to HPOA at acidic pH and reabsorbed back into the cell where it causes cytoplasmic acidification. The cellar damage and oxidative stress from cytoplasm acidification constitutes the antimicrobial activity of PZA [361, 366]. PZA is also used in synergy with INH, rifampicin and ethambutol in TB management, where is shortens TB treatment from nine to six months [367–369]. This first-line pro-drug is also active against the semi-dormant bacilli in latent TB [361, 370].

3.1.3 Structural Architecture of the MtPncA

MtPncA is encoded by the *pncA* gene and it assumes a typical Rossman fold structure consisting of six parallel β-strands interlaced by α-helices forming a single α/β domain [371] (**Figure 3.1**). The metalloenzyme contains Fe^{2+} as a cofactor which is important in the enzyme activity [372]. The metal binding site (MBS) is located besides the active site, where the Fe^{2+} is coordinated by Asp49, His51, His57 and His71 [371]. The MtPncA crystal structure [221, 313] (PBD ID: 3PL1) also shows that Fe^{2+} is held in position by two water molecules, HOH220 and HOH221.

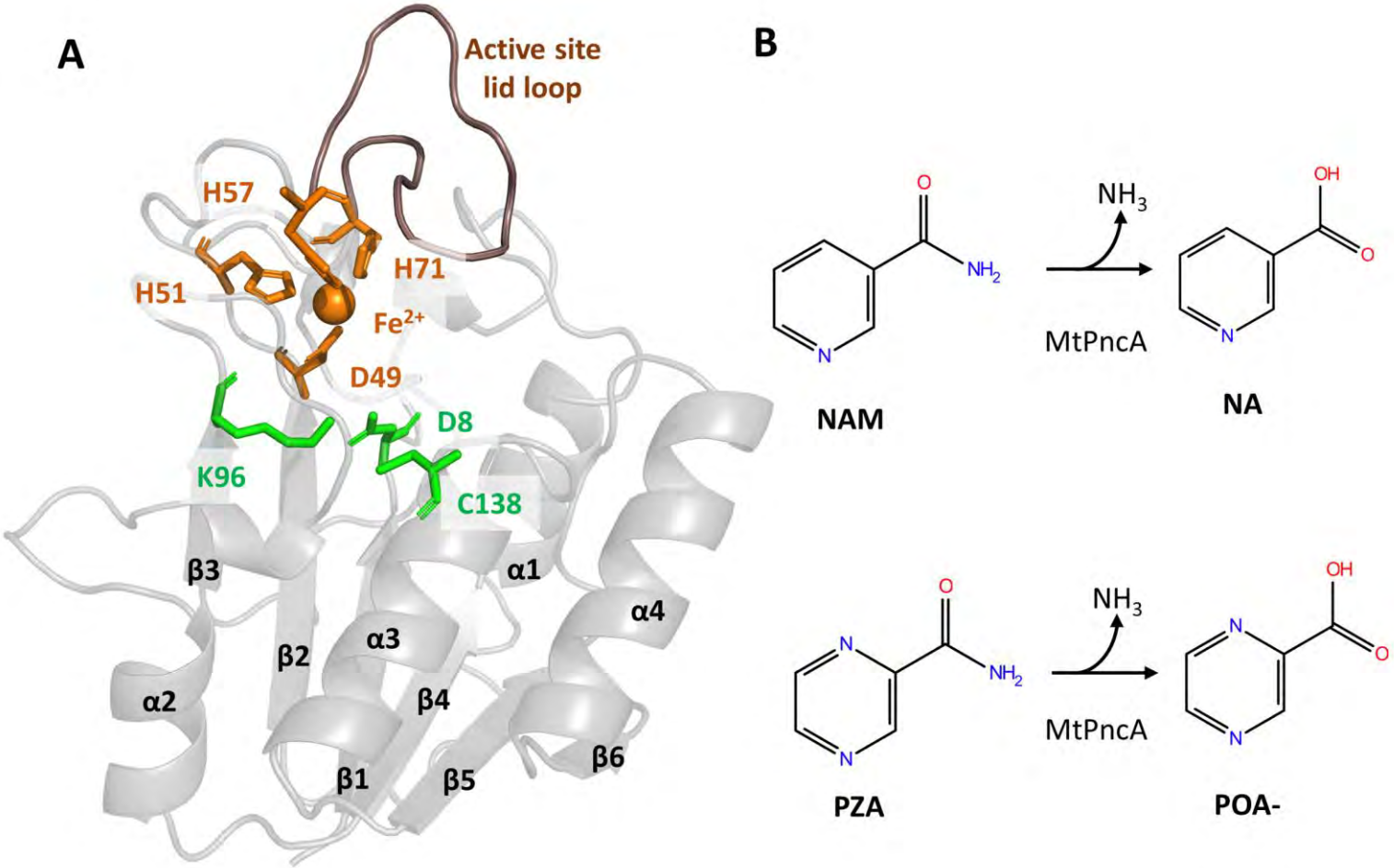

Figure 3. 1: A is the cartoon representation of MtPncA showing the Fe^{2+} (orange sphere) coordinating residues (orange sticks) and the active site residues (green sticks). The active site lid loop region is colored in brown. The Rossman fold α-helix and β-strands are annotated. B shows the structural differences between NAM and PZA and their corresponding active compounds post activation by MtPncA.

The MtPncA main active site residue, Cys138, is opposite the MBS at the N-terminus end of the α3-helix, where it forms H-bonds with PZA [371, 373] and is involved in the substrate hydrolysis [374]. The other active site residues making up the MtPncA catalytic triad include Asp8 and Lys96. In the active site, Ile133 forms a cis peptide bond with Ala134 which in turn reorients Ala134 in place to form an oxyanion hole with Cys138. Upon PZA binding, its carbonyl oxygen atom is placed in the oxyanion hole where the thiolate of Cys138 interacts with the carbonyl atom from PZA to form ammonia and an acyl-enzyme intermediate. The acyl-enzyme intermediate is later hydrolyzed by a water molecule to form POA^- [371].

Furthermore, MBS residues, His57 and His71 also act as anchor points for the active site lid loop extending from His57 to His71. Structurally, this region is more displaced in the *Mtb* MtPncA compared to the PncA in *Pyrococcus horikoshii* and *Acinetobacter baumanii* [112]. The other difference between MtPncA and PZAse of *P. horikoshii* and *A. baumanii* is the insertion of five amino acids *viz*: Val, Asp, Glu, Asn and Gly between residues G108 and T114 [371]. *P. horikoshii* and *A. baumanii* are also 36% and 37% identical to the MtPncA.

The functional importance of the MtPncA especially in cofactor biosynthesis makes it a great drug target for antitubercular therapy. Since MtPncA also activates PZA, resistance SNPs in the protein gene renders the enzyme ineffective in PZA activation [302, 365, 375, 376].

3.1.4 MtPncA Mutations and PZA Resistance

Besides the synthesis of NAD+ cofactor in *Mtb*, MtPncA also fortuitously activates PZA, a TB first-line prodrug, which is active against latent TB (LTB). Consequently, one of the ways *Mtb* escapes the bactericidal activity of PZA is through drug resistance mutations in the MtPncA gene. The mutations impede activation of the prodrug, rendering it inactive against the *Mtb*. Additionally, *Mtb* can afford mutations in MtPncA due to the presence of the *de novo* pathway for NAD+ cofactor synthesis.

Mutations in the *pncA* gene change the MtPncA-PZA interaction profile and enzyme behaviour leading to an inactive MtPncA. There are a number of MtPncA-PZA drug resistance SNPs with the majority of them between residue positons 3 to 17, 61 to 85 and 132 to 142 [134, 301, 377]. Over time, a large body of research has focused on deciphering the PZA drug resistance mutations in the MtPncA enzyme. A study by Khan et al focused on the L19R, R140H, and E144K MtPncA mutations where they observed a tendency of PZA to form less H-bonds and a general reduction in the drug binding ability [134]. Similarly, Nangraj et al used computational techniques to show that the MtPncA mutations, D49N, Y64S, W68G, and F94A affect the orientation of PZA in the active site and that the mutations also affect Fe^{2+} coordination [378]. Seemingly, majority of the MtPncA mutations affect the active site and the MBS environment dynamics which in turn influences drug and ion binding [112, 130, 131, 134, 379–382].

Another body of research shows that mutations in other genes like the *rpsA* gene, that encodes for the ribosomal protein S1 (rpsA), also affect the activity of PZA in *Mtb* [113, 369, 383, 384]. The rpsA enzyme is vital in protein translation and ribosomal sparing trans-translation processes [385]. This suggests that the activity of POA^- is not just limited to the MtPncA enzyme but may have other target proteins. In one of our previous studies, we focused on the conserved MtPncA motions associated PZA unbinding in the 82 MtPncA mutants [135]. We observed that ligand unbinding was associated with conserved protein motions which consisted of antiparallel lid and side flap movements; and contractions of a flanking region towards the protein core [135]. This suggested that there could be common dynamics in the mutant systems leading up to the ligand release.

Interestingly, NAM has previously been used as a substitute for PZA in most susceptibility studies and this is because of the acidic pH requirements for PZA action [386–388]. Despite the structural similarities between PZA and NAM, previous studies have shown that MtPncA

hydrolyzes NAM more efficiently than PZA [364, 365] and that proteins interact and bind different ligands with varying affinities [389, 390]. Based on that, in this chapter we sought to further explore the differences in the NAM and PZA interactions and dynamic profiles in the MtPncA especially in presence of the 82 previously studied PZA resistance conferring mutations [135]. This was in a bid to contribute to the existing knowledge on PZA resistance and to inform future drug design.

3.2 Chapter Aims and Objectives

The work in this chapter is continuation of the 82 previously investigated PZA resistance mutations by the research group [135], and is aimed at deciphering the differences between NAM and PZA interactions within the wildtype MtPncA and the 82 PZA resistance mutants [135]. Identifying the key differences between NAM and PZA interaction will inform drug design efforts and also shed light on effects of ligand structure on ligand-protein interactions. To achieve this, i) 82 previously modeled MtPncA-NAM complex mutants were subjected to 150 ns molecular dynamic simulations, ii) trajectories analyzed in comparison to the same ensemble of MtPncA-PZA complex mutants and iii) systems analyzed using DRN analysis.

3.3 Methods and Materials

3.3.1 Docking of NAM to MtPncA and Modeling of MtPncA Mutants

The WT MtPncA structure (PBD ID: 3PL1) was used as the receptor for docking of NAM retrieved from the PubChem database [391] (compound ID: 936). Prior to docking, all the water molecules were removed from the MtPncA structure except for the two water molecules, HOH220 and HOH221, that are involved in Fe^{2+} coordination. The protein structure was then protonated at pH 6.7 using the H++ web server [392] and the NAM structure minimized. AutoDock [393] tools (version 1.5.6) were used to add gasteiger charges and non-polar hydrogens to the protein (3PL1) and NAM structure to generate *pdbqt* files. Targeted docking was done by centering the docking cubic box of 40x40x40 grid points about the active site (coordinates 9.88, −26.6, 0.35) with a grid space of 0.375 Å. A 100 conformational searches were done in AutoDock4.2 using the Lamarckian genetic algorithm (LGA) [394]. The best ligand pose was selected based on the MtPncA-NAM interaction energy as calculated by the AutoDock semi-empirical algorithm and literature.

The established MtPncA-NAM complex was used as a template for modeling 82 MtPncA-NAM mutant systems using MODELLER [242] version 9.18. The 82 select mutations were previously identified as high confidence mutations by genomic and PZA susceptibility studies (**Table S5**), and investigated in relation to PZA [135]. In MODELLER, the slow refinement method was used to generate 100 models for each MtPncA-NAM complex and the models with the lowest z-Dope score [246] were selected as the best. Ligand orientation and interaction was analyzed in Discovery Studio (DS) [395]. Furthermore, mutant stability was accessed using the mutation Cut-off Scanning Matrix (mSCM) which uses graph-based signatures to predict protein stability [396].

3.3.2 All-atom Molecular Dynamic Simulations

The *in-vivo* WT MtPncA-NAM complex behavior and that of the 82 mutants was simulated using MD simulations. Due to the presence of the iron group in the MtPncA structure, Fe2+ parameters previously generated in [135] were used and protein system topologies generated using Leap modeling of the AMBER ff14SB force field [397]. The GROMACS [315] compatible topology files were generated using the ACPYPE (AnteChamber Python Parser interface) tool [398]. The systems were then placed in a cubic box while maintaining a distance of 1 nm between the box boundary and the solute. Water molecules were added to the cubic box to mimic the *in-vivo* environment using the Transferable Intermolecular Potential with 3 Points (TIP3P) water model [257]. With the systems solvated, the overall charge was neutralized using NaCl ions of 0.15 M concentration prior to minimization. Minimization was performed to remove any steric clashes using no constraints and the steepest descent algorithm until an energy threshold of 1000.0 kJ/mol/nm. The systems were then equilibrated for both temperature and pressure using the NVT and NPT ensembles, respectively. Like in **chapter 2**, temperature was equilibrated at 300 K for 100 ps using the Berendsen thermostat with a time step of 2 fs. Similarly, the pressure was equilibrated for 100 ps at 1 atm using a 2 fs time step and the Parrinello-Rahman barostat [318]. All-atom MD simulations were then run for 150 ns for the WT and mutant MtPncA-NAM complexes at Centre for High Performance Computing (CHPC). For both the equilibration and MD runs, all bonds were constrained using LINCS algorithm [319] and, a cut-off distance of 1.0 nm was used for short-range Coulomb and van der Waals electrostatics. Furthermore, PME electrostatics [320] with a Fourier spacing of 0.16 nm were used for long-distance electrostatics. Trajectories from MD simulations were analyzed for RMSD, RMSF, and Rg using the GROMACS *gmx rms*, *gmx rmsf* and *gmx gyrate* tools, respectively. The trajectory analysis data was presented and plotted using the same Python tools as mentioned in **chapter**

2 section 4.3.3. The MD runs were restricted to 150 ns due to the limitation on the available computational resources. This analysis entailed MD runs of 83 protein systems surmounting to 12450 ns and a HPC time of 2822 hours.

3.3.3 H-bond and Contact Map Analysis

The propensity of NAM to form H-bonds in the WT and mutant systems was investigated through H-bond analysis. Here, the GROMACS *gmx hbond* tool which analyses all possible donors and acceptors to check for existence of H-bonds was used. The frequency of H-bonds formed over the 150 ns was presented as a heat map to identify the NAM retaining and ejecting systems. The propensity of NAM to form other interactions including van der Waals interactions within a Euclidian distance of 6.7 Å [268] was also determined using the *contact_map.py* Python script from MDM-TASK-web server [264, 265]. The NAM and PZA (data from [135]) contact frequency was compared and presented as heat maps.

3.3.4 Dynamic Residue Network Analysis

Like in the previous chapter, DRN analysis was done using the *calc_network.py* Python script (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) from the MDM-TASK web server [263, 264]. Five DRN metrics were used to describe the network patterns in both the WT and mutant MtPncA-NAM systems i.e., averaged *BC*, *CC*, *DC*, *EC* and *KC*. DRN analysis was only computed for the NAM retaining systems. To that effect, DRN analysis was computed for the equilibrated last 50 ns of the simulation. Furthermore, high centrality residues, global top 5%, of the protein ensemble were determined for each DRN metric. The high centrality residues together with the homologous residues from all the other protein systems were presented as heat maps using seaborn [321].

3.4 Results and Discussion

3.4.1 NAM Structural Contacts in MtPncA Compared to PZA

From the 100 NAM conformational searches, the best MtPncA-NAM complex was selected based on the lowest binding energy, -4.12 kcal/mol. Additional structural validation was done using DOPE score, ProSA, and VERIFY3D all of which showed a good quality structure with z-DOPE score of -1.69, ProSA z-score of -7.28, and VERIFY3D score of 98.82%.

To gain a general insight into the NAM interaction profile in relation to PZA [135], the ligand molecular interactions were analyzed in DS. Here, NAM formed four H-bonds with Asp8, Ile133, Ala134 and Cys138; eight van der Waals interactions with Val7, Lue19, Asp49, His57, Trp68, His71, Lys96 and His137; one Pi-Pi interaction with Phe13 and a water bond with HOH220 (**Figure 3.2A**). In total, NAM formed fourteen interactions in MtPncA which is one more interaction (van der Waals interaction with His57) compared to PZA in [135] (**Figure 3.2B**).

Furthermore, the mutant MtPncA-NAM complexes were modeled, and they consisted of mutations distributed across the whole MtPncA protein structure with ~48% of them involving similar phytochemical residue substitutions especially for the catalytic triad residues (**Figure 3.2C**).

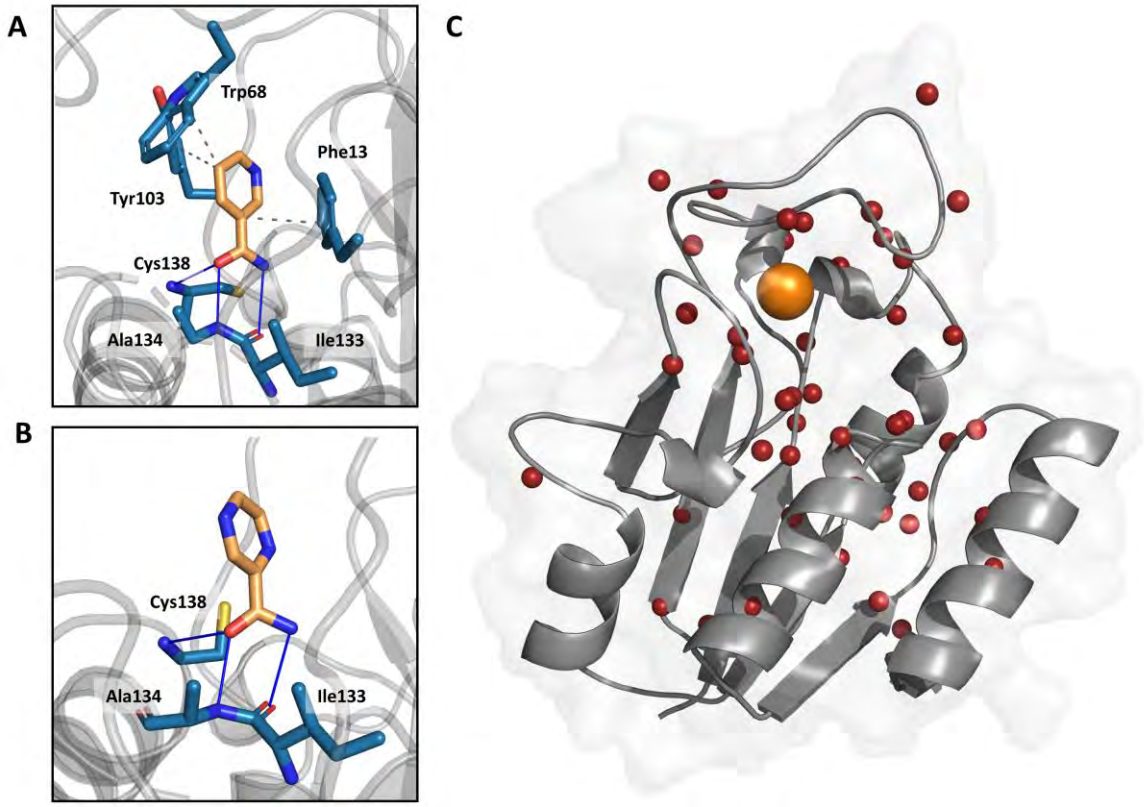

Figure 3. 2: A and B show the docked NAM and PZA H-bonds (blue) and hydrophobic interactions (dotted grey) in the MtPncA. C is the cartoon representation of the MtPncA structure highlighting, in firebrick spheres, the positions of the 82 high confidence mutations. The position of the Fe^{2+} group is shown as an orange sphere.

3.4.2 Global MtPncA-NAM Dynamics in comparison to MtPncA-PZA

The global changes in the MtPncA-NAM and MtPncA-PZA dynamics consequent to the HC mutations were analyzed using RMSD, RMSF and Rg. For the rest of this chapter, the MtPncA-NAM structural dynamics and interaction results are discussed in comparison to the MtPncA-PZA data from [135].

3.4.2.1 Mutation Effects on the Global MtPncA Stability

Complex stability was assessed through the RMSD calculations where the MtPncA-NAM ensemble of proteins (WT and mutant systems) showed early equilibration at ~ 50 ns (**Figure S11**). All the systems showed steady RMSD progression over the simulation period except for G17D, A46V, D49G, D49A, D49G, H51Y, W68R, W68G, C72R, L85R and K96N which

had a pronounced root mean deviation from the initial structure compared to the WT. Furthermore, the RMSD range for the MtPncA-NAM ensemble was between 0 and 0.344 nm signifying less global structural variation even in the presence of mutations. Of these systems L35R, S66P, S67P, L85P, G97S, A102V, V139G, A146V and V155G had lower RMSDs compared to the WT as per the median RMSD. A statistical comparison of the mean WT MtPncA-NAM RMSD to that of the mutants using the Mann-Whitney U-test showed no significant differences between the WT and all mutants systems (p-value > 0.05). As mentioned in chapter 2 section 2.4.1.2, given the small sample size used in this test (1 datapoint per sample) the results were not interpreted in insolation but rather contributed to the general analysis. Most of the MtPncA-NAM mutant systems experienced a unimodal RMSD distribution like in the WT except for A3P, L19R, L19P, D49G, H51P, P54T, W68R, C72R, H82R, T114P, G132D, V139G, Q141P and V155G which had a bi-modal distribution (**Figure 3.3A**). The bi-modal distribution in these systems implies that at least two main structural conformations were sampled throughout the simulation. Furthermore, mutant stability prediction using the mCSM web server [396] implicated, A3E, V7G, V9A, V9G, Y34S, V45G, W68G, W68L, G78D, G97D, W119R, V128G, V130G, G132D, V139G, V155A, V155G and A171E as highly destabilizing; D12A, H51Y, H57Y, Y103S and Q141P as stabilising and the rest as destabilizing (**Table S6**).

In the MtPncA-PZA ensemble, RMSD violin plot representation showed the WT system as the most stable system with the least RMSD as per median RMSD values (**Figure 3.3B**). In addition, the overall RMSD range, 0 to 0.422 nm, was higher than that observed in the MtPncA-NAM ensemble. Here, A3E, V9A, G17D, L19P, V21G, Y34S, H57P, T61P, S66P, W68R, G132D, T142P and V155G explored bi-modal RMSD distribution compared to the unimodal distribution in the WT MtPncA.

Rg calculation further highlighted that the MtPncA-NAM systems experienced slightly less degree of gyration (maximum ensemble Rg: 1.619 nm) compared to the MtPncA-PZA ensemble (maximum ensemble Rg: 1.643 nm).

3.4.2.2 MtPncA Residue Dynamics as Described by RMSF Calculations

RMSF analysis informed on the structural dynamics and characteristics of particular MtPncA regions. The delta RMSF between the MtPncA-NAM and MtPncA-PZA ensembles was determined to access the differences in complex dynamics, by subtracting the MtPncA-NAM RMSF values from the MtPncA-PZA values per system. This analysis showed nominal differences between the WT MtPncA-NAM and WT MtPncA-PZA residue flexibility (**Figure 3.3C**).

In the MtPncA-NAM systems, majority of the mutant systems experienced marginally higher residue flexibility in the loop regions, 16-22, 33-39, 58-69, 162-167 and 183-185, compared to the WT. Naturally, protein loop regions are more flexible compared to the β-strands or α-helices [399].

Focusing on the MtPncA-NAM mutations around the active site region, His51 of H51Y lost the H-bond with Cys72 and also had a 32-fold reduction of the His51-Asp49 H-bond compared to the WT. In W68R, there was a reduction in the number of H-bonds formed by Arg68 from 4 in the WT to 3 in W68R. Trp68 forms part of the active site lid loop where is sits overhead the active site cleft and forms H-bonds with His57 [371]. The noted reduction in number of H-bonds around the active site loop region could explain the observed loop flexibility. Furthermore, W68R mutation has an effect on protein function based on the residue conservation (positioning and interaction) in *Pyrococcus horikoshii* (PhPncA), *Acinetobacter baumanii* (AbPncA) and MtPncA. In C72R, residue contact map analysis highlighted a 6-fold reduction in H-bonds formed between Arg72 and Gln10 compared to the

WT MtPncA-NAM. It is important to note that Cys72 is at the fringe of the active site lip loop where it interacts one of the loop-anchoring residues, His71. A reduction in Cys72 interactions explains the observed active site loop flexibility compared to the WT MtPncA-NAM system. The other mutants with marked flexibility of the lid loop included, L85R, K96N and L159R, which are located distal to the lid loop however, mCSM analysis implicated them as destabilizing mutations (**Table S6**). Destabilizing residues have the potential to disrupt the protein integrity, structure, dynamics which in turn affect protein stability, folding and functioning.

Compared to the MtPncA-PZA ensemble, higher residue flexibility was noted for the active site lid loop region of H51Y, W68R, C72R, L85R, K96N, L159N and L159R in the MtPncA-NAM ensemble. On the other hand, only D49A, D49G and H71R of the MtPncA-PZA ensemble had more active site lid loop flexibility compared to the MtPncA-NAM ensemble.

This analysis highlights the destabilizing effect of MtPncA mutations, marked by increased active site lid loop flexibility, especially for the SNPs in proximity to the active site. Even though the RMSF analysis was on a global scale, the marked flexibility of the lid loop region was more evident in the MtPncA-NAM mutants compared to the MtPncA-PZA mutants signifying differences in binding between the two compounds.

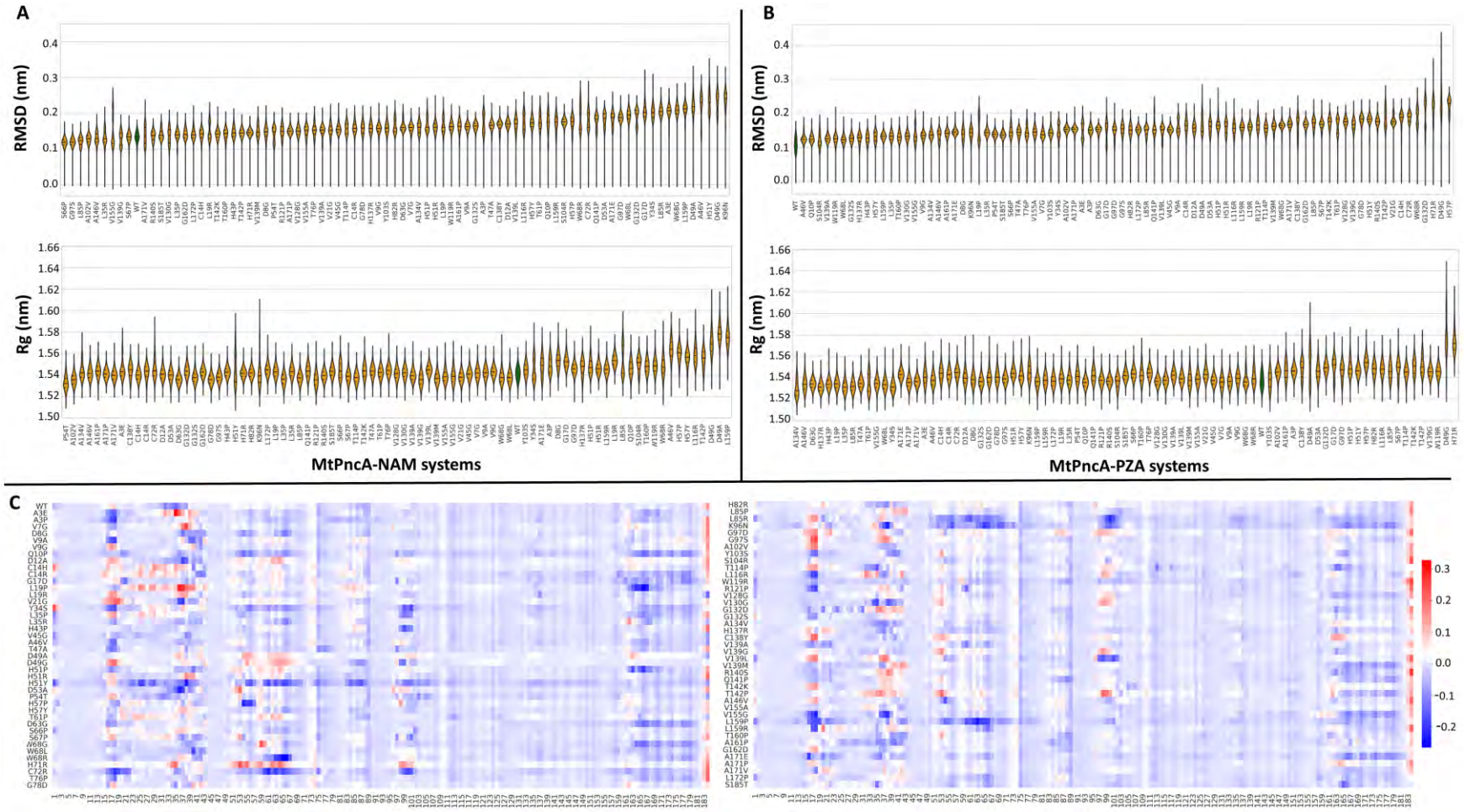

Figure 3. 3: A and B show the RMSD and Rg violin plots for the MtPncA-NAM and MtPncA-PZA systems, respectively. In both subplots, the WT is shown in green and the mutant systems in orange. C is a heat map presentation of the delta RMSF between the MtPncA-NAM and MtPncA-PZA systems (MtPncA-PZA - MtPncA-NAM). Blue shows more residue flexibility in the MtPncA-NAM systems, whereas red shows more flexibility in MtPncA-PZA systems.

3.4.3 Mutation Effect on NAM Interactions

The effects of the marked active site lid loop flexibility in the MtPncA mutants on NAM coordination were investigated here through visualization, H-bond analysis and contact map analysis. Visualization of the trajectories in VMD highlighted three groups of MtPncA-NAM complexes based on the NAM behavior *viz*: NAM stable complexes (34), NAM unstable complexes (31) and NAM disassociated complexes (17), **Table S7**. In comparison to the MtPncA-PZA system, ref [135], NAM was stable in more mutant systems (34) than PZA (21). The stability of NAM in the mutant systems was further confirmed through NAM RMSD calculations which showed high RMSD values in the unstable and NAM disassociating systems compared to the WT (**Figure S12**). In the NAM unstable systems, the ligand had a bimodal RMSD distribution especially in V7G, D8G, C14R, G17D, Y34S, V45G, D49A, L85R, K96N, C138Y, T142K and A171E compared to the NAM stable systems (**Figure 3.4A**).

Like in the our previous work [135], ligand H-bond analysis was performed to uncover the changes in ligand interactions due to the mutations. Here, MtPncA-NAM systems, V9A (average H-bond over 150 ns: 0.101), Q10P (0.632), T47A (0.432), D49G (0.228), H51Y (0.129), P54T (0.524), S66P (0.932), S67P (0.092), W68R (0.556), C72R (0.217), G97D (0.087), L116R (0.157), W119R (1.180), R121P (0.143), V139G (0.909), V155G (0.332) and L159P (0.171) progressively formed fewer H-bonds with NAM than the WT-MtPncA (2.320) which resulted in NAM ejection from the enzyme (**Figure 3.4B**). Of these systems, V9A, L35R, D49G, W68R, L116R, and L159P also portrayed PZA disassociation from the MtPncA, where fewer H-bonds were formed compared to the WT coupled with conserved motions involving the opening of the binding cavity lid loop leading to PZA ejection [135].

In addition to H-bond analysis, NAM contact map analysis was performed to quantify the ligand contact profile in the WT and mutant MtPncA systems. Here, all the critical intermolecular interactions, including van der Waals interactions between the NAM and MtPncA were measured and ranked on a scale of 0 to 1, where 1 is a perfectly maintained contact throughout the simulation and 0 is no contact at all. **Figure 3.4C** is a heat map representation of NAM contact frequency, where NAM maintained contact with nine primary residues, including, Asp8, Phe13, His57, Trp68, His71, Ile133, Ala134, His137, and Cys138 in both the WT and the majority of the mutant MtPncA-NAM systems. As expected, the systems with NAM disassociation showed significantly reduced intermolecular interactions between the NAM and the MtPncA. Interestingly, a couple of new NAM interactions were identified in the MtPncA mutant systems compared to the WT, and these included Ala102 and Tyr103 in A3P, G17D, W68L, G97S, S104R, T114P, A134V, and A161P. These additional interactions were attributed to the NAM motions in the ligand unstable systems.

Previous studies has shown that PZA resistance-conferring mutations disrupt the active site environment, affecting drug binding [112, 131, 301, 379–381]. Likewise, NAM in the V7G system had reduced contact frequency with the main active site residue Cys138 (contact average over 150 ns: 0.242), whereas in D8G and C14R, NAM interaction with active site residues Ile133 (D8G: 0.027, C14R: 0.0) and Ala134 (D8G: 0.0, C14R: 0.025) was significantly reduced compared to the WT-MtPncA (Cys138: 0.42, Ile133: 0.39, Ala134: 0.287) (**Figure 3.4C**). Other systems with reduced NAM interactions at the active site compared to the WT (Phe13: 0.369, Trp68: 0.421, Ile133: 0.309, Ala134: 0.287, Cys138: 0.462) included, G17D for residues Trp68 (0.05), Ile133 (0.0) and Ala134 (0.0) ; Y34S and V45G for residues Trp68 (Y34S: 0.024, V45G: 0.129), Ile133 (Y34S: 0.0, V45G: 0.025), Ala134 (Y34S: 0.025, V45G: 0.0) and Cys138 (Y34S: 0.076, V45G: 0.0); L85R and K96N for residues Phe13 (L85R: 0.098, K96N: 0.0), Ile133 (L85R: 0.0, K96N: 0.143), Ala134

(L85R: 0.0, K96N: 0.091) and Cys138 (L85R: 0.0, K96N: 0.039); T142K for Trp68 (0.0) and V155G at residues Trp68 (0.0) and Cys138 (0.0).

Based on the global structural stability, it is evident that the PZA resistance-conferring mutations affect the enzyme dynamics resulting in a more unstable MtPncA-NAM complex than the WT. This effect seems to be more prevalent in the MtPncA-PZA complexes compared to the MtPncA-NAM systems based on the number of MtPncA systems stably coordinating NAM (34) to PZA (21) and the NAM/PZA interaction profile as discussed in the next section. Consequently, these effects are translated to the enzyme active site, where drug interactions are compromised, resulting in drug disassociation in some mutant systems.

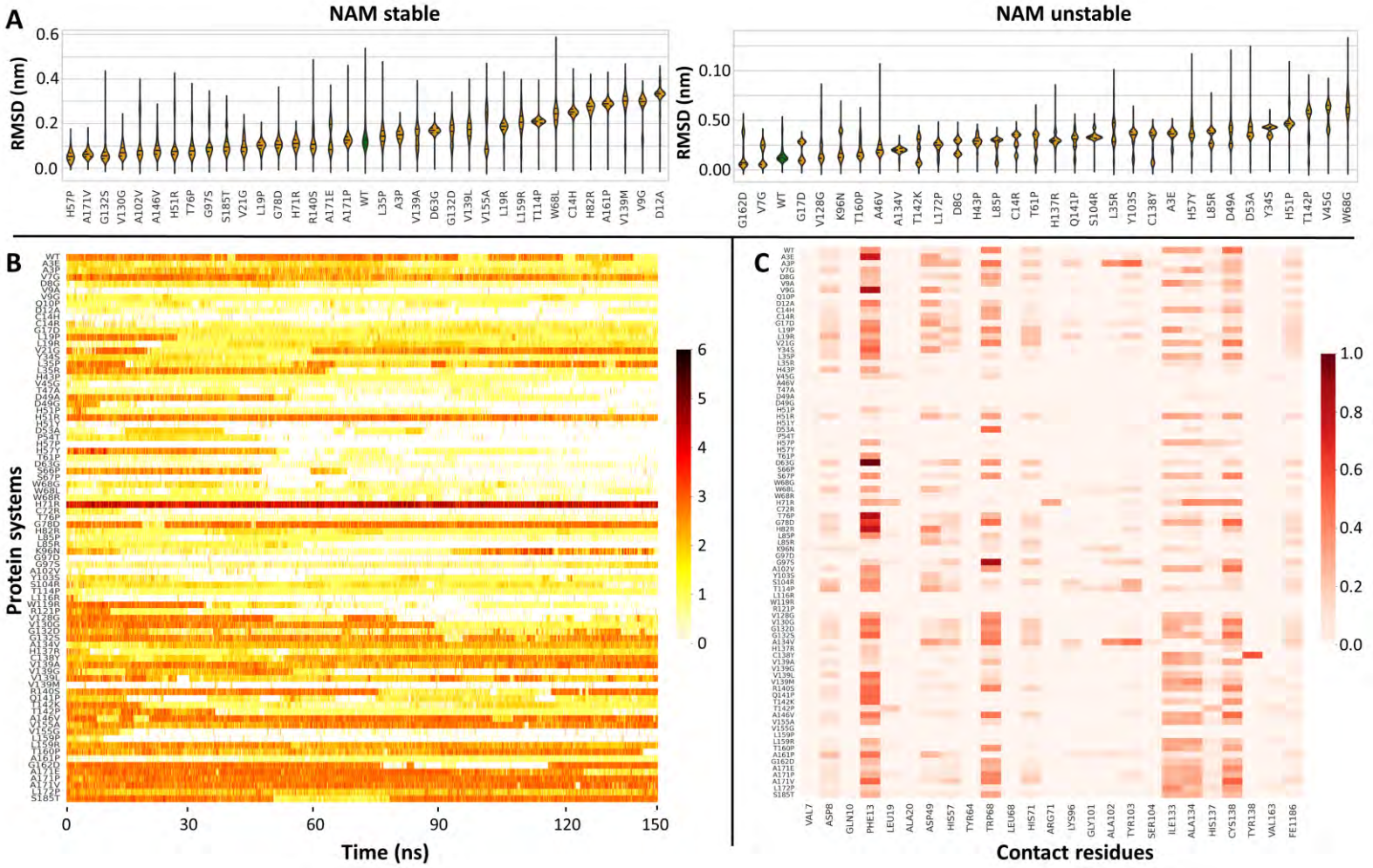

Figure 3. 4: A shows both the MtPncA-NAM stable and unstable RMSD violin plots with the WT in green and mutant systems in orange. In B, MtPncA-NAM H-bond frequency is shown as a heat map over the 150 ns. C shows the NAM interaction frequency ranked from 0-1 with dark red indicating a high interaction frequency and light brown showing no interaction.

3.4.4 A Comparison of the NAM and PZA Interaction Profile in MtPncA

Given the structural similarity between NAM and PZA, their intermolecular interactions in MtPncA WT and mutant systems were compared to identify 1) the critical ligand contacts and 2) the differences between MtPncA-NAM and MtPncA-PZA interaction characteristics in a bid to identify favorable binding properties. We calculated the delta between the PZA, and NAM residue contact frequencies in MtPncA over the simulation time (PZA residue interaction frequency - NAM residue interaction frequency). For each ensemble, the ligand contacts were calculated using the *contact_map.py* tool from MDM-TASK web. The results were presented as a heat map, **Figure 3.5A**, where blue implies higher ligand contact frequency for NAM compared to PZA and red the vice versa.

Interestingly, no major differences in ligand contact frequencies were observed between NAM and PZA in the WT systems. Similarly, there were trivial contact differences between the NAM and PZA in the mutant MtPncA systems that steadily coordinated both ligands *viz*: G132D, V139A, R140S, A171V and S185T with the exception of D63G (**Figure 3.5A**). In D63G, NAM had marked contact frequency with Phe13 (17.01) compared to PZA (4.36). Phe13 forms Pi-Pi stacked interactions with NAM in the active pocket. NAM also had a higher contact frequency with Trp68 than PZA (NAM: 0.369, PZA: 0.229) in majority of the systems. Besides these residues, PZA generally showed slightly higher interactions with Tyr103 and His137 compared to NAM. Across the mutant systems, the ligand (NAM and/or PZA) predominantly maintained interaction with residues, Asp8, Phe13, Asp49, Trp68, Ile133, Ala134 and Cys138. Consequently, mutations at any of these main ligand contacts resulted in either unstable ligand coordination or disassociation.

It is also important to note that for the main active site residues, Ile133, Ala134 and Cys138, majority of the MtPncA-NAM mutant systems had higher ligand contact frequency with

these residues as compared to the MtPncA-PZA systems (**Figure 3.5A**). In addition, more contacts were noted in between the low energy WT MtPncA and NAM (7) compared to PZA (3) for the (**Figure 3.4B and 3.4C**)

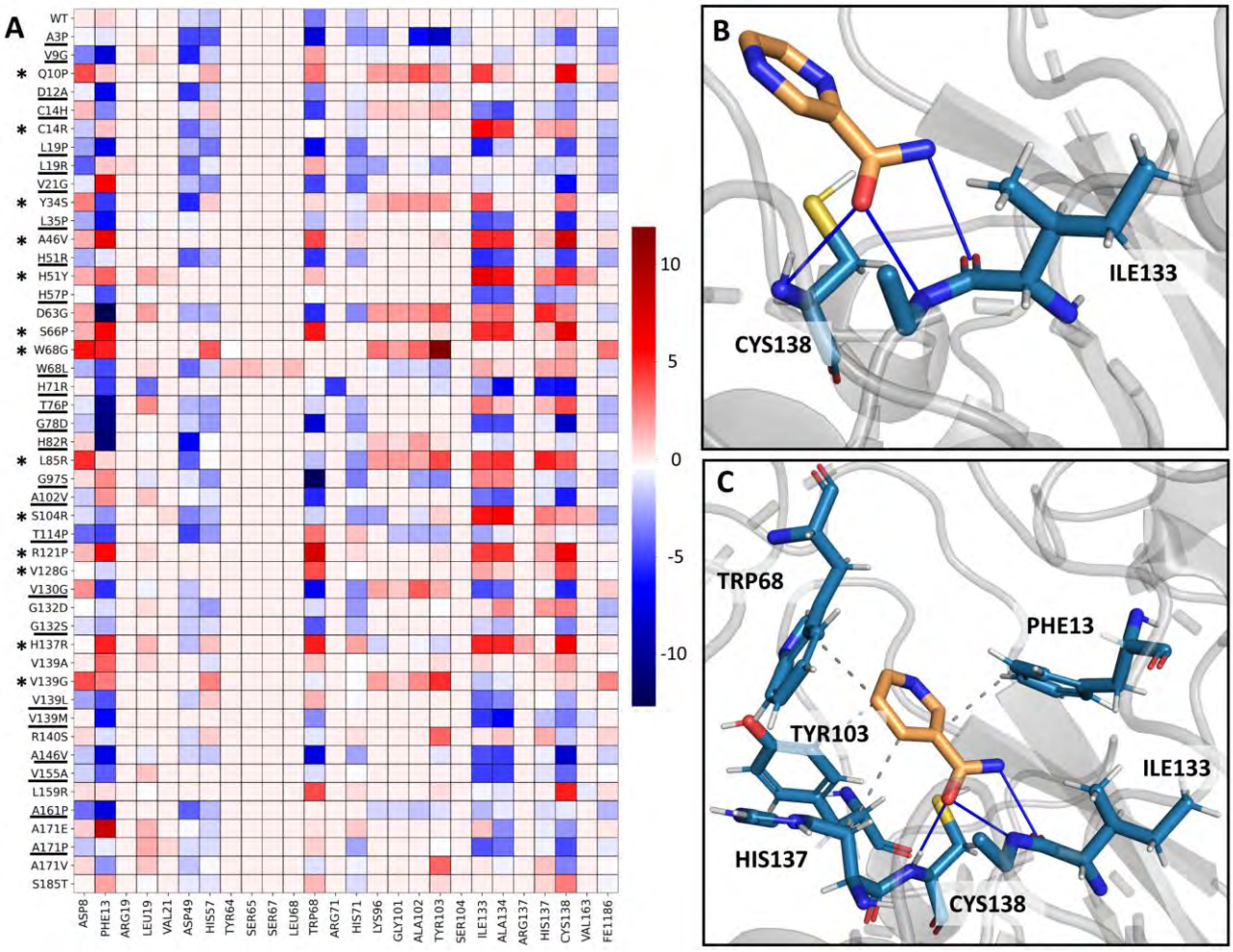

Figure 3. 5: A shows differences in the interaction frequency between PZA and NAM (PZA-NAM) in the WT and mutant systems. The mutant systems that retained NAM and not PZA are underlined whereas those that retained PZA and not NAM are marked with an Asterix. Blue shows more interaction frequency for NAM while red shows more interaction frequency for PZA. B and C show the PZA and NAM interactions, respectively in the minima WT structures from MD simulations.

Even though there were minor differences in MtPncA contact frequencies between NAM and PZA, the high ligand retention for the MtPncA-NAM systems compared to PZA is an indication of a more favourable NAM interaction compared to PZA probably due to the lack of the extra nitrogen atom on the PZA aromatic ring.

3.4.5 Dynamic Residue Network Analysis of the NAM Retaining Systems

Averaged DRN analysis was computed for the equilibrated last 50 ns of the NAM retaining systems to ascertain the changes in residue network patterns consequent to the drug resistance mutations. Like in the previous chapter the analysis involved description of protein network patterns using the averaged *BC*, *CC*, *DC* and *EC* metrics. Here, a cut-off of top 5% was used to establish the high centrality residues/hubs across the NAM retaining protein ensemble.

3.4.5.1 Betweenness centrality

From a global perspective, particular regions of the MtPncA were more involved in signal transduction along the shortest path in the WT and mutant systems. These regions included residues at positions, 5-10, 13, 19, 21, 31, 47, 49, 71, 72, and 94 (**Figure S13)**. The residues are concentrated at the protein core within β1, β 2, β3-strands, and α-helix 1. They also include the catalytic triad residue, Asp8 and Fe^{2+} coordinating residues, Asp49, and His71.

Gln10 in H57P had the highest *BC* value across the MtPncA ensemble, with a centrality value of 0.108. A mutagenesis study by Junaid et al. showed that the Q10P mutation interrupts the communication between the catalytic triad residue, affecting enzyme action [400]. Furthermore, in the WT, Gln10 forms H-bonds with catalytic triad residue Asp8. The marked centrality at Gln10 highlights its importance as a catalytic residue, especially in the presence of mutations. Pro83 and Ser179, located in the loop regions, had the least *BC* values of 0 across all the systems.

The global top 5% centrality analysis identified Val7, Asp8, Val9, and Gln10 as *persistent hubs* across all the NAM retaining systems (**Figure 3.6**). Besides the *persistent hubs*, Ile5, Val139, and Thr142 also had hub status in at least 45 of the 66 NAM retaining systems. Ile5 is located at the MtPncA β1 strand and is involved in interactions linking the β-strands i.e., it forms H-bonds with Val45 of β2, and Val30 and Val128 of the β4-strand. Besides the H-

bonds, Ile5 also forms van der Waals interactions with Ile6, Val7, and Thr142 of α-helix 3. Val7 of the β1 strand forms covalent and alkyl interactions with catalytic triad residues Asp8 and Cys138, respectively. Ile5 also interacts with Thr47, Val130, and Ile133 from the different β-strands, stabilizing the protein core. The Val7 involvement in ligand interactions [135] and coordination of catalytic triad residues explains its marked *BC* values.

Additionally, the catalytic triad residue, Asp8, which facilitates the formation of a nucleophilic thiolate by Cys138 during the activation of PZA to form ammonia and an acyl-enzyme intermediate [371, 401] was also identified as a *persistent hub*. The residue forms a diverse number of interactions with Val7, Val9, Gln10, Thr47, Lys48, Asp49, and Ile133 at the active site which justifies the observed high centrality values. Other *persistent hubs*; Val9, Gln10, Val139, and Thr142. Val9 and Gln10 are located at the loop region around the substrate binding site. Val9 and Gln10 form part of the loop region between β1 and α1, where Val9 forms a hydrogen bond with His82, covalent bonds with Gn10 and Asp8, alkyl interaction with Ile133 and van der Waals interaction with Gly132. Gln10 shares H-bonds with Asp8, Phe13, Cys14, Asp49, and Ala79, van der Waals interaction with Asn11, Asp12, Lys48, Cys72, Asp80, Phe81 and His82 and has covalent bonds with Asn11 and Val9 as per contact maps. Furthermore, Val139 and Thr142 form part of the substrate binding pocket and interact with catalytic residue Cys138. *BC persistent hubs* formed a network of interactions linking the MtPncA Rossman-fold core (β1-α1-β2) to the active site.

Overall, the high *BC* residues were concentrated at the protein core (**Figure S14**) characterized by the β-α-β interaction which is critical for the reaction with adenosine diphosphate (ADP) portion of dinucleotides. The marked centrality of these residues across all the NAM retaining systems, irrespective of the presence of mutations, highlights their importance in enzyme function.

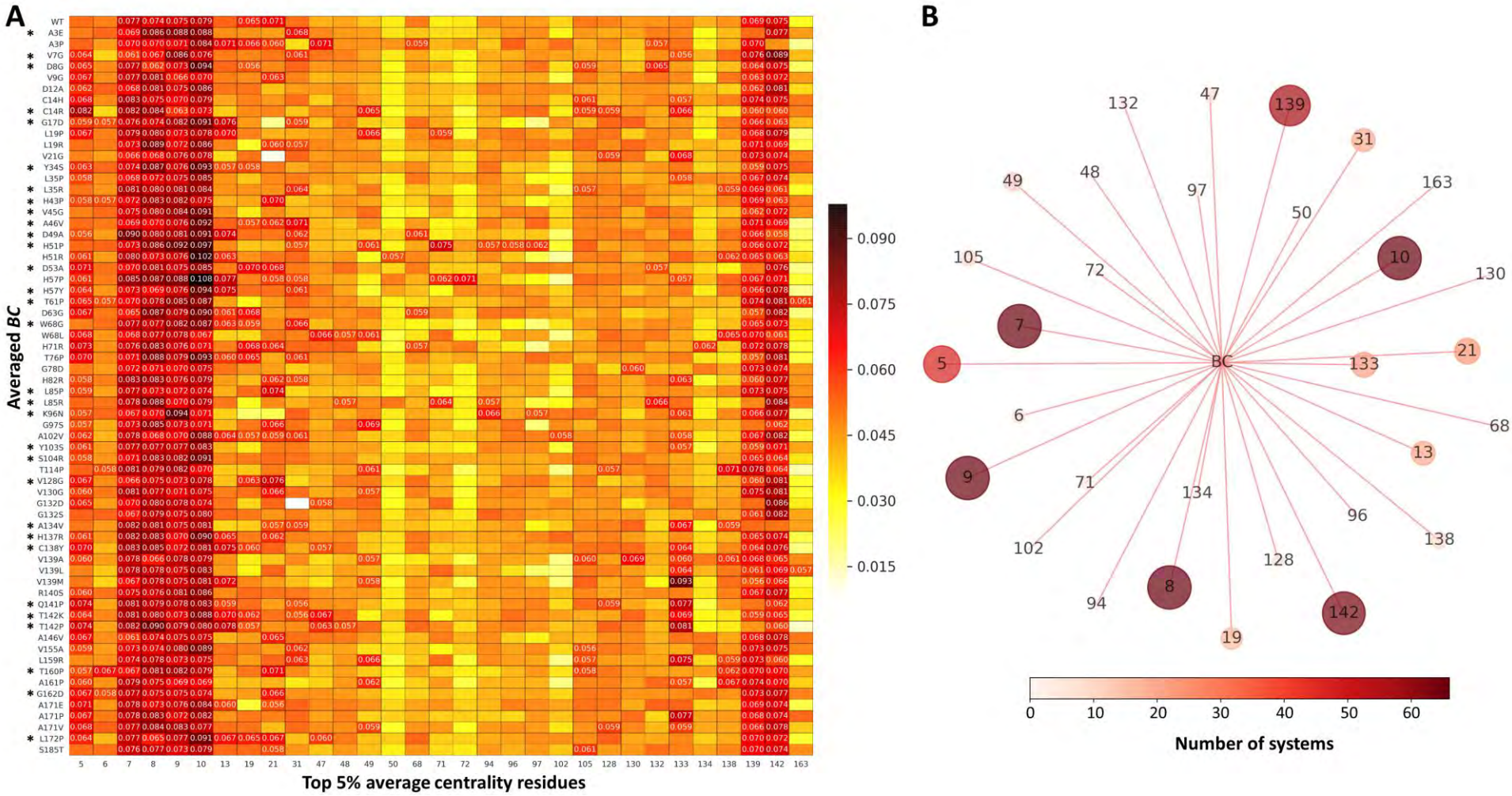

Figure 3. 6: A shows the heat maps of the global top 5% *BC* hubs (annotated with centrality values) for the WT and NAM retaining mutant systems. Systems that unstably coordinated NAM are marked with an Asterix. B is Spring plot showing the frequency of *BC* hub residues across the systems. Larger and dark red nodes correspond to more frequent hubs while small and white nodes to less frequent hubs.

3.4.5.2 CC Distribution in the MtPncA WT and Mutant Systems

As noted from the chapter 2, residues with high *CC* values are commonly positioned at the protein core or protein interface regions, close to all the other residues in the network. The data from the averaged *CC* calculation showed a uniform distribution of centrality values across all MtPncA systems (**Figure S15**), implying that the resistance mutations do not drastically change the enzyme network.

Furthermore, the high centrality residues across the WT and mutant systems were identified using a global top 5% cut-off (**Figure 3.7**). This analysis identified residues, Thr7, Asp8, and Gly132 as *persistent hubs*, whereas Val9, Thr47, Ile133, Cys138, Val139, and Thr142 were hubs in at least 44 of the 66 NAM retaining systems. Most of these residues were previously identified as hubs under the *BC* analysis. No significant differences were noted between the distribution of *CC* hubs in the WT and mutant MtPncA systems (**Figure S15**). Here, the *CC* hubs were clustered at the β-α-β enzyme core consisting of the catalytic triad and active site residues. Residue Thr7 of V139A had the overall highest *CC* value of 0.310. Similarly, Thr7 had high *CC* values in all the systems, making it the most central residue in the enzyme.

The correlation between the *BC* and *CC* hubs across the MtPncA systems emphasizes the importance of these catalytic and core residues in the protein communication network.

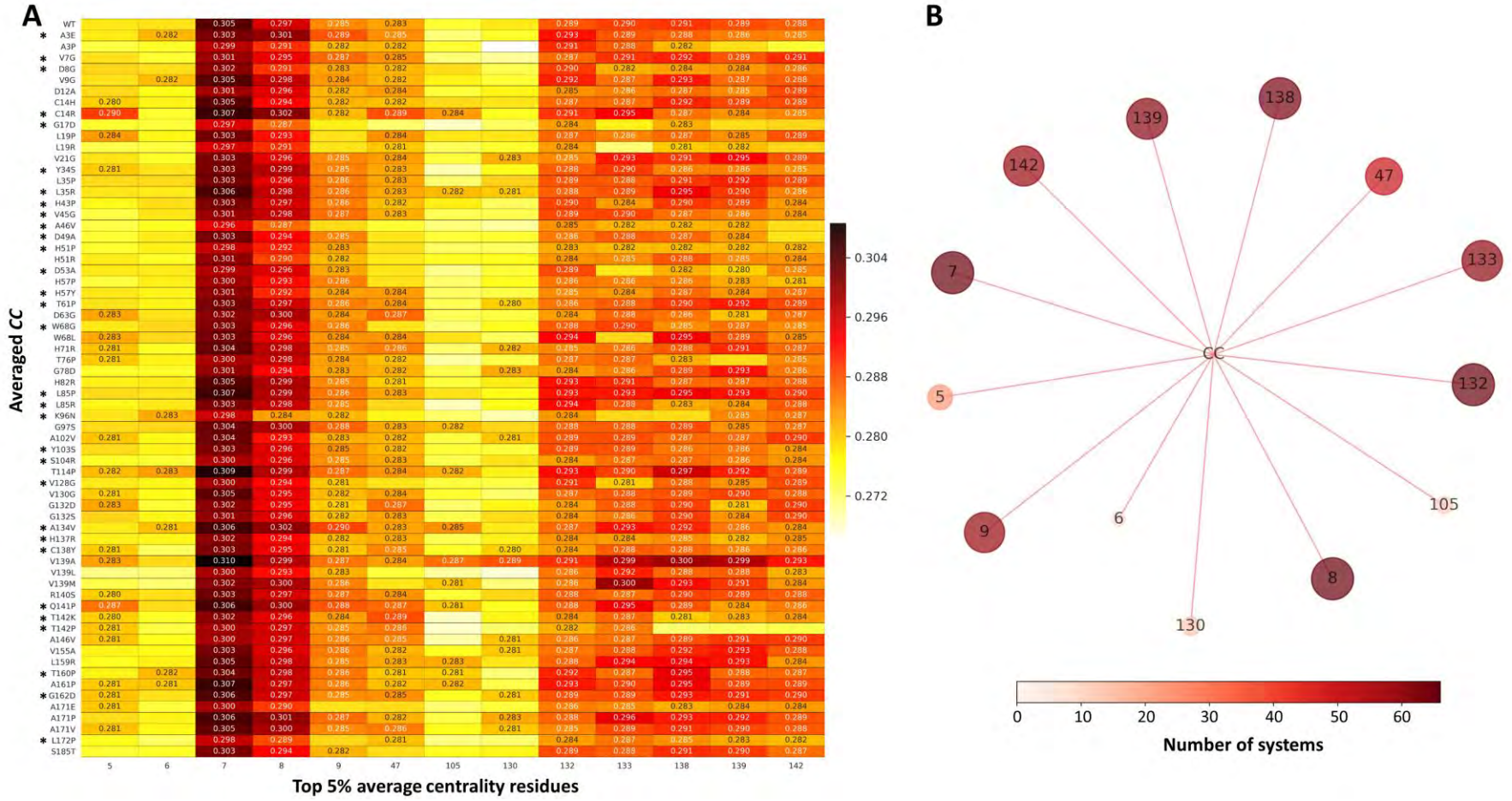

Figure 3. 7: A shows the heat maps of the global top 5% *CC* hubs (annotated with centrality values) for the WT and NAM retaining mutant systems. Systems that unstably coordinated NAM are marked with an Asterix. B is Spring plot showing the frequency of *CC* hub residues across the systems. Larger and dark red nodes correspond to more frequent hubs while small and white nodes to less frequent hubs.

3.4.5.3 Degree of Centrality and Eigenvector centrality

DC and *EC* are related since both metrics derive the node centrality from the neighboring nodes. The global top 5% did not identify any *DC persistent hubs* however, residues Val7, Val9, Gln10, ILe139, Thr142 and Ala143 had hub status in at least 45 out of 66 systems (**Figure S16**). The *BC*, *CC* and *DC* hub status of these residues emphasizes their importance in the enzyme functionality even in the presence of resistance conferring mutations.

Furthermore, the changes in residue communication networks around the enzyme active site consequent to SNPs were highlighted by *DC* analysis, particularly in C14H, G17G, K96N, T142P and T142K (**Figure 3.8**). Compared to the WT, the network of high centrality residues linking the enzyme core to the MBS was interrupted in the mutant systems. For instance, in C14H, residues Val9, Ala28 and Gly132 which connect the β-α-β core to the MBS were lost as *DC* hubs. Furthermore, F17D and K96N also lost Val9 and Gly132 as *DC* hubs. In T142P and T142K, we observed divergent effects of mutations at the same residue position. Here, the T142P mutation interrupted the network of residue hubs connecting enzyme core to the MBS, whereas the same *DC* hub connection was maintained in T142K (**Figure 3.8**).

The *EC* metric further identified the protein core region as highly influential in protein communication and function specifically residues Val7, Asp8, Val9, Gly132, Ile133, Val139 and Thr142 (**Figure S17**). Though not persistent, these residues were hubs in majority of the MtPncA systems studied here.

Collectively, DRN analysis identified the functionally important and influential residues linking the MtPncA protein core to the active site and the MBS. Furthermore, the analysis highlighted the changes in network of hubs around the active site and MBS in the presence of resistance mutations. These high centrality residues present a new perspective on MtPncA functionality.

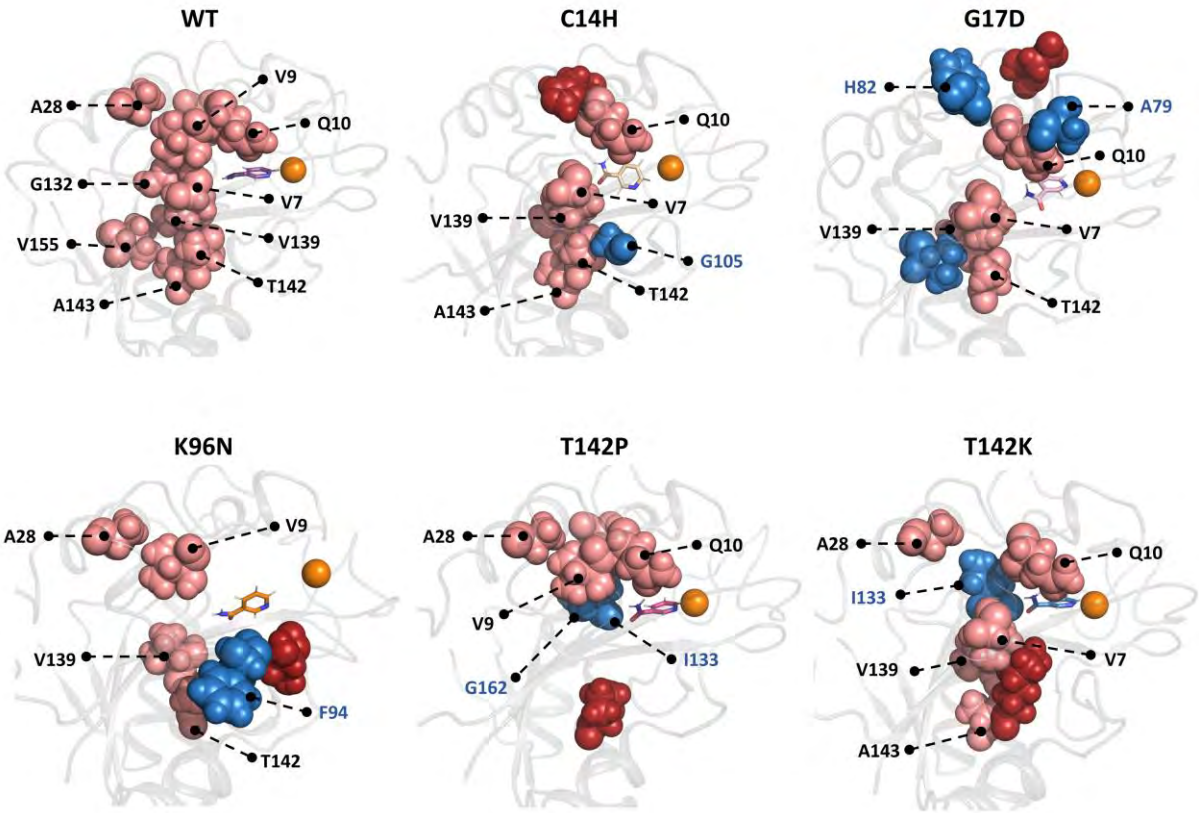

Figure 3. 8: Cartoon representation of the WT and mutant MtPncA systems showing the distribution of *DC* hubs. Hubs common to both WT and mutants are shown in salmon spheres whereas hubs unique to the mutants are shown as sky-blue spheres. Mutation positions are shown as firebrick spheres. NAM is shown in stick presentation.

3.4.6 A Comparison of NAM and PZA MtPncA Anchoring Residue Centrality

In the previous chapter we showed the versatility of DRN metrics at describing the behavior and characteristics of a protein networks especially when used in combination. Here, DRN analysis was used to highlight the nature of interactions of the main drug coordinating residues, *viz*: Asp8, Phe13, Asp49, Trp68, Ile133, Ala134 and Cys138 as identified from section 3.4.3. The objective was to identify the differences in centrality between the PZA and NAM coordinating residues in all systems irrespective of their ability to retain or eject the drug. To achieve this, the difference in NAM/PZA coordinating residue centrality for *BC*, *CC*, *DC* and *EC* between the PZA and NAM systems was obtained i.e., MtPncA-PZA - MtPncA-NAM, and the results were presented as a heat map (**Figure 3.9**). Based on the color

scale in the heat map, we identified which residues had more centrality between the two ensembles. From **Figure 3.9**, blue shows the residues with high centrality in the NAM system, whereas red shows high centrality in the PZA systems.

For the *BC* metric, nominal differences were observed between the MtPncA-NAM and MtPncA-PZA WT systems however, majority of the MtPncA-NAM systems displayed higher *BC* values compared to the MtPncA-PZA homologs especially for residues, Asp8, Phe13, Asp49, Trp68, Ile133 and Ala134. Previous studies involving *BC* and RMSF indicated that residues with high *BC* values undergo more stable conformational changes compared to those with low *BC* values [152]. High *BC* values for the NAM systems is indicative of the stability of NAM binding residues versus PZA binding residues and the effect it can have on drug coordination. Similar observations were noted for the *CC* metric where the NAM coordinating systems had high centrality values compared to PZA. In chapter 2, we showed the relationship between high *CC* values and the interaction distance between protein dimers/protein systems. High *CC* values for NAM coordinating residues could imply a more compact and stable binding cavity in the MtPncA-NAM systems compared to MtPncA-PZA. Of all the drug coordinating residues, Asp49 had the biggest difference between the two ensembles in favor of the MtPncA-NAM systems in favor of the MtPncA-NAM system (**Figure 3.9**). Asp49 is positioned in the substrate binding cavity of MtPncA and coordinated Fe^{2+} which is important in activation of PZA to pyrazinoic acid [371]. Furthermore, the *DC* and *EC* results were also in favour of the MtPncA-NAM systems with a number of drug coordinating residues having high centrality values in the NAM ensemble. Intriguingly, the main catalytic residue, Cys138, maintained high centrality values in the MtPncA-PZA systems compared to the MtPncA-NAM systems across all the DRN metrics. With the exception of Cys138, the high centrality values for drug coordinating residues suggests a more compact, stable and closely linked NAM environment in MtPncA compared to PZA.

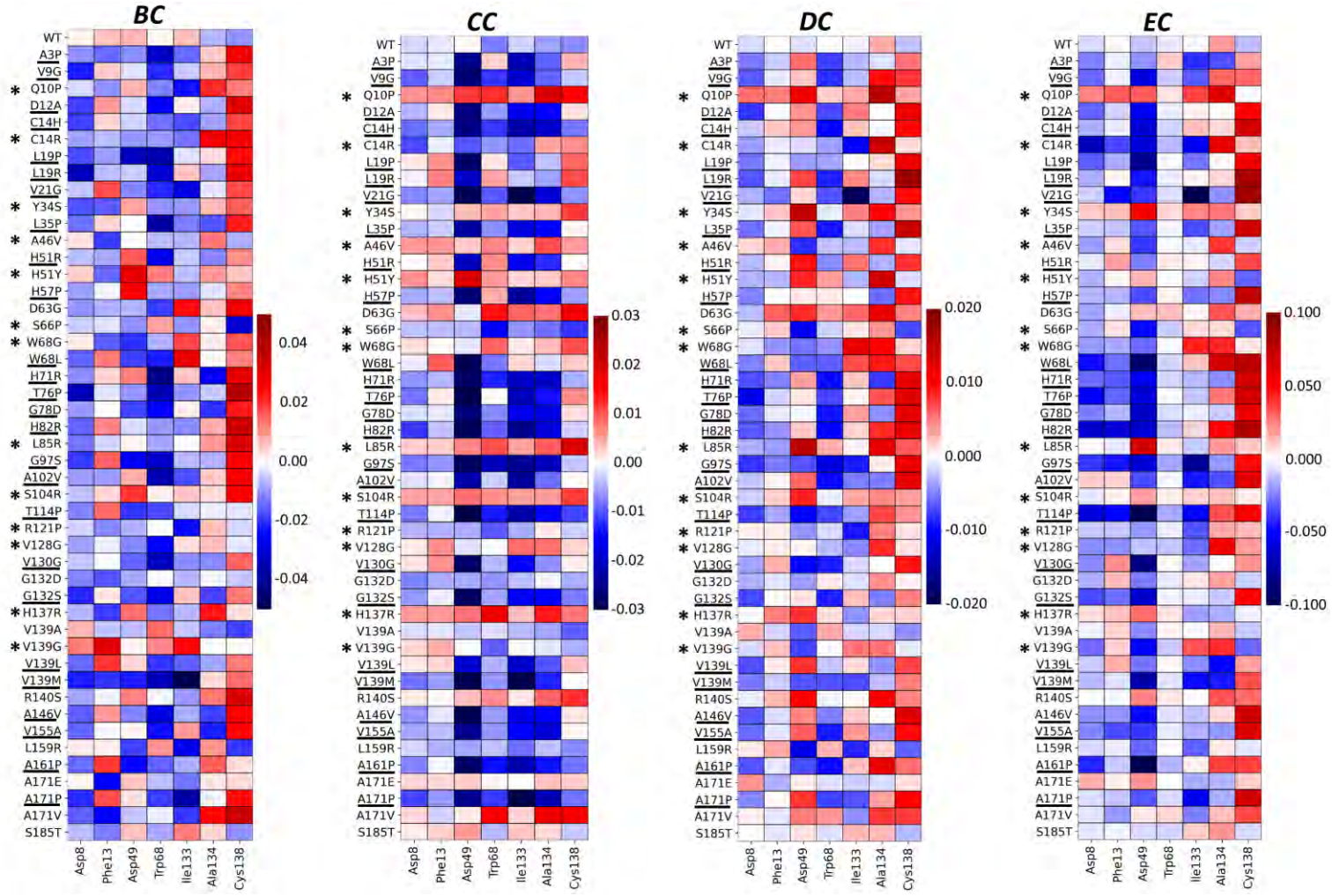

Figure 3. 9: Heat maps of the differences in DRN centrality values between the PZA and NAM coordinating residue. Systems that retained NAM and not PZA for 150 ns are underlined whereas those that retained PZA and not NAM are marked with an Asterix. High centrality values in MtPncA-NAM systems are shown as blue whereas those in the MtPncA-PZA systems as red.

3.5 Chapter Conclusion

This chapter sought to shed light on 1) the effect of PZA drug resistance mutations on the MtPncA-NAM complex dynamics and 2) establish the differences, if any, between the MtPncA-NAM and MtPncA-PZA interactions in both the WT and the drug resistance conferring mutations. This involved docking of NAM to the MtPncA protease structure, modeling of mutant structures in complex with NAM while using the MtPncA-NAM complex structure as a template and subjecting them to 150 ns MD simulations. Inferences on the MtPncA-NAM behavior were made in comparison to the MtPncA-PZA data from our previous work [135]. The work in this chapter was inspired by the ever-increasing antimicrobial drug resistance especially in TB first-line drugs.

From the global perspective, the stability of the MtPncA-NAM complex varied depending on the presence and location of the mutation. Here, the mutations at positions G17D, A46V, D49A, D49G, H51Y, W68R, W68G, C72R, L85R and K96N were associated with a more structurally unstable complex compared to the WT and other mutant systems. Interestingly, these mutations are in proximity to MtPncA metal binding site. RMSD analysis identified the structurally unstable mutant systems characterized by multi-modal RMSD distribution compared to the unimodal distribution in the WT. Furthermore, mutant stability analysis using the mCSM web server identified the same systems as highly destabilizing. As an ensemble of 82 systems, the MtPncA-NAM systems had an RMSD range of 0 - 0.390 Å whereas the MtPncA-PZA systems had an RMSD range of 0 - 0.422 Å. It is evident form **Figure 3.3** that the D49G mutant system was the one with the most effect on MtPncA stability in the MtPncA-PZA ensemble.

The active site lip loop region, previously identified as “opening” leading up to PZA release of in MtPncA mutants [135], had more flexibility in the MtPncA-NAM systems compared to

the MtPncA-PZA ensemble especially for H51Y, W68R, C72R, L82R, K96N, L159N and L159R systems. This was attributed to the reduction in residue contacts consequent to the mutations at these residue positions.

Generally, more stable protein-ligand complexes were observed for MtPncA-NAM complex ensemble (34) compared to the MtPncA-PZA ensemble (24). In the NAM stable systems, we observed strong and persistent NAM interactions with Asp8, Phe13, His57, Trp68, His71, Ile133, Ala134, His137 and the main catalytic residue Cys138. The MtPncA mutations, V9A, L35R, D49G, W68R, L116R and L159P resulted in ejection of both NAM and PZA. This was characterized by the progressive loss of H-bonds and non-bonded interactions.

DRN analysis identified the protein core as highly important and influential based on network analysis. The MtPncA core has previously been identified as critical for the protein interaction with ADP [360]. Furthermore, network analysis highlighted some of the mutational effects on protein network patterns characterized by the interruption of the *DC* hubs linking the protein core to the active site especially in C14H, F17D and T412P. The main drug coordinating residues, Asp8, Phe13, Asp49, Trp68, Ile133 and Ala134 generally displayed higher centrality values across *BC*, *CC* and *EC* in the NAM systems compared to PZA systems. The disparity in NAM and PZA coordinating residue centrality insinuates a more closely bound NAM environment leading to better binding compared to PZA, conceivably due to the structural differences between NAM and PZA.

The MtPncA main catalytic residue Cys138 maintained more contact with NAM than PZA in the mutant system however, the DRN centrality metrics, *BC*, *DC* and *EC* indicated that the active site residue had high centrality in more MtPncA-PZA mutant systems than the MtPncA-NAM systems except for *CC*. The high *CC* in the NAM retaining systems could imply the catalytic residue is in close proximity to other active site residues and hence a more

tightly bound active site in the MtPncA systems. Alternatively, the low *BC*, *DC* and *EC* values in the NAM systems could mean that Cys138 strong interaction with NAM influence changes in the protein communication patterns and information flow.

Collectively this chapter highlights the effects of PZA drug resistance mutations from the NAM perspective while pointing out the critical differences in drug coordination and networks, knowledge which could inform revised drug discovery.

PART II: EFFECTS OF NATURALLY OCCURING SPIKE RBD AND HUMAN ACE2 MISSENSE MUTATIONS ON PROTEIN-PROTEIN INTERACTIONS

CHAPTERS 4 AND 5

CHAPTER 4

4. *In silico* Analysis of the Influence of Progressive SARS-CoV-2 Omicron Sub-lineage Mutations on Spike RBD-hACE2 Interactions.

This chapter reports on the characterization of the SARS-CoV-2 S RBD interactions with the human receptor, hACE2 in presence of the Omicron sub-lineage mutations. The research focused on the Omicron sub-lineages, BA.1, BA.2, BA.3 and BA.4.

> **Contribution**: Mutation identification, homology modeling, MD simulations, trajectory analysis and all data analysis including script writing was done by Victor Barozi. The Python scripts used for DRN top x% calculation and heat maps were generated by Olivier Sheik Amamuddy. Hub visualization scripts in PyMOL were generated by Olivier Sheik Amamuddy, Victor Barozi and Özlem Tastan Bishop.

The work in chapter 4 is reproduced in part from the following publication

- **Barozi V**, Adrienne L. Edkins, Tastan Bishop Ö (2022). *Evolutionary progression of collective mutations of Omicron sub-lineages towards efficient RBD-hACE2: allosteric communications between and within viral and human proteins*. https://doi.org/10.1016/j.csbj.2022.08.015

Authors Contributions: Ö.T.B. conceived the project. V.B. performed the calculations and data analysis under the guidance of Ö.T.B. A.L.E contributed to writing part of the article.

4.1 Introduction

At the end of 2019, there was an outbreak of a severe acute respiratory syndrome known as coronavirus disease 2019 in Wuhan city of China [402–404]. The outbreak was later declared a global pandemic in March 2020 by the WHO [11]. Currently, February 2023, COVID-19 is responsible for ~700 million infections and ~6.8 million deaths worldwide [405]. Of these infections, the WHO regions; Europe, Americas, Western Pacific, South-East Asia, Eastern Mediterranean, and Africa account for approximately 36%, 25%, 26.5%, 8%, 3% and 1.3% of the cases, respectively [405] (February 2022).

Similar to severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS), COVID-19 mainly affects the lower respiratory system where it manifests with symptoms like fever, cough, fatigue, dyspnoea, shortness of breath, and sometimes sputum production [406–408]. More adverse effects are observed in patients with pre-existing comorbidities like hypertension, diabetes, and cardiovascular diseases, and these include, kidney failure, respiratory failure, cardiovascular damage, liver damage and arrhythmias [406, 409, 410]. Interestingly, the COVID-19 case fatality rate (CRF), which is the ratio of infection-associated deaths to the number of infections, is higher in the older and people with pre-existing comorbidities who end up admitted to the intensive care unit (ICU), compared to younger and healthier individuals [411–413]. This is attributed to the strong adaptive immunity in the younger individuals and the absence of predisposing conditions.

4.1.1 Etiology and Transmission of COVID-19

The infectious disease is caused by a severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) belonging to the *beta* genus of the *Coronaviridae* family [161, 162, 414]. Similar coronaviruses in the *Coronaviridae* family have in the past caused SARS and MERS

[415, 416]. Coronaviruses are typically enveloped, single-stranded and positive-sense RNA organisms mainly responsible for zoonotic infections in humans [417–419].

Human-to-human transmission of SARS-CoV-2 occurs through inhalation of SARS-CoV-2-infested aerosol droplets from and an infected person [420, 421] emitted through sneezing, coughing, and talking. Studies have shown that access of infected aerosol to any mucosal membranes, nose, mouth, or eyes of a susceptible individual can result in infection [421]. Other documented modes of SARS-CoV-2 transmission include fomites [422], airborne [423], tears [424, 425], faecal-oral [426] and saliva [427, 428]. From an epidemiological point of view, SARS-CoV-2 has an average R_0 value of 3 [429], which is the average number of susceptible people infected by sick person [420]; however, this is also dependent on social behavior and season.

Upon inhalation/contact with SARS-CoV-2 contaminated aerosols, the viral particles traverse both the upper and lower respiratory mucosal membranes to the lungs where SARS-CoV-2 binds to the human receptor, hACE2 via the RBD of the viral S protein [6, 430]. Following viral attachment to the hACE2, the virus gains access to the receptor cytoplasmic matrix through the S protein cleavage via the S_1-S_2 site and activation by the host proteases, transmembrane protease serine 2 (TMPRSS2), and cathepsin L [431–433], resulting in both endosome and TMPRSS2 mediated viral fusion (**Figure 4.1**)

4.1.2 Structure of the SARS-CoV-2 S Protein

The enveloped SARS-CoV-2 consists of transmembrane spike glycoprotein protuberances covering the entire surface of the viral lipid bilayer, giving it a crown-like appearance, hence the name coronavirus [432, 434]. The S protein plays a crucial role in SARS-CoV-2 pathogenesis by mediating viral envelope binding to the host hACE2 receptor via the S RBD [435]. Besides the S protein, SARS-CoV-2 consists of three other structural proteins,

envelope (E), membrane (M), and nucleocapsid, and 16 non-structural proteins (nsp1–nsp16) [436, 437]

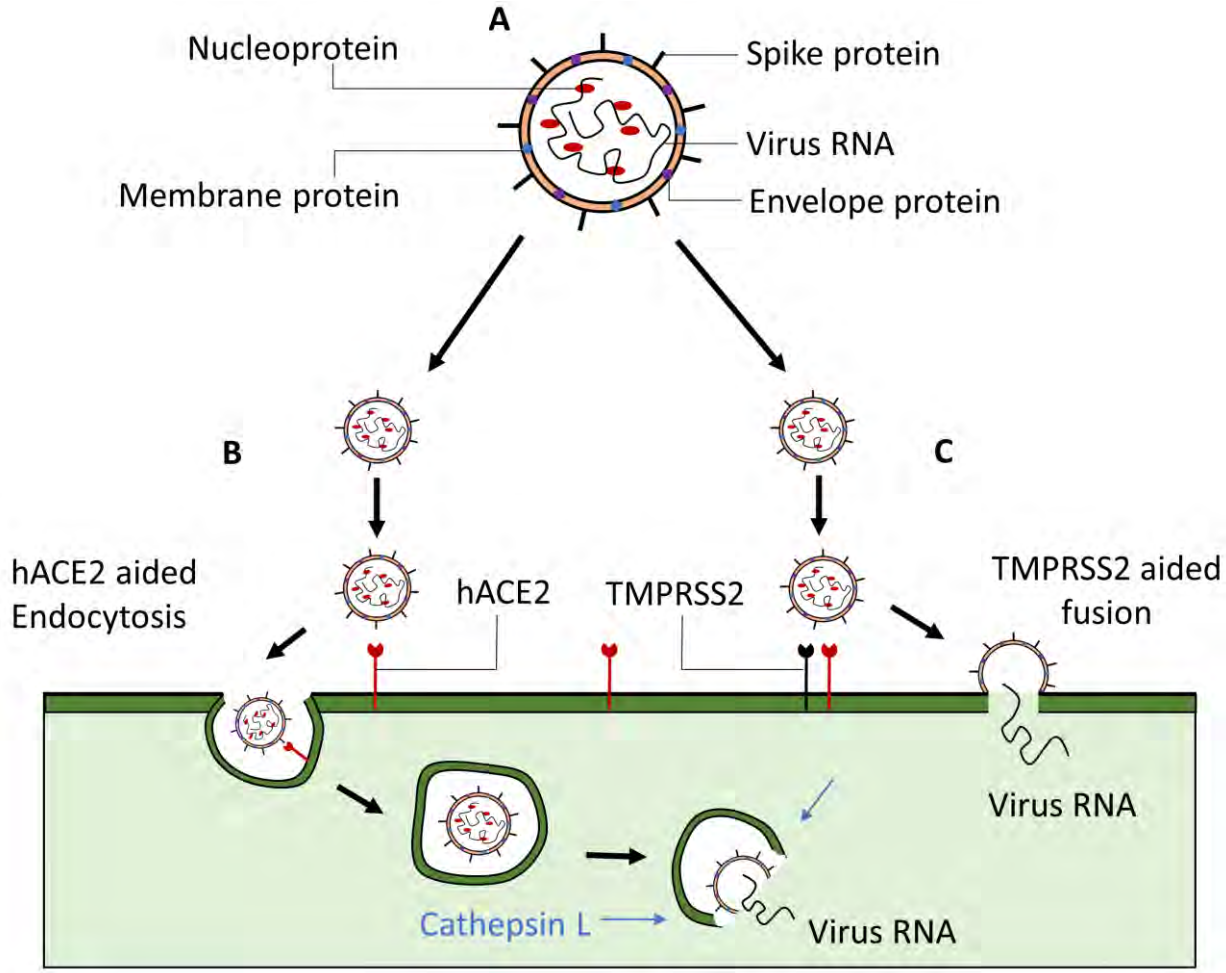

Figure 4. 1: A shows the different structural components of the SARS-CoV-2 virus whereas, B and C are schematic representations of the different routes of the viral fusion process into the host.

The SARS-CoV-2 S protein is homotrimeric, where each protomer is composed of the S_1 and S_2 subunits responsible for viral attachment and fusion, respectively [435]. The S_1 sub-unit comprises the RBD and the N-terminal domain (NTD). At the same time, the S_2 subunit is composed of a fusion peptide (FP), heptad repeat 1 (HR1), heptad repeat 2 (HR2), transmembrane domain (TM), and the cytoplasmic domain (CM) [414, 432, 435] (**Figure 4.2**). Furthermore, the S protein protomers are riddled with N-linked glycans, which are carbohydrate-based polymers, responsible for facilitating protein folding, serving as a

protective barrier [435, 438, 439] and potentially modulating antibody recognition [440]. Besides sharing 76% protein sequence identity, both SARS-CoV-2 and SARS-CoV utilize the hACE2 carboxypeptidase protein as a receptor for viral fusion into the host cells [432, 435, 441]. This is achieved through binding of the S RBD (residues 319–541) to the sub-domain I helix of the hACE2 via the receptor binding motif (RBM: residues 438-506) [414]. The RBM is concave shaped and lies between β-strands 4 and 7 of the RBD (**Figure 4.2**). The rest of the RBD comprises of anti-parallel β-strands 1, 2, 3, 5 and 6 interconnected by short α-helices and loops forming the RBD core. Due to the S_1 subunit involvement in human receptor interaction, both the RBD and NTD are the major targets for neutralizing antibodies and inhibitor binding [442–447].

Further down the S protein is the furin cleavage site characterized by the insertion of four amino acid residues (PRRA) unique to only the SARS-CoV-2, delimiting the S_1 and S_2 subunits of the protein [435]. Following the binding of S RBD to the hACE2, the S protein is firstly cleaved at the S_1/S_2 site [448], a process sometimes referred to as priming, and secondly at the S_2` site located at the N-terminus of the fusion peptide (FP) [449, 450]. Viral priming at the $S_1/S_2$ site affords the S protein conformational flexibility required for membrane fusion. At the same time the subsequent $S_2$`cleavage is believed to generate a mature FP required for viral insertion into the host membrane [450–452]. One exciting features of the SARS-CoV-2 S protein is the insertion of amino acids at the S_1/S_2 site that is absent in other coronaviruses. Jaimes et al. showed that this residue insertion predisposes the S_1/S_2 site to cleavage by enzymes other than furin, i.e., cathepsin L located in the endosomes and TMPRSS2 enzyme located on the membrane surface [448]. This structural modification of SARS-CoV-2 affords it a higher membrane fusion potential than SARS-CoV. Furthermore, the S protein D614G mutation that emerged early in the COVID-19 outbreak is associated with an S protein conformational change that modulates S_1/S_2 cleavage and

increases viral infectivity [453–457]. It is also important to mention that there is a possibility that the S RBD binding may also be influenced by hACE2 polymorphisms, as discussed in **chapter 5**.

4.1.3 S RBD Mutations and SARS-CoV-2 Infectivity

SARS-CoV-2, an RNA (ribonucleic acid) virus, is prone to multiple genome mutations involving missense mutations, insertions and indels to adapt to the environment [70, 163, 458]. Since the advent of the COVID-19 pandemic, a plethora of SARS-CoV-2 structural protein mutations have been documented, influencing both viral binding and antibody recognition patterns. More importantly, most of these mutations are localized at the RBD and NTD of the S protein, which are the critical targets of neutralizing antibodies and vaccine immunogens [444–446, 459]. Subsequently, the RBD and NTD mutations have been linked to SARS-CoV-2 antibody escape through impaired binding [460–466].

The WHO classifies the SARS-CoV-2 variations based on their transmissibility, disease severity, risk of reinfection, and their impact on diagnostics and vaccine performance [467, 468]. Under this criteria, variants of concern (VOC), which contain mutations associated with increased transmissibility or virulence or reduced effectiveness of diagnostics, include the Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2) and Omicron variant (B.1.1.529) [25–28]. The variants of interest (VOI) constitute mutations with predicted or known effects on transmissibility, immune escape, therapeutic escape and cause significant community transmission [467]. As of August 2023, the Omicron SARS-CoV-2 variant is the most prevalent [467]. Of all the VOC, the Omicron variant constitutes the most mutations in the S protein RBD (**Figure 4.2F**).

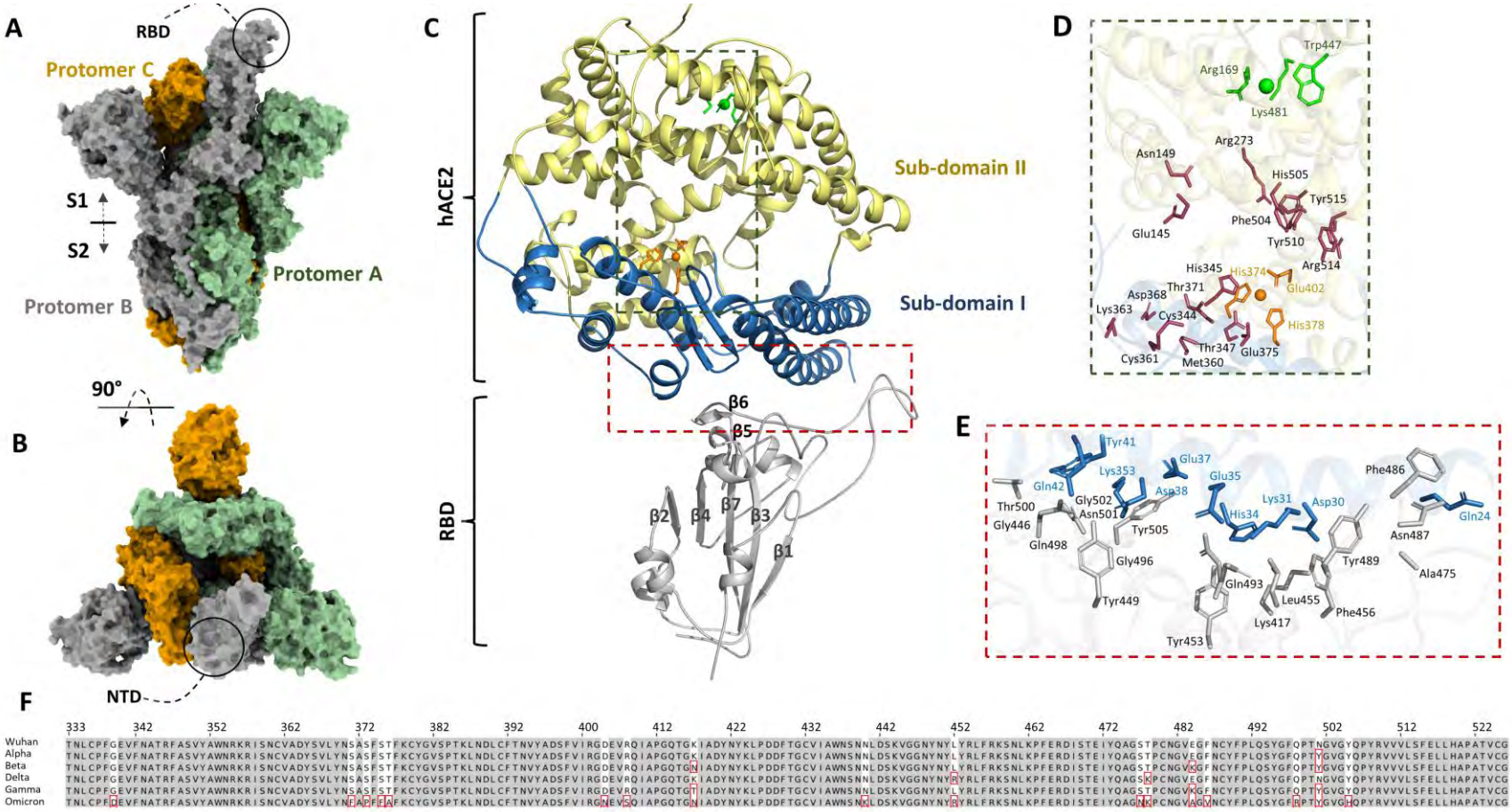

Figure 4. 2: A and B are surface presentations of the S protein in different orientations. C shows RBD-hACE2 complex highlighting the different regions of the complex. The hACE2 sub-domains I and II are colored in blue and yellow, respectively. Zinc and chloride ions are shown as orange and green spheres, whereas their coordinating residues as sticks in the same colors. The S RBD is shown in grey. D focuses on the active site, zinc and chloride coordinating residues, whereas E shows the RBD-hACE2 interface residues. F shows the RBD protein sequence alignment of the Wuhan and all VOC. The red boxes show the positions of mutation. Only the VOC most frequent mutations are shown.

4.1.3.1 The SARS-CoV-2 Omicron Variant and Sub-lineages

The SARS-CoV-2 Omicron variant (B.1.1.529) was first identified in Gauteng province, South Africa, and has since spread to all corners of the world [469, 470]. At the time of identification, the B.1.1.529 Omicron variant consisted of 30 S protein mutations, 15 of which were in the RBD, the most for any variant [471–473]. Over time, the Omicron variant has accumulated more mutations, especially in S protein resulting in the emergence of Omicron BA.1, BA.2, BA.3. BA.4, and BA.5 sub-lineages [474]. BA.1 contains 15 RBD mutations, whereas BA.2 harbors 16 RBD mutations. The S371F, T376A, D405N, and R408S SNPs of the BA.2 are absent in BA.1 [475, 476]. Like BA.1, BA.3 has a maximum of 15 RBD mutations, whereas BA.4 harbors 17 RBD mutations (**Figure 4.3**). Genome sequencing studies have shown that the BA.4 and BA.5 Omicron sub-lineages share the same S protein mutations [201, 477, 478].

Consequent to the critical role of the RBD in SARS-CoV-2 binding, a large body of research has focused on deciphering and understanding the effect of the RBD mutations on the S protein binding and antibody escape [188, 479–481]. Furthermore, studies have shown that RBD mutations, G339D, S375F, K417N, N440K, G446S, L452R, S477N, T478K, F486V and Q493R are associated with neutralizing antibody escape [465, 482–485]. The progressive emergence of Omicron sub-lineages with both familiar and exclusive RBD mutations hints to the evolution towards a more robust and immune viral structure, a situation that requires investigation. This chapter of the thesis therefore focused on profiling the dynamics and interaction between the S RBD and hACE2 across the different Omicron sub-lineages in a bid of informing future therapeutic measures. The chapter employs a dynamic residue network approach of analyzing RBD-hACE2 interactions for the first time.

4.2 Chapter Aims and Objectives

The unprecedented spread of COVID-19 coupled with the frequently evolving genome of SARS-CoV-2 underpins the research on SARS-CoV-2 mutations and their effects on hACE2 receptor interactions. Considering the elevated infectivity and transmissibility exhibited by the SARS-CoV-2 Omicron sub-lineage in contrast to other variants of concern (VOC), elucidating the sub-lineage's molecular interactions with the human angiotensin-converting enzyme 2 (hACE2) is of utmost significance in the pursuit of advancing virus management strategies and therapeutic interventions. To that effect, this chapter is focused on describing the effects and progressive change in the interaction between the RBD, harboring the different Omicron sub-lineage mutations (BA.1, BA.2, BA.3 and BA.4), and the human receptor, ACE2. The sub-lineage effects were characterized through i) identification of the Omicron sub-lineage mutations, ii) modeling of the RBD-hACE2 complex 3D structures, iii) MD simulation of the RBD-hACE2 WT and Omicron sub-lineage RBD-hACE2 complexes, iv) A comparison of the global changes in RBD-hACE2 dynamics with the progression of the Omicron sub-lineages through RMSD, RMSF and Rg calculations, and v) Network analysis of the RBD-hACE2 complexes through DRN analysis to establish the modifications in communication patterns in the various Omicron sub-lineage RBD and the hACE2.

4.3 Methods and Materials

4.3.1 Retrieval of SARS-CoV-2 Omicron Sub-lineage Mutations and Modeling of the RBD-hACE2 Complex Structures

The Omicron sub-lineage sequences were obtained from GISAID [157], a database of all influenza and coronavirus sequences, by searching using sub-lineage IDs *viz*: BA.1, BA.2, BA.3 and BA.4. This Omicron sub-lineage nomenclature is assigned by the phylogenetic assignment of named global outbreak lineages (PANGOLIN: https://cov-lineages.org/lineage_list.html) scheme [486]. Only complete human sequences of African origin and with patient status were retrieved from GISAID. Subsequently, CoVsurver [487], a GISAID application, was used to compare the identified sequences to the SARS-CoV-2 Wuhan reference sequence (hCoV-19/Wuhan/WIV04/2019) in order to identify specific protein SNPs. The RBD localized mutations for each sub-lineage were filtered out using an ad hoc Python script.

The identified RBD mutations for each Omicron sub-lineage were introduced into the RBD of the RBD-hACE2 template structure (PBD ID 6M0J) using PyMOL (version 2.5) [488]. The template (6M0J) had a resolution of 2.45 Å, R-value free of 0.227 and an R-value work of 0.192. Validation with MolProbity showed <0% bad bonds, <0.1% bad angles, 97.8% Ramachandran favored residues and 90.95% favored rotamers. The 6M0J was preferred as the template because it had better structure resolution compared to other existing Omicron structures like 7T9J (2.79 Å) and 7WBP (3.00 Å). The higher resolution for 6M0J meant more accurate atomic positions, allowing for detailed analysis of bonding interactions, protein folding, and active site geometry. In the rest of this chapter, the RBD-hACE2 complexes with Omicron sub-lineage RBD mutations are referred to by the Omicron sub-lineage name, i.e., BA.1, BA.2, BA.3 and BA.4 The six generated RBD-hACE2 complex

structures (for BA.2, 3 different mutation sequences were retrieved) were validated using VERIFY3D [247] and ProSA [248]. N-linked glycan free structures were used for MD calculations and, the titratable residues for each system were protonated at pH 7.0 using PROPKA, a tool from PDB2QR (version 2.1.1) [316]. N-linked glycan free structures were used to reduce the computational complexity of the of MD simulations and to focus on inter-protein interactions.

4.3.2 Molecular Dynamic Simulation

The protonated coordinate information for each Omicron sub-lineage and WT (reference structure) were converted to *gro* and *top* format using the GROMOS54a7 force fields [317] from GROMACS [315] (version 2019.4). Afterward, the systems were subjected to periodic boundary conditions (PBC) using a cubic box of 1 nm clearance prior to solvation using a SPC216 [257] water model. The solvated system charge was neutralized using NaCl ions of 0.15 M concentration. With the system charge neutralized, minimization commenced to correct the dihedral angles and remove any steric clashes using the steepest descent algorithm with an energy step size of 0.01. No constraints were used for minimization until an energy threshold of 1000.0 kJ/mol/nm was reached. Minimization was followed by temperature equilibration at 300K for 100 ps using Berendsen temperature coupling and pressure equilibration at 1 atm and 300 K using the Parrinello–Rahman barostat [318]. With both temperature and pressure equilibration achieved, the systems were subjected to 100 ns all-atom MD simulations with a step size of 2 fs. Importantly, all bonds were constrained for the equilibration and production runs under the LINCS algorithm [319] and, Particle Mesh Ewald (PME) electrostatics [320] were used for long-range electrostatic calculations using a Fourier spacing of 0.16 nm. A cut-off distance of 1.4 nm was used for the short-range coulomb and van der Waals interactions electrostatics.

Trajectory files from all-atom MD simulations were fitted to the reference frame using the *gmx trjconv* tool before removing the PBC. Solvent-free trajectories were then analyzed for RMSD, RMSF, Rg using the GROMACS tools *gmx rms*, *gmx rmsf* and *gmx gyrate*, respectively. The trajectory analysis results were further analyzed and presented using Seaborn [321], Pandas [322], pytraj [323], matplotlib [324], Numpy [325], and NGLview [326] Python libraries. System dynamics were visualized using the VMD tool [327].

4.3.3 Comparative Essential Dynamics

The GROMACS generated Omicron sub-lineage trajectories were each individually compared to that of the WT (reference) to decipher the most dominant motions using the *compare_essential_dynamics.py* Python script (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) from the MDM-TASK web server [264]. Here, the individual Omicron sub-lineage trajectories were aligned to that of the WT via the *Cα* atoms before decomposing the variance-covariance matrix. The advantage of comparative ED over traditional PCA is that for any given ensemble of systems, dominant motions are assessed within the same eigenspace [264]. The last three highly flexible terminal residues were excluded from the calculation to avoid noise and the results were presented as scatter plots along principal components 1 and 2 (PC1 and PC2).

4.3.4 Dynamic Cross Correlation

The RBD-hACE2 inter-protein and intra-protein atomic motions were appraised using the *calc_correlation.py* Python script (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) from MDM-TASK web server [264]. The last 20 ns of the equilibrated and coarse-grained (*Cα* atoms) trajectories were used for DCC calculation. Here, the degree of atomic correlation between the RBD and hACE2 proteins and within each protein per system were

ranked from -1 to 1, where -1 implied complete anti-correlation, 1 meant complete correlation and 0 meant no correlation at all. Results were presented as heat maps.

4.3.5 Dynamic Residue Network Analysis

The residue level changes in inter-protein and intra-protein network patterns were deciphered using DRN analysis. For each complex system, residue characteristics as described by five averaged DRN metrics, *BC*, *CC*, *DC*, *EC* and *KC* were determined using the *calc-network.py* (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) Python script from the MDM-TASK [264, 265] web server. Additionally, high centrality residues (hubs) per the global top 4% (hACE2) and 5% (RBD) were determined for each complex using an ad hoc Python script. Results from the global top 4% and 5% were presented as a heat map in Seaborn [321] to enable comparison across the WT and Omicron sub-lineage systems.

4.3.6 Contact Map Analysis

The *contact_map.py* Python script (https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web) from the MDM-TASK web server [264, 265] was used to calculate contact frequencies for RBD interface residues of each complex system. Interface residues were determined using the ROBETTA web server [330, 352] and the low-energy ED structure from **section 4.2.3**. A delta heat map of the residue pair RBD-hACE2 interactions between the WT and the Omicron sub-lineages was created in seaborn [321] to highlight the changes in complex interactions. Furthermore, bipartite plots of the RBD-hACE2 residue pair interactions, were plotted for each system using NetworkX [329], where the weight of the edges corresponded to the interaction frequency.

4.4 Results and Discussion

4.4.1 Omicron Sub-lineage Mutation Characteristics

The GISAID search yielded 56 SARS-CoV-2 sequences in total, accounting for the BA.1, BA.2, BA.3 and BA.4 Omicron sub-lineages (**Table S8**). At the time of the search (April 2022), no BA.5 sequences were retrieved under the specified search criteria (**section 4.2.1**). Of these sequences, BA.1, BA.2 and BA.4 comprised of 15, 16 and 17 RBD mutations. The BA.3 Omicron sub-lineage comprised of sequences with 10, 12 and 15 RBD mutations hence they are referred to here as BA.3_10, BA.3_12 and BA.3_15 for the rest of this chapter.

Alignment of the Omicron sub-lineage sequences using the Clustal Omega web tool [489] identified the common RBD mutations across all the sub-lineages *viz*:, G339D, S373P, S375F, S477N, T478K, E484A, Q498R, N501Y and Y505H (**Figure 4.3**). Furthermore, S protein mutation studies have previously shown that, among the shared mutations, G339D, S375F, S477N and T478K are associated with neutralizing antibody escape [465, 482, 485]. Of the unique mutations, L452R and F486V are limited to the BA.4 sub-lineage. Interesting, the number of RBD mutations increased with the progression of the Omicron sub-lineages suggesting an evolutionary adaptation of the virus for effective attachment and survival. The RBD mutations K417N, G446S, N501Y and Y505H are positioned at the RBD interface where they are involved and affect the RBD-hACE2 interactions.

It is important to note that not all the Omicron sub-lineage mutations improve the RBD-hACE2 interactions, studies have shown that the individual mutations i.e., S375F, K417N, G446S, G496S and Y505H are linked to decreased RBD-hACE2 binding [195, 490]. In contrast, some single mutations like G339D, N440K, T478K, S477N and N501Y are linked to the increased RBD affinity for hACE2. Other single mutations like S371L, S373P, E484A, Q493R and Q498R don’t have any documented effect of the inter-protein interaction. It is

also important to note that the change in residue physicochemical properties (**Table S9**) as a result of mutations influences RBD-hACE2 interaction. For example, the change from non-polar/uncharged in the WT to polar/positively charged residue in the Omicron sub-lineages for residues at positions 339, 452, 478, 493 and 498 favors inter-protein electrostatic interactions since hACE2 is predominantly negatively charged [491].

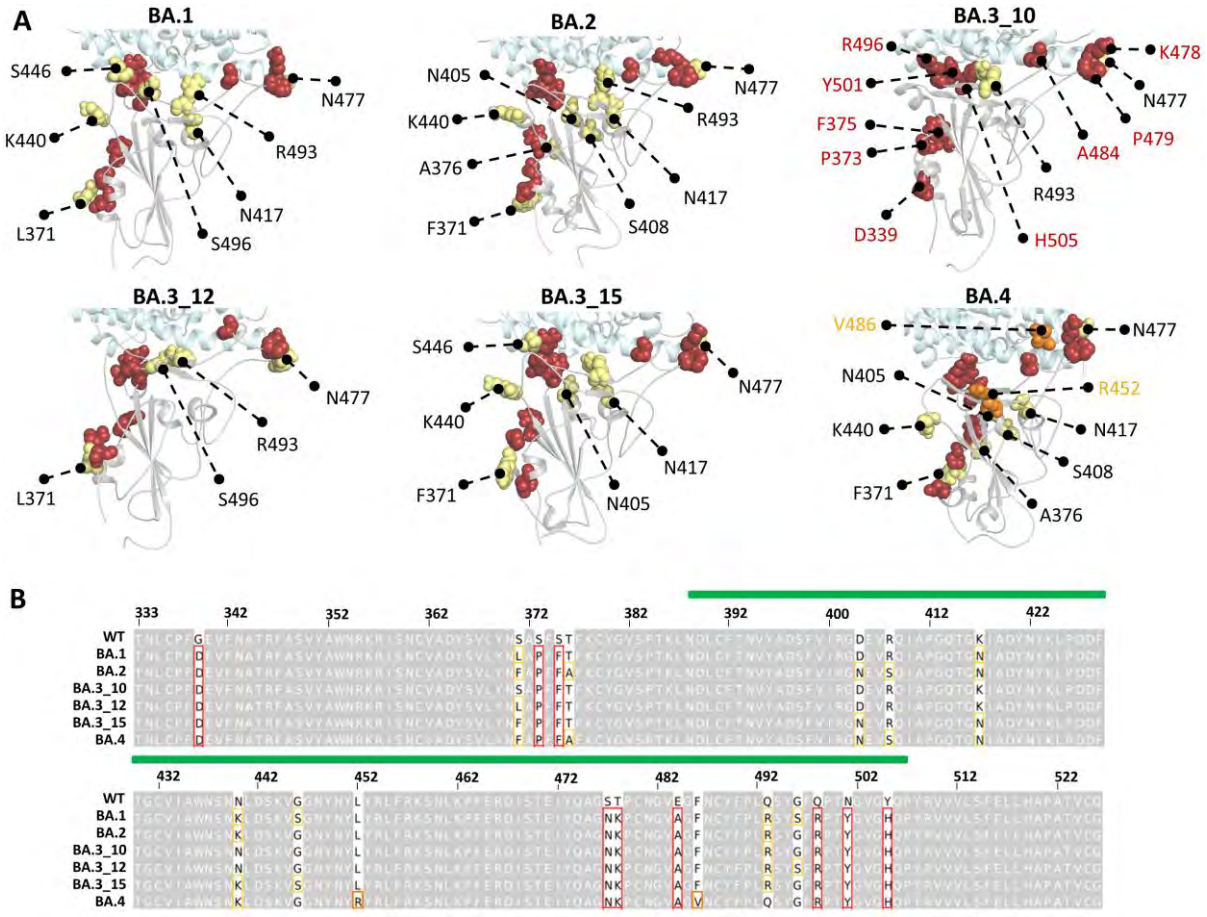

Figure 4. 3: A shows the distribution of RBD mutations for each Omicron sub-lineage. Mutations common to all sub-lineages are shown as firebrick spheres and annotated in BA.3_10, whereas the others are yellow spheres. The BA.4 unique mutations are annotated in orange. B shows the Omicron sub-lineage RBD protein sequence alignment highlighting the common, BA.4 unique, and other mutations in red, orange, and yellow, respectively. Adapted from Barozi et al., **[271]**. RBM residue positons are marked in green.

Furthermore, introduction of the tyrosine ring in the N501Y RBD mutation potentially alters the interface interactions through Pi-Pi stacking and pi-cation interactions. Ultimately, the

Omicron sub-lineages mutations have varied effects including increased affinity, decreased affinity, neutral and antibody escape characteristics which work in synergy for the most effective "induced fit" virus. Previous studies have illustrated that in some protein systems, a combination of mutations induce an allosteric effects that are not elicited in individual mutations [86]. These observations highlight the need to study co-occurring mutations as a collective, as discussed in next sections.

4.4.2 Effects of Omicron Sub-lineage Mutation on RBD-hACE2 Complex Dynamics

The effects of the Omicron sub-lineage RBD mutations on the complex and individual proteins were investigated through RMSD, RMSF and Rg analysis. For the WT complex, duplicate MD runs were performed and analyzed for experimental confirmation. Both WT runs achieved equilibration and displayed stable conformational evolution (**Figure S18A**). Comparative RMSD line plots between the WT and Omicron sub-lineages also showed similar and stable structural evolution except for BA.3_12 that experienced higher structural variability than the WT (**Figure S18B**). The RBD RMSD violin plot representation highlighted how both the WT and Omicron sub-lineage mutants sampled at least two main conformations except for unimodal distribution in BA.2 (**Figure 4.4A).** In the hACE2, the RMSD violin plots showed a stable conformational distribution with all systems experiencing a unimodal distribution except for BA.3_15 (**Figure 4.4B**). From a global perspective, the RBD mutations did not directly influence hACE2 dynamics except for BA.3_15.

Further analysis of RBD residue level behavior through RMSF highlighted regions 348-393 and 423-453 as having significant fluctuation compared to the WT, especially for BA.2 and BA.3_12 (**Figure 4.4C**). It is important to note that these highly flexible RBD regions constitute the antigenic sites that are targets for neutralizing antibodies i.e., 357-374, 381-394, 426-429, 438-450, 455-470, 469-491 and 495-506 [492].

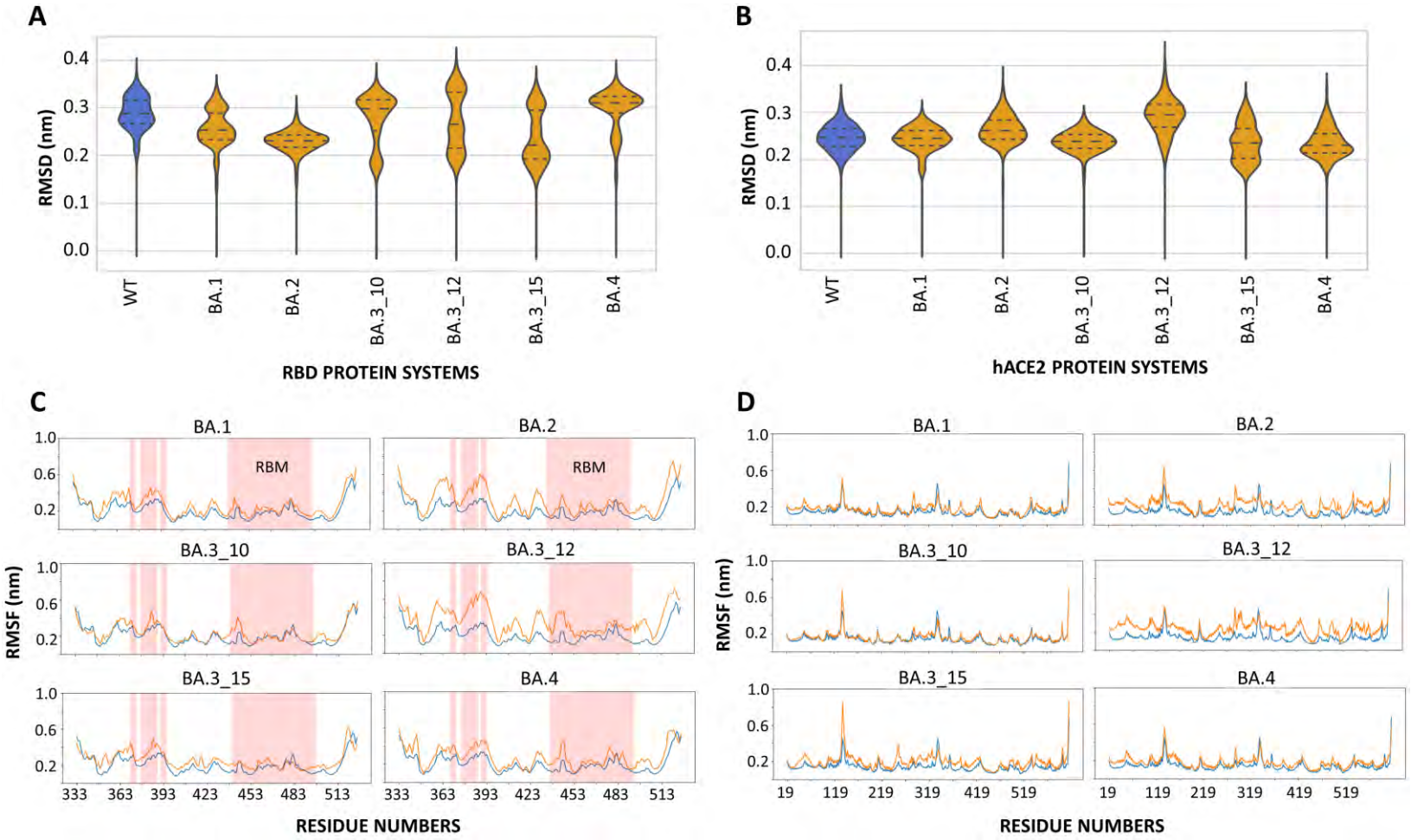

Figure 4. 4: A and B show the violin plot representation of RMSD for the WT (blue) and Omicron sub-lineage (orange) for the RBD and hACE2 proteins, respectively. C and D show the comparative line plots between the WT (blue) and Omicron sub-lineage (orange) residue flexibility for the RBD and hACE2 systems, respectively. RBD regions with an observable increase in the RMSF compared to the WT are highlighted in red. Adapted from Barozi et al., **[271]**.

This effect of Omicron sub-lineage mutations on RBD dynamics could potentially aid antibody evasion through preventing both antibody binding and recognition [493, 494]. Furthermore, the increased flexibility of the RBD interface loop region in the Omicron sub-lineages fosters more interactions with the hACE2 facilitating better binding compared to the WT [479]. Analysis of the protein Rg showed minimal differences between the WT and the Omicron sub-lineages with BA.4 and BA.3_12 experiencing the highest gyration in the RBD and hACE2 proteins, respectively (**Figure S18C and D**).

4.4.3 Comparative Essential Dynamics of the WT and Omicron Sub-lineage Systems

The most dominant motions of the Omicron RBD-hACE2 systems were determined in comparison to WT within same eigenspace using comparative ED as described in **section 1.10.3.5.** Scatter plots of protein dynamics described by PC1 and PC2 were used to analyze the extent of protein motions. Here, ED analysis was performed for RBD, hACE2 and the hACE2 active site pocket.

In the RBD, all the systems sampled a more diverse conformational space compared to the WT (**Figure 4.5A**). In the BA.1 system, conformational diversity in comparison to the WT was along PC2, and both PC1 and PC2 accounted for ~44% of the ED. Likewise, all other Omicron sub-lineage RBD systems experienced a more diverse conformational evolution along PC2 compared to the WT. The total variance based on PC1 and PC2 for BA.2, BA.3_10, BA.3_12, BA.3_15 and BA.4 was approximately 57, 52, 51 and 60%, respectively. Alignment of the Omicron sub-lineage low energy (basin) RBD structures as determined by ED to that of the WT identified the loop regions including the RBM as the most variant in structure compared to the WT. The RBD loop and helix region: 360-372 also showed pronounced conformational variation between the Omicron RBD and the WT systems as depicted from the basin structure RMSD (BA.1: 1.588, BA.2: 1.932, BA.3_10: 2.612,

BA.3_12: 2.411, BA.3_15: 1.737 and BA.4: 2.343). These observations are in agreement with the RMSF results (section 4.4.2, **Figure 4.4**) which highlighted the loop regions including the RBM as highly flexible Omicron sub-lineages. Previous studies have highlighted how increased flexibility, particularly of antigenic sites, can affect antibody recognition and can be a mechanism of antibody escape [493, 494]. We believe that the increased flexibility of the Omicron sub-lineage RBD including the regions identified as antigenic targets: 370, 375-386, 390, and 444-456 is an antibody escape strategy to prevent neutralizing antibody binding to the S protein RBD.

In the same way, the effects of Omicron RBD mutations translated to the hACE2 where the mutant proteins experienced more diverse hACE2 conformational dynamics compared to the WT (**Figure 4.5B**). Interestingly, the collective variance as accounted by PC1 and PC2 was higher in the hACE2 systems compared to the RBD systems, probably due to minimal changes in conformational variation of the hACE2. Here, the total PC1 and PC2 variance was ~62, 52, 50, 61, 57 and 51% for BA.1, BA.2, BA.3_10, BA.3_12, BA.3_15 and BA.4 systems, respectively. Like in the RBD, the differences in the sampled conformational space between the hACE bound to the WT RBD and that of the sub-lineages was attributed to the highly flexible conformational poses of the loop regions especially 128-144 and 333-347. The aligned hACE basin structures had the following RMSDs to the WT; BA.1: 2.726, BA.2: 1.993, BA.3_10: 2.411, BA.3_12: 2.412, BA.3_15: 2.407 and BA.4: 1.958.

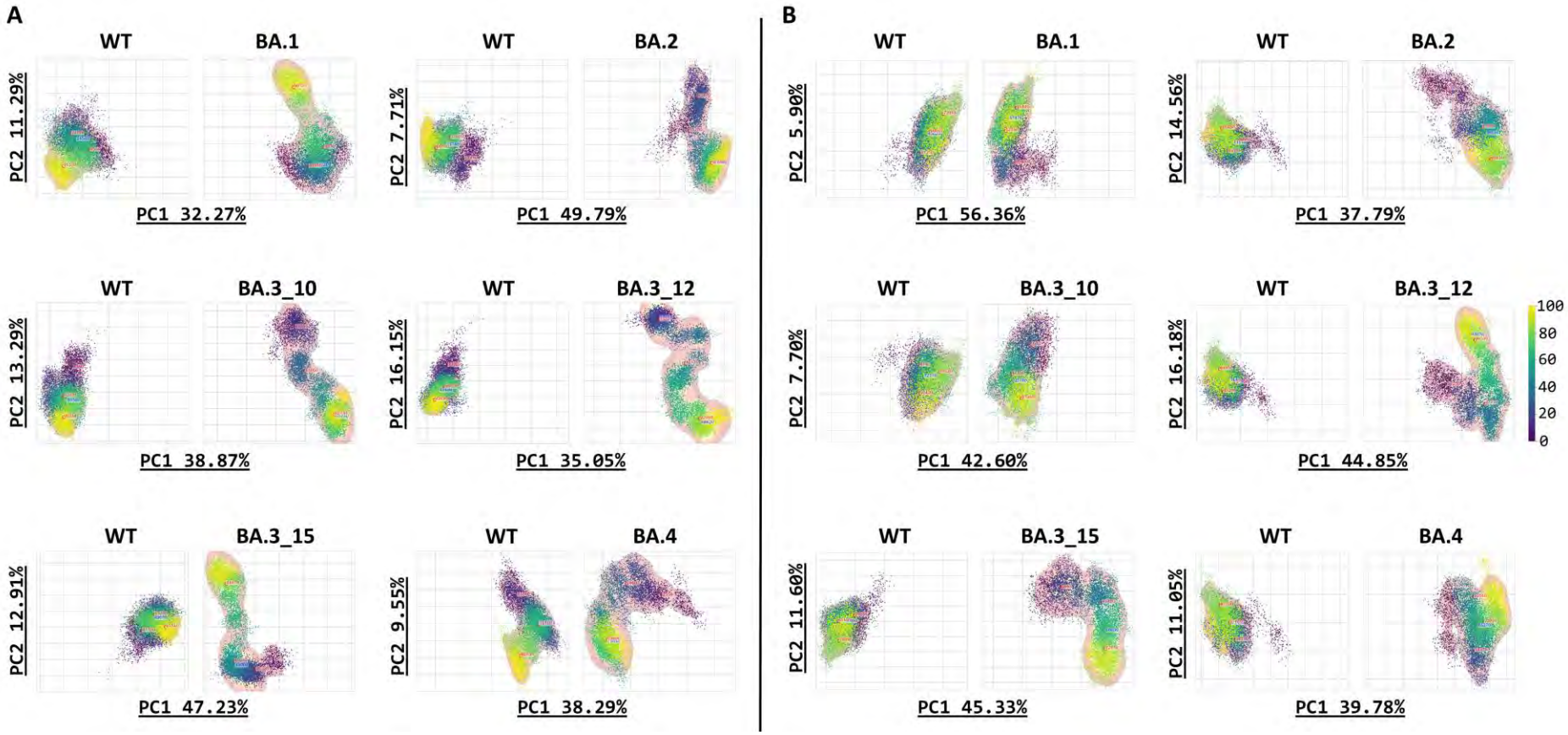

Figure 4. 5: A and B are comparative ED scatter plots as described by PC1 and PC2 for the RBD and hACE2 proteins, respectively. The x and y-axes in each subplot show the percentage variance as described by PC1 and PC2, respectively. Adapted from Barozi et al., **[271]**.

Zooming in on the hACE2 protein systems, we sought to decipher the effects of the RBD Omicron mutations on the protein active pocket. Previous studies have shown that binding of RBD to the human receptor influences the protein functionality resulting in increased carboxypeptidase activity against peptide substrates such as caspase-1 and Bradykinin-analog [495–497]. The hACE2 active pocket is located in the groove between sub-domain I and II where ligand binding triggers hinge movement of sub-domain I towards sub-domain II, enclosing the ligand in place [498]. Here, we determined the comparative ED for the active site groove region of the of the Omicron sub-lineages in comparison to the WT along both PC1 and PC2.

The active cavity of all the Omicron RBD complexed hACE2 proteins sampled a higher conformational space compared to the WT (**Figure S19A**). The most conformational variation was observed in the BA.2 system (along PC1) where the collective variance was ~56%. Interestingly, the hACE2 groove ED weren't significantly higher than the WT as confirmed by the groove Rg calculations which showed minimal changes between the systems (**Figure S19B**).

Further investigation into the hACE2 hinge movements through COM distance calculation between sub-domain 1 (defined by active site residues, His345, Asp368, Thr371, Glu375, His378 and Glu402) and sub-domain II (defined by active site residues, Phe274, Leu278, His505, Tyr510, Arg514 and Tyr515) showed trivial changes in sub-domain movements compared to the WT throughout the simulation (**Figure S20A**). Further still, Rg calculations of these residues hinted to a similar conclusion with an Rg value range of 1.00–1.4 nm (**Figure S20B**). The lack of significant hACE2 hinge movement here could be due to the absence of a hACE2 substrate.

In summary, the comparative ED results highlighted 1) the effects of RBD omicron sub-lineage mutations on the dynamics of the domain characterized by a higher conformational variation in the mutants compared to the WT. This was in agreement with the RMSD and RMSF results that showed pronounced dynamics especially for the RBM region (**section 4.3.2**). 2) Even though RBD binding to hACE2 is associated with increased carboxypeptidase activity, no significant structural changes of the hACE2 active site groove region were observed in the Omicron sub-lineages compared to the WT and this was attributed to the absence of a hACE2 substrate which tiggers the hinge movement upon binding.

Both sections **4.3.2** and **4.3.3** focused on the effects of RBD Omicron sub-lineage mutations on individual proteins, RBD and hACE2. To gain a full picture on both inter-protein and intra-protein atomic relation, DCC was computed for each complex in the next section.

4.4.4 Analysis of the RBD-hACE2 Relationship Through Dynamic Cross-Correlation

Protein functionality involves an ensemble of conformational changes at varying times scales which facilitate molecular interactions, ligand (un)binding and protein (un)folding [499]. Analysis of RBD-hACE2 protein conformational changes, through DCC, provided insight into the protein interrelationship and functionality. DCC analysis identified anti-correlated atomic motions between the RBD and hACE2 proteins in BA.1, BA.3_10, BA.3_15, and BA.4 contrary to the WT which showed correlated motions (**Figure S21A**). Furthermore, analysis of the RBD-hACE2 inter-protein COM also highlighted marginal increase in the inter-protein interaction distance in BA.1, BA.3_10, BA.3_15, and BA.4 compared to the WT (**Figure S22**). This suggests that the RBD mutations result in diametric protein movements in some Omicron sub-lineages. Analysis of the protein binding energy using the low energy complex structures from ED (section 4.4.3) and the HawkDock web server [500] showed that the WT experienced lower binding energy, -80.6 kcal/mol, compared to BA.1 (-46.45

kcal/mol), BA.2 (-55.94 kcal/mol) and BA.4 (-74.4 kcal/mol), whereas BA.3_10, BA.3_12 and BA.3_15 had energies of -82.72, -82.72 and -93.78 kcal/mol, respectively.

At an individual protein level, anti-correlated atomic motions were observed in the RBD of BA.1, BA.2 and BA.3_12 defined by protein region 440-508 (**Figure S21B**). Interestingly, this region also encompasses the RBM which as discussed in **section 4.4.2**, underwent greater residue fluctuations in these mutant systems compared to the WT. The anti-correlated motions of the RBD systems, BA.1, BA.2 and BA.3_12 translated to the hACE2 protein as observed from the DCC results. Here, the hACE2 experienced anti-correlated atomic motions in BA.1, BA.2 and BA.3_12 systems, specifically around residues 119-315 and 419-519 (**Figure S21C**). The X-ray crystal structure of the hACE2 [498] identifies this region as part of the sub-domain I (**Figure 4.2**) which interacts with the RBD of the S protein. Seemingly, RBD binding influences the dynamics of hACE2, an observation that is in collaboration with Lu & Sun, 2020 [497] who showed that RBD binding is associated with conformational changes in the hACE2 sub-domain II [498].

Therefore, DCC analysis highlights the changes in atomic motions of the viral RBD due to Omicron sub-lineage mutations and in the human receptor, hACE2, as a result of RBD binding. Furthermore, DCC analysis emphasized the effects of mutations of protein dynamics through both correlated and anti-correlated intra-protein motions. The consequential effects of these dynamics on protein residue networks were investigated in the next section.

4.4.5 Dynamic Residue Network Analysis of the RBD-hACE2 Complexes

Here we applied five metrics, averaged *BC*, *CC*, *DC*, *EC* and *KC* to identify the functional/central residues and characterize mutation effects to the enzyme communication patterns. Previous application of these matrices identified key allosteric residues [76] and

described the changes in network communication paths consequent to mutations and/or ligand binding [152, 153, 155, 501]. In these studies, we defined high centrality residues, hubs, as residues with high centrality values per DRN metric based on a set global threshold of the protein ensemble under study. Furthermore, we defined *persistent hubs* as the common high centrality residues across all the protein ensembles (**Figure 4.6**).

4.4.5.1 High BC Residue Networks Connected the RBD core to the hACE2

As explained in chapter 1, **section 1.10.4.1**, the *BC* metric identifies highly central residues based on their participation in the shortest paths between any given pair of network residues [502, 503]. At a 5% centrality cut-off, *BC* analysis identified residues Tyr508 and Val510 as *persistent hubs* [76] in the RBD. In as far as the RBD-hACE2 interactions are concerned, mutations on either Tyr508 or Val510 have been documented to reduce the RBD-hACE2 binding affinity [195]. In the RBD, Tyr508 is responsible for binding neutralizing camel nanobodies [504], natural inhibitory compounds [505], standard therapeutic drugs [506, 507], metal complexes [508] and inhibitory peptides [509]. Similarly, Val510 is functionally vital in binding antiviral compounds [506, 510, 511]. The high centrality at these residue positions emphasizes their importance in RBD functionality.

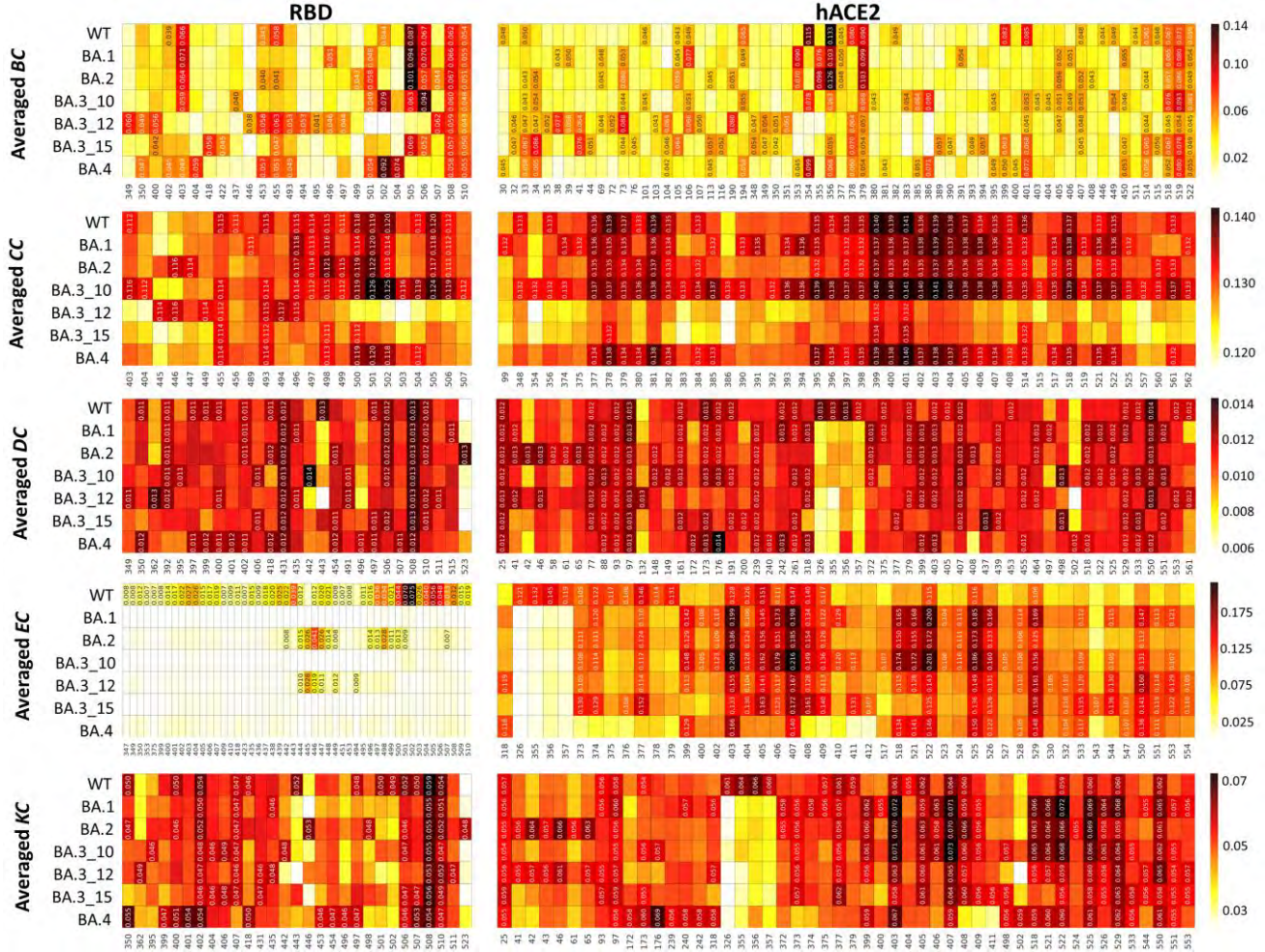

Figure 4. 6: Heat map of top 5% and 4% high centrality residues for the RBD and hACE2 proteins, respectively. Residue hubs are annotated with centrality values, whereas homologous residues from other systems are not. The color scale from white to dark red shows the degree of centrality. Adapted from Barozi et al., **[271]**.

Furthermore, a 4% global top centrality cut-off was used to define hACE2 hubs and this analysis identified Ile379, Arg518, Thr519 and Gln522 as the *persistent hubs*. Ile379 forms van der Waals interactions with His378 known to coordinate the zinc ion, which is necessary for the carboxypeptidase activity. Other hACE2 residues with high *BC* values in at least five systems included Asn33, His34, Leu73, Phe356, His401, Ile407 and Arg514. Asn33 and His34 are involved in the hACE2-RBD inter-protein interactions, whereas His401 forms part of the zinc-binding site [498]. Identification of these residues as important by this approach of *BC* analysis validates our centrality analysis approach given the residue importance/functional role in the protein.

Interestingly, the assortment of *BC* hubs in the WT RBD and hACE2 formed a pattern interlacing the two protein systems along two paths, I and II, when mapped on the 3D complex structure (**Figure 4.7**). Path I was the most prominent consisting of majority of *BC* hubs originating from the RBD core through the interface region connecting the RBD to the hACE2 zinc-binding pocket. The second path (II), on the other end of the structure, consisted of fewer residues. The paths also consisted of the five highest centrality residues, shown in dark grey (RBD) and bark blue (hACE2) that bridged communication between the two protein systems.

Gains and losses of *BC* hub residues in the Omicron sub-lineage systems were observed in comparison to the WT (**Figure 4.7**). The key differences between WT and Omicron sub-lineages included residue Tyr505 in the RBD, which lost centrality in BA.3_12 and BA.4 when mutated to His505. Previous studies have shown that the substitution of tyrosine to histidine at position 505 enhances the RBD binding to hACE2 [512, 513]; however, another school of thought argues that the residue substitution reduces the RBD-hACE2 interactions [195, 514]. We are of the view that the change to a positively charged histidine might increase the electrostatic interactions with the primarily negatively charged hACE2. This

view is further supported by Starr et al., 2020 who used deep mutational scanning to show that mutations of polar amino acids enhance expression at interface residues Y449, L455, F486, and Y505 which are destabilizing region however, the from essential contacts with hACE2 necessary for S protein binding [195].

The BA.1 system lost interface residue Gly502 as a hub in the RBD and Asp30 and Gly354 in the hACE2. Compensatory gains in *BC* hubs were noted at residues Thr496 and Thr501 (RBD) and Asp355 (hACE2). Similarly, the loss of RBD Gly502 and hACE2 Asp30 as interface *BC* hubs was also observed in BA.2, BA.3_10, BA.3_12, BA.3_15 and BA.4 (except Asp30). These systems also had gains in *BC* hubs along the communication path and at the protein interface. Gained hubs included residues, Pro499 and Thr501 in BA.2; Gly446, Arg493, Ser494 and Ser496 in BA.3_12 and, Arg493, and Tyr501 in BA.4. The *BC* hub residue Thr501 is linked to increased RBD-hACE2 interactions in the mutant systems compared to the WT [194, 515, 516].

In addition to the observed variations in the quantity of *BC* hubs comprising the communication pathways within the Omicron sub-lineages, there existed inconsequential disparities in the degree of *BC* exhibited by these hubs between the WT and Omicron sub-lineages, both in relation to the RBD and the hACE2 protein. Notably, in the context of the RBD, the BA.1 sub-lineage displayed the highest mean *BC* among the sub-lineages, with values as follows: WT: 0.058, BA.1: 0.065, BA.2: 0.057, BA.3_10: 0.062, BA.3_12: 0.052, BA.3_15: 0.053, and BA.4: 0.057. Conversely, within the hACE2 protein, the highest mean *BC* among the hubs was exhibited by the WT, as indicated by the following values: WT: 0.064, BA.1: 0.063, BA.2: 0.062, BA.3_10: 0.057, BA.3_12: 0.055, BA.3_15: 0.057, and BA.4: 0.057.

It is also important to note that this analysis identified, for the first time, the key residues facilitating communication between the RBD and hACE2 proteins. Secondly, the Omicron sub-lineage systems BA.3_10 (33 hubs), BA.3_12 (42 hubs), BA.3_15 (38 hubs), and BA.4 (35 hubs) had more *BC* hubs compared to the WT (32 hubs). The gain in *BC* hubs at the interface and in the rest of the mutant structures is indicative of a transformative adaptation of the Omicron sub-lineages to facilitate efficient inter-protein communication. Furthermore, the results also highlight the effects of the sub-lineage RBD mutations on the hACE2 centrality, implying an allosteric effect on complex communication patterns.

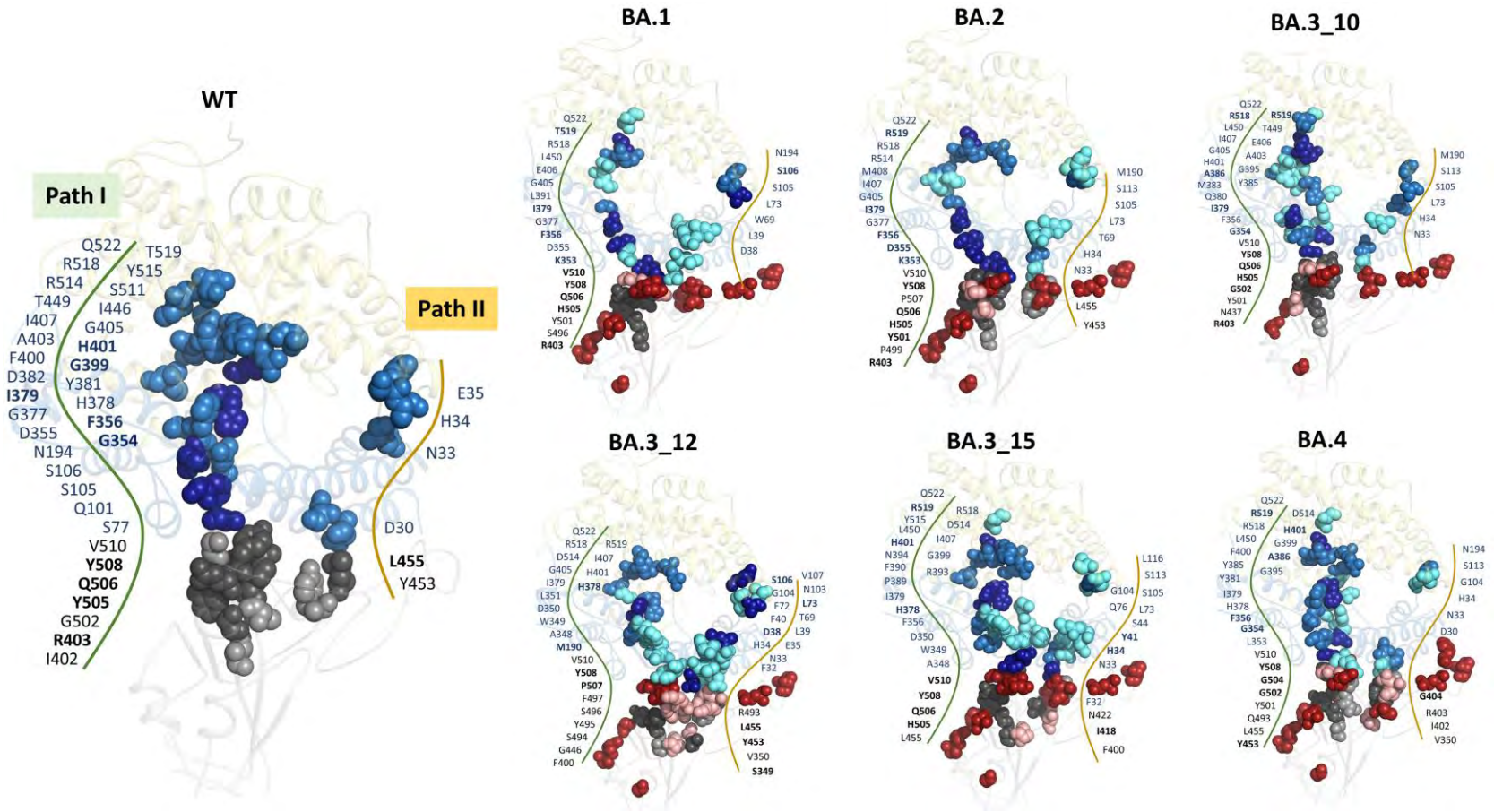

Figure 4. 7: Cartoon representation of the RBD-hACE2 complex showing the distribution of *BC* hubs in the RBD and hACE2 as identified by the global top 5 and 4% cut-off, respectively. The hACE2 sub-domains I and II are shown as sky blue and yellow, respectively, while the RBD is grey. *BC* hubs common to the WT are shown as grey (RBD) and blue (hACE2) spheres. Hubs unique to the Omicron sub-lineages are shown salmon (RBD) and teal (hACE2) spheres. The five highest centrality residues are in dark grey (RBD) and dark blue (hACE2) whereas, the mutation positions are shown as firebrick spheres. Adapted from Barozi et al., **[271]**.

4.4.5.2 CC Hub distribution Correlated with Inter-protein Interaction Distance

The *CC* metric assigns centrality based on how centrally located a residue is in relation to the rest of the network. Residues with high centrality are essential for information dissemination in the network [268]. In the RBD, no *persistent hubs* were observed across the protein systems; however, residues Leu455, Gln493, Gly496, Gln498, Thr500, Asn501, and Gly502 maintained high centrality values in at least five RBD systems (**Figure 4.6**). Interestingly, these residues are positioned at the RBD interface with the hACE2, where they are involved in inter-protein interactions [414, 517–519]. In addition, the assortment of Omicron RBD mutations generally curtailed the distribution of *CC* in the RBD. Here, the WT had the highest number of *CC* hubs compared to other systems except for BA.3_10, i.e., WT: 14 *CC* hubs, BA.1: 9, BA.2: 11, BA.3_10: 16, BA.3_12: 7, BA.3_15: 4 and BA.4: 7 (**Figure 4.8**).

The reduction of Omicron sub-lineage *CC* hubs compared to the WT was further investigated through RBD-hACE2 COM distance measurement to ascertain the complex behavior. The reduction in interface residue centrality was associated with increased inter-protein interaction distance, specifically in BA.3_12 (**Figure S22**). This was attributed to the increased residue flexibility of the RBM, as earlier observed in **section 4.3.2**.

The zinc and active site pocket region dominated the distribution of *CC* hubs in the hACE2 protein system. Here, BA.3_12 and BA.3_15 had the least number of *CC* hubs consisting of residues His378, Gly399 (*persistent hub*), His401 (*persistent hub*), and Arg514 (**Figure 4.6**). The progression of RBD mutations from 10 to 17 in the omicron sub-lineages studied here, displayed an intriguing distribution of *CC* hubs in the protein complexes (**Figure 4.8**). BA.1 lacks the D405N and R408S RBD mutations present in BA.2. In presence of these mutations, the BA.2 system lost the *CC* hubs at positions Tyr489 (RBD) and, His374 and Glu375 (hACE2). The RBD Tyr489 forms hydrogen bonds with Tyr83 (hACE2), His374 coordinates

the zinc ion and Glu375 forms part of the active site in the hACE2. Progression from 16 RBD mutations in BA.2 to 10 in BA.3_10 was characterized by an increase in both the RBD and hACE2 *CC* hubs. Here, the RBD interface residue Gln493, which forms H-bonds with Asp30 of hACE2, gained *CC* hub status in BA.3_10. BA.3_10 also gained hubs at the hACE2 residues His374 and Tyr515 at the active site. Peculiarly, the evolution of Omicron sub-lineages from BA.3_10 to BA.3_12 and further to BA.3_15 resulted in further reduction in *CC* hubs in both RBD and hACE2 proteins, specifically at the interface *viz*: residues Arg403, Gly404, Phe497, Gln498, Pro499, Asn501, Gly502, Val503, Gly504, Tyr505, Gln506 and Pro507. Seemingly, the increase in RBD mutations is associated with a decrease in residue *closeness centrality* until the BA.4 sub-lineage (**Figure 4.8**). The BA.4 sub-lineage with 17 RBD mutations was characterized by a gain in both RBD and hACE2 *CC* hubs compared to BA.3_12 and BA.3_15. It is worth noting that there was a reversion of amino acid residues Gly446, Gln493, and Gly496 back to the same residues as in the WT sequence in the BA.4 Omicron sub-lineage. The progressive change of these Omicron sub-lineage mutations and the change in network patterns represents a trial-and-error-based evolution in the pursuit of the most efficient RBD-hACE2 fit.

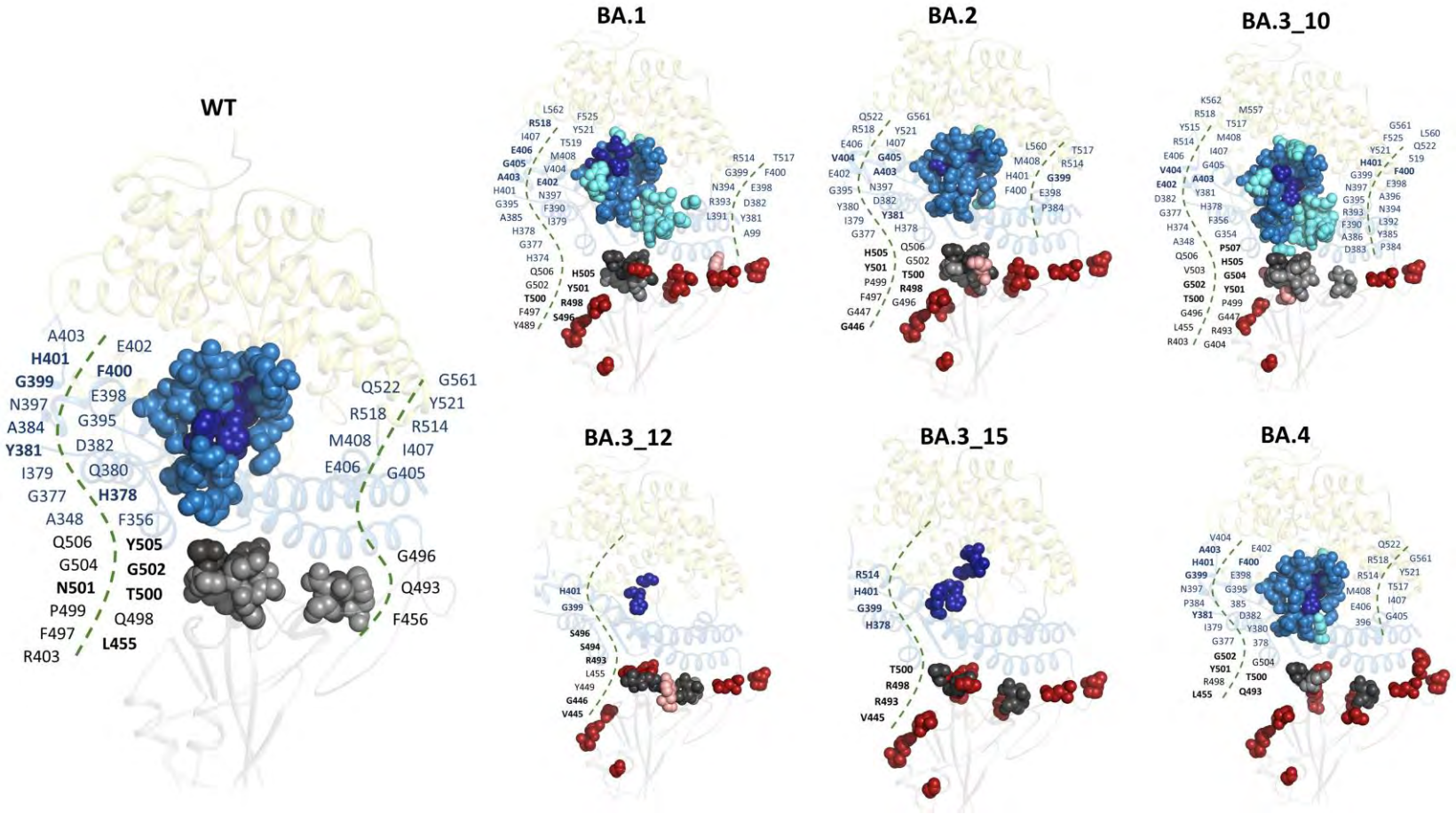

Figure 4. 8: Cartoon representation of the RBD-hACE2 complex showing the distribution of *CC* hubs in the RBD and hACE2 as identified by the global top 5 and 4% cut-off, respectively. The hACE2 sub-domains I and II are shown as sky blue and yellow, respectively while the RBD is grey. *CC* hubs common to the WT are shown as grey (RBD) and blue (hACE2) spheres. *CC* hubs unique to the Omicron sub-lineages are shown as salmon (RBD) and teal (hACE2) spheres. The five highest centrality residues are in dark grey (RBD) and dark blue (hACE2) whereas, the mutation positions are shown as firebrick spheres. Adapted from Barozi et al., **[271]**.

4.4.5.3 EC Analysis Identified an Interrupted hACE2 Allosteric Path in Omicron Sub-lineages

As discussed in **section 1.10.4.4**, the *EC* metric characterizes network properties based on the node/residue influence in the network. Peculiarly, the RBD global top 5% centrality cut-off identified *EC* hubs mainly in the WT system and partly in BA.2 and BA.3_12 Omicron sub-lineage systems (**Figure 4.6**). Here, the collection of RBD mutations significantly changed the connectivity/interaction properties of the RBD network by generally reducing the degree of residue connectivity.

In the absence of RBD mutations, the WT had the highest number of *EC* hubs distributed throughout the β2, β3, β4, and β7 strands of the RBD core and extending to the RBM at the protein interface (**Figure 4.9**). Unlike the WT, the BA.2 and BA.3_12 RBD *EC* hubs were limited to residue positions 422, 444–449, 496–501 and 507 (BA.2), and 444–447, 449, and 494 (BA.3_12). The decline in residue centrality in the Omicron sub-lineages was attributed to increased RBM flexibility in the sub-lineages as discussed in **section 4.4.2**.

Furthermore, the RBD interface residues with significantly high *EC* values unique to the WT system are documented neutralizing antibody epitopes *viz*: Ala403, Ile436, Leu444, Thr449, Cys498, Asn501 and His505 [483, 520, 521]. Therefore, the increase in residue flexibility and subsequent loss of *EC* at these residue positions signals to a neutralizing antibody escape strategy of the Omicron sub-lineages.

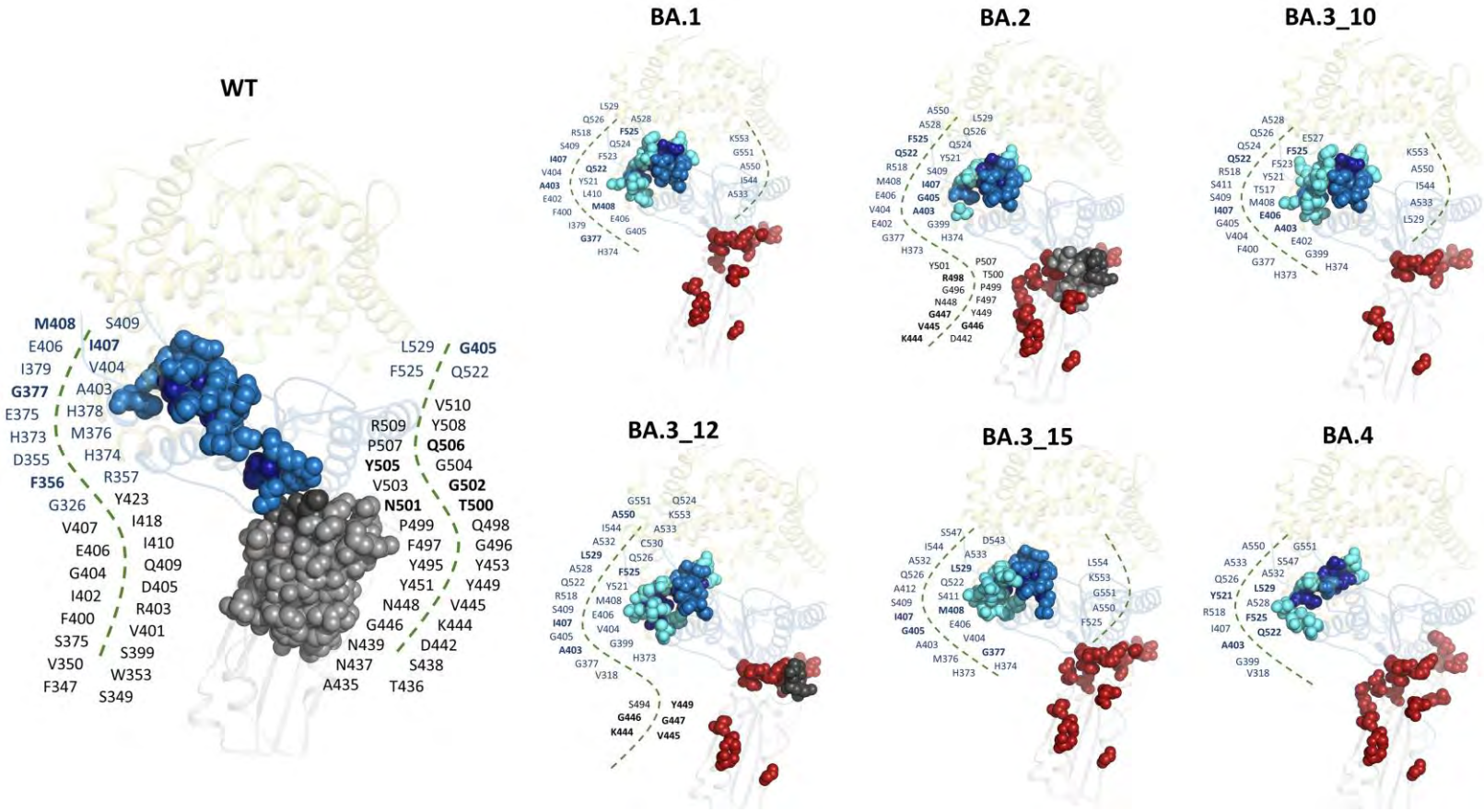

Figure 4. 9: Cartoon representation of the RBD-hACE2 complex showing the distribution of *EC* hubs in the RBD and hACE2 as identified by the global top 5 % and 4% cut-off, respectively. The hACE2 sub-domains I and II are shown as sky blue and yellow, respectively while the RBD is grey. *EC* hubs common to the WT are shown as grey (RBD) and blue (hACE2) spheres. *EC* hubs unique to the Omicron sub-lineages are shown as salmon (RBD) and teal (hACE2) spheres. The five highest centrality residues are in dark grey (RBD) and dark blue (hACE2) whereas, the mutation positions are shown as firebrick spheres. Adapted from Barozi et al., **[271]**.

RBD mutations also allosterically affected the distribution of *EC* in the hACE2 protein. Here, we observed a general increase in the number of hACE2 *EC* hubs in the Omicron sub-lineage systems compared to the WT i.e., BA.3_10: 33 hubs > BA.3_15: 32 > BA.3_12: 30 > WT: 28 > BA.1: 26 > BA.2: 25 > BA.4: 25. The marked increase in hACE2 *EC* hubs especially around the active site and zinc-binding region is believed to be a response to the mutations in the RBD. This is based on the interruption of the high *EC* residue path (residue positions 326, 355–357, 375, 378 and 379) connecting the RBD to the hACE2 in the Omicron sub-lineage systems compared to the WT. The lost *EC* hubs in the Omicron sub-lineages also experienced higher RMSF compared to the WT, highlighting the relationship between *EC* and residue flexibility (**Table S10**).

Based on the relationship between S RBD binding and increased hACE2 carboxypeptidase activity [497], we hypothesize that the observed *EC* path interruption in the sub-lineages hints on the influence of the Omicron variant on hACE2 activity. Given the physiological role of hACE2, the observed changes in the residue communication paths at the enzyme active site could potentially influence its activity in the body.

The findings from *EC* analysis further accentuated the effects of the Omicron RBD mutations on the intra-protein and inter-protein communication patterns in the RBD-hACE2 complex. The decreased centrality of the antigenic hot spots observed here is in agreement with the findings of Cerutti and the team, who described a more flexible S protein RBD in the Cryo-EM structure, a behavior they linked to neutralizing antibody escape [522].

4.4.5.4 Analysis of the RBD-hACE2 DC and KC

Analysis of the RBD-hACE2 network characteristics through the *DC* metric identified RBD residues, Gly431 and Tyr508, as *persistent hubs* (**Figure 4.6**). A general reduction in the number of RBD *DC* hubs was noted in the Omicron sub-lineage systems compared to the WT

i.e., WT: 13 hubs, BA.1: 8, BA.2: 9, BA.3_10: 8, BA.3_12: 10, BA.3_15: 7 and BA.4: 12. The RBD hubs were predominantly distributed at the protein core within the α1, α3, β3 and β7 regions (**Figure S23**).

In the hACE2, residues Ala25, Val93 and Leu97 were the *persistent hubs* (**Figure 4.6).** *DC* hubs were pervasively distributed in the WT and hACE2 systems. Of particular interest in *DC* analysis were hACE2 interface residues Gly326, Asp355, Phe356, and Arg357, which had significantly reduced centrality in all the Omicron sub-lineages compared to the WT. In the RBD-hACE2 complex [414], hACE2 residues Asp355 and Arg357 interact with RBD residue Thr500 via H-bonds and van der Waals interactions, respectively. The loss of *DC* means reduced residue interactions at these positions consequently affecting the RBD-hACE2 interplay in the Omicron sub-lineages.

On the other hand, *KC* analysis identified Ile402, Tyr508, and Val510 as *persistent hubs* in the RBD, and like in *DC*, most Omicron sub-lineages had fewer *KC* hubs than the WT i.e., WT: 13 hubs, BA.2: 10, BA.3_10: 9, BA.3_12: 8 and BA.3_15: 9 (**Figure S24**). Furthermore, the RBD interface residues, Asn501 and Gly502 that are involved in RBD-hACE2 inter-protein interactions, lost the *KC* hub status in the sub-lineages.

In the hACE2 protein, Ala25, Leu97, Ala403, Ile407, Gln522, Phe525, Leu529, and Ala550 were the *persistent hubs*. *DC* and *KC* hub distribution was somewhat the same in the RBD-hACE2 complex, as both metrics derive their centrality from neighbor connectivity [523].

In summary, the RBD-hACE2 network analysis highlighted the effects of Omicron sub-lineage mutations on the inter-protein communication patterns and interactions at the interface. Furthermore, the RBD mutations seem to have an allosteric effect on the hACE2 behavior through influencing centrality distribution in the functional residues. Based on the dynamics of the RBM (section 4.4.2), the inter-protein interaction distance (section 4.4.5.2),

and the distribution of residue centrality of both the RBD and hACE2 interface residues, it is evident that the Omicron sub-lineage mutations affect RBD-hACE2 interactions. The next section delves further into the complex interaction changes due to Omicron mutations.

4.4.6 Effect of the RBD Omicron Mutations On RBD-hACE2 Inter-protein Interactions

The RBD-hACE2 global dynamics highlighted a highly dynamic RBM which accommodates majority of residues involved in inter-protein interaction (**section 4.4.2**). This observation was reenforced by *CC* analysis which identified fewer hubs at the sub-lineage interface region compared to the WT probably due to the RBM dynamics. Additionally, the BA.3_12 system, which had fewer RBD-hACE2 interface *CC* hubs compared to the WT displayed higher inter-protein interaction distance than the WT.

Motivated by these observations, we next analyzed the interaction frequency between RBD and hACE2 interface residues through contact map analysis. The RBD interface residues were identified using the ROBETTA web server [352] as: Lys417, Ile418, Tyr449, Tyr453, Leu455, Phe456, Ser477, Phe486, Asn487, Tyr489, Gln493, Gln498, Thr500, Asn501 and Val503. Their respective contacts were determined using the *contact_map.py* Python script from the MDM-TASK web server [264, 265]. For comparative purposes, the interaction frequency for each sub-lineage RBD-hACE2 interface residue pair was subtracted from that of the WT (WT contact frequency – sub-lineage contact frequency) and presented as a heat map (**Figure 4.10**).

The heat map, **Figure 4.10A**, highlights the differences in the inter-protein residue contact frequencies between the WT and the respective Omicron sub-lineages based in the color spectrum. Of interest, was the general decrease in the inter-protein residue contact frequency in the Omicron sub-lineages compared to the WT for the following RBD-hACE2 residue pairs, Gln493-Glu35, Gln498-Tyr41, Thr500-Asp30, Asn501-Lys353, Asn501-Gly354,

Asn501-Asp355, Val503-Gln325, Tyr505-Lys353, and Tyr505-Gly354. Majority of the lost/reduced contact frequencies observed here are at the mutation positions which explains the changes. In turn, compensatory gains at the mutated positions were observed at the following RBD-hACE2 residue pairs, Gly447-Met82, Thr500-Thr324, Thr500-Gly326, Asn501-Lys353, Asn501-Asp355, Asn501-Gly354, Tyr505-Gly354 and Tyr505-Lys353, especially in BA.4. Furthermore, the BA.4 system didn't show any decrease in the RBD-hACE2 interactions for Leu455-Asp30, Phe456-Thr27, Phe456-Asp30, Gln493-His34, and Thr500-Asp355 compared to the other Omicron sub-lineages. Quantitatively, the BA.4 system also maintained the highest number of RBD-hACE2 interactions compared to all the other systems i.e., BA.4: 32 > BA.3_15: 30 > BA.3_10: 30 > BA.3_12: 26 > BA.2: 25 > WT: 24 > BA.1: 22 (**Figure 4.10B**).

Many studies have shown that the RBD N501Y mutation is associated with strengthening RBD binding to hACE2 [196, 515, 516]. The results here were in agreement with these studies as more interactions were noted for Tyr501 in the Omicron sub-lineage systems compared to Asn501 in the WT i.e., WT: **3** (Lys353, Gly354, Asp355); BA.1 and BA.2: **5** (Tyr41, Leu351, Lys353, Gly354, Asp355); BA.3_10: **6** (Tyr41, Gly326, Gly352, Lys353, Gly354, Asp355); BA.3_12: **4** (Tyr41, Lys353, Gly354, Asp355); BA.3_15: **4** (Leu45, Gly352, Lys353, Asp355) and BA.4: **6** (Tyr41, Gln325, Gly352, Lys353, Gly354 and Asp355). It is important to note that not all the Omicron sub-lineages RBD are associated with efficient receptor binding. A couple of mutations especially at the antibody binding sites and are aimed at antibody escape.

RBD-hACE2 interface residues interaction analysis shows that the RBD mutations result in enhanced RBD-hACE2 interactions especially with the progression of the Omicron sub-lineages. The latest Omicron sub-lineage, BA.4 (as of April 2022), had the highest number of RBD-hACE2 interactions than the WT and all the other Omicron sub-lineages. This

observation, coupled with the DRN analysis, highlight the progressive effect of the Omicron sub-lineages towards more efficient receptor binding in the Omicron sub-lineages.

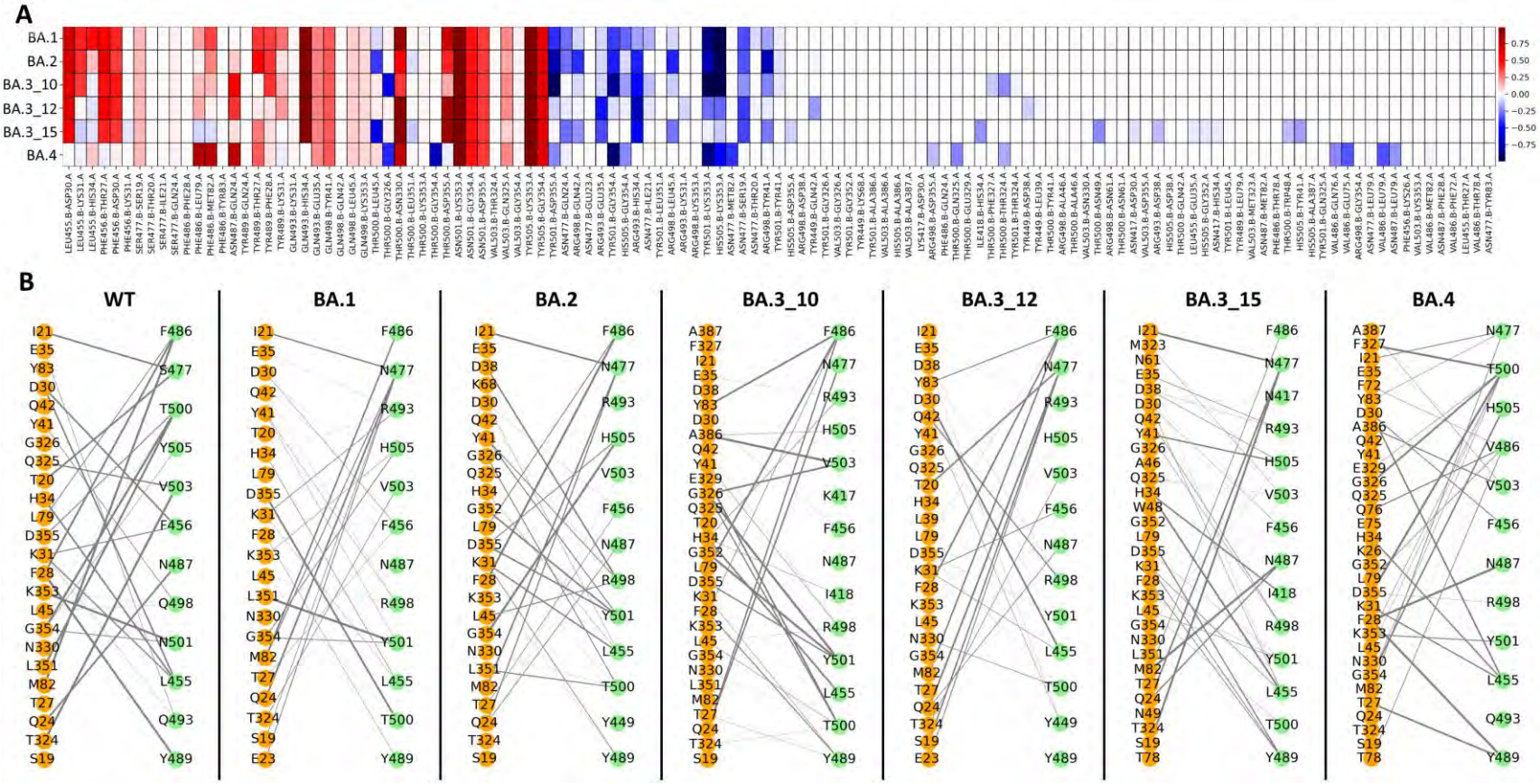

Figure 4. 10: A shows the delta heat map of the RBD-hACE2 residue pair contact frequencies between the WT and Omicron sub-lineages (WT contact frequency – Omicron sub-lineage contact frequency). Blue implies higher contact frequency in the sub-lineage system, white means no contact difference and red implies reduced contact frequency in the sub-lineage system compared to the WT. B shows the bipartite interaction plots between the RBD (green) and hACE2 (orange) residues. The thickness of the edges corresponds to the contact frequency. Adapted from Barozi et al., **[271]**.

4.5 Chapter Conclusion

Chapter 4 describes the global and residue level effects of progressive Omicron RBD mutations (BA.1, BA.2, BA.3_10, BA.3_12, BA.3_15 and BA.4) on RBD-hACE2 dynamics, interaction, and communication patterns in relation to the WT (Wuhan RBD-hACE2).

Comparative global analysis using RMSD, RMSF, and Rg identified a more dynamic RBD in the Omicron sub-lineages compared to the WT. Of particular interest was that the RBM, residues 438-506, which contains majority of the RBD interface residues involved in RBD-hACE2 interactions. This region experienced significantly higher residue fluctuations in the Omicron sub-lineages compared to the WT. Furthermore, known antigenic sites in this region registered significantly high residue fluctuation in the Omicron sub-lineages compared to the WT, all of which point to neutralizing antibody escape mechanisms. The RBD-hACE2 DCC calculation further implicated the RBM dynamics as influencing complex dynamics resulting in anti-correlated motions, especially in BA.1, BA.2 and BA.3_12 Omicron sub-lineages. DCC also showed that the RBD dynamics influence the conformational changes in the hACE2 protein system.

The chapter also documents changes in the RBD-hACE2 inter-protein interaction distance as described by COM distance measurement, which links *CC* hub distribution at the interface to the proximity of protein interaction. BA.3_12, with the least number of *CC* hubs, had the highest RBD-hACE2 inter-protein interaction distance.

Furthermore, residue level analysis focused on the RBD-hACE2 interaction patterns highlighted an increase in the inter-protein interactions in the Omicron sub-lineages compared to the WT. The RBD of the BA.4 Omicron sub-lineage displayed more interactions with hACE2 than all the other sub-lineages and the WT.

Interestingly, DRN analysis through *BC* highlighted two novel high centrality residue communication paths connecting the RBD to the hACE2. The most prominent, Path I, originated from the RBD core, through the RBD interface and hACE2 sub-domain I to the zinc-binding region of the human receptor. Identification of these residues goes a long way in isolating key residue targets for RBD binding inhibitor studies.

More importantly, each Omicron sub-lineage ensemble of mutations uniquely affected the consistency and integrity of the communication paths; however, the BA.4 system displayed similar path characteristics to the WT. The BA.4 sub-lineage is unique to the other Omicron sub-lineages as given its reversion of mutations at positions 446, 496 and 493 back to WT sequence residues. This reversion could imply an explorative viral evolution for a variant of best fit.

Unique high *EC* allosteric paths were also observed connecting the RBD and hACE2 in the WT some of which were lost in some Omicron sub-lineages. Previous studies have indicated that RBD binding to hACE2 enhances the receptor peptidase activity by 10-fold [497] compared to the reference complex. Hence, the observed changes in residue centrality reflect the collective effect of the sub-lineage mutations on hACE2 functionality.

In summary, the chapter highlights the effects of progressive Omicron sub-lineages, tailored towards a more effective RBD-hACE2 complex, elusive to targeted therapy. Additionally, network analysis identified critical residues in the RBD-hACE2 communication with potential of influencing viral binding.

CHAPTER 5

5. Effects of Naturally Occurring African hACE2 Polymorphisms on the BA.4/5 Omicron Sub-lineage RBD-hACE2 Interaction: A Dynamic Residue Network Analysis

This chapter reports on the DRN analysis of the polymorphism harboring angiotensin converting enzyme-2 in complex with the BA.4 Omicron sub-lineage receptor binding domain. The chapter focuses on the effect of the naturally occurring hACE2 polymorphisms on RBD-hACE2 interactions and dynamics.

Contribution: Polymorphism identification, homology modeling, MD simulations, trajectory analysis and all data analysis including script writing was done by Victor Barozi. The scripts used for DRN top x% calculation and heat maps were generated by Olivier Sheik Amamuddy. Hub visualization scripts in PyMOL were generated by Olivier Sheik Amamuddy, Victor Barozi and Özlem Tastan Bishop

5.1 Introduction

In chapter 4, we discussed the prevalence of COVID-19, etiology, transmission and the mutation effects of the S protein RBD, which acts as the mediator for viral binding and fusion [414]. The chapter also introduced the hACE2 as the SARS-CoV-2 receptor in the human host. In this chapter, we delve into the structure, function and effects of naturally occurring hACE2 polymorphisms on SARS-CoV-2 interaction.

5.1.1 The hACE2 and Receptor Structural Architecture

The hACE2 is a transmembrane I protein made up of a membrane bound and extracellular region [498]. The receptor is expressed in the kidney, heart, liver, colon, small intestines, lungs and the testis [524–527] where it is involved in blood pressure homeostasis under renin angiotensin systems (RAS), discussed further in **section 5.1.2**.

Structurally, the hACE2 extracellular region consists of 1) a metallopeptidase domain (residues 19-615) which harbors a zinc ion and has a ~42% sequence identity to the catalytic domain of ACE, and 2) a collectrin domain (residues 615-740) with ~48% sequence identity to the human collectrin. The metallopeptidase domain which is also the catalytic domain is further sub-divided into two sub-domains, I and II at the N and C-terminal ends of the protein, respectively [498] (**Figure 5.1**). Structurally, sub-domains I and II form two sides of a cleft, inter-connected at the base by an α-helix (residues 511-531). The groove formed by the sub-domains is partly the substrate binding site. Furthermore, the hACE2 is homologous to ACE and both contain a conserved motif, HEXXH + E motif, whose residues coordinate the zinc at the catalytic domain [498, 528, 529]. The X-ray crystal structure of hACE2 (PDB ID: 1R42) identifies His374, His378 and Glu402 as the zinc coordinating residues in the sub-domain II of the catalytic domain. Furthermore, the receptor also harbors a chlorine ion in sub-domain II, which it is coordinated by Arg169, Trp447 and Lys481.

The peculiar design of the hACE2 substrate binding cleft is believed to prevent the hydrolysis of correctly folded and functional peptides [326, 327]. Furthermore, substrate binding to hACE2 causes ~16° hinge movement of sub-domain II towards sub-domain I [498].

Additionally, the hACE2 N-terminal α-helix also binds both SARS-CoV-2 and SARS-CoV [532] via the S protein RBD. The cascade of events following SARS-CoV-2 binding to hACE2 has previously been discussed in **chapter 4**.

5.1.2 Physiological Function of hACE2

Human ACE2 a type I transmembrane protein that belongs to the carboxypeptidase family of enzymes and is predominantly expressed in the kidney, heart, liver, colon, small intestines, lungs and the testis [524–527]. The carboxypeptidase activity of hACE2 involves hydrolysis of a single peptide at the C-terminal end of the substrate [527, 529]. Protein/substrate hydrolysis by the hACE2 is essential in catabolism, post-translational protein modification and regulation of biological processes. Some of the known hACE2 substrates include apelin-13, apelin-36, des-Arg9-bradykinin and angiotensin II [498, 528, 529, 533]. The hACE2, like the ACE is part of the renin-angiotensin systems (RAS) which is responsible for blood pressure homeostasis through modulation of blood volume, sodium retention, vascular tone and production of inflammatory factors [534, 535]. In the RAS, hACE2 predominantly decomposes Ang II to Ang (1-7) and Ang I to Ang (1-9). Besides being a vasodilator, Ang (1-7) also improves coronary perfusion, lowers blood pressure and inhibits cardiomyocyte hypertrophy, which is the enlargement of individual heart muscle cells [534]. Despite acting on both Ang I and II, hACE2 decomposes Ang II more efficiently than Ang I [534]. Contrary to the hACE2 activity, its homolog ACE activates Ang I to Ang II, which is a vasoconstrictor, therefore, hACE2 and ACE are antagonistic and together they regulate the RAS.

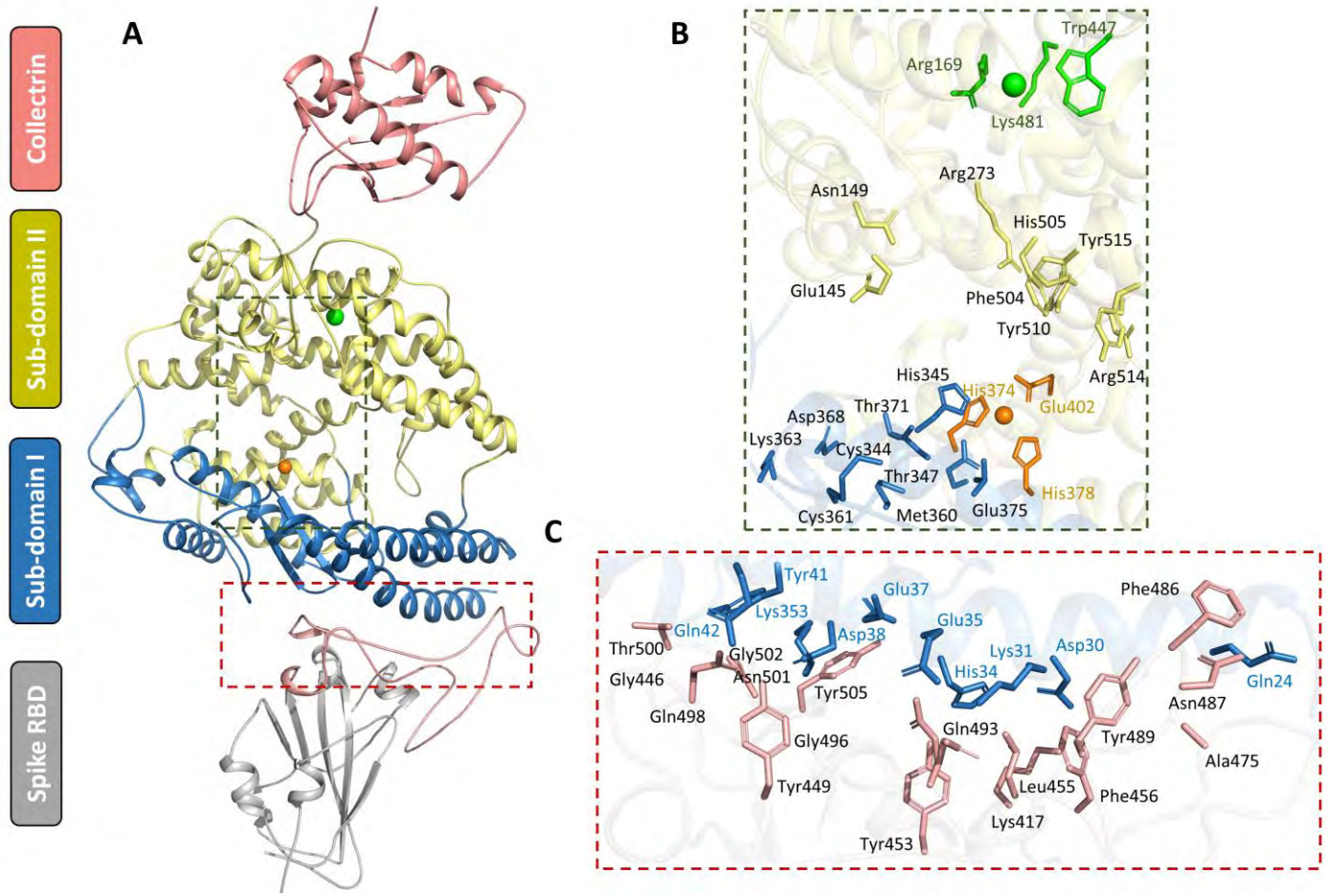

Figure 5. 1: A shows the cartoon presentation of the extracellular domains of the hACE2, and S RBD. The RBD receptor binding motif (RBM) is shown in salmon. The zinc (orange sphere) binding residues (orange sticks) in the catalytic domain and the chloride (green sphere) coordinating residues (green sticks) in sub-domain II are highlighted in B. C shows the hACE2 (blue sticks) and RBD (salmon sticks) interface residues involved in inter-protein interaction.

Due to the activity of hACE2 on Ang II and blood pressure homeostasis, its expression and is important in the cardiovascular function and homeostasis. Deficiency of hACE2 activity through inhibitors or disease associated polymorphisms is linked to cardiovascular diseases [536–538].

5.1.3 Naturally Occurring hACE2 Polymorphisms and SARS-CoV-2 Activity

Despite the critical physiological role of hACE2 in blood pressure homeostasis, the *ACE2* gene still undergoes genetic variation leading to insertions/deletions and SNPs in the hACE2 protein sequence [535, 537, 539]. Previous studies have linked hACE2 polymorphisms to elevated cardiovascular risk, obesity and hypertension [537, 538, 540–542].

Furthermore, different studies have also shown an association between the hACE2 polymorphisms and the variability of infection and manifestation of the recently discovered SARS-CoV-2, a virus responsible for the COVID-19 global pandemic [207, 212, 213, 543–545]. Hu et al., 2022, through transduction studies showed that some hACE2 polymorphisms i.e., S19P, I21V, E23K, K26R, K31R, T27A, N33I, H34R, D38V, Y83H, T92I, V184A, S257N, G326E, G352V, I468V and N368S modulate SARS-CoV-2 infectivity [213]. Similarly, Suryamohan et al., used biochemical assays to show a higher binding affinity for the S RBD in the hACE2 bearing the K26R and T92I polymorphisms [207]. *In silico* studies have indicated that certain hACE2 polymorphisms influence the S protein RBD binding to the hACE2 [204–206, 215]. Notwithstanding other factors influencing the SARS-CoV-2 infectivity i.e., age, sex, existing comorbidities and season, the hACE2 genetic variations remain a key factor in viral infectivity. Based on the hACE2 involvement in SARS-CoV-2 binding, and the effect of hACE2 polymorphisms on cardiovascular function, blood pressure and vascular tone, there is a need for more research on hACE2 polymorphisms and their effects on protein function especially amidst the ever-evolving SARS-CoV-2 variants.

5.2 Chapter Aims and Objectives

The variation of the hACE2 genome among individuals and populations holds an important piece of the puzzle to explaining the inter-individual and inter-population variability in SARS-CoV-2 infectivity and severity. With the emergence of new SARS-CoV-2 variants and accrual of RBD mutations, an understanding of the RBD-hACE2 interactions is required especially in relation to the hACE2 polymorphisms. This chapter therefore focuses on characterizing the network properties of the RBD-hACE2 interactions amidst both the BA.4 RBD Omicron sub-lineage mutations and naturally occurring hACE2 African polymorphisms. Here, the influence of hACE2 polymorphisms on Omicron BA.4/5 RBD interaction were accessed through i) identification of the hACE2 polymorphisms and Omicron BA.4 sub-lineage mutations, ii) modeling of the BA.4/5 RBD - variant hACE2 complex 3D structures, iii) MD simulation of the BA.4/5 RBD - WT hACE2 and BA.4/5 RBD- variant hACE2 complexes and iv) post-MD trajectory analysis and DRN analysis.

5.3 Methods and Materials

5.3.1 Retrieval of hACE2 Polymorphisms and Omicron BA.4/5 RBD Mutations

The naturally occurring hACE2 polymorphisms were retrieved from the genome aggregation database (gnomAD) [546] version 2.1.1. Here, missense SNPs from an African/African American population with an allele frequency of $\geq 2.7 \times 10^{-4}$ were selected. The gnomAD contains exome and genome data from large-scale sequencing projects.

For the Omicron BA.4/5 sub-lineage RBD mutations, GISAID [157] database was queried for high coverage Omicron BA.4/5 sequences of human origin and with patient status. It is important to note that the BA.4 and BA.5 Omicron sub-lineage sequences share the same S protein mutations [201, 477, 478].

Furthermore, the hACE2 SNPs and BA.4 Omicron RBD mutations were introduced into the RBD-hACE2 complex in PyMOL [488]. Here, seven protein structures were created, i.e., BA.4 RBD-WT hACE2 (reference system), BA.4 RBD-S19P hACE2, BA.4 RBD-K26R hACE2, BA.4 RBD-M82I hACE2, BA.4 RBD-K341R hACE2, BA.4 RBD-N546D hACE2 and BA.4 RBD-D597Q hACE2. The generated structures were validated using VERIFY3D [247] and ProSA [248] before any further analysis.

5.3.2 All-Atom Molecular Dynamic Simulation

MD simulations were set up using specific environmental conditions before applying force fields. Here, titratable residues of the modeled structures were first protonated at pH 7.2 using the PROPKA tool from PDB2PQR (version 2.1.1) [316]. With protonation complete, *top* and *gro* files were generated for each complex system using the GROMOS54a7 force field [317] from GROMACS [315] version 2019.4. The structures were placed in a cubic box of 1 nm clearance distance before solvation using SPC216 water model [257]. Subsequently, the

system charge was neutralized by adding NaCl ions of 0.15 M concentration to enable long-range electrostatics calculation using the PME method. Once neutralized, structural steric clashes were eliminated, and dihedral angles corrected through structure minimization using the steepest descent algorithm with no constraints and an energy step size of 0.01 until an energy threshold of 1000.0 kJ/mol/nm. Minimization was followed by temperature (NVT) equilibration at 300K and pressure (NPT) equilibration at 1 atm for 100 ps each using Berendsen temperature coupling and Parrinello–Rahman barostat [318], respectively.

The equilibrated systems were subjected to 400 ns all-atom MD simulations with a time step of 2 fs. Here, like in chapter 4, the bonds were constrained under the LINCS algorithm [319]. Furthermore, the long-distance electrostatics were computed under the PME electrostatics [320] using a Fourier spacing of 0.16 nm. For the short-range coulomb and van der Waals interactions, a threshold distance of 1.4 nm was used.

Post MD trajectory analysis followed the same techniques, steps and tools as discussed in **chapter 4, section 4.3.2**.

5.3.3 Dynamic Residue Network Analysis

Since the same reference protein structure, PDB ID: 6M0J, was used here as in chapter 4, the same DRN protocol; **chapter 4 section 4.3.5** was used here.

5.3.4 Contact Map Analysis

The MDM-TASK web server [264, 265] was used to calculate RBD interface and *EC* hub residue contact frequencies using the same protocol as described in **chapter 4 section 4.3.6.** For the *EC* hubs, the network and centrality distribution was plotted using the NetworkX spring layout [329].

5.4 Results and Discussion

5.4.1 Distribution and Characteristics of the hACE2 Polymorphisms

The search for naturally occurring hACE2 polymorphisms in African population from gnomAD yielded six SNPs with an allele frequency of ≥ 2.7 x 10^{-4} i.e., S19P, K26R, M82I, K341R, N546D and D597Q. An allele frequency of ≥ 2.7 x 10^{-4} equated to an allele count of 5 and above per given population in gnomAD. Half of the SNPs i.e., S19P, K26R and M82I are positioned in the α-helix of sub-domain I at the hACE2 interface (**Figure 5.2**). The hACE2 interface region interacts with the RBM of the S RBD, where hACE2 residues 19, 26 and 82 are involved in inter-protein interactions. K341R is positioned in sub-domain I away from the interface region. The other SNPs; N546D and D597Q are located in sub-domain II region of the protein, where D546D forms part of the hinge axis between the two sub-domain clefts. All the SNPs are distal to the active site and zinc binding region however, position 546 has been previously identified as a glycosylation spot [414]. Furthermore, all the hACE2 SNPs within the scope of this study involved similar physicochemical residue substitutions, meaning that the residue properties in the protein were maintained.

For the RBD, the same Omicron BA.4 sub-lineage RBD mutations investigated in chapter 4 were retrieved from GISAID [157]. Here, the BA.4 and BA.5 mutation search yielded the same 17 RBD mutations i.e., G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, L452R, S477N, T478K, E484A, F486V, Q498R, N501Y and Y505H [201, 477, 478]. The L452R and F486V mutations are unique to BA.4 when compared to BA.2 Omicron sub-lineage. As discussed in chapter 4, both L452R and F486V are linked to neutralizing antibody escape [485].

For the rest of this chapter, the BA.4 RBD-WT hACE2 (reference) system will be referred to as WThACE2 whereas, BA.4 RBD-S19P hACE2 as S19P, BA.4 RBD-K26R hACE2 as

K26R, BA.4 RBD-M82I hACE2 as M82I, BA.4 RBD-K341R hACE2 as K341R, BA.4 RBD-N546D hACE2 as N546D and BA.4 RBD-D597Q as D597Q.

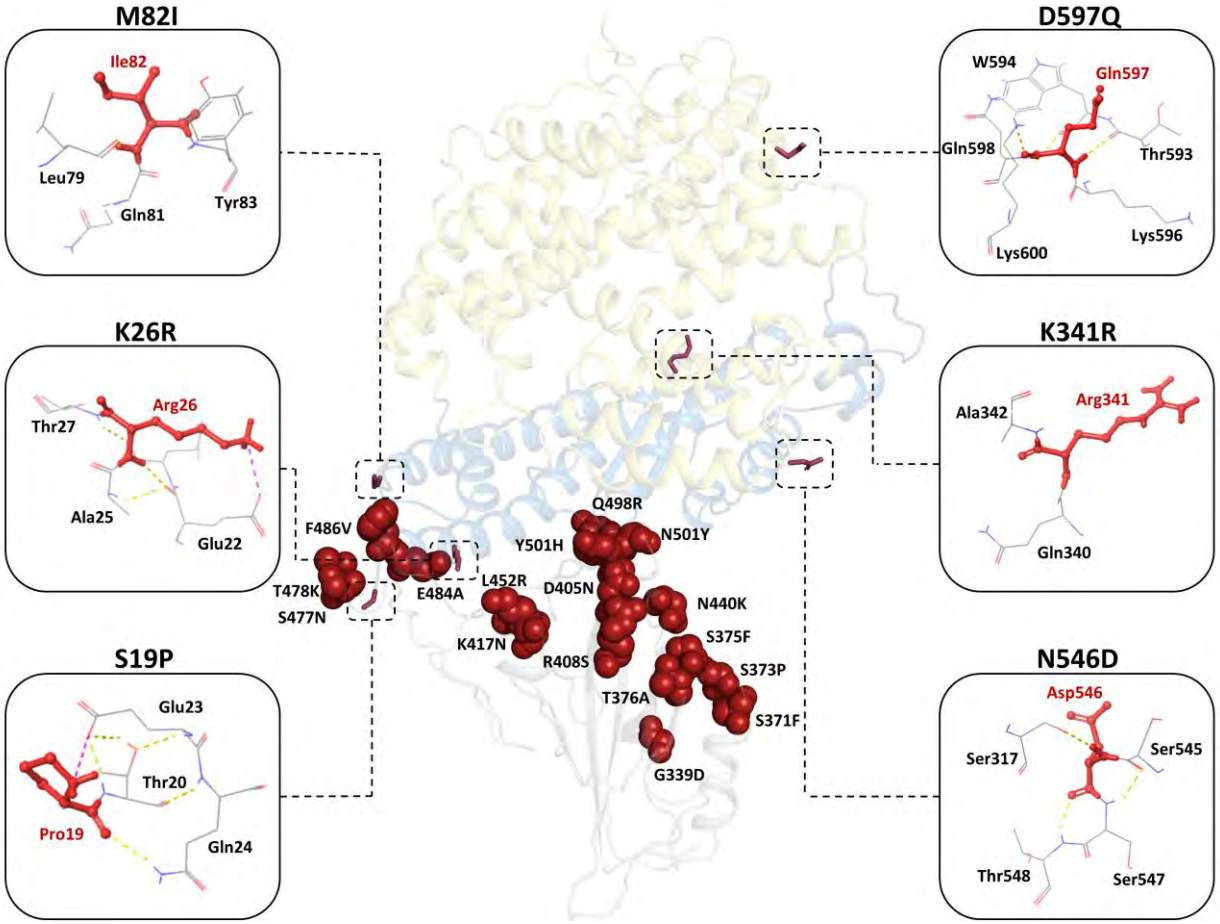

Figure 5. 2: Cartoon representation of the RBD-hACE2 complex showing the distribution of the Omicron BA.4 RBD mutations (firebrick spheres) and hACE2 SNPs (firebrick sticks). The subplots highlight the hACE2 mutated residue interactions within 3 Å. H-bonds, Pi-Pi stacking interactions and salt bridges are shown in yellow, blue and purple, respectively.

5.4.2 Global Effects of BA.4 RBD and hACE2 SNPs on RBD-hACE2 Dynamics

To assess the global mutation effect, trajectories from the MD runs were analyzed using RMSD, RMSF and Rg. For the reference system (WThACE2), duplicate MD runs were computed to check for reproducibility. RMSD line plots of the duplicate WThACE2 runs showed strong agreement between the reference structures with an RMSD range of 0.1 Å between them (**Figure S25A**). Furthermore, comparative RMSD line plots of the WThACE2 (run 1) and hACE2-SNP bearing systems indicated that the WThACE2 system converged within the 400 ns simulation (**Figure S25B**). The mutational effects on individual protein dynamics were investigated through per system RMSD calculations i.e., RBD and hACE2 separately. Here, the RBD RMSD violin plots showed similar RMSD distribution for the duplicate reference system runs where, two closely related conformations (based on data distribution) were sampled in each run (**Figure 5.3A**). For S19P, M82I, K341R, N546D and D597Q the RBD RMSD distribution comprised of one main conformation and other small ones attributed to pre-equilibration dynamics.

In the case of the hACE2, all the systems explored a unimodal RMSD distribution with the exception of M82I where at least two main conformations were sampled throughout the MD simulation.

Furthermore, RMSF calculations showed trivial differences in residue flexibility between the WThACE2 and hACE2 SNP bearing systems for both the RBD and hACE2 proteins. The implication of this being that the phytochemically similar hACE2 polymorphisms do not dramatically change the complex global dynamics (**Figure 5.3C and D**). Rg analysis also showed minimal differences between the WThACE2 and hACE2 polymorphs in both the RBD and hACE2 proteins (**Figure S25C and D**).

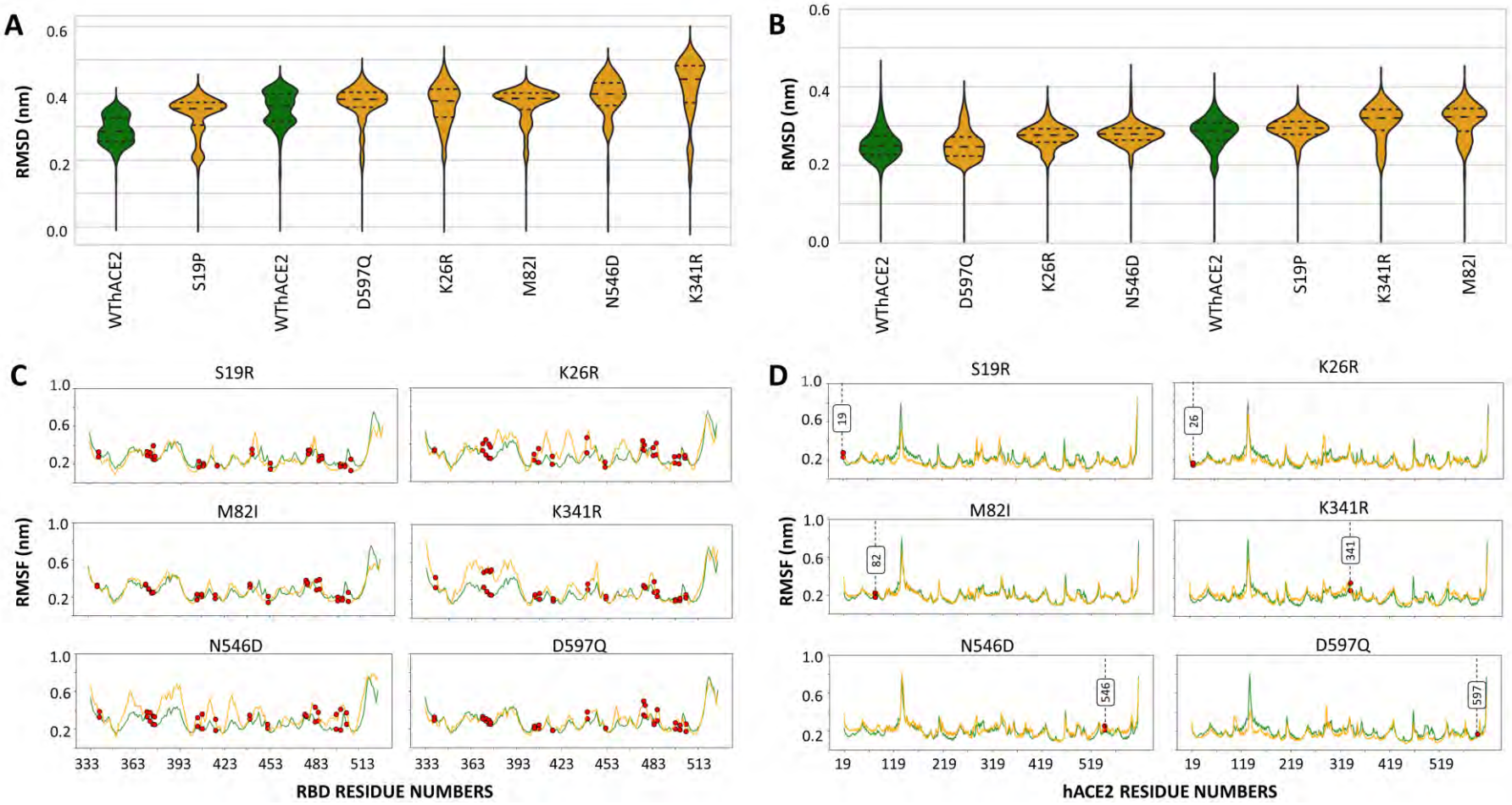

Figure 5. 3: A and B show violin plots of the RMSD distribution in the RBD and hACE2 systems, respectively. The WThACE2 is in green and hACE2 systems in orange. The duplicate WThACE2 runs are labelled. RMSD violin plots are arranged in ascending order of median RMSD. C and D show the comparative RMSF line plots between the WThACE2 (green) and hACE2 systems (orange) for the RBD and hACE2 respectively. SNP positions in the RMSF plots are indicated with red markers in both proteins.

5.4.3 RBD-hACE2 DRN Analysis

In **chapter 4**, we introduced protein network analysis of the RBD-hACE2 complex using graph theory principles as applied through DRN analysis. The approach characterized the relationship between the S RBD and hACE2 proteins through, *BC*, *CC*, *DC* and *EC* analysis. *BC* analysis particularly showed two high centrality residue paths bridging the RBD and hACE2 proteins through which we hypothesize fast and efficient communication occurs.

The same principles were applied in this chapter while trying to decipher the effects of the hACE2 polymorphisms on the RBD-hACE2 communication patterns in the presence of Omicron BA.4/5 RBD mutations. Here, two approaches were followed for DRN analysis i.e., the local/per protein system analysis and the global/ensemble centrality analysis.

The local/individual protein system DRN analysis focused on individual protein centrality distribution. Even though the global DRN approach was informative as proven in chapter 4, the high centrality cut-off value is dictated by the ensemble centrality which in some cases excludes some system-specific high centrality residues. To account for this, the top 5% (RBD) and 4% (hACE2) high centrality residues were determined for each individual system separately. This approach further highlighted the inter-system differences in centrality.

For the global perspective, the same approach as in **chapter 4** was employed, where the ensemble top 5% centrality cut-off was used for the RBD, and 4% for the hACE2 ensemble based on the protein sizes. Heat map representation of the data aided in the inter-protein centrality comparability.

5.4.3.1 Local/per System DRN Analysis

5.4.3.1.1 Betweenness Centrality

WThACE2 duplicate runs from MD simulations were both subjected to DRN *BC* analysis as a measure of reproducibility. Subsequently, the WThACE2 runs had similar *BC* distribution as shown in **Figure 5.4A**. Furthermore, system specific *BC* analysis for the RBD and hACE2 proteins showed similar centrality distribution patterns across the systems i.e., the WThACE2 and hACE2 polymorphs. In the RBD, residues 396-409, 452-457 and 494-513 had high centrality values compared to other residues across the systems (**Figure 5.4A**). These residues make up part of the RBD core (β3 and β7 strands) and the RBM at the RBD interface. Existing literature suggests that the stability of protein structures is dependent on residue organization at the core [547–550]. Evidently, the Omicron RBD core was averse to mutations as none of the 17 BA.4/5 RBD mutations occurred at the protein core. On the other hand, the less hydrophobic interface region is prone to residue variation through mutations. Interestingly, the RBD regions with the lowest *BC* i.e., positions 355-390 and 465-480 were mutation hot spots with at least six Omicron BA.4/5 RBD mutations (S371F, S373P, S375F, T376A, S477N and T478K). The high centrality observed at the RBD core highlights its importance in the S protein functionality.

Interestingly, the RBD heat map representation of the individual system high centrality residues (top 5%), **Figure 5.4B**, identified interface residues, Leu455, His505, Gln506 and Tyr508 with high centrality across all RBD systems. These residues are part of the RBD-hACE2 interface interactions which are further discussed in **section 5.4.4**. Furthermore, residue Gln506 had highest centrality of all the residues across the systems. Gln506 is positioned at the terminal end of the RBM, where it forms H-bonds with Asn437, Gly404, Gly502 and Val503. Gln506 also connects the RBM to the β7 strand. The other residue of interest was Tyr501 which had high centrality in all systems except S19P. As seen from

chapter 4, the N501Y mutation resulted in a gain in centrality for the Omicron sub-lineages compared to the WT. Here, Tyr501 maintained high centrality values in the WThACE2 however, the presence of hACE2 polymorphisms significantly curtailed the residue centrality especially in S19P. Chapter 4 also discussed the importance of the N501Y mutation especially in strengthening the RBD-hACE2 interactions, an effect that seems to be affected by hACE2 polymorphisms. The lowest centrality value was observed at the loop residue Gly526 in K341R.

Like in the RBD, the hACE2 *BC* distribution followed a similar pattern defined by specific regions of high centrality *viz*: 23-26, 72-135, 177-210, 347-364, 371-423, 436-468 and 509-527 (**Figure 5.5A**). The top 4% centrality classification identified Asp355, Phe356, Ile379, Arg518 and Thr519 with high *BC* values across all hACE2 systems (**Figure S26**). It is evident from residue mapping that the hACE2 core, consisting of interconnected α-helices, is important for the carboxypeptidase activity of the protein (**Figure 5.5B**). Additionally, the high centrality hACE2 regions also makeup the highly specific receptor binding cavity. WThACE2 interface residue Phe356 had the highest *BC* value in the hACE2 systems. This residue is implicated in RBD interaction [205] and binding of hACE2 inhibitors [551, 552].

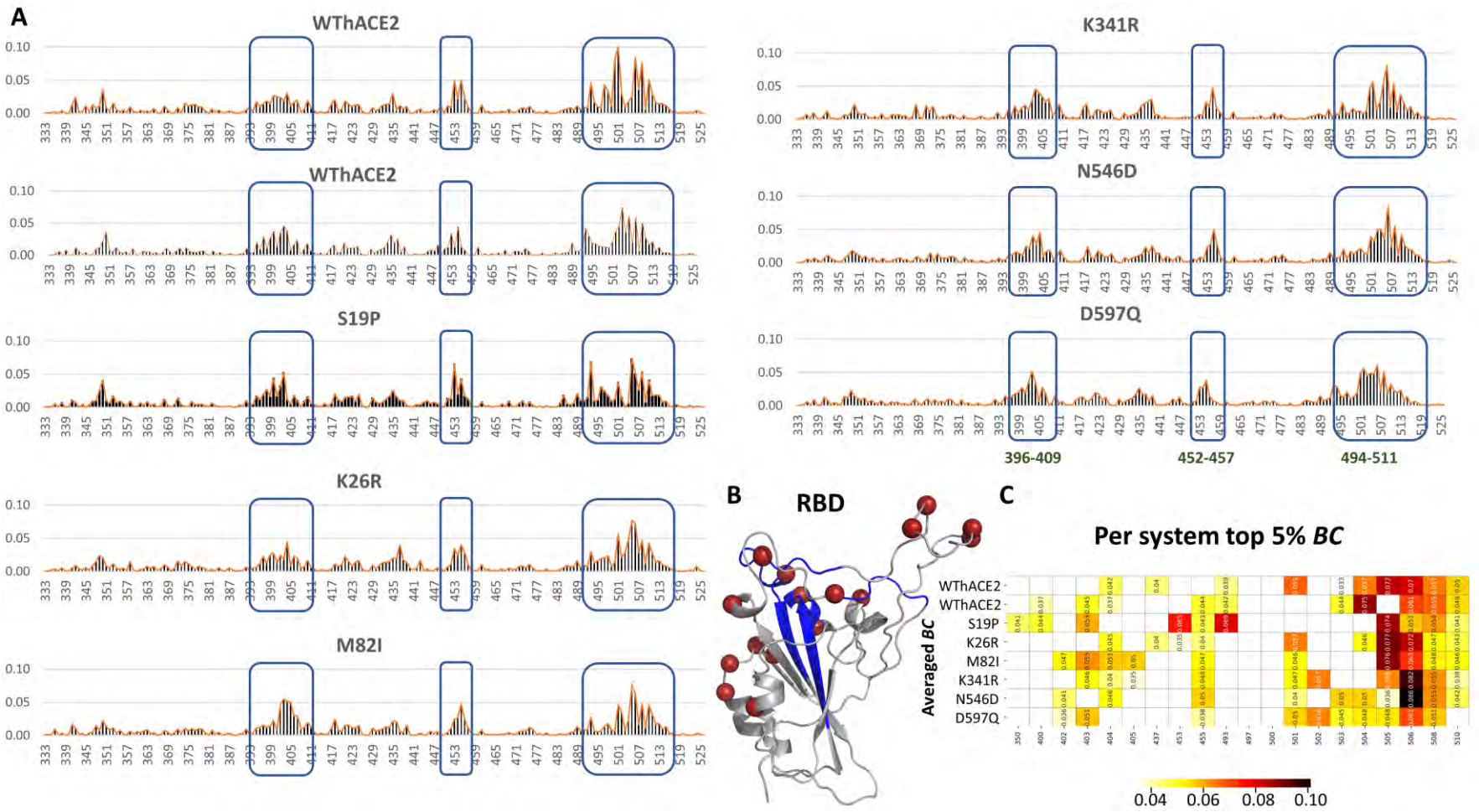

Figure 5. 4: A shows RBD *BC* distribution in individual protein system. Regions with high centrality are highlighted with blue boxes. Residue numbers are on the x-axis. The high centrality regions are colored in blue on the RBD structure in B. The BA.4/5 mutation positions are shown as firebrick spheres. The heat map in C shows system specific high centrality residues (annotated with centrality values).

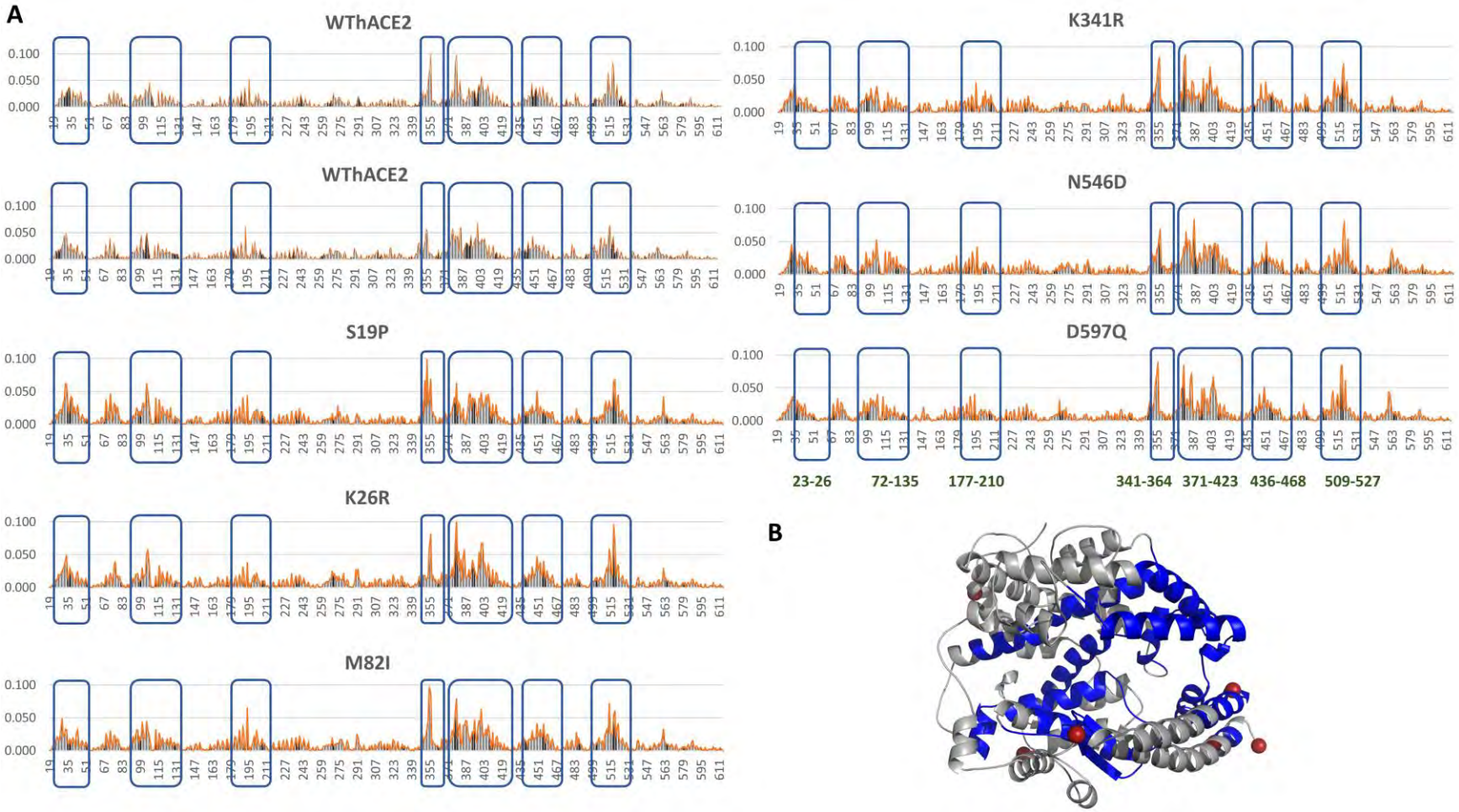

Figure 5. 5: Bar plots of the hACE2 system specific *BC* distribution are shown in A. Residues with high centrality as per top 4% are highlighted with blue boxes. The x and y-axes show the residue number and centrality values, respectively. B is the cartoon representation of the hACE2 protein marking the high centrality regions in blue.

5.4.3.1.2 Eigenvector centrality

Like in *BC*, the WThACE2 duplicate runs had similar *EC* distribution characterized by a few residue differences attributed to molecular dynamics.

Despite a fairly stable RBD as seen from the RMSF analysis (**section 5.4.2**), *eigenvector centrality* was not uniformly distributed across the mutant systems as observed in *BC*. In the WThACE2 and M82I, the high centrality regions were noted at positions 493-514. In S19P, K26R, and N546D the high *EC* regions were at positions 434-456 and 493-514. Finally, the D597Q high *EC* regions included regions 398-412, 434-456 and 493-514.

The hACE2 SNPs also had varying effects on the *EC* distribution in the RBD characterized by more regions with high centrality compared to the WThACE2 system (**Figure S27A**). All systems had high centrality values for RBD residues 493-514 which makes up part of the RBM extending to the β7 strand at the protein core. The high *BC* and *EC* values of the RBD core highlight its importance in S protein. Individually, all the systems had high centrality for residues 501-506 based on the top 5% cut-off except for S19P (**Figure S27C**). This region contains key RBD interface residues, Tyr501 and His505, that are involved in RBD binding to the hACE2.

A fairly similar *EC* distribution as seen in the RBD was observed across the hACE2 systems where the high EC was limited to specific regions i.e., the zinc binding domain region and the protein core i.e., 309-323, 367-386, 395-414, 511-536 and 546-589. Generally, *EC* ranged from 0 to 0.212 where, residues Gly377, Ala403, Val404, Gly405, Glu406, Ile407, Met408, Ser409 Arg518, Tyr521, Gln522, Phe525, Gln526, Leu529 and Ala550 maintained high centrality across all the hACE2 systems (**Figure S28**). These residues make up the α-helical structures at the base of the groove cavity encompassing the active site. Active site and

surrounding areas are expected to have high centrality values based on their importance in the protein.

5.4.3.1.2 Closeness centrality and Degree of centrality

CC was closely distributed within both the RBD and hACE2 proteins with a centrality range of 0.064-0.125 and 0.072-0.134 in the RBD and hACE2 systems, respectively. Structurally this means that both protein structures maintained a very tight folding degree across the systems. This was earlier shown by the degree of gyration (**section 5.4.2, Figure S25D).** Similarly, the range between the highest and lowest *DC* was trivial in both the RBD (range: 0.003-0.013) and hACE2 (range: 0.002-0.013), **Figure S29**.

5.4.3.2 Global DRN Analysis

As discussed in **chapter 4**, the global DRN analysis is highly effective at identification of key functional residues and also for inter-system comparability. Based on the similarity of DRN centrality distribution between the WThACE2 runs, the global analysis proceeded with one WThACE2 run.

5.4.3.2.1 The hACE2 Polymorphs Maintained High BC Paths Between the RBD and hACE2

The RBD global top 5% centrality analysis identified both unique and shared hubs between the WThACE2 and the hACE2 variant systems. Interestingly, the gains and losses in centrality between the systems weren't significant. The total number of hubs per system ranged from nine to eleven i.e., WThACE2: 9 hubs, S19P: 11, K26R: 9, M82I: 10, K341R: 9, N546D: 10 and D597Q: 10. The heat map representation, **Figure 5.6** (averaged *BC*), identified residues Tyr501, His505, Gln506 and Tyr508 as the *persistent hubs* in the RBD ensemble. Tyr501, His505 and Gln506 are located in the RBM at the protein interface where they interact with the hACE2 receptor [414, 517–519, 553, 554]. Even though Tyr501 was a

persistent hub under the global DRN analysis, local centrality analysis (section 5.4.3.1.2) showed a reduction in centrality at this residue position in the hACE2 polymorphs especially S19P compared to the WThACE2 system. Given its importance in RBD binding, the loss of centrality can be attributed to the hACE2 SNPs.

Interestingly, both the WThACE2 here and in chapter 4 (BA.4) maintained His505, Val506, Tyr508 and Val510 as *BC* hubs. The other *BC* hubs in WThACE2 included Gly404, Asn437, Gln493, Tyr501 and Gly504. These residues are positioned at the RBD interface and are involved in inter-protein interactions as previously discussed. A couple of *BC* hubs were exclusive to the mutant systems *viz*: Ile402 (M82I, N546D and D597Q), Arg403 (S19P, M82I, K341R and D597Q) and Leu455 (S19P, K26R, M82I, K341R, N546D and D597Q). Some of these residues are involved in hACE2 interactions i.e., the side chains of Arg403 extend towards hACE2 where they form salt bridges and H-bonds with Glu37 of hACE2, respectively. Similarly, Leu455 interacts with Asp30, Lys31 and His34 of hACE2. The centrality in the variants implies increased usage in these systems rather than the WThACE2.

Residues Asp355, Phe356, Ile379, Arg518, and Thr519 were identified as *BC persistent hubs* in the hACE2 systems based on the global top 4% centrality analysis (**Figure 5.6**). Asp355 is positioned at the hACE2 interface where is interacts with Asn501 from the RBD. Structurally, Phe356 is located at the hACE2 substrate binding pocket where it is known to coordinate natural flavonoids with hACE2 inhibitory characteristics [551], and the prodrug Alacepril [552]. Similarly, Ile379 and Arg518 located at the hACE2 substate binding cleft bind verteporfin and isovitexin which have potential inhibitory activity on SARS-CoV-2 [555, 556]. The identification of active site and protein binding residues that are important in the protein function using this approach further validates the effectiveness of the DRN analysis approach in addition to identifying novel residues with no documented protein functionality yet.

Like in the RBD, there were slight differences in hubs between the WThACE2 and hACE2 variant systems. The WThACE2 had 18 hACE2 *BC* hubs whereas S19P: 28 hubs, K26R: 23, N82I: 22, K341R: 24, N546D: 26 and D597Q: 26. Collectively, S19P had the most *BC* hubs and the WThACE2 the least *viz*: (RBD 11 + hACE2 28 = 39) > D597Q (10 + 26 = 36) > N546D (10 + 26 = 36) > K341R (9 + 24 = 33) > K26R (9 + 23 = 32) > M82I (10 + 22 = 32) > WThACE2 (9 + 18 = 27).

Furthermore, the allosteric paths formed by *BC* hubs connecting the RBD core to the hACE2 substrate binding pocket, formerly identified in chapter 4 (**section 4.4.5.1**), were generally maintained even in the presence of hACE2 polymorphisms. In the WThACE2, the main path, Path I, also consisted of the five highest centrality residues in each protein connecting the RBD core to the hACE2 sub-domain I. These included RBD residues Tyr501 (*persistent hub*), Gly504, His505 (*persistent hub*), Gln506 (*persistent hub*) and Tyr508 (*persistent hub*); and hACE2 residues, Phe356 (*persistent hub*), His378, Ile379 (*persistent hub*), Arg518 (*persistent hub*) and Thr519 (*persistent hub*) (**Figure 5.7**). Nominal differences in the distribution of *BC* hubs at the interface were noted between the WThACE2 and hACE2 variant and these mainly involved the interface residues.

Although the observed differences were subtle, a discernible distinction in the mean *BC* of path hubs emerged between the hACE2 variant systems and the WThACE2 systems, encompassing both the RBD and hACE2 proteins. For the RBD mean hub centrality, the values were as follows: WThACE2: 0.054, S19P: 0.052, K26R: 0.052, M82I: 0.053, K641R: 0.052, N546D: 0.050, and D597Q: 0.048. Similarly, the mean hub centrality values for hACE2 were: WThACE2: 0.061, S19P: 0.052, K26R: 0.058, M82I: 0.055, K341R: 0.053, N546D: 0.051, and D597Q: 0.057.

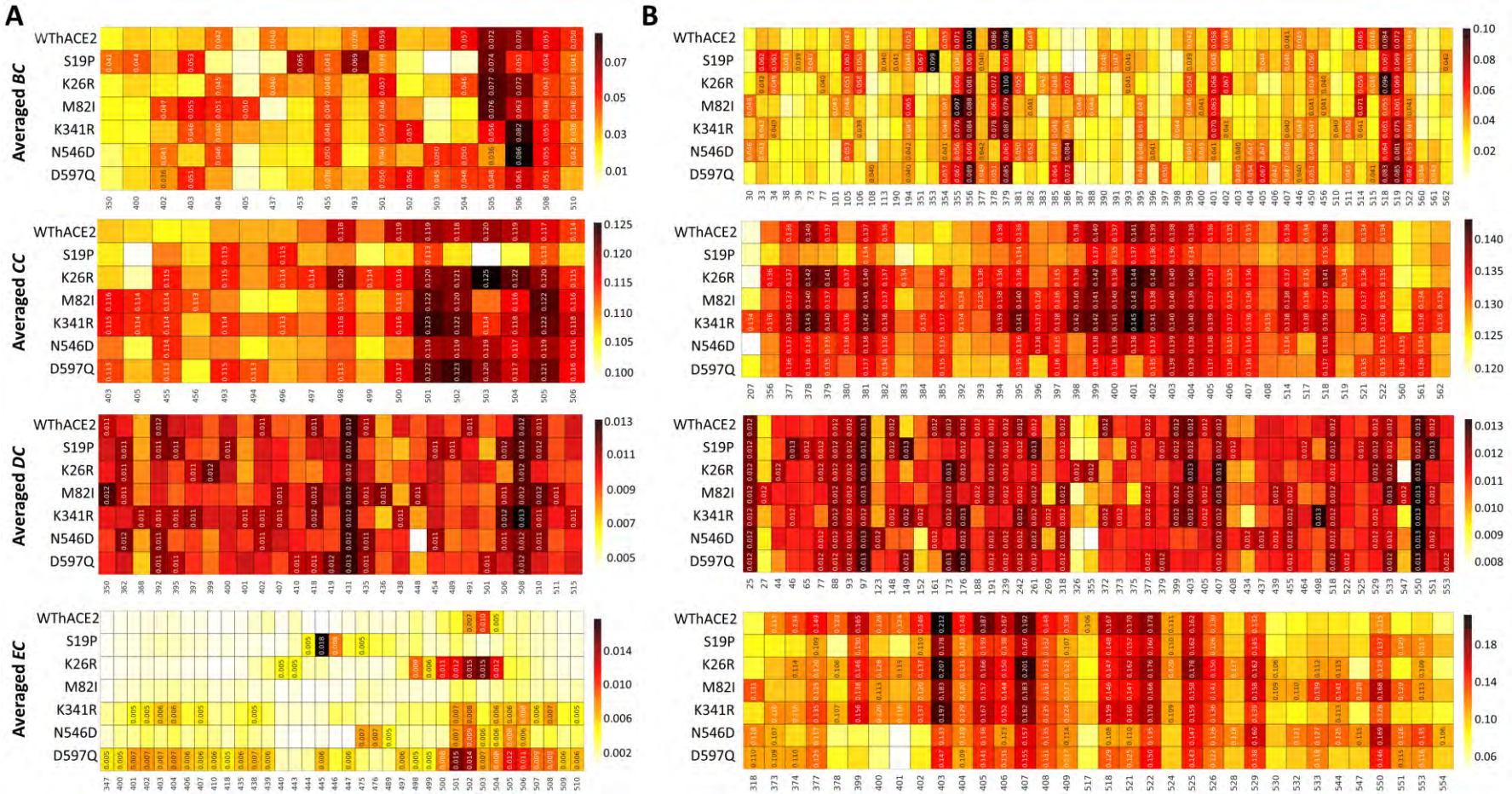

Figure 5. 6: A and B show the heat maps of the top 5% and 4% high centrality residues (hubs) across all the RBD and hACE2 proteins, respectively. The x and y-axes represent the residue numbers and protein systems, respectively. The color scale from white to dark red illustrated the degree of centrality.

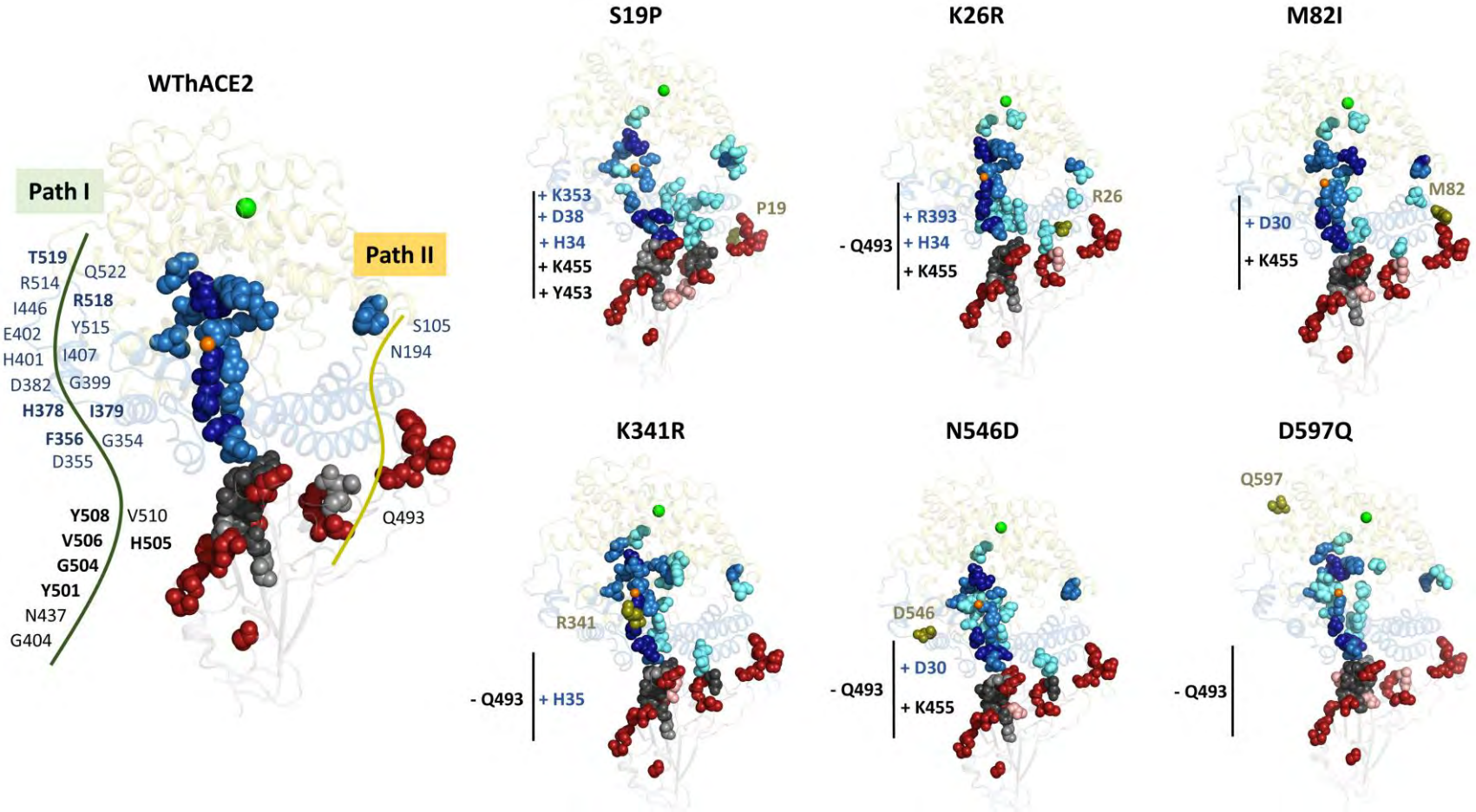

Figure 5. 7: Cartoon representation of the RBD-hACE2 complexes showing the distribution of the *BC* hubs. The RBD is shown in grey while the hACE2 sub-domain I and II in blue and yellow, respectively. WThACE2 hubs are shown as grey (RBD) and blue (hACE2) spheres, whereas the RBD mutations as firebrick spheres. Hubs unique to the hACE2 polymorphs are shown as salmon (RBD) and cyan (hACE2) spheres. SNP positions in the hACE2 are shown as olive spheres while the gained and lost variant interface hubs are annotated with + and - symbols.

The current scientific body of research on the effects of the hACE2 S19P SNP on RBD binding to hACE2 is contradictory with some researchers like Suryamohan et al., 2021 [207] and Hu et al., 2022 [213] suggesting that the SNP favours binding of the wildtype SARS-CoV-2. However, other research shows that the interface SNP reduces the hACE2 affinity for S protein characterized by high binding energies in S19P hACE2 [557, 558]. Here, *BC* analysis showed a reinforcement of the main *BC* Path I that connects the RBD to the hACE2 in S19P characterized by hub residue gains at positions Tyr453 (RBD), Lys455 (RBD), His34 (hACE2), Arg38 (hACE2) and Lys353 (hACE2). These residues are strategically positioned at the interface region to facilitate inter-protein interactions. Another important observation was the loss of *BC* for all of the zinc coordinating and chloride coordinating residues in S19P except for Lys481 (**Table 5.1**). Zinc is a critical cofactor in the metalloprotein activity as it facilitates substrate hydrolysis [559]. From this analysis, it is possible that the S19P SNP influences the interaction between the hACE2 and S RBD and might even affect hACE2 functionality. The extent to which S19P polymorphism affects RBD-hACE2 interaction is further analyzed under **section 5.4.4**.

Table 5.1: *BC* values for hACE2 zinc and chloride coordinating residues per system. The ↑ and ↓ arrows show the gain and loss in centrality, respectively compared to WThACE2 system.

Residue	WThACE2	S19P	K26R	M82I	K341R	N546D	D597Q
Arg169 (Cl)	0.009	0.009	0.010 ↑	0.008 ↓	0.010 ↑	0.014 ↑	0.014 ↑
His374 (Zn)	0.020	0.015 ↓	0.019 ↓	0.016 ↓	0.020	0.014 ↓	0.018 ↓
His378 (Zn)	0.086	0.032 ↓	0.072 ↓	0.063 ↓	0.078 ↓	0.034 ↓	0.051 ↓
Glu402 (Zn)	0.049	0.025 ↓	0.067 ↑	0.021 ↓	0.041 ↓	0.027 ↓	0.029 ↓
Trp447 (Cl)	0.023	0.018 ↓	0.019 ↓	0.017 ↓	0.021 ↓	0.019 ↓	0.018 ↓
Lys481 (Cl)	0.009	0.012 ↑	0.014 ↑	0.015 ↑	0.014 ↑	0.010 ↑	0.013 ↑

For the K26R polymorph, the zinc and chloride coordinating residues had higher *BC* compared to S19P and WThACE2 for some residues (**Table 5.1**). Furthermore, previous research has shown that the K26R mutation results in a higher receptor affinity for the RBD compared to the WT hACE2 [208, 560]. Mapping of the K26R *BC* hubs revealed both gains and losses in residue centrality at the interface compared to WThACE2 *viz*: the RBD lost Tyr453 and gained Lys455, whereas hACE2 gained His34 and Arg393 as hubs. Of these residues Tyr453 forms polar contacts with His34 of the hACE2; Lys455 forms van der Waals interactions with Asp30 of the hACE2, and His34 of the hACE2 forms polar contacts with the side of chains of Tyr453 and Leu455. Arg393 forms contact interactions with His505 of the RBD. The high centrality at these positions suggests favorable RBD-hACE2 interactions in the variants.

In the M82I system, *BC* hub gains were for interface residues Lys455 (RBD) and Asp30 (hACE2). A reduction in centrality for the RBM residues Tyr501, Gln506 and Tyr508 was noted in comparison to the WThACE2. Literature shows that the hACE2 M82I polymorphism disrupts receptor interaction with S RBD at the interface [210], and previous docking studies have showed a reduced affinity for S RBD compared to the WT hACE2 [206]. Like in the WThACE2, S19P and K26R the highest centrality residues in M82I linked the RBD core to the hACE2 zinc domain.

The other variants, K341R, N546D and D597Q all registered gains and losses in *BC* at the interface region. Despite being distal to the active site region, these variants also showed a general reduction in *BC* for the zinc and chloride coordinating residues compared to the WThACE2. In contrast, the chloride coordinating residues, Arg169 and Lys481 had high centrality in all systems compared to the WThACE2. It is likely that the hACE2 polymorphisms have allosteric effects on the protease centrality distribution especially at the active site.

5.4.3.2.2 The hACE2 Polymorphs had more Interface CC hubs Compared to WThACE2

Residues with high *CC* values act as information dissemination points and are usually clustered at the protein interface or core. The RBD global *CC* analysis showed a rather scarce distribution of hubs at the RBD core and interface (**Figure 5.6**). Residues Tyr501 and His505 were the only persistent hubs in the RBD. Other high centrality residues included interface residues Leu455 and Val510, of which Leu455 participates in inter-protein communication [414, 561].

Compared to the WThACE2, S19P experienced the most significant loss of *CC* at the RBD interface involving residues Gln498, Thr500 and Gly502 (**Figure 5.8**). The loss of *CC* hubs at the interface disrupts communication between proteins through reduction of the information dissemination points (*CC* hubs). In the **chapter 4** we showed a relationship between the frequency of *CC* hubs at the interface and the closeness of interaction between the RBD and hACE2 proteins. Similarly, the RBD-hACE2 inter-protein COM distance results further emphasized this relationship. For instance, the S19P system with the least number of *CC* hubs at the interface had the highest interaction distance between the RBD and hACE2 as per COM distance measurement (**Figure 5.8**). In contrast, systems with a high *CC* hub concentration at the complex interface i.e., K341R, K26R and M82I, had closer inter-protein interaction compared to the WThACE (**Figure 5.8**). The global top 5% analysis meant that K341R and K26R had the most RBD *CC* hubs *viz*: K341R: 13 hubs, K26R: 13, D597Q: 12, M82I: 11, WThACE2: 8, N546D: 7 and lastly S19P with 4 hubs.

In comparison to the WThACE2, there was an increase of *CC* hubs in the hACE2 systems especially at the protein core encompassing the active site domain in all hACE2 polymorphs except S19P. The high centrality residues included the zinc coordinating residue Glu402. The hACE2 *persistent CC* hubs included, Tyr381, Gly395, Gly399, Phe400, Glu402, Ala403,

Val404 and Arg518, some of which were among the top five high centrality residues in the hACE2 i.e., Gly399, Glu402 and Ala403. Structurally, the hACE2 *CC* hubs were positioned at the hACE2 active site cleft and substrate binding site where they are involved in both zinc and substrate coordination [498].

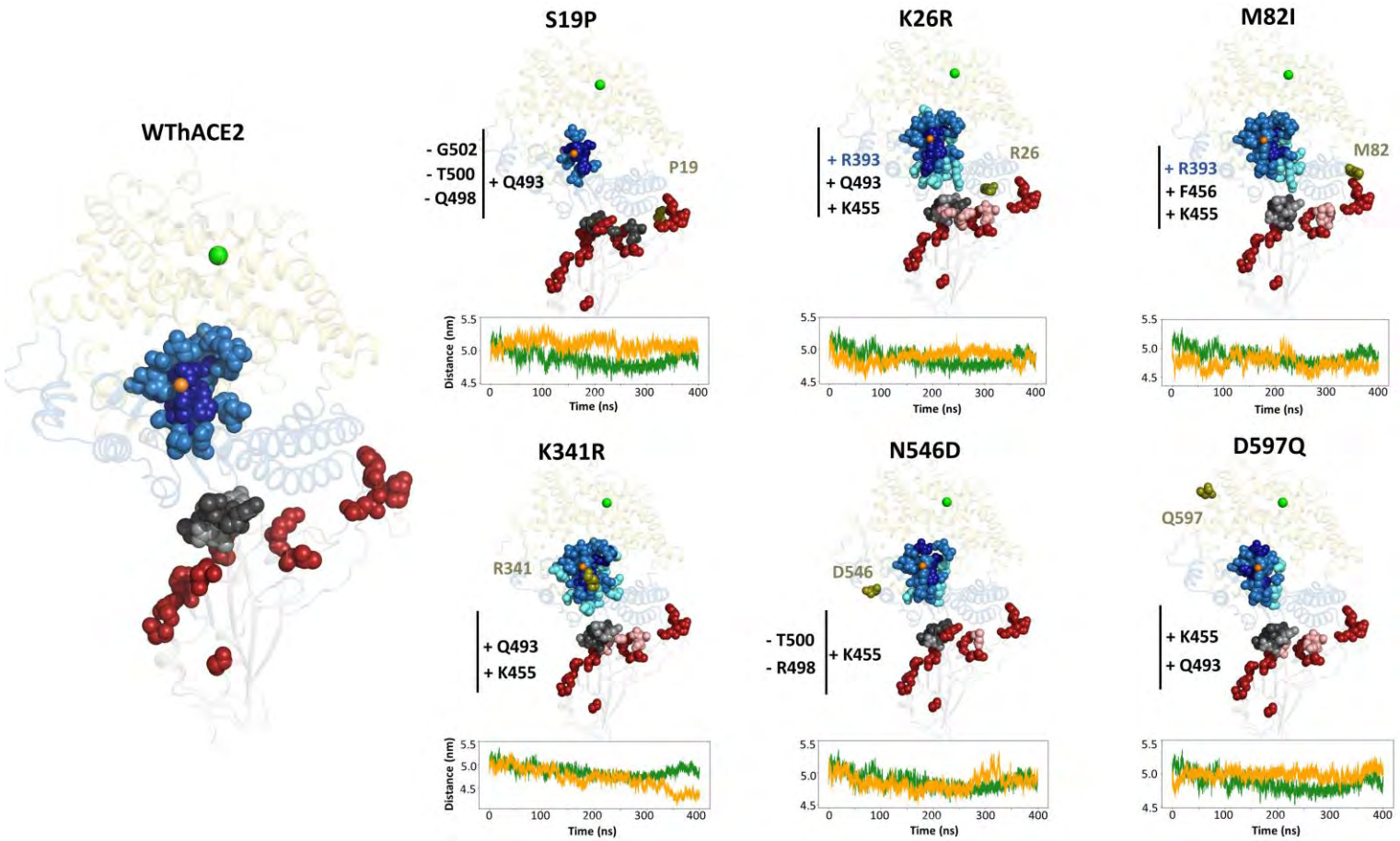

Figure 5. 8: RBD-hACE2 structures showing *CC* hub distribution per global 5% and 4% in the RBD (grey) and hACE2 (sub-domain I: blue, sub-domain II: yellow) systems, respectively. WThACE2 hubs are shown as grey (RBD) and blue (hACE2) spheres. Hubs unique to the to the mutant systems are in salmon (RBD) and cyan (hACE2) sphere colors. Each subplot is annotated with interface hub gains and losses marked as + and -, respectively. The subplots also show the RBD-hACE2 inter-protein COM distance for the variants (yellow) compared to the WThACE2 (green).

The high *closeness centrality* at the substrate site implies a closely bound network of substrate/active pocket residues in the hACE2 polymorphs compared to the WThACE2 system, which influences protein activity. Collectively K341R had the highest number of *CC* hubs whereas, S19P had the least i.e., K341R (RBD: 13 + hACE2: 33 = 46) > K26R (13 + 29 = 42) > M82I (11 + 29 = 40) > D597Q (12 + 21 = 33) > N546D (7 + 24 = 31) > WThACE2 (9 + 23 = 30) > S19P (4 + 9 = 13). Evidently, some the hACE2 polymorphisms i.e., K341R, K26R and S19P seem to have more pronounced effects on the RBD-hACE2 behavior compared to the rest.

5.4.3.2.3 The hACE2 Zinc Binding Domain Harbored Most EC hubs

As earlier noted in the local DRN analysis (**section 5.4.3.1.2**), the global *EC* hub distribution wasn't consistent across all the RBD systems. No RBD *EC persistent hubs* were identified from the global top 5% centrality analysis (**Figure 5.6**). Interestingly, D579Q had the highest number of *EC* hubs in the RBD *viz*: WThACE2: 3 hubs, S19P: 4, K26R: 9, M82I: 0, K341R: 14, N546D: 9 and D597Q: 29. Even in the absence of *persistent hubs*, residue Gly502 and Gly504 had hub status in at least five of the seven systems. These interface residues are part of the RBM that connects the RBD to hACE2. *EC* hubs in the RBD were mainly concentrated at the interface region, where they are involved in the inter-protein interaction (**Figure 5.9**).

Unlike the RBD, the hACE2 had a uniform distribution of *EC* hubs across the protein systems with majority if the hubs confined at the zinc and substrate/active site (**Figure 5.9**). The *persistent hubs* included, Gly377, Val403-Gln409, Arg518, Tyr521, Gln522, Phe525, Gln526, Leu529 and Ala550. These *persistent hubs* also included some of the top five high centrality residues i.e., Val403, Gly405, Gln522 and Phe525 in majority of the hACE2 polymorph systems. Collectively these residues made up part the of the α-helix region at the

base of the hACE2 receptor binding cleft and part of the substrate binding site. Furthermore, residue Arg518 binds bioactive compounds in the hACE2 active site [562] whereas, Gln522 coordinates Hesperidin which has SARS-CoV-2 inhibiting potential [563]. Generally, all the hACE2 variant systems had more hACE2 *EC* hubs compared to the WThACE2 i.e., D597Q (RBD 29 + hACE2 21 = 50) > K341R (14 + 24 = 38) > K26R (9 + 27 = 36) > N546D (9 + 25 = 34) > M82I (0 + 26 = 26) > WThACE2 (3 + 24 = 27) > S19P (4 + 20 = 24). The hubs formed a network of interactions linking the active site and zinc coordinating residues (**Figure 5.10 and S30**).

RMSD, RMSF and Rg analysis of the hACE2 *EC* hubs informed on the relationship between *EC* and residue dynamics. For most hACE2 variant systems, the *EC* hub residues experienced less residue fluctuation compared to the WThACE2 (**Figure S31A**). Furthermore, the RMSD and Rg analysis showed an all-round stable zinc and substrate domain especially in K26R (**Figure S31 B, C**). It is evident that stable protein regions are associated with high *EC* values.

Furthermore, the differences in the hACE2 *EC* hub distribution and behaviour at the zinc binding site between the WThACE2 and hACE2 SNPs highlights the effect of the naturally occurring hACE2 SNPs on protein communication and network patterns. The high centrality at these regions could imply more efficient protein activity, this based on studies linking RBD binding to increased carboxypeptidase activity [497].

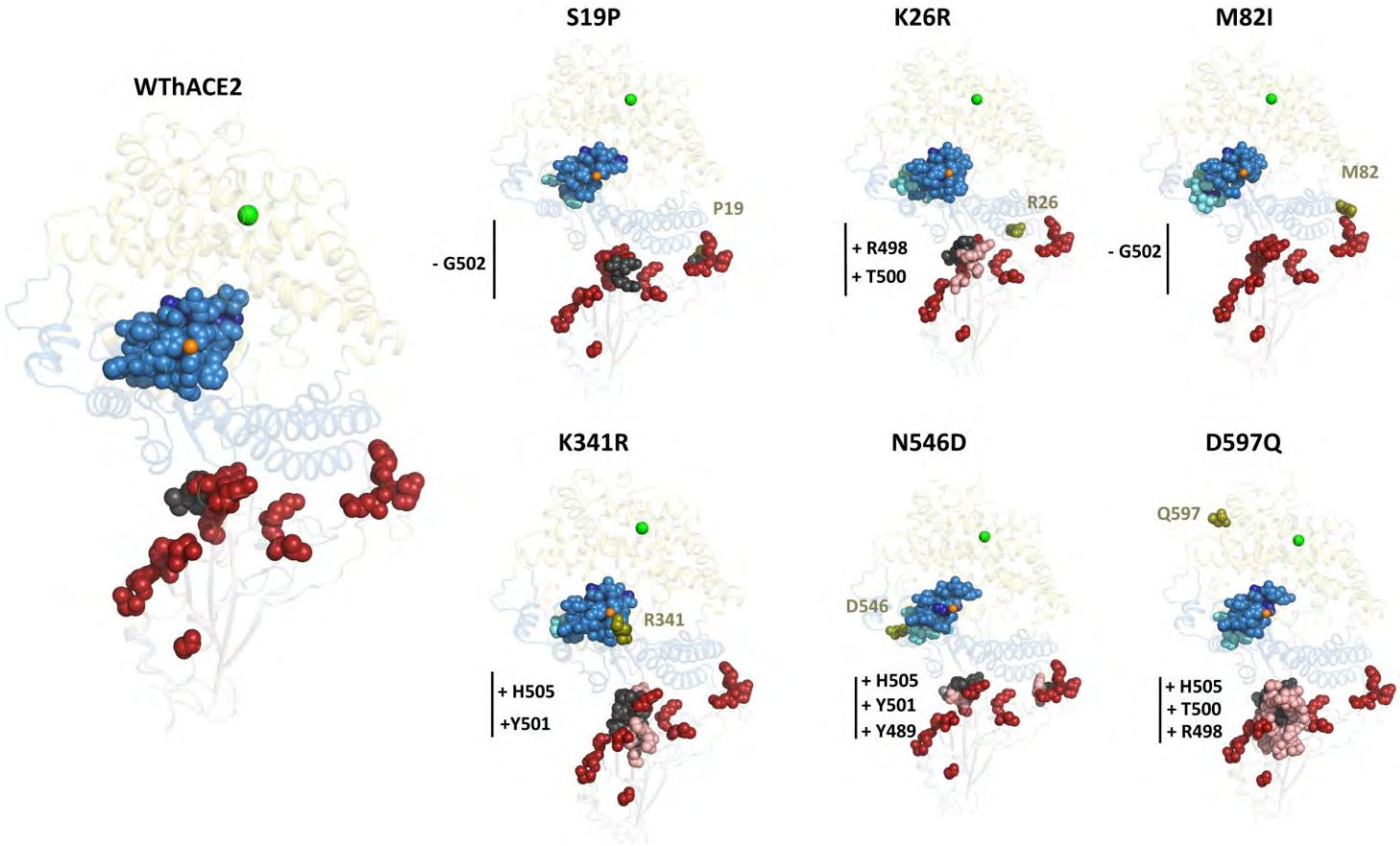

Figure 5. 9: RBD-hACE2 structures showing the distribution of the *EC* hubs. The RBD is shown in grey color while the hACE2 sub-domain I and II in blue and yellow, respectively. WThACE2 hubs are shown as grey (RBD) and blue (hACE2) spheres, whereas the RBD mutations as firebrick. Hubs unique to the hACE2 SNPs are shown as boron spheres (RBD) and cyan spheres (hACE2). SNP positions in the hACE2 are shown as olive spheres while the gained and lost variant interface hubs annotated with + and - symbols.

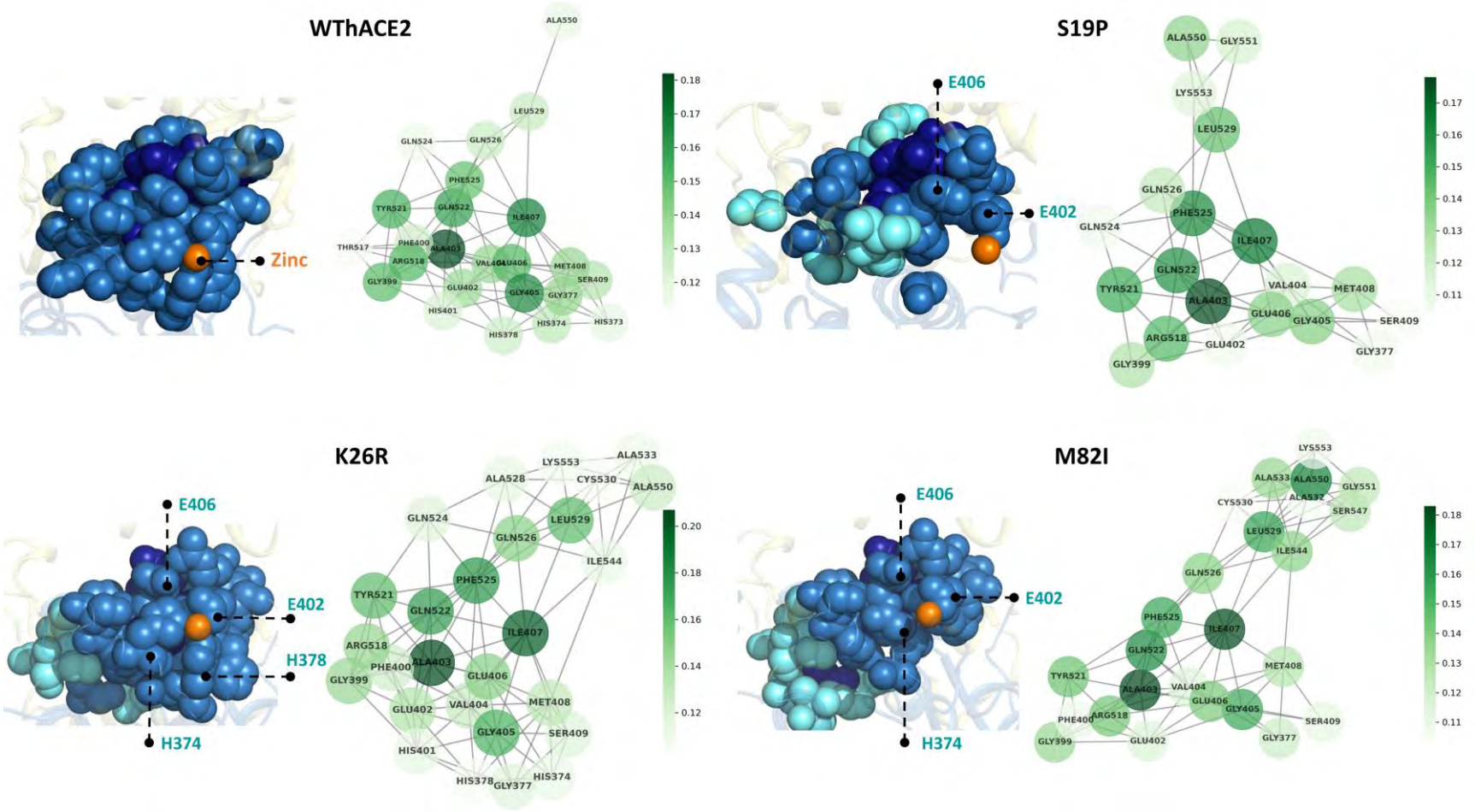

Figure 5. 10: Zoomed in view of the *EC* distribution in the hACE2 systems. In each subplot, hubs common in both the WThACE2 and hACE2 variant systems are shown as blue spheres and those unique to the hACE2 variants as cyan spheres. The zinc ion is shown as an orange sphere and the coordinating residues with hub status annotated. Each subplot contains a network graph generated with NetworkX showing hub interaction. The node color corresponds to the to the degree of centrality with dark green indicating high *EC* and light green low *EC*.

5.4.3.2.4 Degree of Centrality

There were nominal differences in the RBD and hACE2 *DC* hub distribution in the WThACE2 and hACE2 variant systems. However, only Gly431 and Tyr508 had *persistent hub* status in the RBD across all systems (**Figure 5.6**). Structurally both residues are positioned at the RBD interface, where Gly431 binds remdesivir and natural compounds with SARS-CoV-2 inhibiting activity [506, 564]. Tyr508 binds inhibitors and RBD neutralizing antibodies [506].

The hACE2 system had more *persistent hubs* than the RBD *viz*: Ala25, Ile88, Val93, Leu97, Cys261, Ile407, Arg518 and Ala550. Some of these residues are at the hACE2 interface where they facilitate RBD binding i.e., Ala550 which stabilizes Tyr83 in the hACE2 which in turn forms H-bonds with Asn487 and Tyr489 [414]. Furthermore, S19P, K26R, M82I and D597Q gained interface hubs at positions Tyr489 (RBD), Asp355 (hACE2), Thr27 (hACE2) and Tyr501 (RBD), respectively.

Collectively, DRN analysis through *BC* showed the conservation of high centrality residue networks between the RBD and hACE2 irrespective of the presence of hACE2 polymorphisms. However, key zinc and chloride coordinating residues in the hACE2 protein displayed lower centrality values compared the WThACE2 especially for S19P. *EC* analysis identified differences in centrality distribution at the zinc and substrate binding site consequent to the naturally occurring hACE2 polymorphisms. The high concentration of *EC* hubs at the hACE2 active site in the variants suggest a mutational influence on the carboxypeptidase activity of the protein. *CC* network analysis suggested an unfavourable RBD binding to the S19P hACE2 polymorph based on the reduced number of *CC* hubs at the interface which is associated with increased inter-protein interaction distance. The eventual

effect of increased inter-protein distance on protein interactions is further investigated through contact map analysis in the next section.

5.4.4 Changes in the RBD-hACE2 Interactions Over the MD Simulation

The contact map analysis approach previously discussed in **chapter 4** was applied here to investigate the RBD-hACE2 interaction changes in the presence of both Omicron BA.4/5 and hACE2 polymorphisms. This involved the identification of RBD interface residues, and using the *contact_map.py* Python script from MDM-web server [264, 265] to determine their contact frequencies over the MD simulation. The changes in residue interactions between the WThACE2 and hACE2 variants was identified by calculating the difference between the WThACE2 and hACE2 system contacts (WThACE2 interface residue contact frequency – hACE2 variant interface residue contact frequency) and results presented as a heat map (**Figure 5.11A**). From **Figure 5.11A**, blue means a higher residue contact frequency in the hACE2 variants, whereas red means higher contact frequency in the WThACE2.

Contact map analysis showed a general reduction in the residue contact frequencies of the following RBD-hACE2 residue pairs in the hACE2 variants compared to the WThACE2 i.e., Ala475-Ser19, Arg498-Tyr41, Thr500-Asn330 (except K26R), Thr500-Leu351, Val503-Thr324 (except K26R), Val503-Phe327, Val503-Gly354 (except K26R and N546D), Val503-Phe356, Val503-Met383 (except K26R). With the exception of residue position 19, none of the hACE2 residues listed above have polymorphs, meaning that the observed interaction disparity between the hACE2 variants and the WThACE2 weas related to the RBM dynamics.

On a systems level, the following RBD-hACE2 residues pairs had more contact frequency in S19P variant compared to the WThACE2 i.e., Leu455-Lys31, Leu455-His34, Ala475-Glu23,

Val486-Met82, Asn487-Gln24, Thr489-Phe28, Gln493-Glu35, Gly496-Asp38, Thr501-Lys353 and His505-Lys353.

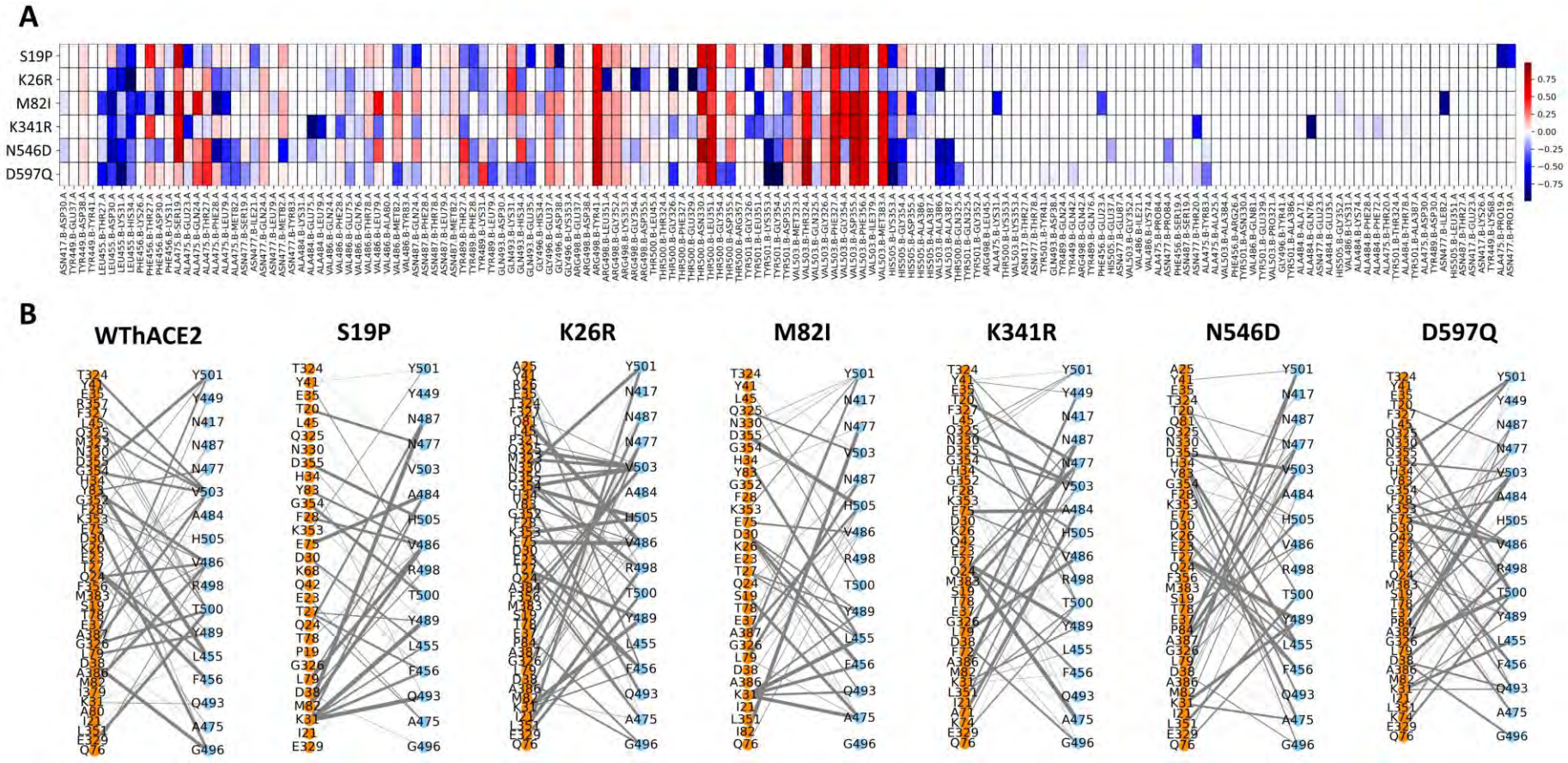

Figure 5. 11: A is a heat map showing the contact frequency differences between the RBD-hACE2 residue pairs in the WThACE2 and hACE2 variants for each RBD-hACE2 complex. Blue means more residue contact frequency in the hACE2 variants compared the WThACE2 and red the reverse. Residue pair numbers are on the x-axis and protein systems on the y-axis. B shows bipartite network graphs of the RBD-hACE2 interactions. RBD and hACE2 residues are shown as blue and orange nodes, respectively. The edge thickness corresponds to the contact frequency.

Additionally, proline in the S19P SNP had a 1.7-fold and 12-fold increase in contact frequency with Gly476 and Asn477, respectively, compared to the serine in WThACE2 (Gly476-Pro19: 00.977, Gly476-Ser19: 0.059; Asn477-Pro19: 0.485, Asn477-Ser19: 0.04). Besides the mentioned gains, significant loses of RBD-hACE2 interactions were noted in S19P compared to the reference i.e., Thr27 of hACE2 had a 1.39-fold reduction in interaction with Phe456 of the RBD. Similarly, Asp355 of the hACE2 had 1.6-fold reduction in interaction with Tyr501 of the RBD. Additionally, the RBD-hACE2 network bipartite graphs (**Figure 5.11B**) showed that S19P had significantly less inter-protein (RBD-hACE2) interactions/contacts (58) compared to the WThACE2 (92) over the MD simulation. Structurally, the proline amino acid residue substituted for the serine is more hydrophobic as it is usually buried inside the protein structures. Its presence at the interface would imply less interactions. Furthermore, its unique cyclic property restricts its flexibility and interactions compared to the WT serine amino acid which is polar, hydrophilic and the hydroxyl group in serine's side chain allows it to participate in hydrogen bonding, contributing to protein stability and interactions. These unique amino acid properties could explain the observed differences in interactions between S19P complex and WThACE2 complex. More interestingly, network analysis metrics like *CC* which inform on residue proximity to other residues are able to pick up these residue level changes which are portrayed in the centrality distribution. The less *CC* hubs at the interface in the S19P systems compared to the WThACE2 hints on increased RBD-hACE2 interaction distance and hence less inter-protein interactions. In a biophysical study by Suryamohan and colleagues [207], they showed that S19P only had marginal increase in affinity to the SARS-CoV-2 S protein and that K26R had marked increase in affinity to the S protein. It is important to note that molecular interaction is a complex mechanism influenced by many factors, furthermore, the study by Suryamohan

et al [207] used the WT S protein unlike here where the BA.4 RBD was used with an array of 17 mutations contributing to the factors affecting RBD-hACE2 interactions.

The K26R SNP at the hACE2 interface resulted in a 1.5-fold gain in the Arg26 (hACE2)-Phe456 (RBD) interaction compared to the WThACE2. Further gains in the RBD-hACE2 interactions were noted for the following K26R residue pairs i.e., Lys455-Asp30, Leu455-Lys31, Leu455-His34, Ag498-Leu351, Arg498-Gly354, Thr500-Gly326, Thr500-Glu329, Tyr501-Gly326, His505-Ala386, His505-Ala387 and Val503-Ala386. Besides the gains in contact frequency, the total number of K26R inter-protein interactions were 87 compared to the 92 in WThACE2. Even though more contacts were observed in the reference system, the hACE2 K26R polymorphism seems to favor RBD binding compared S19P, M82I and N546D. The arginine residue at position 26 in the hACE2 polymorph typically forms more interactions due to its guanidinium group, which provides it with a greater positive charge and multiple hydrogen bonding sites compared to lysine. This enhanced charge and structural arrangement make arginine particularly effective at forming interactions with negatively charged molecules and other polar entities. Arginine's guanidinium group contains three nitrogen atoms, allowing it to potentially form multiple hydrogen bonds with various partners. This property enables arginine to engage in diverse interactions, such as salt bridges, hydrogen bonds, and electrostatic interactions [565, 566]. These changes in atomic interactions possibly influence RBD-hACE2 interactions at the interface as depicted from the network analysis metrics.

Like in K26R, the M82I hACE2 SNP also resulted in a 5.11-fold and 6-fold increase in interactions with Asn477 and Gly498, respectively (**Figure 5.11**). Even though less interactions were noted between the RBD and hACE2 proteins in M82I (71) compared to the WThACE2 (92), distinct gains in M82I complex interactions were noted for following RBD-hACE2 residue pairs, Leu455-Thr27, Leu455-Asp30, Leu455-His34, Phe456-Lys26, Ala475-

Phe28, Ala475-Leu79, Tyr501-Leu351, Tyr501-Asp355, His505-Lys353, His505-Asp355, Ala475Lys31, Phe456-Glu23 and Asn477-Ile82.

Bipartite plots showed K341R as having the most RBD-hACE2 interactions, 94, of all the hACE2 variants under study (**Figure 5.11B**). This was characterized by contact gains at for the following RBD-hACE2 residue pairs, Leu455-Asp30, Leu455-His34, Ala484-Glu75, Ala484-Leu79, Val486-Phe28, Tyr501-GLy326, Tyr501-Leu351, His505-Lys353, His505-Asp355, Asn477-Thr20 and Ala484-Gln76. Of interest here was the allosteric effect of the K341R SNP on the hACE2 interface interactions.

Lastly, N546D and D597Q, located distal to the interface, had more inter-protein interactions, 88 and 92, respectively, compared to the S19P system. Some of the gained interface interactions involved the following RBD-hACE2 residue pairs, Leu455-Asp30, Leu455-Lys31, Leu455-His34, Ala475-Leu79, Ala475-Met82, Tyr501-Lys353, Tyr501- Gly354, His505-Lys353, His505-Gly354, Val503-Ala386 and Val503-Gln325.

Furthermore, we determined the binding energy of the low energy structures for each system using the HawkDock web server [500] to get an idea of the polymorphism effect on RBD affinity. The K26R system had the least energy of all the systems *viz*: K26R (-136.69 kcal/mol) > WThACE2 (-98.54) > K341R (-82.11) > S19P (-81.79) > M82I (-78.95) > D597Q (-69.75) > N546D (-69.37).

In summary, the inter-protein interaction analysis shows that the hACE2 polymorphisms have variable effects on complex interactions irrespective of their location in the protein. The results are also in agreement with the previous research on the S19P [208, 210, 211, 557, 567, 568] and K26R [204, 207, 208, 211, 568, 569] hACE2 polymorphisms which are associated with decreased and increased binding affinity of hACE2 for the S RBD, respectively.

5.5 Chapter Conclusion

The global spread of the COVID-19 pandemic was characterized by intermittent waves of outbreaks with a disproportionate spread across different populations. In part, this was due to the different population densities and urbanization which goes hand in hand with community interconnectedness all of which facilitate disease spread. On the other hand, the genetic variability of the hACE2 receptor among different populations [207, 570–572], in addition to ACE2 expression levels [573, 574] are also predicted to influence S RBD binding and SARS-CoV-2 infectivity.

Chapter 5 concentrated on deciphering the effects of the naturally occurring African hACE2 polymorphisms on the interaction between the Omicron BA.4 sub-variant RBD and the hACE2. To this end, the hACE2 SNPs with high allele frequency were retrieved from the gnomAD database i.e., S19P, K26R, M82I, K341R, N546D and D597Q. Each of these hACE2 polymorph structures in complex with the BA.4/5 RBD were modeled and subjected to 400 ns MD simulations.

Trajectory analysis highlighted how the phytochemically similar hACE2 polymorphisms have minimal effects on the hACE2 dynamics compared to the RBD. This was further supported by the Rg and RMSF analysis that showed minimal differences in global dynamics between the WThACE2 reference and hACE2 variant systems.

Residue level analysis of the hACE2 SNP effects was completed using DRN analysis from the MDM webserver tools [264, 265] which applies graph theory to describe protein network characteristics. *BC* analysis showed that the phytochemically akin hACE2 polymorphisms maintain a network of high centrality residues linking the S RBD core to the hACE2 zinc binding site as previously seen in chapter 4. The zinc and substrate binding site were also identified as highly influential in the hACE2 variants compared to the WThACE2. This was

highlighted by the *EC* metric which identified a cluster of residues around the zinc binding site with high *eigenvector centrality* most especially in the hACE2 variant systems. We hypothesize that the concentration of *EC* hubs in the hACE2 is a compensatory mechanism in the variants, which in turn affects the hACE2 activity in presence of the RBD. Furthermore, the SNP effects on inter-protein interactions were shown by the *CC* analysis, where less *CC* hubs were noted at the RBD-hACE2 interface of some hACE2 variant systems compared to the WThACE2. The number of *CC* hubs at the interface is associated with the inter-protein interaction distance as previously shown in **chapter 4**. Here, the S19P variant with the least number of interface *CC* hubs had the most distant RBD-hACE2 interaction compared to the WThACE2. Subsequently, contact map analysis of the RBD interface residues showed the S19P variant as having the least number of interactions connecting the two proteins, whereas WThACE2 had the most inter-protein interactions. Finally, binding energy calculations also identified S19P as requiring more energy for RBD-hACE2 binding while K26R required the least energy. Furthermore, binding energy analysis also highlighted how the hACE2 SNPs distal to the interface region affect the hACE2 affinity for the RBD. For instance, protein binding energy requirements of K341R were close to that of the WThACE2 however, K341R had the most RBD-hACE2 interactions across all the systems implying that the K341R polymorphism may favor RBD binding.

In conclusion, the naturally occurring hACE2 polymorphism having varying effects on the RBD-hACE2 interactions with some polymorphisms like S19P associated with decreased binding affinity and inter-protein interactions as alluded else where[208, 210, 211, 557, 567, 568], whereas K26R and K341R are associated with favorable RBD-hACE2 binding [207, 208, 558, 560, 575]. The work in this chapter further validates the network analysis approach used to investigate SNP effects on protein communication patterns through DRN analysis.

Concluding Remarks

AMR has morphed into one of the greatest hurdles in infectious disease management which poses a great challenge to public health. The development of antimicrobial resistance is associated with longer hospitalizations, increased health-care cost and increased mortalities especially in low-income countries [576, 577]. A 2019 systematic review of the global AMR burden attributed approximately 4.95 million mortalities to AMR with at least 1.27 million deaths directly linked to AMR [100]. These statistics are particularly alarming since there hasn't been any major discoveries of novel antimicrobial classes since the 1980s [578], let alone discovery of drugs active against resistant bacterial strains. Therefore, exploration and deciphering drug resistance mechanisms especially in SNPs remains one of most promising avenues of addressing AMR.

Besides AMR, new emerging infectious diseases that are new to humans have recently increased in incidence and geographic range. These diseases such as COVID-19 arise from the introduction of new pathogens into human populations, the evolution or adaptation of existing pathogens, or changes in the environment that promote the transmission of infectious agents. One of the key features of emerging infectious diseases is their ability to mutate rapidly and adapt to changing conditions, which can lead to the emergence of new strains that are more virulent, more transmissible, or more resistant to treatment. A recent example is the rapidly evolving SARS-CoV-2 pathogen responsible for COVID-19. Understanding the mechanisms of pathogen evolution and the factors that contribute to the emergence of infectious diseases is critical for developing effective strategies to prevent and control these threats to global health.

This work discusses the consequences of *Mycobacterium tuberculosis* drug resistance mutations in Part I of the thesis (chapters 2 and 3), and in Part II (chapters 4 and 5), the

effects of naturally occurring mutations in both SAR-CoV2 RBD and hACE2 on the inter-protein interactions. The discussion is premised on the structural, dynamic and network changes in the target proteins based on *in silico* approaches including, MD simulations, comparative ED and DRN analysis. This piece of research adds to the scientific body on mutational studies and the approaches therein.

Chapter 1 introduced AMR and its burden on global health focusing mainly on drug resistance in infectious diseases like *Mycobacterium tuberculosis*. Through literature review the chapter justifies the need for new studies with novel approaches of studying drug resistance in a bid to inform future drug design.

The chapter also highlights the emergence of new infectious diseases consequent to the encroachment of wildlife habitats resulting in inter-species cross infections. This was in light with the outbreak of the COVID-19 infectious disease caused by SARS-CoV-2. In the case of new infectious pathogens, the chapter points out the need for mutational studies due to address the continuous evolution of SARS-CoV-2 affording it higher transmissibility and infectivity. Gaps in literature on mutational studies in infectious diseases are highlighted in this chapter together with the detailed explanation of the computational approaches employed in the thesis.

Chapter 2 concentrated on TB, where we described the global and local changes associated with eleven resistance-conferring mutations in the KatG enzyme. Global analysis revealed diverse conformational sampling in the mutant systems compared to the wildtype characterized by asymmetric protomer behavior. Most importantly, the unstable protein dynamics in the mutant systems cascaded to the active site region resulting in an unstable heme-binding environment. Cofactor analysis highlighted the effects of the assorted active site motions in the mutants characterized by a reduction in the number of H-bonds and non-

bonded interactions between heme and the mutant systems. The compromise of heme coordination in the mutants ultimately affects the catalase activity of the enzyme including isoniazid activation. Furthermore, DRN analysis identified the dimerization domain as highly central and influential in both information flow and dissemination based on *BC* and *CC* analysis. This was further emphasized by the calculated binding energy contribution of the dimerization domain to KatG inter-protomer interactions. Ultimately the chapter sheds light on how the mutation induced protein dynamics affect the conformational stability and network patterns around the heme environment. The work also further highlights the asymmetric protomer behaviour in dimeric proteins.

Chapter 3 focused on the TB drug resistance mutations in the MtPncA enzyme responsible for activation of the first-line drug, PZA, to pyrazinoic acid. The 82 PZA-resistance mutations in MtPncA, previously investigated in [135], were analyzed in the presence of NAM an analogue of PZA. Global MD trajectory analysis showed that the mutations, particularly those around the MBS, are associated with an unstable MtPncA structure *viz*: G17D, A46V, D49A, D49G, H51Y, W68R, W68G, C72R, L85R and K96N. Furthermore, stability analysis using the mCSM web server identified the PZA-resistance mutations with a destabilizing effect on the MtPncA structure. Interestingly, MtPncA had a higher NAM retaining ability compared to PZA with 34 MtPncA-NAM complexes stably coordinating NAM compared to the 24 in MtPncA-PZA complexes. This was further shown to be a consequence of more interactions formed by NAM in the substrate pocket compared to PZA due to the absence of the extra nitrogen atom in NAM. The extra nitrogen atom in PZA is believed to result in a less energetically favorable MtPncA-PZA complex. Like in chapter 2, the DRN analysis technique identified the critical MtPncA regions at the protein core consisting of the active site and MBS. DRN also highlighted the implications of mutations on the protein signal transduction.

Chapter 4 focused on characterizing the effects of RBD mutations in the SARS-CoV-2 Omicron sub-lineages, BA.1, BA.2, BA.3 and BA.4 on RBD binding to the hACE2. Trajectory analysis through RMSD, RMSF and DCC revealed a flexible RBM region in the Omicron sub-lineages which influenced both the RBD and hACE2 dynamics. *BC* analysis showed novel high centrality residue paths bridging the RBD core to the hACE2 zinc binding domain which were modified in the Omicron sub-lineages. Furthermore, the concentration of high *CC* hubs at the RBD-hACE2 complex interface influenced the inter-protein interaction distance, where higher *CC* hub numbers were associated with closer protein-protein interaction. This was further emphasized by residue interaction analysis which identified pronounced RBD-hACE2 interactions in the Omicron sub-lineages compared to the wildtype. *EC* analysis also showed the effect of the RBD mutations on the hACE2 functionality.

Chapter 5 compliments the discoveries from chapter 4 by focusing on the effects of the naturally occurring hACE2 polymorphisms on the receptor affinity and interaction with the S RBD of the Omicron BA.4/5 sub-lineage. Global trajectory analysis showed minimal changes in the hACE2 dynamics, and this was attributed to the preserved phytochemical properties within hACE2 polymorphisms. Residue level DRN analysis affirmed the maintenance of two *BC* hub communication paths between the S RBD and hACE2 as earlier observed in chapter 4 even with the introduction of hACE2 SNPs. *CC* analysis highlighted the effects of hACE2 interface SNPs *viz*: S19P and K26R, on the RBD-hACE2 interaction and complex formation, based *CC* distribution. This was also collaborated by interface residue contact map analysis, which identified S19P having significantly less inter-protein interactions compared to the WThACE2. Conversely, K26R had significantly more RDB-hACE2 protein-protein interactions and required less RBD-hACE2 binding energy compared to the WThACE2. Furthermore, *EC* analysis showed that even though the hACE2 polymorphisms are naturally occurring, they affect the hACE2 functionality, in relation to the BA.4 RBD, based on the

concentration of *EC* hubs around hACE2 zinc and active site. Collectively, DRN analysis effectively characterized the changes in RBD-hACE2 communication patterns in the presence of hACE2 polymorphisms.

In summary, the thesis employs integrative *in silico* approaches to investigate the impacts of drug resistance and naturally occurring protein mutations in infectious diseases. Specifically, the analysis of drug resistance mutations highlights alterations in protein behavior and characteristics that confer the ability to tolerate or resist the effects of drugs, elucidating the molecular mechanisms underlying of drug-resistant strains.

DRN analysis was the standout *in silico* analysis approach in the thesis as it was extensively used to highlight previously known and new key protein residues in the various protein systems discussed. Graph theory and social networks analysis techniques have provided a mathematical model for the study and better understanding of protein/molecular communication mechanisms [270, 579]. The study of dynamic protein networks through the combination of MD-simulations and DRN analysis [263, 264] exposes the intricate patterns in protein communication and how they change with protein perturbations [76, 155, 271]. Even though network analysis consists of a number of metrics, this thesis mainly highlights the importance and applicability of *BC*, *CC*, *EC* and *DC* in analysis of protein structures. The *BC* metric consistently highlighted important functional regions of the protein systems based on the metric values. Here, high *BC* values coincided with functional regions of the proteins (active site, substrate binding site and the protein core). High *CC* values were mostly concentrated at protein interfaces and a correlation between *CC* values and intra/inter-protein interaction distances was established in chapter 2 section 2.4.4.3.2 and chapter 4 section 4.4.5.2. This potentially makes *CC* as a descriptor of molecular interaction behaviour. The *EC* and *DC* metric characteristics highlighted the local effects of mutations on residue interactions.

Furthermore, the mutational analysis here alluded to a negative correlation between high DRN metric values (especially *BC*) and mutation hotspots implying that protein network analysis has the potential to be used as a predictor of mutational hot and cold spots. It is also important to note that the combinatory use DRN metrics here, especially *BC*, *CC* and *EC*, provided better resolution of the wholesome protein behaviour.

Finally, because of the dependency of DRN analysis on MD-simulations, care should be taken to use only representative MD trajectories and where possible an n >1 used for each protein system for comparative and quality control purposes, where n is the number of MD runs.

Limitations

Molecular level changes happen over varying time scales some of which take longer than the nanosecond scale. This means that with longer simulations, more information can be obtained however, this can be limited by the availability of computational resources like computing power. In this work, single MD runs (n = 1) were used for most WT systems and all mutant systems due to the limitations on available computing resources.

Furthermore, MD computations produce large volumes of data files requiring external and cluster storage which was a limiting factor on the number and length of the simulations. Because of this, only one run was done (using a single seed) for most system analysis. The MD trajectory data from these experiments was also not uploaded to any community repository.

Additionally, the statistical comparison of the MD-simulation runs (using the RMSD and RG results) was limited by number of sample-runs for each sample (n = 1). It would have been ideal to have multiple runs per sample for a representative comparison. The use of one run per sample was a trade-off between getting as much information about the systems, meaning longer simulation time, and having multiple runs for each system with shorter simulation time.

Future work

The ability of DRN analysis to consistently identify key functional residues in various protein systems and, the observed inverse relation between DRN centrality values (*BC*, *CC* and *EC*) and residue mutability implies that DRN analysis can be applied to mutation prediction. The next steps in will be using MD-simulations and data from DRN analysis to model predictors of mutational hot and cold spots in various protein systems.

Furthermore, future work will involve developing better models for the inter-system DRN metric comparison besides hubs distribution. This would enable robust statistical comparison between systems.

References

1. Barozi V (2020) Understanding of the underlying resistance mechanism of the Kat-G protein against isoniazid in Mycobacterium tuberculosis using bioinformatics approaches

2. WHO EMRO | Infectious diseases | Health topics. In: World Health Organ. - Reg. Off. East. Mediterr. http://www.emro.who.int/health-topics/infectious-diseases/index.html. Accessed 6 Jul 2022

3. Nii-Trebi NI (2017) Emerging and Neglected Infectious Diseases: Insights, Advances, and Challenges. BioMed Res Int 2017:e5245021. https://doi.org/10.1155/2017/5245021

4. Rahman MdT, Sobur MdA, Islam MdS, et al (2020) Zoonotic Diseases: Etiology, Impact, and Control. Microorganisms 8:1405. https://doi.org/10.3390/microorganisms8091405

5. Global tuberculosis report 2021. https://www.who.int/publications-detail-redirect/9789240037021. Accessed 28 Oct 2021

6. Shereen MA, Khan S, Kazmi A, et al (2020) COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses. J Adv Res 24:91–98. https://doi.org/10.1016/j.jare.2020.03.005

7. Singh D, Yi SV (2021) On the origin and evolution of SARS-CoV-2. Exp Mol Med 53:537–547. https://doi.org/10.1038/s12276-021-00604-z

8. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int. Accessed 15 Apr 2022

9. HIV data and statistics. https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/strategic-information/hiv-data-and-statistics. Accessed 6 Jul 2022

10. World malaria report 2021. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021. Accessed 6 Jul 2022

11. Cucinotta D, Vanelli M (2020) WHO Declares COVID-19 a Pandemic. Acta Bio Medica Atenei Parm 91:157–160. https://doi.org/10.23750/abm.v91i1.9397

12. Rawat D, Dixit V, Gulati S, et al (2021) Impact of COVID-19 outbreak on lifestyle behaviour: A review of studies published in India. Diabetes Metab Syndr Clin Res Rev 15:331–336. https://doi.org/10.1016/j.dsx.2020.12.038

13. Baena-Díez JM, Barroso M, Cordeiro-Coelho SI, et al (2020) Impact of COVID-19 outbreak by income: hitting hardest the most deprived. J Public Health 42:698–703. https://doi.org/10.1093/pubmed/fdaa136

14. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A (2020) The outbreak of COVID-19 coronavirus and its impact on global mental health. Int J Soc Psychiatry 66:317–320. https://doi.org/10.1177/0020764020915212

15. Salian VS, Wright JA, Vedell PT, et al (2021) COVID-19 Transmission, Current Treatment, and Future Therapeutic Strategies. Mol Pharm 18:754–771. https://doi.org/10.1021/acs.molpharmaceut.0c00608

16. Khan AH, Tirth V, Fawzy M, et al (2021) COVID-19 transmission, vulnerability, persistence and nanotherapy: a review. Environ Chem Lett 19:2773–2787. https://doi.org/10.1007/s10311-021-01229-4

17. Edelson PJ, Phypers M (2011) TB transmission on public transportation: A review of published studies and recommendations for contact tracing. Travel Med Infect Dis 9:27–31. https://doi.org/10.1016/j.tmaid.2010.11.001

18. Lygizos M, Shenoi SV, Brooks RP, et al (2013) Natural ventilation reduces high TB transmission risk in traditional homes in rural KwaZulu-Natal, South Africa. BMC Infect Dis 13:300. https://doi.org/10.1186/1471-2334-13-300

19. Rahman MA, Sarkar A (2017) Extensively Drug-resistant Tuberculosis (XDR-TB): A daunting challenge to the current End TB Strategy and policy recommendations. Indian J Tuberc 64:153–160. https://doi.org/10.1016/j.ijtb.2017.03.006

20. Shibabaw A, Gelaw B, Gebreyes W, et al (2020) The burden of pre-extensively and extensively drug-resistant tuberculosis among MDR-TB patients in the Amhara region, Ethiopia. PLOS ONE 15:e0229040. https://doi.org/10.1371/journal.pone.0229040

21. Kaniga K, Hasan R, Jou R, et al (2022) Bedaquiline Drug Resistance Emergence Assessment in Multidrug-Resistant Tuberculosis (MDR-TB): a 5-Year Prospective In Vitro Surveillance Study of Bedaquiline and Other Second-Line Drug Susceptibility Testing in MDR-TB Isolates. J Clin Microbiol 60:e02919-20. https://doi.org/10.1128/JCM.02919-20

22. Ushtanit A, Kulagina E, Lyubimova A, et al (2021) Selective pressure diversity and phenotype-genotype correlation in clinical isolates of Mycobacterium tuberculosis. Int J Mycobacteriology 10:16. https://doi.org/10.4103/2212-5531.307065

23. Mortazavi A, Mortazavi SMJ, Sihver L (2021) Selective Pressure-Free Treatments for COVID-19. Radiation 1:18–32. https://doi.org/10.3390/radiation1010003

24. Schön T, Miotto P, Köser CU, et al (2017) Mycobacterium tuberculosis drug-resistance testing: challenges, recent developments and perspectives. Clin Microbiol Infect 23:154–160. https://doi.org/10.1016/j.cmi.2016.10.022

25. Krause PR, Fleming TR, Longini IM, et al (2021) SARS-CoV-2 Variants and Vaccines. N Engl J Med 385:179–186. https://doi.org/10.1056/NEJMsr2105280

26. Konings F, Perkins MD, Kuhn JH, et al (2021) SARS-CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse. Nat Microbiol 6:821–823. https://doi.org/10.1038/s41564-021-00932-w

27. Khateeb J, Li Y, Zhang H (2021) Emerging SARS-CoV-2 variants of concern and potential intervention approaches. Crit Care 25:244. https://doi.org/10.1186/s13054-021-03662-x

28. Thakur V, Ratho RK (2022) OMICRON (B.1.1.529): A new SARS-CoV-2 variant of concern mounting worldwide fear. J Med Virol 94:1821–1824. https://doi.org/10.1002/jmv.27541

29. Bedford J, Farrar J, Ihekweazu C, et al (2019) A new twenty-first century science for effective epidemic response. Nature 575:130–136. https://doi.org/10.1038/s41586-019-1717-y

30. Huremović D (2019) Brief History of Pandemics (Pandemics Throughout History). In: Huremović D (ed) Psychiatry of Pandemics: A Mental Health Response to Infection Outbreak. Springer International Publishing, Cham, pp 7–35

31. Petersen E, Petrosillo N, Koopmans M, et al (2018) Emerging infections—an increasingly important topic: review by the Emerging Infections Task Force. Clin Microbiol Infect 24:369–375. https://doi.org/10.1016/j.cmi.2017.10.035

32. Bloom DE, Cadarette D (2019) Infectious Disease Threats in the Twenty-First Century: Strengthening the Global Response. Front Immunol 10:

33. Piret J, Boivin G (2021) Pandemics Throughout History. Front Microbiol 11:

34. Hall CM, Scott D, Gössling S (2020) Pandemics, transformations and tourism: be careful what you wish for. Tour Geogr 22:577–598. https://doi.org/10.1080/14616688.2020.1759131

35. da Costa VG, Moreli ML, Saivish MV (2020) The emergence of SARS, MERS and novel SARS-2 coronaviruses in the 21st century. Arch Virol 165:1517–1526. https://doi.org/10.1007/s00705-020-04628-0

36. Saleh A, Qamar S, Tekin A, et al Vaccine Development Throughout History. Cureus 13:e16635. https://doi.org/10.7759/cureus.16635

37. Voigt EA, Kennedy RB, Poland GA Defending against smallpox: a focus on vaccines. Expert Rev Vaccines 15:1197–1211. https://doi.org/10.1080/14760584.2016.1175305

38. Desmond A, Offit PA (2021) On the Shoulders of Giants — From Jenner's Cowpox to mRNA Covid Vaccines. N Engl J Med 384:1081–1083. https://doi.org/10.1056/NEJMp2034334

39. Global HIV & AIDS statistics — Fact sheet. https://www.unaids.org/en/resources/fact-sheet. Accessed 6 Feb 2023

40. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int. Accessed 6 Feb 2023

41. Smallpox vaccines. https://www.who.int/news-room/feature-stories/detail/smallpox-vaccines. Accessed 6 Feb 2023

42. Sáinz MP, Pelayo R, Laxe S, et al (2022) Describing post-polio syndrome. Neurol Engl Ed 37:346–354. https://doi.org/10.1016/j.nrleng.2019.03.023

43. Rachlin A, Patel JC, Burns CC, et al (2022) Progress Toward Polio Eradication — Worldwide, January 2020–April 2022. Morb Mortal Wkly Rep 71:650–655. https://doi.org/10.15585/mmwr.mm7119a2

44. The Lancet (2022) Polio eradication: falling at the final hurdle? The Lancet 400:1079. https://doi.org/10.1016/S0140-6736(22)01875-X

45. Lee SE, Greene SA, Burns CC, et al (2023) Progress Toward Poliomyelitis Eradication — Worldwide, January 2021–March 2023. Morb Mortal Wkly Rep 72:517–522. https://doi.org/10.15585/mmwr.mm7219a3

46. Bigouette JP, Henderson E, Traoré MA, et al (2023) Update on Vaccine-Derived Poliovirus Outbreaks - Worldwide, January 2021-December 2022. MMWR Morb Mortal Wkly Rep 72:366–371. https://doi.org/10.15585/mmwr.mm7214a3

47. Saeed M, Ahmad M, Iram S, et al (2017) GeneXpert technology. Saudi Med J 38:699–705. https://doi.org/10.15537/smj.2017.7.17694

48. Pardi N, Hogan MJ, Porter FW, Weissman D (2018) mRNA vaccines — a new era in vaccinology. Nat Rev Drug Discov 17:261–279. https://doi.org/10.1038/nrd.2017.243

49. Extance A (2021) mRNA vaccines: hope beneath the hype. BMJ 375:n2744. https://doi.org/10.1136/bmj.n2744

50. Fackelmann G, Gillingham MAF, Schmid J, et al (2021) Human encroachment into wildlife gut microbiomes. Commun Biol 4:1–11. https://doi.org/10.1038/s42003-021-02315-7

51. Hussain S, Ram MS, Kumar A, et al (2013) Human Presence Increases Parasitic Load in Endangered Lion-Tailed Macaques (Macaca silenus) in Its Fragmented Rainforest Habitats in Southern India. PLOS ONE 8:e63685. https://doi.org/10.1371/journal.pone.0063685

52. San Juan PA, Hendershot JN, Daily GC, Fukami T (2020) Land-use change has host-specific influences on avian gut microbiomes. ISME J 14:318–321. https://doi.org/10.1038/s41396-019-0535-4

53. Cunningham AA, Daszak P, Wood JLN (2017) One Health, emerging infectious diseases and wildlife: two decades of progress? Philos Trans R Soc Lond B Biol Sci 372:20160167. https://doi.org/10.1098/rstb.2016.0167

54. Baker RE, Mahmud AS, Miller IF, et al (2022) Infectious disease in an era of global change. Nat Rev Microbiol 20:193–205. https://doi.org/10.1038/s41579-021-00639-z

55. Mora C, McKenzie T, Gaw IM, et al (2022) Over half of known human pathogenic diseases can be aggravated by climate change. Nat Clim Change 12:869–875. https://doi.org/10.1038/s41558-022-01426-1

56. Carlson CJ, Albery GF, Merow C, et al (2022) Climate change increases cross-species viral transmission risk. Nature 607:555–562. https://doi.org/10.1038/s41586-022-04788-w

57. Khan MD, Thi Vu HH, Lai QT, Ahn JW (2019) Aggravation of Human Diseases and Climate Change Nexus. Int J Environ Res Public Health 16:2799. https://doi.org/10.3390/ijerph16152799

58. Gh ES (2019) CLIMATE CHANGE AND HUMAN INFECTIOUS DISEASES (REVIEW). Egypt J Occup Med 43:33–56. https://doi.org/10.21608/ejom.2019.25106

59. Caminade C, McIntyre KM, Jones AE (2019) Impact of recent and future climate change on vector-borne diseases. Ann N Y Acad Sci 1436:157–173. https://doi.org/10.1111/nyas.13950

60. Liang L, Gong P (2017) Climate change and human infectious diseases: A synthesis of research findings from global and spatio-temporal perspectives. Environ Int 103:99–108. https://doi.org/10.1016/j.envint.2017.03.011

61. Sun X, Wandelt S, Zheng C, Zhang A (2021) COVID-19 pandemic and air transportation: Successfully navigating the paper hurricane. J Air Transp Manag 94:102062. https://doi.org/10.1016/j.jairtraman.2021.102062

62. Ayouni I, Maatoug J, Dhouib W, et al (2021) Effective public health measures to mitigate the spread of COVID-19: a systematic review. BMC Public Health 21:1015. https://doi.org/10.1186/s12889-021-11111-1

63. Hebbani AV, Pulakuntla S, Pannuru P, et al (2021) COVID-19: comprehensive review on mutations and current vaccines. Arch Microbiol 204:8. https://doi.org/10.1007/s00203-021-02606-x

64. Tiwari V, Kumar M, Tiwari A, et al (2021) Current trends in diagnosis and treatment strategies of COVID-19 infection. Environ Sci Pollut Res 28:64987–65013. https://doi.org/10.1007/s11356-021-16715-z

65. Madhi SA, Baillie V, Cutland CL, et al (2021) Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. N Engl J Med 384:1885–1898. https://doi.org/10.1056/NEJMoa2102214

66. Heinz FX, Stiasny K (2021) Profiles of current COVID-19 vaccines. Wien Klin Wochenschr 133:271–283. https://doi.org/10.1007/s00508-021-01835-w

67. Malik JA, Ahmed S, Mir A, et al (2022) The SARS-CoV-2 mutations versus vaccine effectiveness: New opportunities to new challenges. J Infect Public Health 15:228–240. https://doi.org/10.1016/j.jiph.2021.12.014

68. Vasireddy D, Vanaparthy R, Mohan G, et al (2021) Review of COVID-19 Variants and COVID-19 Vaccine Efficacy: What the Clinician Should Know? J Clin Med Res 13:317–325. https://doi.org/10.14740/jocmr4518

69. Tenaillon O, Matic I (2020) The Impact of Neutral Mutations on Genome Evolvability. Curr Biol 30:R527–R534. https://doi.org/10.1016/j.cub.2020.03.056

70. Duffy S (2018) Why are RNA virus mutation rates so damn high? PLoS Biol 16:e3000003. https://doi.org/10.1371/journal.pbio.3000003

71. Chaplin A, Blundell T, Asma M, et al (2020) Using Cryo-EM to understand antimycobacterial resistance in the catalase-peroxidase (KatG) from Mycobacterium tuberculosis. https://doi.org/10.17863/CAM.62351

72. Li M, Lu J, Lu Y, et al (2021) rpoB Mutations and Effects on Rifampin Resistance in Mycobacterium tuberculosis. Infect Drug Resist 14:4119–4128. https://doi.org/10.2147/IDR.S333433

73. Hillen HS (2021) Structure and function of SARS-CoV-2 polymerase. Curr Opin Virol 48:82–90. https://doi.org/10.1016/j.coviro.2021.03.010

74. Hu Y, Lewandowski EM, Tan H, et al Naturally occurring mutations of SARS-CoV-2 main protease confer drug resistance to nirmatrelvir. bioRxiv. https://doi.org/10.1101/2022.06.28.497978

75. Ullrich S, Ekanayake KB, Otting G, Nitsche C (2022) Main protease mutants of SARS-CoV-2 variants remain susceptible to nirmatrelvir. Bioorg Med Chem Lett 62:128629. https://doi.org/10.1016/j.bmcl.2022.128629

76. Sheik Amamuddy O, Afriyie Boateng R, Barozi V, et al (2021) Novel dynamic residue network analysis approaches to study allosteric modulation: SARS-CoV-2 Mpro and its evolutionary mutations as a case study. Comput Struct Biotechnol J 19:6431–6455. https://doi.org/10.1016/j.csbj.2021.11.016

77. Karim SSA, Karim QA (2021) Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. The Lancet 398:2126–2128. https://doi.org/10.1016/S0140-6736(21)02758-6

78. Dhawan M, Priyanka, Choudhary OP (2022) Omicron SARS-CoV-2 variant: Reasons of emergence and lessons learnt. Int J Surg Lond Engl 97:106198. https://doi.org/10.1016/j.ijsu.2021.106198

79. Kannan SR, Spratt AN, Sharma K, et al (2022) Omicron SARS-CoV-2 variant: Unique features and their impact on pre-existing antibodies. J Autoimmun 126:102779. https://doi.org/10.1016/j.jaut.2021.102779

80. He X, Hong W, Pan X, et al (2021) SARS-CoV-2 Omicron variant: Characteristics and prevention. MedComm 2:838–845. https://doi.org/10.1002/mco2.110

81. Fan Y, Li X, Zhang L, et al (2022) SARS-CoV-2 Omicron variant: recent progress and future perspectives. Signal Transduct Target Ther 7:1–11. https://doi.org/10.1038/s41392-022-00997-x

82. Seera S, Nagarajaram HA (2022) Effect of Disease Causing Missense Mutations on Intrinsically Disordered Regions in Proteins. Protein Pept Lett 29:254–267. https://doi.org/10.2174/0929866528666211126161200

83. Kumar A, Biswas P (2019) Effect of Correlated Pair Mutations in Protein Misfolding. J Phys Chem B 123:5069–5078. https://doi.org/10.1021/acs.jpcb.9b03533

84. Shanthirabalan S, Chomilier J, Carpentier M (2018) Structural effects of point mutations in proteins. Proteins 86:853–867. https://doi.org/10.1002/prot.25499

85. Tee W-V, Guarnera E, Berezovsky IN (2019) On the Allosteric Effect of nsSNPs and the Emerging Importance of Allosteric Polymorphism. J Mol Biol 431:3933–3942. https://doi.org/10.1016/j.jmb.2019.07.012

86. Tastan Bishop Ö, Musyoka TM, Barozi V (2022) Allostery and Missense Mutations as Intermittently Linked Promising Aspects of Modern Computational Drug Discovery. J Mol Biol 167610. https://doi.org/10.1016/j.jmb.2022.167610

87. Portelli S, Phelan JE, Ascher DB, et al (2018) Understanding molecular consequences of putative drug resistant mutations in Mycobacterium tuberculosis. Sci Rep 8:15356. https://doi.org/10.1038/s41598-018-33370-6

88. Ando H, Kondo Y, Suetake T, et al (2010) Identification of katG Mutations Associated with High-Level Isoniazid Resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 54:1793–1799. https://doi.org/10.1128/AAC.01691-09

89. Das JK, Thakuri B, MohanKumar K, et al (2021) Mutation-Induced Long-Range Allosteric Interactions in the Spike Protein Determine the Infectivity of SARS-CoV-2 Emerging Variants. ACS Omega 6:31305–31320. https://doi.org/10.1021/acsomega.1c05155

90. Verkhivker GM (2020) Molecular Simulations and Network Modeling Reveal an Allosteric Signaling in the SARS-CoV-2 Spike Proteins. J Proteome Res 19:4587–4608. https://doi.org/10.1021/acs.jproteome.0c00654

91. Liu J, Nussinov R (2016) Allostery: An Overview of Its History, Concepts, Methods, and Applications. PLoS Comput Biol 12:e1004966. https://doi.org/10.1371/journal.pcbi.1004966

92. Zhang F, Yuan Y, Xiang M, et al (2019) Molecular Mechanism Regarding Allosteric Modulation of Ligand Binding and the Impact of Mutations on Dimerization for CCR5 Homodimer. J Chem Inf Model 59:1965–1976. https://doi.org/10.1021/acs.jcim.8b00850

93. Nussinov R, Tsai C-J (2013) Allostery in Disease and in Drug Discovery. Cell 153:293–305. https://doi.org/10.1016/j.cell.2013.03.034

94. Guarnera E, Berezovsky IN (2020) Allosteric drugs and mutations: chances, challenges, and necessity. Curr Opin Struct Biol 62:149–157. https://doi.org/10.1016/j.sbi.2020.01.010

95. Sheik Amamuddy O, Bishop NT, Tastan Bishop Ö (2018) Characterizing early drug resistance-related events using geometric ensembles from HIV protease dynamics. Sci Rep 8:17938. https://doi.org/10.1038/s41598-018-36041-8

96. Tan ZW, Tee W-V, Samsudin F, et al (2022) Allosteric perspective on the mutability and druggability of the SARS-CoV-2 Spike protein. Structure 30:590-607.e4. https://doi.org/10.1016/j.str.2021.12.011

97. Tan ZW, Guarnera E, Tee W-V, Berezovsky IN (2020) AlloSigMA 2: paving the way to designing allosteric effectors and to exploring allosteric effects of mutations. Nucleic Acids Res 48:W116–W124. https://doi.org/10.1093/nar/gkaa338

98. Goncearenco A, Mitternacht S, Yong T, et al (2013) SPACER: Server for predicting allosteric communication and effects of regulation. Nucleic Acids Res 41:W266-272. https://doi.org/10.1093/nar/gkt460

99. Tan ZW, Tee W-V, Guarnera E, et al (2019) AlloMAPS: allosteric mutation analysis and polymorphism of signaling database. Nucleic Acids Res 47:D265–D270. https://doi.org/10.1093/nar/gky1028

100. Murray CJ, Ikuta KS, Sharara F, et al (2022) Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet 399:629–655. https://doi.org/10.1016/S0140-6736(21)02724-0

101. Antimicrobial resistance. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance. Accessed 7 Jul 2022

102. Tiberi S, Utjesanovic N, Galvin J, et al (2022) Drug resistant TB – latest developments in epidemiology, diagnostics and management. Int J Infect Dis. https://doi.org/10.1016/j.ijid.2022.03.026

103. Wallis CL, Godfrey C, Fitzgibbon JE, Mellors JW (2017) Key Factors Influencing the Emergence of Human Immunodeficiency Virus Drug Resistance in Low- and Middle-Income Countries. J Infect Dis 216:S851–S856. https://doi.org/10.1093/infdis/jix409

104. Kurt Yilmaz N, Schiffer CA (2021) Introduction: Drug Resistance. Chem Rev 121:3235–3237. https://doi.org/10.1021/acs.chemrev.1c00118

105. Chokshi A, Sifri Z, Cennimo D, Horng H (2019) Global Contributors to Antibiotic Resistance. J Glob Infect Dis 11:36–42. https://doi.org/10.4103/jgid.jgid_110_18

106. Sulis G, Sayood S, Gandra S (2022) Antimicrobial resistance in low- and middle-income countries: current status and future directions. Expert Rev Anti Infect Ther 20:147–160. https://doi.org/10.1080/14787210.2021.1951705

107. Singh S, Charani E, Devi S, et al (2021) A road-map for addressing antimicrobial resistance in low- and middle-income countries: lessons learnt from the public private participation and co-designed antimicrobial stewardship programme in the State of Kerala, India. Antimicrob Resist Infect Control 10:32. https://doi.org/10.1186/s13756-020-00873-9

108. Iskandar K, Molinier L, Hallit S, et al (2020) Drivers of Antibiotic Resistance Transmission in Low- and Middle-Income Countries from a "One Health" Perspective—A Review. Antibiotics 9:372. https://doi.org/10.3390/antibiotics9070372

109. Hameed HMA, Islam MM, Chhotaray C, et al (2018) Molecular Targets Related Drug Resistance Mechanisms in MDR-, XDR-, and TDR-Mycobacterium tuberculosis Strains. Front Cell Infect Microbiol 8:

110. Wong-Sam A, Wang Y-F, Zhang Y, et al (2018) Drug Resistance Mutation L76V Alters Nonpolar Interactions at the Flap–Core Interface of HIV-1 Protease. ACS Omega 3:12132–12140. https://doi.org/10.1021/acsomega.8b01683

111. Nayak C, Chandra I, Singh SK (2019) An in silico pharmacological approach toward the discovery of potent inhibitors to combat drug resistance HIV-1 protease variants. J Cell Biochem 120:9063–9081. https://doi.org/10.1002/jcb.28181

112. Junaid M, Khan MT, Malik SI, Wei D-Q (2019) Insights into the Mechanisms of the Pyrazinamide Resistance of Three Pyrazinamidase Mutants N11K, P69T, and D126N. J Chem Inf Model 59:498–508. https://doi.org/10.1021/acs.jcim.8b00525

113. Khan MT, Khan A, Rehman AU, et al (2019) Structural and free energy landscape of novel mutations in ribosomal protein S1 (rpsA) associated with pyrazinamide resistance. Sci Rep 9:7482. https://doi.org/10.1038/s41598-019-44013-9

114. Macalino SJY, Billones JB, Organo VG, Carrillo MCO (2020) In Silico Strategies in Tuberculosis Drug Discovery. Molecules 25:665. https://doi.org/10.3390/molecules25030665

115. Barozi V, Musyoka TM, Sheik Amamuddy O, Tastan Bishop Ö (2022) Deciphering Isoniazid Drug Resistance Mechanisms on Dimeric Mycobacterium tuberculosis KatG via Post-molecular Dynamics Analyses Including Combined Dynamic Residue Network Metrics. ACS Omega. https://doi.org/10.1021/acsomega.2c01036

116. Lin X, Li X, Lin X (2020) A Review on Applications of Computational Methods in Drug Screening and Design. Molecules 25:1375. https://doi.org/10.3390/molecules25061375

117. Malathi K, Ramaiah S (2018) Bioinformatics approaches for new drug discovery: a review. Biotechnol Genet Eng Rev 34:243–260. https://doi.org/10.1080/02648725.2018.1502984

118. Agamah FE, Mazandu GK, Hassan R, et al (2020) Computational/in silico methods in drug target and lead prediction. Brief Bioinform 21:1663–1675. https://doi.org/10.1093/bib/bbz103

119. Shahbaaz M, Qari SH, Abdellattif MH, Hussien MA (2022) Structural analyses and classification of novel isoniazid resistance coupled mutational landscapes in Mycobacterium tuberculosis: a combined molecular docking and MD simulation study. J Biomol Struct Dyn 40:4791–4800. https://doi.org/10.1080/07391102.2020.1861986

120. Mugumbate G, Nyathi B, Zindoga A, Munyuki G (2021) Application of Computational Methods in Understanding Mutations in Mycobacterium tuberculosis Drug Resistance. Front Mol Biosci 8:

121. Vidossich P, Loewen PC, Carpena X, et al (2014) Binding of the Antitubercular Pro-Drug Isoniazid in the Heme Access Channel of Catalase-Peroxidase (KatG). A Combined Structural and Metadynamics Investigation. J Phys Chem B 118:2924–2931. https://doi.org/10.1021/jp4123425

122. Kapetanaki S, Chouchane S, Girotto S, et al (2003) Conformational Differences in Mycobacterium tuberculosis Catalase-Peroxidase KatG and Its S315T Mutant Revealed by Resonance Raman Spectroscopy. Biochemistry 42:3835–3845. https://doi.org/10.1021/bi026992y

123. Ramasubban G, Therese KL, Vetrivel U, et al (2011) Detection of novel coupled mutations in the katG gene (His276Met, Gln295His and Ser315Thr) in a multidrug-resistant Mycobacterium tuberculosis strain from Chennai, India, and insight into the molecular mechanism of isoniazid resistance using structural bioinformatics approaches. Int J Antimicrob Agents 37:368–372. https://doi.org/10.1016/j.ijantimicag.2010.11.023

124. da Cunha EFF, Ramalho TC, de Alencastro RB, Maia ER (2007) Docking simulations and QM/MM studies between isoniazid prodrug, catalase-peroxidase (KatG) and S315T mutant from Mycobacterium tuberculosis. Comput Math Methods Med 8:113–124. https://doi.org/10.1080/17486700701374292

125. Alame Emane AK, Guo X, Takiff HE, Liu S (2021) Drug resistance, fitness and compensatory mutations in Mycobacterium tuberculosis. Tuberculosis 129:102091. https://doi.org/10.1016/j.tube.2021.102091

126. Marney MW, Metzger RP, Hecht D, Valafar F (2018) Modeling the structural origins of drug resistance to isoniazid via key mutations in Mycobacterium tuberculosis catalase-peroxidase, KatG. Tuberculosis 108:155–162. https://doi.org/10.1016/j.tube.2017.11.007

127. Purkan P, Ihsanawati I, Natalia D, et al (2018) Molecular Analysis of katG Encoding Catalase-Peroxidase from Clinical Isolate of Isoniazid-Resistant Mycobacterium tuberculosis. J Med Life 11:160–167

128. Srivastava G, Tripathi S, Kumar A, Sharma A (2017) Molecular investigation of active binding site of isoniazid (INH) and insight into resistance mechanism of S315T-MtKatG in Mycobacterium tuberculosis. Tuberculosis 105:18–27. https://doi.org/10.1016/j.tube.2017.04.002

129. Pimentel AL, de Lima Scodro RB, Caleffi-Ferracioli KR, et al (2017) Mutations in catalase-peroxidase KatG from isoniazid resistant Mycobacterium tuberculosis clinical isolates: insights from molecular dynamics simulations. J Mol Model 23:121. https://doi.org/10.1007/s00894-017-3290-3

130. Vats C, Dhanjal JK, Goyal S, et al (2015) Mechanistic analysis elucidating the relationship between Lys96 mutation in Mycobacterium tuberculosispyrazinamidase enzyme and pyrazinamide susceptibility. BMC Genomics 16:S14. https://doi.org/10.1186/1471-2164-16-S2-S14

131. Mehmood A, Khan MT, Kaushik AC, et al (2019) Structural Dynamics Behind Clinical Mutants of PncA-Asp12Ala, Pro54Leu, and His57Pro of Mycobacterium tuberculosis Associated With Pyrazinamide Resistance. Front Bioeng Biotechnol 7:. https://doi.org/10.3389/fbioe.2019.00404

132. Unissa AN, Doss C GP, Kumar T, et al (2018) Significance of catalase-peroxidase (KatG) mutations in mediating isoniazid resistance in clinical strains of Mycobacterium tuberculosis. J Glob Antimicrob Resist 15:111–120. https://doi.org/10.1016/j.jgar.2018.07.001

133. Singh A, Singh A, Grover S, et al (2018) Wild-type catalase peroxidase vs G279D mutant type: Molecular basis of Isoniazid drug resistance in Mycobacterium tuberculosis. Gene 641:226–234. https://doi.org/10.1016/j.gene.2017.10.047

134. Ali A, Khan M, Khan A, et al (2020) Pyrazinamide resistance of novel mutations in pncA and their dynamic behavior. RSC Adv 10:. https://doi.org/10.1039/d0ra06072k

135. Sheik Amamuddy O, Musyoka TM, Boateng RA, et al (2020) Determining the unbinding events and conserved motions associated with the pyrazinamide release due to resistance mutations of Mycobacterium tuberculosis pyrazinamidase. Comput Struct Biotechnol J 18:1103–1120. https://doi.org/10.1016/j.csbj.2020.05.009

136. Kavvas ES, Catoiu E, Mih N, et al (2018) Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. Nat Commun 9:4306. https://doi.org/10.1038/s41467-018-06634-y

137. Boolchandani M, D'Souza AW, Dantas G (2019) Sequencing-based methods and resources to study antimicrobial resistance. Nat Rev Genet 20:356–370. https://doi.org/10.1038/s41576-019-0108-4

138. Davis JJ, Boisvert S, Brettin T, et al (2016) Antimicrobial Resistance Prediction in PATRIC and RAST. Sci Rep 6:27930. https://doi.org/10.1038/srep27930

139. Arango-Argoty G, Garner E, Pruden A, et al (2018) DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome 6:23. https://doi.org/10.1186/s40168-018-0401-z

140. Meehan CJ, Goig GA, Kohl TA, et al (2019) Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. Nat Rev Microbiol 17:533–545. https://doi.org/10.1038/s41579-019-0214-5

141. Deelder W, Christakoudi S, Phelan J, et al (2019) Machine Learning Predicts Accurately Mycobacterium tuberculosis Drug Resistance From Whole Genome Sequencing Data. Front Genet 10:

142. Carter JJ, Walker TM, Walker AS, et al (2019) Prediction of pyrazinamide resistance in Mycobacterium tuberculosis using structure-based machine learning approaches. 518142

143. Deelder W, Napier G, Campino S, et al (2022) A modified decision tree approach to improve the prediction and mutation discovery for drug resistance in Mycobacterium tuberculosis. BMC Genomics 23:46. https://doi.org/10.1186/s12864-022-08291-4

144. Green AG, Yoon CH, Chen ML, et al (2022) A convolutional neural network highlights mutations relevant to antimicrobial resistance in Mycobacterium tuberculosis. Nat Commun 13:3817. https://doi.org/10.1038/s41467-022-31236-0

145. Jiang Z, Lu Y, Liu Z, et al (2022) Drug resistance prediction and resistance genes identification in Mycobacterium tuberculosis based on a hierarchical attentive neural network utilizing genome-wide variants. Brief Bioinform 23:bbac041. https://doi.org/10.1093/bib/bbac041

146. Jamal S, Khubaib M, Gangwar R, et al (2020) Artificial Intelligence and Machine learning based prediction of resistant and susceptible mutations in Mycobacterium tuberculosis. Sci Rep 10:5487. https://doi.org/10.1038/s41598-020-62368-2

147. Yang Y, Walker TM, Walker AS, et al (2019) DeepAMR for predicting co-occurrent resistance of Mycobacterium tuberculosis. Bioinformatics 35:3240–3249. https://doi.org/10.1093/bioinformatics/btz067

148. Camacho DM, Collins KM, Powers RK, et al (2018) Next-Generation Machine Learning for Biological Networks. Cell 173:1581–1592. https://doi.org/10.1016/j.cell.2018.05.015

149. Yang Y, Walker TM, Kouchaki S, et al (2021) An end-to-end heterogeneous graph attention network for Mycobacterium tuberculosis drug-resistance prediction. Brief Bioinform 22:bbab299. https://doi.org/10.1093/bib/bbab299

150. Gröschel MI, Owens M, Freschi L, et al (2021) GenTB: A user-friendly genome-based predictor for tuberculosis resistance powered by machine learning. Genome Med 13:138. https://doi.org/10.1186/s13073-021-00953-4

151. Muzondiwa D, Mutshembele A, Pierneef RE, Reva ON (2020) Resistance Sniffer: An online tool for prediction of drug resistance patterns of Mycobacterium tuberculosis isolates using next generation sequencing data. Int J Med Microbiol 310:151399. https://doi.org/10.1016/j.ijmm.2020.151399

152. Penkler DL, Atilgan C, Tastan Bishop Ö (2018) Allosteric Modulation of Human Hsp90α Conformational Dynamics. J Chem Inf Model 58:383–404. https://doi.org/10.1021/acs.jcim.7b00630

153. Amusengeri A, Tata RB, Tastan Bishop Ö (2020) Understanding the Pyrimethamine Drug Resistance Mechanism via Combined Molecular Dynamics and Dynamic Residue Network Analysis. Molecules 25:904. https://doi.org/10.3390/molecules25040904

154. Sanyanga TA, Nizami B, Tastan Bishop Ö (2019) Mechanism of Action of Non-Synonymous Single Nucleotide Variations Associated with α-Carbonic Anhydrase II Deficiency. Molecules 24:3987. https://doi.org/10.3390/molecules24213987

155. Okeke CJ, Musyoka TM, Sheik Amamuddy O, et al (2021) Allosteric pockets and dynamic residue network hubs of falcipain 2 in mutations including those linked to artemisinin resistance. Comput Struct Biotechnol J 19:5647–5666. https://doi.org/10.1016/j.csbj.2021.10.011

156. Sanyanga TA, Tastan Bishop Ö (2020) Structural Characterization of Carbonic Anhydrase VIII and Effects of Missense Single Nucleotide Variations to Protein Structure and Function. Int J Mol Sci 21:2764. https://doi.org/10.3390/ijms21082764

157. Shu Y, McCauley J (2017) GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance 22:30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

158. Zhang Y-Z, Holmes EC (2020) A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. Cell 181:223–227. https://doi.org/10.1016/j.cell.2020.03.035

159. Lu R, Zhao X, Li J, et al (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 395:565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

160. Zhu N, Zhang D, Wang W, et al (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017

161. Wu F, Zhao S, Yu B, et al (2020) A new coronavirus associated with human respiratory disease in China. Nature 579:265–269. https://doi.org/10.1038/s41586-020-2008-3

162. Zhou P, Yang X-L, Wang X-G, et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7

163. Akkiz H (2021) Implications of the Novel Mutations in the SARS-CoV-2 Genome for Transmission, Disease Severity, and the Vaccine Development. Front Med 8:636532. https://doi.org/10.3389/fmed.2021.636532

164. Escalera A, Gonzalez-Reiche AS, Aslam S, et al (2022) Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage and virus transmission. Cell Host Microbe 30:373-387.e7. https://doi.org/10.1016/j.chom.2022.01.006

165. Ren S-Y, Wang W-B, Gao R-D, Zhou A-M (2022) Omicron variant (B.1.1.529) of SARS-CoV-2: Mutation, infectivity, transmission, and vaccine resistance. World J Clin Cases 10:1–11. https://doi.org/10.12998/wjcc.v10.i1.1

166. Franceschi VB, Caldana GD, de Menezes Mayer A, et al (2021) Genomic epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil. BMC Genomics 22:371. https://doi.org/10.1186/s12864-021-07708-w

167. Zhou W, Xu C, Wang P, et al (2022) Impact of mutations in SARS-COV-2 spike on viral infectivity and antigenicity. Brief Bioinform 23:bbab375. https://doi.org/10.1093/bib/bbab375

168. Niu Z, Zhang Z, Gao X, et al (2021) N501Y mutation imparts cross-species transmission of SARS-CoV-2 to mice by enhancing receptor binding. Signal Transduct Target Ther 6:1–3. https://doi.org/10.1038/s41392-021-00704-2

169. Korber B, Fischer WM, Gnanakaran S, et al (2020) Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 182:812-827.e19. https://doi.org/10.1016/j.cell.2020.06.043

170. Jiang X, Zhang Z, Wang C, et al (2020) Bimodular effects of D614G mutation on the spike glycoprotein of SARS-CoV-2 enhance protein processing, membrane fusion, and viral infectivity. Signal Transduct Target Ther 5:1–3. https://doi.org/10.1038/s41392-020-00392-4

171. Loucera C, Perez-Florido J, Casimiro-Soriguer CS, et al (2022) Assessing the Impact of SARS-CoV-2 Lineages and Mutations on Patient Survival. Viruses 14:1893. https://doi.org/10.3390/v14091893

172. Nagy Á, Pongor S, Győrffy B (2021) Different mutations in SARS-CoV-2 associate with severe and mild outcome. Int J Antimicrob Agents 57:106272. https://doi.org/10.1016/j.ijantimicag.2020.106272

173. Dubey A, Choudhary S, Kumar P, Tomar S (2021) Emerging SARS-CoV-2 Variants: Genetic Variability and Clinical Implications. Curr Microbiol 79:20. https://doi.org/10.1007/s00284-021-02724-1

174. Volz E, Hill V, McCrone JT, et al (2021) Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell 184:64-75.e11. https://doi.org/10.1016/j.cell.2020.11.020

175. Frampton D, Rampling T, Cross A, et al (2021) Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. Lancet Infect Dis 21:1246–1256. https://doi.org/10.1016/S1473-3099(21)00170-5

176. Akkız H (2022) The Biological Functions and Clinical Significance of SARS-CoV-2 Variants of Corcern. Front Med 9:849217. https://doi.org/10.3389/fmed.2022.849217

177. Becerra-Flores M, Cardozo T (2020) SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. Int J Clin Pract 74:e13525. https://doi.org/10.1111/ijcp.13525

178. Qin L, Ding X, Li Y, et al (2021) Co-mutation modules capture the evolution and transmission patterns of SARS-CoV-2. Brief Bioinform 22:bbab222. https://doi.org/10.1093/bib/bbab222

179. Freer G, Lai M, Quaranta P, et al Evolution of viruses and the emergence of SARS-CoV-2 variants

180. Giovanetti M, Benedetti F, Campisi G, et al (2021) Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. Biochem Biophys Res Commun 538:88–91. https://doi.org/10.1016/j.bbrc.2020.10.102

181. Subissi L, von Gottberg A, Thukral L, et al (2022) An early warning system for emerging SARS-CoV-2 variants. Nat Med 28:1110–1115. https://doi.org/10.1038/s41591-022-01836-w

182. Markov PV, Katzourakis A, Stilianakis NI (2022) Antigenic evolution will lead to new SARS-CoV-2 variants with unpredictable severity. Nat Rev Microbiol 20:251–252. https://doi.org/10.1038/s41579-022-00722-z

183. da Costa CHS, de Freitas CAB, Alves CN, Lameira J (2022) Assessment of mutations on RBD in the Spike protein of SARS-CoV-2 Alpha, Delta and Omicron variants. Sci Rep 12:8540. https://doi.org/10.1038/s41598-022-12479-9

184. Celik I, Khan A, Dwivany FM, et al (2022) Computational prediction of the effect of mutations in the receptor-binding domain on the interaction between SARS-CoV-2 and human ACE2. Mol Divers 26:3309–3324. https://doi.org/10.1007/s11030-022-10392-x

185. Wozney AJ, Smith MA, Abdrabbo M, et al (2022) Evolution of Stronger SARS-CoV-2 Variants as Revealed Through the Lens of Molecular Dynamics Simulations. Protein J 41:444–456. https://doi.org/10.1007/s10930-022-10065-6

186. Cheng MH, Krieger JM, Banerjee A, et al (2022) Impact of new variants on SARS-CoV-2 infectivity and neutralization: A molecular assessment of the alterations in the spike-host protein interactions. iScience 25:103939. https://doi.org/10.1016/j.isci.2022.103939

187. Verma J, Subbarao N (2021) Insilico study on the effect of SARS-CoV-2 RBD hotspot mutants' interaction with ACE2 to understand the binding affinity and stability. Virology 561:107–116. https://doi.org/10.1016/j.virol.2021.06.009

188. Aljindan RY, Al-Subaie AM, Al-Ohali AI, et al (2021) Investigation of nonsynonymous mutations in the spike protein of SARS-CoV-2 and its interaction with the ACE2 receptor by molecular docking and MM/GBSA approach. Comput Biol Med 135:104654. https://doi.org/10.1016/j.compbiomed.2021.104654

189. Celik I, Yadav R, Duzgun Z, et al (2021) Interactions of the Receptor Binding Domain of SARS-CoV-2 Variants with hACE2: Insights from Molecular Docking Analysis and Molecular Dynamic Simulation. Biology 10:880. https://doi.org/10.3390/biology10090880

190. Dong R, Hu T, Zhang Y, et al (2022) Assessing the Transmissibility of the New SARS-CoV-2 Variants: From Delta to Omicron. Vaccines 10:496. https://doi.org/10.3390/vaccines10040496

191. Saxena SK, Kumar S, Ansari S, et al (2022) Characterization of the novel SARS-CoV-2 Omicron (B.1.1.529) variant of concern and its global perspective. J Med Virol 94:1738–1744. https://doi.org/10.1002/jmv.27524

192. Araf Y, Akter F, Tang Y, et al (2022) Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. J Med Virol 94:1825–1832. https://doi.org/10.1002/jmv.27588

193. Willett BJ, Grove J, MacLean OA, et al (2022) SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. Nat Microbiol 7:1161–1179. https://doi.org/10.1038/s41564-022-01143-7

194. Luan B, Wang H, Huynh T (2021) Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. FEBS Lett 595:1454–1461. https://doi.org/10.1002/1873-3468.14076

195. Starr TN, Greaney AJ, Hilton SK, et al (2020) Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell 182:1295-1310.e20. https://doi.org/10.1016/j.cell.2020.08.012

196. Tian F, Tong B, Sun L, et al (2021) N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. eLife 10:e69091. https://doi.org/10.7554/eLife.69091

197. Deng X, Garcia-Knight MA, Khalid MM, et al (2021) Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. Cell 184:3426-3437.e8. https://doi.org/10.1016/j.cell.2021.04.025

198. Chakraborty C, Sharma AR, Bhattacharya M, Lee S-S (2022) A Detailed Overview of Immune Escape, Antibody Escape, Partial Vaccine Escape of SARS-CoV-2 and Their Emerging Variants With Escape Mutations. Front Immunol 13:

199. Wu L, Peng C, Yang Y, et al (2022) Exploring the immune evasion of SARS-CoV-2 variant harboring E484K by molecular dynamics simulations. Brief Bioinform 23:bbab383. https://doi.org/10.1093/bib/bbab383

200. Kullappan M, Mary U, Ambrose JM, et al (2021) Elucidating the role of N440K mutation in SARS-CoV-2 spike – ACE-2 binding affinity and COVID-19 severity by virtual screening, molecular docking and dynamics approach. J Biomol Struct Dyn 0:1–18. https://doi.org/10.1080/07391102.2021.2014973

201. Tuekprakhon A, Nutalai R, Dijokaite-Guraliuc A, et al (2022) Antibody escape of SARS-CoV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. Cell 185:2422-2433.e13. https://doi.org/10.1016/j.cell.2022.06.005

202. Pitsillou E, Liang JJ, Beh RC, et al (2022) Molecular dynamics simulations highlight the altered binding landscape at the spike-ACE2 interface between the Delta and Omicron variants compared to the SARS-CoV-2 original strain. Comput Biol Med 149:106035. https://doi.org/10.1016/j.compbiomed.2022.106035

203. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G (2022) Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. J Med Virol 94:1641–1649. https://doi.org/10.1002/jmv.27526

204. Li Q, Cao Z, Rahman P (2020) Genetic variability of human angiotensin-converting enzyme 2 (hACE2) among various ethnic populations. Mol Genet Genomic Med 8:e1344. https://doi.org/10.1002/mgg3.1344

205. Tahsin A, Ahmed R, Bhattacharjee P, et al (2022) Most frequently harboured missense variants of hACE2 across different populations exhibit varying patterns of binding interaction with spike

glycoproteins of emerging SARS-CoV-2 of different lineages. Comput Biol Med 148:105903. https://doi.org/10.1016/j.compbiomed.2022.105903

206. Hadi-Alijanvand H, Rouhani M (2020) Studying the Effects of ACE2 Mutations on the Stability, Dynamics, and Dissociation Process of SARS-CoV-2 S1/hACE2 Complexes. J Proteome Res 19:4609–4623. https://doi.org/10.1021/acs.jproteome.0c00348

207. Suryamohan K, Diwanji D, Stawiski EW, et al (2021) Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2. Commun Biol 4:1–11. https://doi.org/10.1038/s42003-021-02030-3

208. Calcagnile M, Forgez P, Iannelli A, et al (2021) Molecular docking simulation reveals ACE2 polymorphisms that may increase the affinity of ACE2 with the SARS-CoV-2 Spike protein. Biochimie 180:143–148. https://doi.org/10.1016/j.biochi.2020.11.004

209. Paik H, Kim J, Seo S (2022) Analysis of the docking property of host variants of hACE2 for SARS-CoV-2 in a large cohort. PLOS Comput Biol 18:e1009834. https://doi.org/10.1371/journal.pcbi.1009834

210. Ashoor D, Ben Khalaf N, Marzouq M, et al (2021) A Computational Approach to Evaluate the Combined Effect of SARS-CoV-2 RBD Mutations and ACE2 Receptor Genetic Variants on Infectivity: The COVID-19 Host-Pathogen Nexus. Front Cell Infect Microbiol 11:

211. Calcagnile M, Forgez P, Iannelli A, et al (2020) ACE2 polymorphisms and individual susceptibility to SARS-CoV-2 infection: insights from an in silico study. 2020.04.23.057042

212. Hattori T, Saito T, Okuya K, et al (2022) Human ACE2 Genetic Polymorphism Affecting SARS-CoV and SARS-CoV-2 Entry into Cells. Microbiol Spectr 10:e00870-22. https://doi.org/10.1128/spectrum.00870-22

213. Hu P, Bauer VL, Sawyer SL, Diaz-Griffero F (2022) Human ACE2 Polymorphisms from Different Human Populations Modulate SARS-CoV-2 Infection. Viruses 14:1451. https://doi.org/10.3390/v14071451

214. Hancock JT, Rouse RC, Stone E, Greenhough A (2021) Interacting Proteins, Polymorphisms and the Susceptibility of Animals to SARS-CoV-2. Animals 11:797. https://doi.org/10.3390/ani11030797

215. Hashizume M, Gonzalez G, Ono C, et al (2021) Population-Specific ACE2 Single-Nucleotide Polymorphisms Have Limited Impact on SARS-CoV-2 Infectivity In Vitro. Viruses 13:67. https://doi.org/10.3390/v13010067

216. Ovsyannikova IG, Haralambieva IH, Crooke SN, et al (2020) The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. Immunol Rev 296:205–219. https://doi.org/10.1111/imr.12897

217. Mphande-Nyasulu FA, Puengpipattrakul P, Praipruksaphan M, et al (2022) Prevalence of tuberculosis (TB), including multi-drug-resistant and extensively-drug-resistant TB, and association with occupation in adults at Sirindhorn Hospital, Bangkok. IJID Reg 2:141–148. https://doi.org/10.1016/j.ijregi.2022.01.004

218. Uffelmann E, Huang QQ, Munung NS, et al (2021) Genome-wide association studies. Nat Rev Methods Primer 1:1–21. https://doi.org/10.1038/s43586-021-00056-9

219. Cano-Gamez E, Trynka G (2020) From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:

220. Tam V, Patel N, Turcotte M, et al (2019) Benefits and limitations of genome-wide association studies. Nat Rev Genet 20:467–484. https://doi.org/10.1038/s41576-019-0127-1

221. Burley SK, Bhikadiya C, Bi C, et al (2021) RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Res 49:D437–D451. https://doi.org/10.1093/nar/gkaa1038

222. Chatzou M, Magis C, Chang J-M, et al (2016) Multiple sequence alignment modeling: methods and applications. Brief Bioinform 17:1009–1023. https://doi.org/10.1093/bib/bbv099

223. Chowdhury B, Garai G (2017) A review on multiple sequence alignment from the perspective of genetic algorithm. Genomics 109:419–431. https://doi.org/10.1016/j.ygeno.2017.06.007

224. Thompson JD, Linard B, Lecompte O, Poch O (2011) A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives. PloS One 6:e18093. https://doi.org/10.1371/journal.pone.0018093

225. Barton C, Flouri T, Iliopoulos CS, Pissis SP (2015) Global and local sequence alignment with a bounded number of gaps. Theor Comput Sci 582:1–16. https://doi.org/10.1016/j.tcs.2015.03.016

226. Huang M, Shah ND, Yao L (2019) Evaluating global and local sequence alignment methods for comparing patient medical records. BMC Med Inform Decis Mak 19:263. https://doi.org/10.1186/s12911-019-0965-y

227. Xiong J (2006) Essential Bioinformatics. Cambridge University Press, Cambridge

228. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

229. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

230. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436

231. Blackshields G, Sievers F, Shi W, et al (2010) Sequence embedding for fast construction of guide trees for multiple sequence alignment. Algorithms Mol Biol AMB 5:21. https://doi.org/10.1186/1748-7188-5-21

232. Söding J (2005) Protein homology detection by HMM-HMM comparison. Bioinforma Oxf Engl 21:951–960. https://doi.org/10.1093/bioinformatics/bti125

233. Eddy SR (1998) Profile hidden Markov models. Bioinforma Oxf Engl 14:755–763. https://doi.org/10.1093/bioinformatics/14.9.755

234. Vyas VK, Ukawala RD, Ghate M, Chintha C (2012) Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. Indian J Pharm Sci 74:1–17. https://doi.org/10.4103/0250-474X.102537

235. Cavasotto CN, Phatak SS (2009) Homology modeling in drug discovery: current trends and applications. Drug Discov Today 14:676–683. https://doi.org/10.1016/j.drudis.2009.04.006

236. Muhammed MT, Aki-Yalcin E (2019) Homology modeling in drug discovery: Overview, current applications, and future perspectives. Chem Biol Drug Des 93:12–20. https://doi.org/10.1111/cbdd.13388

237. França TCC (2015) Homology modeling: an important tool for the drug discovery. J Biomol Struct Dyn 33:1780–1793. https://doi.org/10.1080/07391102.2014.971429

238. Martí-Renom MA, Stuart AC, Fiser A, et al (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325. https://doi.org/10.1146/annurev.biophys.29.1.291

239. Read RJ, Adams PD, Arendall WB, et al (2011) A New Generation of Crystallographic Validation Tools for the Protein Data Bank. Struct England1993 19:1395–1412. https://doi.org/10.1016/j.str.2011.08.006

240. Cooper DR, Porebski PJ, Chruszcz M, Minor W (2011) X-ray crystallography: assessment and validation of protein–small molecule complexes for drug discovery. Expert Opin Drug Discov 6:771–782. https://doi.org/10.1517/17460441.2011.585154

241. Chen VB, Arendall WB, Headd JJ, et al (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21. https://doi.org/10.1107/S0907444909042073

242. Webb B, Sali A (2014) Protein Structure Modeling with MODELLER. In: Kihara D (ed) Protein Structure Prediction. Springer, New York, NY, pp 1–15

243. Hatherley R, Brown DK, Glenister M, Tastan Bishop Ö (2016) PRIMO: An Interactive Homology Modeling Pipeline. PLOS ONE 11:e0166698. https://doi.org/10.1371/journal.pone.0166698

244. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248. https://doi.org/10.1093/nar/gki408

245. Rost B (1999) Twilight zone of protein sequence alignments. Protein Eng 12:85–94. https://doi.org/10.1093/protein/12.2.85

246. Shen M, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15:2507–2524. https://doi.org/10.1110/ps.062416606

247. Eisenberg D, Lüthy R, Bowie JU (1997) [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. In: Methods in Enzymology. Elsevier, pp 396–404

248. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410. https://doi.org/10.1093/nar/gkm290

249. Sippl MJ (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol 213:859–883. https://doi.org/10.1016/s0022-2836(05)80269-4

250. Sippl MJ (1993) Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. J Comput Aided Mol Des 7:473–501. https://doi.org/10.1007/BF02337562

251. Durrant JD, McCammon JA (2011) Molecular dynamics simulations and drug discovery. BMC Biol 9:71. https://doi.org/10.1186/1741-7007-9-71

252. Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. Neuron 99:1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

253. Aduri R, Psciuk BT, Saro P, et al (2007) AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. J Chem Theory Comput 3:1464–1475. https://doi.org/10.1021/ct600329w

254. MacKerell AD, Bashford D, Bellott M, et al (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B 102:3586–3616. https://doi.org/10.1021/jp973084f

255. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656–1676. https://doi.org/10.1002/jcc.20090

256. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc 118:11225–11236. https://doi.org/10.1021/ja9621760

257. Mark P, Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w

258. Wu W, Owino J, Al-Ostaz A, Cai L (2014) Applying Periodic Boundary Conditions in Finite Element Analysis. 13

259. Knapp B, Frantal S, Cibena M, et al (2011) Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? J Comput Biol J Comput Mol Cell Biol 18:997–1005. https://doi.org/10.1089/cmb.2010.0237

260. Allen TW, Andersen OS, Roux B (2004) On the Importance of Atomic Fluctuations, Protein Flexibility, and Solvent in Ion Permeation. J Gen Physiol 124:679–690. https://doi.org/10.1085/jgp.200409111

261. Bro R, K. Smilde A (2014) Principal component analysis. Anal Methods 6:2812–2831. https://doi.org/10.1039/C3AY41907J

262. Kherif F, Latypova A (2020) Chapter 12 - Principal component analysis. In: Mechelli A, Vieira S (eds) Machine Learning. Academic Press, pp 209–225

263. Ross C, Nizami B, Glenister M, et al (2018) MODE-TASK: large-scale protein motion tools. Bioinformatics 34:3759–3763. https://doi.org/10.1093/bioinformatics/bty427

264. Sheik Amamuddy O, Glenister M, Tshabalala T, Tastan Bishop Ö (2021) MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics. Comput Struct Biotechnol J 19:5059–5071. https://doi.org/10.1016/j.csbj.2021.08.043

265. Brown DK, Penkler DL, Sheik Amamuddy O, et al (2017) MD-TASK: a software suite for analyzing molecular dynamics trajectories. Bioinformatics 33:2768–2771. https://doi.org/10.1093/bioinformatics/btx349

266. Amamuddy OS, Glenister M, Bishop ÖT (2021) MDM-TASK-web: A web platform for protein dynamic residue networks and modal analysis

267. Trudeau RJ (2013) Introduction to Graph Theory. Courier Corporation

268. Atilgan AR, Akan P, Baysal C (2004) Small-World Communication of Residues and Significance for Protein Dynamics. Biophys J 86:85–91. https://doi.org/10.1016/S0006-3495(04)74086-2

269. Penkler DL, Tastan Bishop Ö (2019) Modulation of Human Hsp90α Conformational Dynamics by Allosteric Ligand Interaction at the C-Terminal Domain. Sci Rep 9:1600. https://doi.org/10.1038/s41598-018-35835-0

270. Analyzing the Social Web - 1st Edition. https://www.elsevier.com/books/analyzing-the-social-web/golbeck/978-0-12-405531-5. Accessed 29 Jun 2021

271. Barozi V, Edkins AL, 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. Comput Struct Biotechnol J. https://doi.org/10.1016/j.csbj.2022.08.015

272. Zhan J, Gurung S, Parsa SPK (2017) Identification of top-K nodes in large networks using Katz centrality. J Big Data 4:16. https://doi.org/10.1186/s40537-017-0076-5

273. Sol A del, Fujihashi H, Amoros D, Nussinov R (2006) Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families. Protein Sci 15:2120–2128. https://doi.org/10.1110/ps.062249106

274. Global tuberculosis report 2019. https://www.who.int/publications-detail-redirect/9789241565714. Accessed 12 Jul 2022

275. Lange C, Mori T (2010) Advances in the diagnosis of tuberculosis. Respirology 15:220–240. https://doi.org/10.1111/j.1440-1843.2009.01692.x

276. Natarajan A, Beena PM, Devnikar AV, Mali S (2020) A systemic review on tuberculosis. Indian J Tuberc 67:295–311. https://doi.org/10.1016/j.ijtb.2020.02.005

277. Carranza C, Pedraza-Sanchez S, de Oyarzabal-Mendez E, Torres M (2020) Diagnosis for Latent Tuberculosis Infection: New Alternatives. Front Immunol 11:

278. Rodriguez-Takeuchi SY, Renjifo ME, Medina FJ (2019) Extrapulmonary Tuberculosis: Pathophysiology and Imaging Findings. RadioGraphics 39:2023–2037. https://doi.org/10.1148/rg.2019190109

279. Shah MI, Mishra S, Yadav VK, et al (2017) Ziehl–Neelsen sputum smear microscopy image database: a resource to facilitate automated bacilli detection for tuberculosis diagnosis. J Med Imaging 4:027503. https://doi.org/10.1117/1.JMI.4.2.027503

280. Agrawal M, Bajaj A, Bhatia V, Dutt S (2016) Comparative Study of GeneXpert with ZN Stain and Culture in Samples of Suspected Pulmonary Tuberculosis. J Clin Diagn Res JCDR 10:DC09-DC12. https://doi.org/10.7860/JCDR/2016/18837.7755

281. Shi J, Dong W, Ma Y, et al (2018) GeneXpert MTB/RIF Outperforms Mycobacterial Culture in Detecting Mycobacterium tuberculosis from Salivary Sputum. BioMed Res Int 2018:e1514381. https://doi.org/10.1155/2018/1514381

282. Lange C, Abubakar I, Alffenaar J-WC, et al (2014) Management of patients with multidrug-resistant/extensively drug-resistant tuberculosis in Europe: a TBNET consensus statement. Eur Respir J 44:23–63. https://doi.org/10.1183/09031936.00188313

283. Migliori GB, Tiberi S, Zumla A, et al (2020) MDR/XDR-TB management of patients and contacts: Challenges facing the new decade. The 2020 clinical update by the Global Tuberculosis Network. Int J Infect Dis 92:S15–S25. https://doi.org/10.1016/j.ijid.2020.01.042

284. O'Connor C, Brady MF (2022) Isoniazid. In: StatPearls. StatPearls Publishing, Treasure Island (FL)

285. Vilchèze C, Jacobs WR (2019) The Isoniazid Paradigm of Killing, Resistance, and Persistence in Mycobacterium tuberculosis. J Mol Biol 431:3450–3461. https://doi.org/10.1016/j.jmb.2019.02.016

286. Hang NTL, Hijikata M, Maeda S, et al (2019) Drug resistance-conferring mutations in Mycobacterium-tuberculosis isolated from retreated patients in Hanoi, Vietnam. Eur Respir J 54:. https://doi.org/10.1183/13993003.congress-2019.OA2135

287. Charan AS, Gupta N, Dixit R, et al (2020) Pattern of InhA and KatG mutations in isoniazid monoresistant Mycobacterium tuberculosis isolates. Lung India Off Organ Indian Chest Soc 37:227–231. https://doi.org/10.4103/lungindia.lungindia_204_19

288. Ghosh A, N S, Saha S (2020) Survey of drug resistance associated gene mutations in Mycobacterium tuberculosis , ESKAPE and other bacterial species. Sci Rep 10:8957. https://doi.org/10.1038/s41598-020-65766-8

289. Kabir S, Junaid K, Rehman A (2021) Variations in rifampicin and isoniazid resistance associated genetic mutations among drug naïve and recurrence cases of pulmonary tuberculosis. Int J Infect Dis 103:56–61. https://doi.org/10.1016/j.ijid.2020.11.007

290. Hamed Z, Mohajeri P, Farahani A, et al (2021) The frequency of point mutations associated with resistance to isoniazid and rifampin among clinical isolates of multidrug-resistant Mycobacterium tuberculosis in the west of Iran. Gene Rep 22:100981. https://doi.org/10.1016/j.genrep.2020.100981

291. Bertrand T, Eady NAJ, Jones JN, et al (2004) Crystal Structure of *Mycobacterium tuberculosis* Catalase-Peroxidase. J Biol Chem 279:38991–38999. https://doi.org/10.1074/jbc.M402382200

292. Welinder KG (1992) Superfamily of plant, fungal and bacterial peroxidases. Curr Opin Struct Biol 2:388–393. https://doi.org/10.1016/0959-440X(92)90230-5

293. Zhao X, Hersleth H-P, Zhu J, et al (2013) Access channel residues Ser315 and Asp137 in Mycobacterium tuberculosis catalase-peroxidase (KatG) control peroxidatic activation of the pro-drug isoniazid. Chem Commun 49:11650–11652. https://doi.org/10.1039/C3CC47022A

294. Yamada Y, Fujiwara T, Sato T, et al (2002) The 2.0 Å crystal structure of catalase-peroxidase from Haloarcula marismortui. Nat Struct Biol 9:691–695. https://doi.org/10.1038/nsb834

295. Baker RD, Cook CO, Goodwin DC (2004) Properties of catalase–peroxidase lacking its C-terminal domain. Biochem Biophys Res Commun 320:833–839. https://doi.org/10.1016/j.bbrc.2004.06.026

296. Wang Y, Goodwin DC (2013) Integral role of the I′-helix in the function of the "inactive" C-terminal domain of catalase–peroxidase (KatG). Biochim Biophys Acta BBA - Proteins Proteomics 1834:362–371. https://doi.org/10.1016/j.bbapap.2012.08.003

297. DeVito JA, Morris S (2003) Exploring the Structure and Function of the Mycobacterial KatG Protein Using trans-Dominant Mutants. Antimicrob Agents Chemother 47:188–195. https://doi.org/10.1128/AAC.47.1.188-195.2003

298. Wilming M, Johnsson K (2001) Inter- and intramolecular domain interactions of the catalase-peroxidase KatG from M. tuberculosis. FEBS Lett 509:272–276. https://doi.org/10.1016/S0014-5793(01)03179-9

299. Reingewertz TH, Meyer T, McIntosh F, et al (2020) Differential Sensitivity of Mycobacteria to Isoniazid Is Related to Differences in KatG-Mediated Enzymatic Activation of the Drug. Antimicrob Agents Chemother 64:e01899-19. https://doi.org/10.1128/AAC.01899-19

300. Bollela VR, Namburete EI, Feliciano CS, et al (2016) Detection of katG and inhA mutations to guide isoniazid and ethionamide use for drug-resistant tuberculosis. Int J Tuberc Lung Dis 20:1099–1104. https://doi.org/10.5588/ijtld.15.0864

301. Aggarwal M, Singh A, Grover S, et al (2018) Role of pncA gene mutations W68R and W68G in pyrazinamide resistance. J Cell Biochem 119:2567–2578. https://doi.org/10.1002/jcb.26420

302. Allana S, Shashkina E, Mathema B, et al (2017) pncA Gene Mutations Associated with Pyrazinamide Resistance in Drug-Resistant Tuberculosis, South Africa and Georgia. Emerg Infect Dis 23:491–495. https://doi.org/10.3201/eid2303.161034

303. Isakova J, Sovkhozova N, Vinnikov D, et al (2018) Mutations of rpoB, katG, inhA and ahp genes in rifampicin and isoniazid-resistant Mycobacterium tuberculosis in Kyrgyz Republic. BMC Microbiol 18:22. https://doi.org/10.1186/s12866-018-1168-x

304. Motavaf B, Keshavarz N, Ghorbanian F, et al (2021) Detection of genomic mutations in katG and rpoB genes among multidrug-resistant Mycobacterium tuberculosis isolates from Tehran, Iran. New Microbes New Infect 41:100879. https://doi.org/10.1016/j.nmni.2021.100879

305. Pitso L, Potgieter S, van der S van DA (2019) Prevalence of isoniazid resistance-conferring mutations associated with multidrug-resistant tuberculosis in Free State Province, South Africa. S Afr Med J 109:659–664. https://doi.org/10.7196/SAMJ.2019.v109i9.13730

306. Rueda J, Realpe T, Mejia GI, et al (2015) Genotypic Analysis of Genes Associated with Independent Resistance and Cross-Resistance to Isoniazid and Ethionamide in Mycobacterium tuberculosis Clinical Isolates. Antimicrob Agents Chemother 59:7805–7810. https://doi.org/10.1128/AAC.01028-15

307. Gao M, Yang TT, Li GL, et al (2020) [Analysis on drug resistance-associated mutations of multi-drug resistant Mycobacterium tuberculosis based on whole-genome sequencing in China]. Zhonghua Liu Xing Bing Xue Za Zhi Zhonghua Liuxingbingxue Zazhi 41:770–775. https://doi.org/10.3760/cma.j.cn112338-20191111-00800

308. Charoenpak R, Santimaleeworagun W, Suwanpimolkul G, et al (2020) Association Between the Phenotype and Genotype of Isoniazid Resistance Among Mycobacterium tuberculosis Isolates in Thailand. Infect Drug Resist 13:627–634. https://doi.org/10.2147/IDR.S242261

309. Umar FF, Husain DR, Hatta MM, et al (2020) Molecular characterisation of mutations associated with resistance to first- and second-line drugs among Indonesian patients with tuberculosis. J Taibah Univ Med Sci 15:54–58. https://doi.org/10.1016/j.jtumed.2019.12.003

310. Welekidan LN, Skjerve E, Dejene TA, et al (2021) Frequency and patterns of first- and second-line drug resistance-conferring mutations in Mycobacterium tuberculosis isolated from pulmonary tuberculosis patients in a cross-sectional study in Tigray Region, Ethiopia. J Glob Antimicrob Resist 24:6–13. https://doi.org/10.1016/j.jgar.2020.11.017

311. Fathurrachman A, Andriani RL, Ahmad Z (2022) Frequency of Mutation katG S315T Mycobacterium Tuberculosis in Multi Drug Resistance Tuberculosis Patients in South Sumatera. Int J Infect Dis 116:S8. https://doi.org/10.1016/j.ijid.2021.12.020

312. The UniProt Consortium (2007) The Universal Protein Resource (UniProt). Nucleic Acids Res 36:D190–D195. https://doi.org/10.1093/nar/gkm895

313. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235

314. Sandgren A, Strong M, Muthukrishnan P, et al (2009) Tuberculosis Drug Resistance Mutation Database. PLoS Med 6:e1000002. https://doi.org/10.1371/journal.pmed.1000002

315. Abraham MJ, Murtola T, Schulz R, et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001

316. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res 32:W665–W667. https://doi.org/10.1093/nar/gkh381

317. Lin Z, Gunsteren WF van (2013) Refinement of the application of the GROMOS 54A7 force field to β-peptides. J Comput Chem 34:2796–2805. https://doi.org/10.1002/jcc.23459

318. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 52:7182–7190. https://doi.org/10.1063/1.328693

319. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

320. Petersen HG (1995) Accuracy and efficiency of the particle mesh Ewald method. J Chem Phys 103:3668–3679. https://doi.org/10.1063/1.470043

321. Waskom ML (2021) seaborn: statistical data visualization. J Open Source Softw 6:3021. https://doi.org/10.21105/joss.03021

322. McKinney W (2010) Data Structures for Statistical Computing in Python. Proc 9th Python Sci Conf 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

323. Roe DR, Cheatham TE (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput 9:3084–3095. https://doi.org/10.1021/ct400341p

324. Hunter JD (2007) Matplotlib: A 2D Graphics Environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55

325. Walt S van der, Colbert SC, Varoquaux G (2011) The NumPy Array: A Structure for Efficient Numerical Computation. Comput Sci Eng 13:22–30. https://doi.org/10.1109/MCSE.2011.37

326. H N, Da C, As R (2018) NGLview-interactive molecular graphics for Jupyter notebooks. Bioinforma Oxf Engl 34:1241–1242. https://doi.org/10.1093/bioinformatics/btx789

327. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38, 27–28. https://doi.org/10.1016/0263-7855(96)00018-5

328. McGibbon RT, Beauchamp KA, Harrigan MP, et al (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J 109:1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015

329. Hagberg A, Swart P, S Chult D (2008) Exploring network structure, dynamics, and function using networkx. Los Alamos National Lab. (LANL), Los Alamos, NM (United States)

330. Kortemme T, Kim DE, Baker D (2004) Computational Alanine Scanning of Protein-Protein Interfaces. Sci Signal 2004:pl2–pl2. https://doi.org/10.1126/stke.2192004pl2

331. Ross C, Knox C, Tastan Bishop Ö (2017) Interacting motif networks located in hotspots associated with RNA release are conserved in Enterovirus capsids. FEBS Lett 591:1687–1701. https://doi.org/10.1002/1873-3468.12663

332. Lempens P, Meehan CJ, Vandelannoote K, et al (2018) Isoniazid resistance levels of Mycobacterium tuberculosis can largely be predicted by high-confidence resistance-conferring mutations. Sci Rep 8:3246. https://doi.org/10.1038/s41598-018-21378-x

333. Ruesen C, Riza AL, Florescu A, et al (2018) Linking minimum inhibitory concentrations to whole genome sequence-predicted drug resistance in Mycobacterium tuberculosis strains from Romania. Sci Rep 8:9676. https://doi.org/10.1038/s41598-018-27962-5

334. Nonghanphithak D, Kaewprasert O, Chaiyachat P, et al (2020) Whole-genome sequence analysis and comparisons between drug-resistance mutations and minimum inhibitory concentrations of Mycobacterium tuberculosis isolates causing M/XDR-TB. PLOS ONE 15:e0244829. https://doi.org/10.1371/journal.pone.0244829

335. Pinhata JMW, Brandao AP, Mendes F de F, et al (2021) Correlating genetic mutations with isoniazid phenotypic levels of resistance in Mycobacterium tuberculosis isolates from patients with drug-resistant tuberculosis in a high burden setting. Eur J Clin Microbiol Infect Dis 40:2551–2561. https://doi.org/10.1007/s10096-021-04316-0

336. Anukool U, Phunpae P, Tharinjaroen CS, et al (2020) Genotypic Distribution and a Potential Diagnostic Assay of Multidrug-Resistant Tuberculosis in Northern Thailand. Infect Drug Resist 13:3375–3382. https://doi.org/10.2147/IDR.S263082

337. Salvato RS, Reis AJ, Schiefelbein SH, et al (2021) Genomic-based surveillance reveals high ongoing transmission of multi-drug-resistant Mycobacterium tuberculosis in Southern Brazil. Int J Antimicrob Agents 58:106401. https://doi.org/10.1016/j.ijantimicag.2021.106401

338. Seid A, Berhane N, Nureddin S (2022) Frequency of rpoB, katG, and inhA Gene Polymorphisms Associated with Multidrug-Resistant Mycobacterium tuberculosis Complex Isolates among Ethiopian TB Patients: A Systematic Review. Interdiscip Perspect Infect Dis 2022:e1967675. https://doi.org/10.1155/2022/1967675

339. Sheik Amamuddy O, Verkhivker GM, Tastan Bishop Ö (2020) Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 Mpro. J Chem Inf Model 60:5080–5102. https://doi.org/10.1021/acs.jcim.0c00634

340. Daura X, Gademann K, Jaun B, et al (1999) Peptide Folding: When Simulation Meets Experiment. Angew Chem Int Ed 38:236–240. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M

341. Dundas J, Ouyang Z, Tseng J, et al (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 34:W116–W118. https://doi.org/10.1093/nar/gkl282

342. Zhao X, Yu S, Ranguelova K, et al (2009) Role of the Oxyferrous Heme Intermediate and Distal Side Adduct Radical in the Catalase Activity of Mycobacterium tuberculosis KatG Revealed by the W107F Mutant. J Biol Chem 284:7030–7037. https://doi.org/10.1074/jbc.M808107200

343. Salentin S, Schreiber S, Haupt VJ, et al (2015) PLIP: fully automated protein–ligand interaction profiler. Nucleic Acids Res 43:W443–W447. https://doi.org/10.1093/nar/gkv315

344. Svistunenko DA (2005) Reaction of haem containing proteins and enzymes with hydroperoxides: The radical view. Biochim Biophys Acta BBA - Bioenerg 1707:127–155. https://doi.org/10.1016/j.bbabio.2005.01.004

345. Yang YX, Wang P, Zhu BT (2022) Importance of interface and surface areas in protein-protein binding affinity prediction: A machine learning analysis based on linear regression and artificial neural network. Biophys Chem 283:106762. https://doi.org/10.1016/j.bpc.2022.106762

346. Swapna LS, Bhaskara RM, Sharma J, Srinivasan N (2012) Roles of residues in the interface of transient protein-protein complexes before complexation. Sci Rep 2:334. https://doi.org/10.1038/srep00334

347. Banerjee S, Zamocky M, Furtmüller PG, Obinger C (2010) Probing the two-domain structure of homodimeric prokaryotic and eukaryotic catalase–peroxidases. Biochim Biophys Acta BBA - Proteins Proteomics 1804:2136–2145. https://doi.org/10.1016/j.bbapap.2010.07.013

348. Lukat-Rodgers GS, Wengenack NL, Rusnak F, Rodgers KR (2000) Spectroscopic Comparison of the Heme Active Sites in WT KatG and Its S315T Mutant. Biochemistry 39:9984–9993. https://doi.org/10.1021/bi0006870

349. Mumtaz Khan M, S Alves M, Alam S, et al (2020) In Silico Analysis of S315T and S315R Mutations of Multidrug-resistant Mycobacterium tuberculosis Clinical Isolates from Karachi, Pakistan. Jundishapur J Microbiol 13:. https://doi.org/10.5812/jjm.100948

350. Georgoulia PS, Bjelic S (2021) Prediction of Protein–Protein Binding Interactions in Dimeric Coiled Coils by Information Contained in Folding Energy Landscapes. Int J Mol Sci 22:1368. https://doi.org/10.3390/ijms22031368

351. Simões ICM, Costa IPD, Coimbra JTS, et al (2017) New Parameters for Higher Accuracy in the Computation of Binding Free Energy Differences upon Alanine Scanning Mutagenesis on

Protein–Protein Interfaces. J Chem Inf Model 57:60–72. https://doi.org/10.1021/acs.jcim.6b00378

352. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32:W526–W531. https://doi.org/10.1093/nar/gkh468

353. Chernyaeva EN, Shulgina MV, Rotkevich MS, et al (2014) Genome-wide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. BMC Genomics 15:308. https://doi.org/10.1186/1471-2164-15-308

354. Krishnamoorthy N, Fakhro K (2021) Identification of mutation resistance coldspots for targeting the SARS-CoV2 main protease. IUBMB Life 73:670–675. https://doi.org/10.1002/iub.2465

355. Gradmann C (2001) Robert Koch and the pressures of scientific research: tuberculosis and tuberculin. Med Hist 45:1–32

356. Whitfield MG, Soeters HM, Warren RM, et al (2015) A Global Perspective on Pyrazinamide Resistance: Systematic Review and Meta-Analysis. PLOS ONE 10:e0133869. https://doi.org/10.1371/journal.pone.0133869

357. Njire M, Tan Y, Mugweru J, et al (2016) Pyrazinamide resistance in Mycobacterium tuberculosis: Review and update. Adv Med Sci 61:63–71. https://doi.org/10.1016/j.advms.2015.09.007

358. Khawbung JL, Nath D, Chakraborty S (2021) Drug resistant Tuberculosis: A review. Comp Immunol Microbiol Infect Dis 74:101574. https://doi.org/10.1016/j.cimid.2020.101574

359. Gazzaniga F, Stebbins R, Chang SZ, et al (2009) Microbial NAD metabolism: lessons from comparative genomics. Microbiol Mol Biol Rev MMBR 73:529–541, Table of Contents. https://doi.org/10.1128/MMBR.00042-08

360. Vu CQ, Coyle DL, Tai HH, et al (1997) Intramolecular ADP-ribose transfer reactions and calcium signalling. Potential role of 2'-phospho-cyclic ADP-ribose in oxidative stress. Adv Exp Med Biol 419:381–388

361. Stehr M, Elamin AA, Singh M (2015) Pyrazinamide: the importance of uncovering the mechanisms of action in mycobacteria. Expert Rev Anti Infect Ther 13:593–603. https://doi.org/10.1586/14787210.2015.1021784

362. Boshoff HIM, Xu X, Tahlan K, et al (2008) Biosynthesis and Recycling of Nicotinamide Cofactors in Mycobacterium tuberculosis. J Biol Chem 283:19329–19341. https://doi.org/10.1074/jbc.M800694200

363. Murray MF (2003) Nicotinamide: an oral antimicrobial agent with activity against both Mycobacterium tuberculosis and human immunodeficiency virus. Clin Infect Dis Off Publ Infect Dis Soc Am 36:453–460. https://doi.org/10.1086/367544

364. Jarzembska KN, Hoser AA, Kamiński R, et al (2014) Combined Experimental and Computational Studies of Pyrazinamide and Nicotinamide in the Context of Crystal Engineering and Thermodynamics. Cryst Growth Des 14:3453–3465. https://doi.org/10.1021/cg500376z

365. Sun Z, Zhang Y (1999) Reduced Pyrazinamidase Activity and the Natural Resistance of Mycobacterium kansasii to the Antituberculosis Drug Pyrazinamide. Antimicrob Agents Chemother 43:537–542. https://doi.org/10.1128/AAC.43.3.537

366. Zhang Y, Mitchison D (2003) The curious characteristics of pyrazinamide: a review. Int J Tuberc Lung Dis 7:6–21

367. Yang J, Liu Y, Bi J, et al (2015) Structural basis for targeting the ribosomal protein S1 of Mycobacterium tuberculosis by pyrazinamide. Mol Microbiol 95:791–803. https://doi.org/10.1111/mmi.12892

368. Yadon AN, Maharaj K, Adamson JH, et al (2017) A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide. Nat Commun 8:588. https://doi.org/10.1038/s41467-017-00721-2

369. Rehman AU, Khan MT, Liu H, et al (2019) Exploring the Pyrazinamide Drug Resistance Mechanism of Clinical Mutants T370P and W403G in Ribosomal Protein S1 of Mycobacterium tuberculosis. J Chem Inf Model 59:1584–1597. https://doi.org/10.1021/acs.jcim.8b00956

370. Grosset J (1978) The sterilizing value of rifampicin and pyrazinamide in experimental short-course chemotherapy. Bull Int Union Tuberc 53:5–12

371. Petrella S, Gelus-Zietal N, Maudry A, et al (2011) Crystal Structure of the Pyrazinamidase of Mycobacterium tuberculosis: Insights into Natural and Acquired Resistance to Pyrazinamide. PLOS ONE 6:e15785. https://doi.org/10.1371/journal.pone.0015785

372. Sheen P, Ferrer P, Gilman RH, et al (2012) Role of Metal Ions on the Activity of Mycobacterium tuberculosis Pyrazinamidase. Am J Trop Med Hyg 87:153–161. https://doi.org/10.4269/ajtmh.2012.10-0565

373. Khan T, Khan A, Ali SS, et al (2021) A computational perspective on the dynamic behaviour of recurrent drug resistance mutations in the pncA gene from Mycobacterium tuberculosis. RSC Adv 11:2476–2486. https://doi.org/10.1039/D0RA09326B

374. Du X, Wang W, Kim R, et al (2001) Crystal Structure and Mechanism of Catalysis of a Pyrazinamidase from Pyrococcus horikoshii. Biochemistry 40:14166–14172. https://doi.org/10.1021/bi0115479

375. Barco P, Cardoso RF, Hirata RDC, et al (2006) pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis clinical isolates from the southeast region of Brazil. J Antimicrob Chemother 58:930–935. https://doi.org/10.1093/jac/dkl363

376. khan MT, Malik SI, Ali S, et al (2019) Pyrazinamide resistance and mutations in pncA among isolates of Mycobacterium tuberculosis from Khyber Pakhtunkhwa, Pakistan. BMC Infect Dis 19:116. https://doi.org/10.1186/s12879-019-3764-2

377. Ramirez-Busby SM, Valafar F (2015) Systematic Review of Mutations in Pyrazinamidase Associated with Pyrazinamide Resistance in Mycobacterium tuberculosis Clinical Isolates. Antimicrob Agents Chemother 59:5267–5277. https://doi.org/10.1128/AAC.00204-15

378. Nangraj AS, Khan A, Umbreen S, et al (2021) Insights Into Mutations Induced Conformational Changes and Rearrangement of Fe2+ Ion in pncA Gene of Mycobacterium tuberculosis to Decipher the Mechanism of Resistance to Pyrazinamide. Front Mol Biosci 8:633365. https://doi.org/10.3389/fmolb.2021.633365

379. Khan MT, Junaid M, Mao X, et al (2019) Pyrazinamide resistance and mutations L19R, R140H, and E144K in Pyrazinamidase of Mycobacterium tuberculosis. J Cell Biochem 120:7154–7166. https://doi.org/10.1002/jcb.27989

380. Khan MT, Malik SI (2020) Structural dynamics behind variants in pyrazinamidase and pyrazinamide resistance. J Biomol Struct Dyn 38:3003–3017. https://doi.org/10.1080/07391102.2019.1650113

381. Khan MT, Chinnasamy S, Cui Z, et al (2020) Mechanistic analysis of A46V, H57Y, and D129N in pyrazinamidase associated with pyrazinamide resistance. Saudi J Biol Sci 27:3150–3156. https://doi.org/10.1016/j.sjbs.2020.07.015

382. Esmaeeli R, Mehrnejad F, Mir-Derikvand M, Gopalpoor N (2019) Computational insights into pH-dependence of structure and dynamics of pyrazinamidase: A comparison of wild type and mutants. J Cell Biochem 120:2502–2514. https://doi.org/10.1002/jcb.27543

383. Werngren J, Alm E, Mansjö M (2017) Non-pncA Gene-Mutated but Pyrazinamide-Resistant Mycobacterium tuberculosis: Why Is That? J Clin Microbiol 55:1920–1927. https://doi.org/10.1128/JCM.02532-16

384. Simons SO, Mulder A, Ingen J van, et al (2013) Role of rpsA Gene Sequencing in Diagnosis of Pyrazinamide Resistance. J Clin Microbiol 51:382–382. https://doi.org/10.1128/JCM.02739-12

385. Shi W, Zhang X, Jiang X, et al (2011) Pyrazinamide Inhibits Trans-Translation in Mycobacterium tuberculosis. Science 333:1630–1632. https://doi.org/10.1126/science.1208813

386. Martin A, Cubillos-Ruiz A, Von Groll A, et al (2008) Nitrate reductase assay for the rapid detection of pyrazinamide resistance in Mycobacterium tuberculosis using nicotinamide. J Antimicrob Chemother 61:123–127. https://doi.org/10.1093/jac/dkm418

387. Mirabal NC, Yzquierdo SL, Lemus D, et al (2010) Evaluation of Colorimetric Methods Using Nicotinamide for Rapid Detection of Pyrazinamide Resistance in Mycobacterium tuberculosis. J Clin Microbiol 48:2729–2733. https://doi.org/10.1128/JCM.00311-10

388. Hu Y, Wu X, Luo J, et al (2017) Detection of pyrazinamide resistance of Mycobacterium tuberculosis using nicotinamide as a surrogate. Clin Microbiol Infect 23:835–838. https://doi.org/10.1016/j.cmi.2017.03.028

389. Neyfakh AA (2001) The ostensible paradox of multidrug recognition. J Mol Microbiol Biotechnol 3:151–154

390. Vazquez-Laslop N, Zheleznova EE, Markham PN, et al (2000) Recognition of multiple drugs by a single protein: a trivial solution of an old paradox. Biochem Soc Trans 28:517–520

391. Kim S, Thiessen PA, Bolton EE, et al (2016) PubChem Substance and Compound databases. Nucleic Acids Res 44:D1202–D1213. https://doi.org/10.1093/nar/gkv951

392. Gordon JC, Myers JB, Folta T, et al (2005) H++: a server for estimating p Ka s and adding missing hydrogens to macromolecules. Nucleic Acids Res 33:W368–W371. https://doi.org/10.1093/nar/gki464

393. El-Hachem N, Haibe-Kains B, Khalil A, et al (2017) AutoDock and AutoDockTools for Protein-Ligand Docking: Beta-Site Amyloid Precursor Protein Cleaving Enzyme 1(BACE1) as a Case Study. Methods Mol Biol Clifton NJ 1598:391–403. https://doi.org/10.1007/978-1-4939-6952-4_20

394. Morris GM, Goodsell DS, Halliday RS, et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–

1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

395. Inc AS (2012) Discovery studio modeling environment, release 3.5. Accelrys Discov Studio Accelrys Softw Inc San Diego

396. Pires DEV, Ascher DB, Blundell TL (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics 30:335–342. https://doi.org/10.1093/bioinformatics/btt691

397. Maier JA, Martinez C, Kasavajhala K, et al (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput 11:3696–3713. https://doi.org/10.1021/acs.jctc.5b00255

398. Sousa da Silva AW, Vranken WF (2012) ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res Notes 5:367. https://doi.org/10.1186/1756-0500-5-367

399. Teilum K, Olsen JG, Kragelund BB (2009) Functional aspects of protein flexibility. Cell Mol Life Sci 66:2231–2247. https://doi.org/10.1007/s00018-009-0014-6

400. Junaid M, Li C-D, Li J, et al (2021) Structural insights of catalytic mechanism in mutant pyrazinamidase of Mycobacterium tuberculosis. J Biomol Struct Dyn 39:3172–3185. https://doi.org/10.1080/07391102.2020.1761879

401. Hanukoglu I (2015) Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. Biochem Mol Biol Educ 43:206–209. https://doi.org/10.1002/bmb.20849

402. Morens DM, Breman JG, Calisher CH, et al (2020) The Origin of COVID-19 and Why It Matters. Am J Trop Med Hyg 103:955–959. https://doi.org/10.4269/ajtmh.20-0849

403. Guo Y-R, Cao Q-D, Hong Z-S, et al (2020) The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. Mil Med Res 7:11. https://doi.org/10.1186/s40779-020-00240-0

404. Berekaa MM (2020) Insights into the COVID-19 pandemic: Origin, pathogenesis, diagnosis, and therapeutic interventions. Front Biosci-Elite 13:117–139. https://doi.org/10.2741/874

405. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int. Accessed 11 Feb 2023

406. ALIMOHAMADI Y, SEPANDI M, TAGHDIR M, HOSAMIRUDSARI H (2020) Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis. J Prev Med Hyg 61:E304–E312. https://doi.org/10.15167/2421-4248/jpmh2020.61.3.1530

407. Struyf T, Deeks JJ, Dinnes J, et al (2022) Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.CD013665.pub3

408. Ardura-Garcia C, Mallet MC, Pedersen ES, et al (2021) Effect of COVID19 preventive measures on respiratory infections and symptoms in children. Eur Respir J 58:. https://doi.org/10.1183/13993003.congress-2021.OA2962

409. Jutzeler CR, Bourguignon L, Weis CV, et al (2020) Comorbidities, clinical signs and symptoms, laboratory findings, imaging features, treatment strategies, and outcomes in adult and pediatric

patients with COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis 37:101825. https://doi.org/10.1016/j.tmaid.2020.101825

410. Ejaz H, Alsrhani A, Zafar A, et al (2020) COVID-19 and comorbidities: Deleterious impact on infected patients. J Infect Public Health 13:1833–1839. https://doi.org/10.1016/j.jiph.2020.07.014

411. Singhal T (2020) A Review of Coronavirus Disease-2019 (COVID-19). Indian J Pediatr 87:281–286. https://doi.org/10.1007/s12098-020-03263-6

412. ALIMOHAMADI Y, TOLA HH, ABBASI-GHAHRAMANLOO A, et al (2021) Case fatality rate of COVID-19: a systematic review and meta-analysis. J Prev Med Hyg 62:E311–E320. https://doi.org/10.15167/2421-4248/jpmh2021.62.2.1627

413. Singh BB, Ward MP, Lowerison M, et al (2021) Meta-analysis and adjusted estimation of COVID-19 case fatality risk in India and its association with the underlying comorbidities. One Health 13:100283. https://doi.org/10.1016/j.onehlt.2021.100283

414. Lan J, Ge J, Yu J, et al (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581:215–220. https://doi.org/10.1038/s41586-020-2180-5

415. Berber E, Sumbria D, Çanakoğlu N (2021) Meta-analysis and comprehensive study of coronavirus outbreaks: SARS, MERS and COVID-19. J Infect Public Health 14:1051–1064. https://doi.org/10.1016/j.jiph.2021.06.007

416. Zaki AM, van Boheemen S, Bestebroer TM, et al (2012) Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. N Engl J Med 367:1814–1820. https://doi.org/10.1056/NEJMoa1211721

417. Zhou H, Yang J, Zhou C, et al (2021) A Review of SARS-CoV2: Compared With SARS-CoV and MERS-CoV. Front Med 8:628370. https://doi.org/10.3389/fmed.2021.628370

418. Chaudhuri A (2021) Comparative analysis of non structural protein 1 of SARS-CoV2 with SARS-CoV1 and MERS-CoV: An in silico study. J Mol Struct 1243:130854. https://doi.org/10.1016/j.molstruc.2021.130854

419. Dhama K, Patel SK, Sharun K, et al (2020) SARS-CoV-2 jumping the species barrier: Zoonotic lessons from SARS, MERS and recent advances to combat this pandemic virus. Travel Med Infect Dis 37:101830. https://doi.org/10.1016/j.tmaid.2020.101830

420. Chen J (2020) Pathogenicity and transmissibility of 2019-nCoV—A quick overview and comparison with other emerging viruses. Microbes Infect 22:69–71. https://doi.org/10.1016/j.micinf.2020.01.004

421. Kutti-Sridharan G, Vegunta R, Vegunta R, et al (2020) SARS-CoV2 in Different Body Fluids, Risks of Transmission, and Preventing COVID-19: A Comprehensive Evidence-Based Review. Int J Prev Med 11:97. https://doi.org/10.4103/ijpvm.IJPVM_255_20

422. Peng X, Xu X, Li Y, et al (2020) Transmission routes of 2019-nCoV and controls in dental practice. Int J Oral Sci 12:9. https://doi.org/10.1038/s41368-020-0075-9

423. Liu Y, Ning Z, Chen Y, et al (2020) Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. Nature 582:557–560. https://doi.org/10.1038/s41586-020-2271-3

424. Zhang X, Chen X, Chen L, et al (2020) The evidence of SARS-CoV-2 infection on ocular surface. Ocul Surf 18:360–362. https://doi.org/10.1016/j.jtos.2020.03.010

425. Xia J, Tong J, Liu M, et al (2020) Evaluation of coronavirus in tears and conjunctival secretions of patients with SARS-CoV-2 infection. J Med Virol 92:589–594. https://doi.org/10.1002/jmv.25725

426. Xiao F, Tang M, Zheng X, et al (2020) Evidence for Gastrointestinal Infection of SARS-CoV-2. Gastroenterology 158:1831-1833.e3. https://doi.org/10.1053/j.gastro.2020.02.055

427. To KK-W, Tsang OT-Y, Yip CC-Y, et al (2020) Consistent Detection of 2019 Novel Coronavirus in Saliva. Clin Infect Dis Off Publ Infect Dis Soc Am 71:841–843. https://doi.org/10.1093/cid/ciaa149

428. To KK-W, Tsang OT-Y, Leung W-S, et al (2020) Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. Lancet Infect Dis 20:565–574. https://doi.org/10.1016/S1473-3099(20)30196-1

429. Elsaid M, Nasef MA, Huy NT (2021) R0 of COVID-19 and its impact on vaccination coverage: compared with previous outbreaks. Hum Vaccines Immunother 17:3850–3854. https://doi.org/10.1080/21645515.2020.1865046

430. Harrison AG, Lin T, Wang P (2020) Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. Trends Immunol 41:1100–1115. https://doi.org/10.1016/j.it.2020.10.004

431. Fehr AR, Perlman S (2015) Coronaviruses: An Overview of Their Replication and Pathogenesis. Coronaviruses 1282:1–23. https://doi.org/10.1007/978-1-4939-2438-7_1

432. Huang Y, Yang C, Xu X, et al (2020) Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin 41:1141–1149. https://doi.org/10.1038/s41401-020-0485-4

433. Wang Q, Zhang Y, Wu L, et al (2020) Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell 181:894-904.e9. https://doi.org/10.1016/j.cell.2020.03.045

434. Chorba T (2020) The Concept of the Crown and Its Potential Role in the Downfall of Coronavirus. Emerg Infect Dis 26:2302–2305. https://doi.org/10.3201/eid2609.AC2609

435. Walls AC, Park Y-J, Tortorici MA, et al (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181:281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058

436. Mittal A, Manjunath K, Ranjan RK, et al (2020) COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. PLOS Pathog 16:e1008762. https://doi.org/10.1371/journal.ppat.1008762

437. Naqvi AAT, Fatima K, Mohammad T, et al (2020) Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. Biochim Biophys Acta Mol Basis Dis 1866:165878. https://doi.org/10.1016/j.bbadis.2020.165878

438. Gagneux P, Hennet T, Varki A (2022) Biological Functions of Glycans. In: Varki A, Cummings RD, Esko JD, et al (eds) Essentials of Glycobiology, 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)

439. Varki A (2017) Biological roles of glycans. Glycobiology 27:3–49. https://doi.org/10.1093/glycob/cww086

440. Walls AC, Xiong X, Park Y-J, et al (2019) Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. Cell 176:1026-1039.e15. https://doi.org/10.1016/j.cell.2018.12.028

441. Mohammed MEA (2021) The percentages of SARS-CoV-2 protein similarity and identity with SARS-CoV and BatCoV RaTG13 proteins can be used as indicators of virus origin. J Proteins Proteomics 12:81–91. https://doi.org/10.1007/s42485-021-00060-3

442. Liu L, Wang P, Nair MS, et al (2020) Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature 584:450–456. https://doi.org/10.1038/s41586-020-2571-7

443. Jiang S, Zhang X, Yang Y, et al (2020) Neutralizing antibodies for the treatment of COVID-19. Nat Biomed Eng 4:1134–1139. https://doi.org/10.1038/s41551-020-00660-2

444. Wang C, Li W, Drabek D, et al (2020) A human monoclonal antibody blocking SARS-CoV-2 infection. Nat Commun 11:2251. https://doi.org/10.1038/s41467-020-16256-y

445. Shi R, Shan C, Duan X, et al (2020) A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. Nature 584:120–124. https://doi.org/10.1038/s41586-020-2381-y

446. Jiang S, Hillyer C, Du L (2020) Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. Trends Immunol 41:355–359. https://doi.org/10.1016/j.it.2020.03.007

447. Chen P, Nirula A, Heller B, et al (2021) SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19. N Engl J Med 384:229–237. https://doi.org/10.1056/NEJMoa2029849

448. Jaimes JA, Millet JK, Whittaker GR (2020) Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. iScience 23:101212. https://doi.org/10.1016/j.isci.2020.101212

449. Activation of Viruses by Host Proteases

450. Hoffmann M, Hofmann-Winkler H, Pöhlmann S (2018) Priming Time: How Cellular Proteases Arm Coronavirus Spike Proteins. In: Böttcher-Friebertshäuser E, Garten W, Klenk HD (eds) Activation of Viruses by Host Proteases. Springer International Publishing, Cham, pp 71–98

451. Belouzard S, Chu VC, Whittaker GR (2009) Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc Natl Acad Sci U S A 106:5871–5876. https://doi.org/10.1073/pnas.0809524106

452. Millet JK, Whittaker GR (2014) Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc Natl Acad Sci U S A 111:15214–15219. https://doi.org/10.1073/pnas.1407087111

453. Zhang L, Jackson CB, Mou H, et al (2020) SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nat Commun 11:6013. https://doi.org/10.1038/s41467-020-19808-4

454. Gobeil SM-C, Janowska K, McDowell S, et al (2021) D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. Cell Rep 34:108630. https://doi.org/10.1016/j.celrep.2020.108630

455. Cheng Y-W, Chao T-L, Li C-L, et al (2021) D614G Substitution of SARS-CoV-2 Spike Protein Increases Syncytium Formation and Virus Titer via Enhanced Furin-Mediated Spike Cleavage. mBio 12:e00587-21. https://doi.org/10.1128/mBio.00587-21

456. Jackson CB, Zhang L, Farzan M, Choe H (2021) Functional importance of the D614G mutation in the SARS-CoV-2 spike protein. Biochem Biophys Res Commun 538:108–115. https://doi.org/10.1016/j.bbrc.2020.11.026

457. Mohammad A, Alshawaf E, Marafie SK, et al (2021) Higher binding affinity of furin for SARS-CoV-2 spike (S) protein D614G mutant could be associated with higher SARS-CoV-2 infectivity. Int J Infect Dis 103:611–616. https://doi.org/10.1016/j.ijid.2020.10.033

458. Fitzsimmons WJ, Woods RJ, McCrone JT, et al (2018) A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLoS Biol 16:e2006459. https://doi.org/10.1371/journal.pbio.2006459

459. Tuccori M, Ferraro S, Convertino I, et al (2020) Anti-SARS-CoV-2 neutralizing monoclonal antibodies: clinical pipeline. mAbs 12:1854149. https://doi.org/10.1080/19420862.2020.1854149

460. McCallum M, Czudnochowski N, Rosen LE, et al (2022) Structural basis of SARS-CoV-2 Omicron immune evasion and receptor engagement. Science 375:864–868. https://doi.org/10.1126/science.abn8652

461. Cui Z, Liu P, Wang N, et al (2022) Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. Cell 185:860-871.e13. https://doi.org/10.1016/j.cell.2022.01.019

462. Dejnirattisai W, Huo J, Zhou D, et al (2022) SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. Cell 185:467-484.e15. https://doi.org/10.1016/j.cell.2021.12.046

463. Ao D, Lan T, He X, et al (2022) SARS-CoV-2 Omicron variant: Immune escape and vaccine development. MedComm 3:e126. https://doi.org/10.1002/mco2.126

464. Ju B, Zheng Q, Guo H, et al (2022) Immune escape by SARS-CoV-2 Omicron variant and structural basis of its effective neutralization by a broad neutralizing human antibody VacW-209. Cell Res 1–4. https://doi.org/10.1038/s41422-022-00638-6

465. Cao Y, Wang J, Jian F, et al (2022) Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. Nature 602:657–663. https://doi.org/10.1038/s41586-021-04385-3

466. Planas D, Saunders N, Maes P, et al (2022) Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. Nature 602:671–675. https://doi.org/10.1038/s41586-021-04389-z

467. Tracking SARS-CoV-2 variants. https://www.who.int/health-topics/typhoid/tracking-SARS-CoV-2-variants. Accessed 11 Mar 2022

468. Tao K, Tzou PL, Nouhin J, et al (2021) The biological and clinical significance of emerging SARS-CoV-2 variants. Nat Rev Genet 22:757–773. https://doi.org/10.1038/s41576-021-00408-x

469. Madhi SA, Kwatra G, Myers JE, et al (2022) Population Immunity and Covid-19 Severity with Omicron Variant in South Africa. N Engl J Med 386:1314–1326. https://doi.org/10.1056/NEJMoa2119658

470. Du P, Gao GF, Wang Q (2022) The mysterious origins of the Omicron variant of SARS-CoV-2. The Innovation 3:100206. https://doi.org/10.1016/j.xinn.2022.100206

471. Manjunath R, Gaonkar SL, Saleh EAM, Husain K (2022) A comprehensive review on Covid-19 Omicron (B.1.1.529) variant. Saudi J Biol Sci 29:103372. https://doi.org/10.1016/j.sjbs.2022.103372

472. Islam F, Dhawan M, Nafady MH, et al (2022) Understanding the omicron variant (B.1.1.529) of SARS-CoV-2: Mutational impacts, concerns, and the possible solutions. Ann Med Surg 78:103737. https://doi.org/10.1016/j.amsu.2022.103737

473. Shao W, Zhang W, Fang X, et al (2022) Challenges of SARS-CoV-2 Omicron Variant and appropriate countermeasures. J Microbiol Immunol Infect 55:387–394. https://doi.org/10.1016/j.jmii.2022.03.007

474. Shrestha LB, Foster C, Rawlinson W, et al Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission. Rev Med Virol n/a:e2381. https://doi.org/10.1002/rmv.2381

475. Iketani S, Liu L, Guo Y, et al (2022) Antibody evasion properties of SARS-CoV-2 Omicron sublineages. Nature 604:553–556. https://doi.org/10.1038/s41586-022-04594-4

476. Majumdar S, Sarkar R (2022) Mutational and phylogenetic analyses of the two lineages of the Omicron variant. J Med Virol 94:1777–1779. https://doi.org/10.1002/jmv.27558

477. Tegally H, Moir M, Everatt J, et al (2022) Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. Nat Med 1–6. https://doi.org/10.1038/s41591-022-01911-2

478. Mohapatra RK, Kandi V, Sarangi AK, et al (2022) The recently emerged BA.4 and BA.5 lineages of Omicron and their global health concerns amid the ongoing wave of COVID-19 pandemic – Correspondence. Int J Surg Lond Engl 103:106698. https://doi.org/10.1016/j.ijsu.2022.106698

479. Khan A, Gui J, Ahmad W, et al (2021) The SARS-CoV-2 B.1.618 variant slightly alters the spike RBD–ACE2 binding affinity and is an antibody escaping variant: a computational structural perspective. RSC Adv 11:30132–30147. https://doi.org/10.1039/D1RA04694B

480. Ortega JT, Pujol FH, Jastrzebska B, Rangel HR (2021) Mutations in the SARS-CoV-2 spike protein modulate the virus affinity to the human ACE2 receptor, an in silico analysis. EXCLI J 20:585–600. https://doi.org/10.17179/excli2021-3471

481. Verkhivker GM, Agajanian S, Oztas D, Gupta G (2021) Computational analysis of protein stability and allosteric interaction networks in distinct conformational forms of the SARS-CoV-2 spike D614G mutant: reconciling functional mechanisms through allosteric model of spike regulation. J Biomol Struct Dyn 0:1–18. https://doi.org/10.1080/07391102.2021.1933594

482. Greaney AJ, Loes AN, Crawford KHD, et al (2021) Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe 29:463-476.e6. https://doi.org/10.1016/j.chom.2021.02.003

483. Barnes CO, Jette CA, Abernathy ME, et al (2020) SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. Nature 588:682–687. https://doi.org/10.1038/s41586-020-2852-1

484. Weisblum Y, Schmidt F, Zhang F, et al (2020) Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. eLife 9:e61312. https://doi.org/10.7554/eLife.61312

485. Liu Z, VanBlargan LA, Bloyet L-M, et al (2021) Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe 29:477-488.e4. https://doi.org/10.1016/j.chom.2021.01.014

486. O'Toole Á, Scher E, Underwood A, et al (2021) Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol 7:veab064. https://doi.org/10.1093/ve/veab064

487. Khare S, Gurry C, Freitas L, et al (2021) GISAID's Role in Pandemic Response. China CDC Wkly 3:1049–1051. https://doi.org/10.46234/ccdcw2021.255

488. Schrödinger LLC (2015) The PyMOL molecular graphics system, version 1.8

489. McWilliam H, Li W, Uludag M, et al (2013) Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 41:W597–W600. https://doi.org/10.1093/nar/gkt376

490. Barton MI, MacGowan SA, Kutuzov MA, et al (2021) Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. eLife 10:e70658. https://doi.org/10.7554/eLife.70658

491. Xie Y, Guo W, Lopez-Hernadez A, et al (2022) The pH Effects on SARS-CoV and SARS-CoV-2 Spike Proteins in the Process of Binding to hACE2. Pathogens 11:238. https://doi.org/10.3390/pathogens11020238

492. Dakal TC (2021) Antigenic sites in SARS-CoV-2 spike RBD show molecular similarity with pathogenic antigenic determinants and harbors peptides for vaccine development. Immunobiology 226:152091. https://doi.org/10.1016/j.imbio.2021.152091

493. Manivel V, Sahoo NC, Salunke DM, Rao KV (2000) Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. Immunity 13:611–620. https://doi.org/10.1016/s1074-7613(00)00061-3

494. Kong L, Lee DE, Kadam RU, et al (2016) Structural flexibility at a major conserved antibody target on hepatitis C virus E2 antigen. Proc Natl Acad Sci U S A 113:12768–12773. https://doi.org/10.1073/pnas.1609780113

495. Zamai L (2021) Upregulation of the Renin–Angiotensin System Pathways and SARS-CoV-2 Infection: The Rationale for the Administration of Zinc-Chelating Agents in COVID-19 Patients. Cells 10:506. https://doi.org/10.3390/cells10030506

496. Kiseleva AA, Troisi EM, Hensley SE, et al (2021) SARS-CoV-2 spike protein binding selectively accelerates substrate-specific catalytic activity of ACE2. J Biochem (Tokyo) 170:299–306. https://doi.org/10.1093/jb/mvab041

497. Lu J, Sun PD (2020) High affinity binding of SARS-CoV-2 spike protein enhances ACE2 carboxypeptidase activity. J Biol Chem 295:18579–18588. https://doi.org/10.1074/jbc.RA120.015303

498. Towler P, Staker B, Prasad SG, et al (2004) ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis*. J Biol Chem 279:17996–18007. https://doi.org/10.1074/jbc.M311191200

499. Surpeta B, Sequeiros-Borja CE, Brezovsky J (2020) Dynamics, a Powerful Component of Current and Future in Silico Approaches for Protein Design and Engineering. Int J Mol Sci 21:2713. https://doi.org/10.3390/ijms21082713

500. Weng G, Wang E, Wang Z, et al (2019) HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. Nucleic Acids Res 47:W322–W330. https://doi.org/10.1093/nar/gkz397

501. Brown DK, Sheik Amamuddy O, Tastan Bishop Ö (2017) Structure-Based Analysis of Single Nucleotide Variants in the Renin-Angiotensinogen Complex. Glob Heart 12:121–132. https://doi.org/10.1016/j.gheart.2017.01.006

502. Tizghadam A, Leon-Garcia A (2010) Betweenness centrality and resistance distance in communication networks. IEEE Netw 24:10–16. https://doi.org/10.1109/MNET.2010.5634437

503. Barthélemy M (2004) Betweenness centrality in large complex networks. Eur Phys J B 38:163–168. https://doi.org/10.1140/epjb/e2004-00111-4

504. Hong J, Kwon HJ, Cachau R, et al (2021) Camel nanobodies broadly neutralize SARS-CoV-2 variants. bioRxiv 2021.10.27.465996. https://doi.org/10.1101/2021.10.27.465996

505. Mathew SM, Benslimane F, Althani AA, Yassine HM (2021) Identification of potential natural inhibitors of the receptor-binding domain of the SARS-CoV-2 spike protein using a computational docking approach. Qatar Med J 2021:12. https://doi.org/10.5339/qmj.2021.12

506. Kwofie SK, Broni E, Asiedu SO, et al (2021) Cheminformatics-Based Identification of Potential Novel Anti-SARS-CoV-2 Natural Compounds of African Origin. Molecules 26:406. https://doi.org/10.3390/molecules26020406

507. Bharathi M, Sivamaruthi BS, Kesika P, et al (2022) In Silico Screening of Bioactive Compounds of Representative Seaweeds to Inhibit SARS-CoV-2 ACE2-Bound Omicron B.1.1.529 Spike Protein Trimer. Mar Drugs 20:148. https://doi.org/10.3390/md20020148

508. Pal M, Musib D, Zade AJ, et al (2021) Computational Studies of Selected Transition Metal Complexes as Potential Drug Candidates against the SARS-CoV-2 Virus. ChemistrySelect 6:7429–7435. https://doi.org/10.1002/slct.202101852

509. Biswas S, Mahmud S, Mita MA, et al (2022) Molecular Docking and Dynamics Studies to Explore Effective Inhibitory Peptides Against the Spike Receptor Binding Domain of SARS-CoV-2. Front Mol Biosci 8:791642. https://doi.org/10.3389/fmolb.2021.791642

510. Kiran G, Karthik L, Shree Devi MS, et al (2022) In Silico computational screening of Kabasura Kudineer - Official Siddha Formulation and JACOM against SARS-CoV-2 spike protein. J Ayurveda Integr Med 13:100324. https://doi.org/10.1016/j.jaim.2020.05.009

511. Yepes-Pérez AF, Herrera-Calderon O, Quintero-Saumeth J (2022) Uncaria tomentosa (cat's claw): a promising herbal medicine against SARS-CoV-2/ACE-2 junction and SARS-CoV-2 spike protein based on molecular modeling. J Biomol Struct Dyn 40:2227–2243. https://doi.org/10.1080/07391102.2020.1837676

512. Jawad B, Adhikari P, Podgornik R, Ching W-Y (2022) Binding Interactions between Receptor-Binding Domain of Spike Protein and Human Angiotensin Converting Enzyme-2 in Omicron Variant. J Phys Chem Lett 13:3915–3921. https://doi.org/10.1021/acs.jpclett.2c00423

513. Li Z, Zhang JZH (2022) Mutational Effect of Some Major COVID-19 Variants on Binding of the S Protein to ACE2. Biomolecules 12:572. https://doi.org/10.3390/biom12040572

514. Socher E, Heger L, Paulsen F, et al (2022) Molecular dynamics simulations of the delta and omicron SARS-CoV-2 spike – ACE2 complexes reveal distinct changes between both variants. Comput Struct Biotechnol J 20:1168–1176. https://doi.org/10.1016/j.csbj.2022.02.015

515. Khan A, Zia T, Suleman M, et al (2021) Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. J Cell Physiol 236:7045–7057. https://doi.org/10.1002/jcp.30367

516. Liu Y, Liu J, Plante KS, et al (2022) The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. Nature 602:294–299. https://doi.org/10.1038/s41586-021-04245-0

517. Li F, Li W, Farzan M, Harrison SC (2005) Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. Science 309:1864–1868. https://doi.org/10.1126/science.1116480

518. Brielle ES, Schneidman-Duhovny D, Linial M (2020) The SARS-CoV-2 Exerts a Distinctive Strategy for Interacting with the ACE2 Human Receptor. Viruses 12:497. https://doi.org/10.3390/v12050497

519. Yan R, Zhang Y, Li Y, et al (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367:1444–1448. https://doi.org/10.1126/science.abb2762

520. Yi C, Sun X, Lin Y, et al (2021) Comprehensive mapping of binding hot spots of SARS-CoV-2 RBD-specific neutralizing antibodies for tracking immune escape variants. Genome Med 13:164. https://doi.org/10.1186/s13073-021-00985-w

521. Sikora M, Bülow S von, Blanc FEC, et al (2021) Computational epitope map of SARS-CoV-2 spike protein. PLOS Comput Biol 17:e1008790. https://doi.org/10.1371/journal.pcbi.1008790

522. Cerutti G, Guo Y, Liu L, et al (2022) Cryo-EM structure of the SARS-CoV-2 Omicron spike. Cell Rep 38:110428. https://doi.org/10.1016/j.celrep.2022.110428

523. Saldaño TE, Tosatto SCE, Parisi G, Fernandez-Alberti S (2019) Network analysis of dynamically important residues in protein structures mediating ligand-binding conformational changes. Eur Biophys J 48:559–568. https://doi.org/10.1007/s00249-019-01384-1

524. Kuba K, Imai Y, Ohto-Nakanishi T, Penninger JM (2010) Trilogy of ACE2: A peptidase in the renin–angiotensin system, a SARS receptor, and a partner for amino acid transporters. Pharmacol Ther 128:119–128. https://doi.org/10.1016/j.pharmthera.2010.06.003

525. Xu P, Sriramula S, Lazartigues E (2011) ACE2/ANG-(1–7)/Mas pathway in the brain: the axis of good. Am J Physiol-Regul Integr Comp Physiol 300:R804–R817. https://doi.org/10.1152/ajpregu.00222.2010

526. Passos-Silva DG, Verano-Braga T, Santos RAS (2012) Angiotensin-(1–7): beyond the cardio-renal actions. Clin Sci 124:443–456. https://doi.org/10.1042/CS20120461

527. Zamai L (2020) The Yin and Yang of ACE/ACE2 Pathways: The Rationale for the Use of Renin-Angiotensin System Inhibitors in COVID-19 Patients. Cells 9:1704. https://doi.org/10.3390/cells9071704

528. Tipnis SR, Hooper NM, Hyde R, et al (2000) A Human Homolog of Angiotensin-converting Enzyme: CLONING AND FUNCTIONAL EXPRESSION AS A CAPTOPRIL-INSENSITIVE CARBOXYPEPTIDASE*. J Biol Chem 275:33238–33243. https://doi.org/10.1074/jbc.M002615200

529. Donoghue M, Hsieh F, Baronas E, et al (2000) A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. Circ Res 87:E1-9. https://doi.org/10.1161/01.res.87.5.e1

530. Fülöp V, Böcskei Z, Polgár L (1998) Prolyl Oligopeptidase: An Unusual β-Propeller Domain Regulates Proteolysis. Cell 94:161–170. https://doi.org/10.1016/S0092-8674(00)81416-6

531. Rockel B, Peters J, Kühlmorgen B, et al (2002) A giant protease with a twist: the TPP II complex from Drosophila studied by electron microscopy. EMBO J 21:5979–5984. https://doi.org/10.1093/emboj/cdf601

532. Hoffmann M, Kleine-Weber H, Schroeder S, et al (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181:271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052

533. Crackower MA, Sarao R, Oliveira-dos-Santos AJ, et al (2002) Angiotensin-converting enzyme 2 is an essential regulator of heart function. Nature 417:822–828. https://doi.org/10.1038/nature00786

534. Bhushan S, Xiao Z, Gao K, et al (2022) Role and Interaction Between ACE1, ACE2 and Their Related Genes in Cardiovascular Disorders. Curr Probl Cardiol 101162. https://doi.org/10.1016/j.cpcardiol.2022.101162

535. Gaurav M, Vibhanshu G, Shahzad FH, Anwer SK (2011) Effect of angiotensin converting enzyme gene I/D polymorphism in patients with metabolic syndrome in North Indian population. Chin Med J (Engl) 124:45–48. https://doi.org/10.3760/cma.j.issn.0366-6999.2011.01.009

536. Dong B, Yu QT, Dai HY, et al (2012) Angiotensin-Converting Enzyme-2 Overexpression Improves Left Ventricular Remodeling and Function in a Rat Model of Diabetic Cardiomyopathy. J Am Coll Cardiol 59:739–747. https://doi.org/10.1016/j.jacc.2011.09.071

537. Fan Z, Wu G, Yue M, et al (2019) Hypertension and hypertensive left ventricular hypertrophy are associated with ACE2 genetic polymorphism. Life Sci 225:39–45. https://doi.org/10.1016/j.lfs.2019.03.059

538. Yang W, Huang W, Su S, et al (2006) Association study of ACE2 (angiotensin I-converting enzyme 2) gene polymorphisms with coronary heart disease and myocardial infarction in a Chinese Han population. Clin Sci 111:333–340. https://doi.org/10.1042/CS20060020

539. Bellone M, Calvisi SL (2020) ACE polymorphisms and COVID-19-related mortality in Europe. J Mol Med 98:1505–1509. https://doi.org/10.1007/s00109-020-01981-0

540. Berge KE, Berg K (1997) Cardiovascular risk factors in people with different genotypes in the insertion/deletion (I/D) polymorphism at the locus for angiotensin I-converting enzyme (ACE). Clin Genet 52:422–426. https://doi.org/10.1111/j.1399-0004.1997.tb02562.x

541. Riera-Fortuny C, Real JT, Chaves FJ, et al (2005) The relation between obesity, abdominal fat deposit and the angiotensin-converting enzyme gene I/D polymorphism and its association with coronary heart disease. Int J Obes 29:78–84. https://doi.org/10.1038/sj.ijo.0802829

542. He Q, Fan C, Yu M, et al (2013) Associations of ACE gene insertion/deletion polymorphism, ACE activity, and ACE mRNA expression with hypertension in a Chinese population. PloS One 8:e75870. https://doi.org/10.1371/journal.pone.0075870

543. Devaux CA, Rolain J-M, Raoult D (2020) ACE2 receptor polymorphism: Susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. J Microbiol Immunol Infect 53:425–435. https://doi.org/10.1016/j.jmii.2020.04.015

544. Hashemi SMA, Thijssen M, Hosseini SY, et al (2021) Human gene polymorphisms and their possible impact on the clinical outcome of SARS-CoV-2 infection. Arch Virol 166:2089–2108. https://doi.org/10.1007/s00705-021-05070-6

545. Pouladi N, Abdolahi S (2021) Investigating the ACE2 polymorphisms in COVID-19 susceptibility: An in silico analysis. Mol Genet Genomic Med 9:e1672. https://doi.org/10.1002/mgg3.1672

546. Karczewski KJ, Francioli LC, Tiao G, et al (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581:434–443. https://doi.org/10.1038/s41586-020-2308-7

547. Alberts B, Johnson A, Lewis J, et al (2002) The Shape and Structure of Proteins. Mol Biol Cell 4th Ed

548. Sælensminde G, Halskau Ø, Jonassen I (2009) Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role. Extremophiles 13:11–20. https://doi.org/10.1007/s00792-008-0192-4

549. Grigas AT, Mei Z, Treado JD, et al (2020) Using physical features of protein core packing to distinguish real proteins from decoys. Protein Sci 29:1931–1944. https://doi.org/10.1002/pro.3914

550. Regan L, Caballero D, Hinrichsen MR, et al (2015) Protein design: Past, present, and future. Biopolymers 104:334–350. https://doi.org/10.1002/bip.22639

551. Muchtaridi M, Fauzi M, Khairul Ikram NK, et al (2020) Natural Flavonoids as Potential Angiotensin-Converting Enzyme 2 Inhibitors for Anti-SARS-CoV-2. Molecules 25:3980. https://doi.org/10.3390/molecules25173980

552. Al-Karmalawy AA, Dahab MA, Metwaly AM, et al (2021) Molecular Docking and Dynamics Simulation Revealed the Potential Inhibitory Activity of ACEIs Against SARS-CoV-2 Targeting the hACE2 Receptor. Front Chem 9:

553. Panda PK, Arul MN, Patel P, et al Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. Sci Adv 6:eabb8097. https://doi.org/10.1126/sciadv.abb8097

554. Baig MS, Reyaz E, Selvapandiyan A, Krishnan A (2021) Differential binding of SARS-CoV-2 Spike protein variants to its cognate receptor hACE2 using molecular modeling based binding analysis. Bioinformation 17:337–347. https://doi.org/10.6026/97320630017337

555. Gu C, Wu Y, Guo H, et al (2021) Protoporphyrin IX and verteporfin potently inhibit SARS-CoV-2 infection in vitro and in a mouse model expressing human ACE2. Sci Bull 66:925–936. https://doi.org/10.1016/j.scib.2020.12.005

556. Ferdausi N, Islam S, Rimti FH, et al (2022) Point-specific interactions of isovitexin with the neighboring amino acid residues of the hACE2 receptor as a targeted therapeutic agent in suppressing the SARS-CoV-2 influx mechanism. J Adv Vet Anim Res 9:230–240. https://doi.org/10.5455/javar.2022.i588

557. Hussain M, Jabeen N, Raza F, et al (2020) Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. J Med Virol 92:1580–1586. https://doi.org/10.1002/jmv.25832

558. Bakhshandeh B, Sorboni SG, Javanmard A-R, et al (2021) Variants in ACE2; potential influences on virus infection and COVID-19 severity. Infect Genet Evol 90:104773. https://doi.org/10.1016/j.meegid.2021.104773

559. Bünning P, Riordan JF (1985) The functional role of zinc in angiotensin converting enzyme: implications for the enzyme mechanism. J Inorg Biochem 24:183–198. https://doi.org/10.1016/0162-0134(85)85002-9

560. Wang J, Xu X, Zhou X, et al (2020) Molecular simulation of SARS-CoV-2 spike protein binding to pangolin ACE2 or human ACE2 natural variants reveals altered susceptibility to infection. J Gen Virol 101:921–924. https://doi.org/10.1099/jgv.0.001452

561. Laurini E, Marson D, Aulic S, et al (2020) Computational Alanine Scanning and Structural Analysis of the SARS-CoV-2 Spike Protein/Angiotensin-Converting Enzyme 2 Complex. ACS Nano 14:11821–11830. https://doi.org/10.1021/acsnano.0c04674

562. Nipun TS, Ema TI, Mia MdAR, et al (2021) Active site-specific quantum tunneling of hACE2 receptor to assess its complexing poses with selective bioactive compounds in co-suppressing SARS-CoV-2 influx and subsequent cardiac injury. J Adv Vet Anim Res 8:540–556. https://doi.org/10.5455/javar.2021.h544

563. Houchi S, Messasma Z (2022) Exploring the inhibitory potential of Saussurea costus and Saussurea involucrata phytoconstituents against the Spike glycoprotein receptor binding domain of SARS-CoV-2 Delta (B.1.617.2) variant and the main protease (Mpro) as therapeutic candidates, using Molecular docking, DFT, and ADME/Tox studies. J Mol Struct 1263:133032. https://doi.org/10.1016/j.molstruc.2022.133032

564. Uddin MB, Sajib EH, Hoque SF, et al (2022) Genomic diversity and molecular dynamics interaction on mutational variances among RB domains of SARS-CoV-2 interplay drug inactivation. Infect Genet Evol 97:105128. https://doi.org/10.1016/j.meegid.2021.105128

565. Paloni M, Bussi G, Barducci A (2021) Arginine multivalency stabilizes protein/RNA condensates. Protein Sci 30:1418–1426. https://doi.org/10.1002/pro.4109

566. Mackenzie HW, Hansen DF (2019) Arginine Side-Chain Hydrogen Exchange: Quantifying Arginine Side-Chain Interactions in Solution. ChemPhysChem 20:252–259. https://doi.org/10.1002/cphc.201800598

567. Ravaioli S, Tebaldi M, Fonzi E, et al (2020) ACE2 and TMPRSS2 Potential Involvement in Genetic Susceptibility to SARS-COV-2 in Cancer Patients. Cell Transplant 29:0963689720968749. https://doi.org/10.1177/0963689720968749

568. Rodriguez JA, Gonzalez J, Arboleda-Bustos CE, et al (2022) Computational modeling of the effect of five mutations on the structure of the ACE2 receptor and their correlation with infectivity and virulence of some emerged variants of SARS-CoV-2 suggests mechanisms of binding affinity dysregulation. Chem Biol Interact 368:110244. https://doi.org/10.1016/j.cbi.2022.110244

569. Al-Mulla F, Mohammad A, Al Madhoun A, et al (2021) ACE2 and FURIN variants are potential predictors of SARS-CoV-2 outcome: A time to implement precision medicine against COVID-19. Heliyon 7:e06133. https://doi.org/10.1016/j.heliyon.2021.e06133

570. Cao Y, Li L, Feng Z, et al (2020) Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov 6:11. https://doi.org/10.1038/s41421-020-0147-1

571. Devaux CA, Pinault L, Osman IO, Raoult D (2021) Can ACE2 Receptor Polymorphism Predict Species Susceptibility to SARS-CoV-2? Front Public Health 8:

572. Abobaker A, Nagib T, Alsoufi A (2021) The impact of certain genetic variants (single nucleotide polymorphisms) on incidence and severity of COVID-19. J Gene Med 23:e3310. https://doi.org/10.1002/jgm.3310

573. Gagliardi MC, Tieri P, Ortona E, Ruggieri A (2020) ACE2 expression and sex disparity in COVID-19. Cell Death Discov 6:1–2. https://doi.org/10.1038/s41420-020-0276-1

574. Chaudhry F, Lavandero S, Xie X, et al (2020) Manipulation of ACE2 expression in COVID-19. Open Heart 7:e001424. https://doi.org/10.1136/openhrt-2020-001424

575. Huang S-W, Wang S-F (2021) SARS-CoV-2 Entry Related Viral and Host Genetic Variations: Implications on COVID-19 Severity, Immune Escape, and Infectivity. Int J Mol Sci 22:3060. https://doi.org/10.3390/ijms22063060

576. Dadgostar P (2019) Antimicrobial Resistance: Implications and Costs. Infect Drug Resist 12:3903–3910. https://doi.org/10.2147/IDR.S234610

577. Friedman ND, Temkin E, Carmeli Y (2016) The negative impact of antibiotic resistance. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis 22:416–422. https://doi.org/10.1016/j.cmi.2015.12.002

578. Silver LL (2011) Challenges of Antibacterial Discovery. Clin Microbiol Rev 24:71–109. https://doi.org/10.1128/CMR.00030-10

579. Kantelis KF, Asteriou V, Papadimitriou-Tsantarliotou A, et al (2022) Graph theory-based simulation tools for protein structure networks. Simul Model Pract Theory 121:102640. https://doi.org/10.1016/j.simpat.2022.102640

Supplementary Material

Table S 1: RMSDs of lowest energy mutant conformations compared to that of the WT

Protein system	WT - mutant RMSD (Å)
S140N	3.118
S140R	2.702
G279D	2.961
G285D	2.723
S315I	2.388
S315N	3.396
S315R	2.519
S315T	2.175
G316D	2.646
S457I	3.326
G593D	2.775

Table S 2: Analysis of protomer A persistent hub contact fingerprint showing the new/increased and the lost/reduced interactions for every mutant system in comparison to the WT. neighbouring resides (non-hubs) are bold and underlined. Protomer B are shown with an Asterix

Protein	New/increased interaction	Lost/reduced interaction
S140N	N44-**Q36**, N44-**R42**, V47-**Q50**, H49-**P52**, H49-**P52***, H49-**S620**, A555-**A559**, L616-**P589**, A621-H49, A621-**A55***, E703-**F129***, E703-**A130***	N44-**E31***, V47-**L611**, L48-**M624**, H49-H49*, L616-**F483**, L616-R484, E703-**E192***, E703-**P193***
S140R	N44-**G33***, N44-**Q36**, N44-**R42,** V47-**G34***, V47-**Q50**, H49-**K46**, H49-**K46*** H49-**Q50**, A621-**P52***, E703-**F129***, E703-**A130***	P29-V47*, N44-N**41,** V47-**L612**, H49-**S620**, A555-**A559**, L616-**F483**, A621-E703, A621-**L704**,
G279D	N44-**R42**, L45-**E192***, V47-**Q50**, L48-**Q50**, L48-**M164**, H49-**P52***, H49-**A55***, H49-**A621**, H49-**S620**, G547-L459, A621-H49, A621-P**52***, A621-**W191***,	N44-P29*, N44-**G33***, V7-**E195**, H49-H49*, H49-L46, V473-G548, G548-V473, L616-**F483**, E703-A621
G285D	N44-**R42**, V47-**Q50**, H49-**P52**, H49-**P52*,** AH49-621, L616-**K590**, A621-H49, A621-P**52***, A621-**W191***, A621-A**55***	N44-**G33***, V47-**L611**, H49-H49*, A621-E703, A621-**L704**,
S315I	N44-**G33***, L45-**E192***, V47-**Q50**, H49-**P52***, H49-**A55***, H49-**A621**, G547-**V544**, A621-H49, A621-P**52***, A621-**W191***, E703-**R42**, E703-**L43**	N44-**E31***, V47-**V30***, V47-**L611**, H49-H49*, A555-**A559**, A621-E703, A621-**L704**, E703-**E192***, E703-**P193***
S315N	N44-**Q36**, N44-**R42**, L45-H49, V47-**Q50**, H49-**K46**, H49-**K46**, H49-**P52**, H49-**P52***, H49-**A55***, H49-**V54***, H49-**A621**, A551-**L634**, A555-**A559**, L616-**G118**, A621-H49, A621-**A55***	N44-**G33**, N44-V47, L45-**L43**, V47-N44, H49-H49*, L616-**D117**, L616-**F483**, L616-**G485**, A621-E703, A621-**L704**, E703-**P193***
S315R	N44-**R42**, N44-**E192***, N44-**P193***, N44-E703, L45-P52*, L45-**W191***, L45-**E192***, V47-**E31***, V47-**G32***, V47-**Q50**, H49-**P52***, H49-**A55***, H49-**A621**, G547-**V554**, L616-S**486**, L616-D**612**, A621-H49, A621-P**52***, A621-**W191***, A621-**A55***, E703-**L43**, ASN44	N44-**P21***, N44-**V30***, N44-**E31***, N44-**G33***, N44-V47, L45-**L43**, V47-P29*, H49-H49*, H49-L45, L616-**D117**, L616-**F483**, L616-R**484**, A621-E703, A621-**L704**
S315T	N44-**Q36**, N44-**R42**, L45-**W191***, L45-**E192***, L48-**P52***, H49-**K46**, H49-**P52***, H49-A53*, H49-A614, V473-L631, G458-**S475**, N44-**Q36**, A555-**A559**, A621-P**52***, A621-**W191***	N44-**G33***, V47-**G33***, L472-**V517**, A621- E703, A621-**L704**
G316D	N44-**Q36**, N44-**R42**, L45-H49, V47-**Q50**, V47-**D621**, V47-**A614**, H49-L45, H49-**K46**, H49-**N51***, H49-**P52**, H49-P193*, G548-**M732**, A555-**A559**, A621-**V54***, A621-H49	N44-**E31***, N44-V47, L45-V47, L45-**A621**, V47-P29*, V47-N44, H49-H49*, V473-**A636**, G547-**A476**, L616-**F483**
S457I	N44-**Q36**, N44-**R42**, V47-**Q50**, H49-**K46***, G547-**L521**, G547-**V544**, A555-**A559**, L616-**G118**, A621-H49, A621-**N51***, A621-**A55***,	N44-**G33***, V47-E195, V47-**L611**, H49-**S620**, L616-**F483**, L616-R**484**, A621-**L704**
G593D	N44-**Q36**, N44-L48, L45-H49, L45-**E192***, V47-**G32***, V47-**G33***, V47-**K46***, L48-V30*, L48-L43, L48-N44, H49-**N51***, H49-**P52**, H49-**A55**, H49-**A621**, G548-**V544**, L616-V586, A621-**L611**, A621-**A614** E703-**R42**, E703-**L43**	N44-P29*, N44-V30*, N44-E31*, L45-**L43**, V47-P29*, V47-**N615**, L48-**A614**, L48-**N615**, H49-H49*, L616-**F483**, L616-R**484**, A621-E703, E703-**A706**

Table S 3: Analysis of protomer B persistent hub contact fingerprint showing the new/increased and the lost/reduced interactions for every mutant system in comparison to the WT. neighbouring resides (non-hubs) are bold and underlined. Protomer A are shown with an Asterix

Protein	New/increased interaction	Lost/reduced interaction
S140N	V47-**Q50**, H49-**P52**, H49-**S620**, H49-A621, N44-**G33**, L472-**T468***, L616-**F483***, L616-R**484**, A621-H49*, A621-**A55**	L48-**M624**, H49-H49*, L616-**G485***, A621-**P52**, E703-**L43**, E703-**W191***
S140R	L45-**L43**, V47-**G34***, L48-**Q50**, H49-**K46**, H49-**K46***, H49-**Q50**, N44-**G33**, N44-**Q36***, N44-**D37***, L616-**F483***, A621-**A55**	H49-A621, A555-**A559**, N44-**P29**, L616-**D117***, L616-R**484**, L616-**G485***, E703-**L43**, E703-**W191**
G279D	L45-**E192***, V47-**Q50**, L48-**Q50**, L48-**L611**, L48-**M624**, H49-**P52**, H49-**A55**, H49-**S620**, H49-A621, V473-**G635**, G547-**K459**, G547-**V544**, G548-**F567**, N44-**G33**, L616-R**484**	L45-**L43**, V47-**E195**, H49-**K46**, H49-H49*, V473-G551, G547-**A475**, L616-**D117***, A621-**P52**, E703-**L43**, E703-**W191**
G285D	V47-**Q50**, H49-**P52***, H49-**P52**, H49-T618, H49-**S620**, H49-A621, N44-**G33**, L616-**F483***, L616-R**484**	V47-**L611**, H49-H49*, L616-**D117***, A621-**P52**, E703-**W191**
S315I	L45-H49, L45-**W191***, L45-**E192**, V47-**V30***, V47-**Q50**, H49-**P52***, H49-**A55***, H49-**S260**, H49-A621, N44-**G33**, N44-**N35***, L616-R**484***, E703-**P193***	V47-**V30***, V47-**L611**, H49-H49*, G547-**A476**, A555-**A559**, L472-**V517***, L616-**D117***, A621-**P52**, E703-**L43**, E703-**W191**
S315N	L45-H49, H49-L45, H49-**K46**, H49-**P52**, A555-**A559**, N44-**G33**, L616-**G118***, A621-H49*, A621-**A55**	L45-**L43**, V47-N44, V47-**Q50**, H49-H49*, E703-**L43**, E703-**W191**
S315R	L45-**P52***, L45-**W191***, L45-**E192**, V47-**E31**, V47-**G32***, V47-**Q50**, H49-L45, H49-**P52***, H49-**S620**, H49-**P52**, H49-**A55**, N44-**G33**, N44-**G33***, N44-**Q36***, L616-**F483***, L616-R**484**, E703-**A130***	L45-**L43**, V47-P29*, V47-N44, V47-**E195**, V47-**L611**, H49-H49*, H49-A621, L472-**V517***, L616-**G485***, A621-**P52**, E703-**L43**, E703-**W191**
S315T	L45-**E192**, V47-**Q50**, L48-**P52***, L48-**M624**, H49-**K46**, H49-**P52***, A555-**A559**, N44-**G33***, N44-**Q36***, L616-**D117***, A621-H49*	L45-**L43**, L45-**E192***, V47-**G33***, V47-**E195**, V47-**L611**, L472-**V517***, L616-**G485***, A621-**P52**, E703-**L43**, E703-**W191***
G316D	L45-V47, L45-H49, V47-**Q50**, V47-**D612**, V47-**A614**, H49-L45, H49-**K46**, H49-**N51***, H49-**P52**, H49-**P193***, V473-**G635,** G548-**M732**, N44-**G33***, N44-L48*, N44-**E195***, L616-R**484***, A621-H49*, E703-L48	L45-A621, V47-P29*, V47-N44, V47-**E195**, L48-**L619**, H49-H49*, H49-**S620**, G547-**A476**, L616-**D117***, A621-**L45***, E703-**W191**, E703-E192, E703-**A706**
S457I	V47-**Q50**, L48-**K46**, H49-**K46**, H49-**K46***, G547-**V544**, A555-**A559**, N44-**G33***, N44-**G34**, L616-**G118***, L616-R**484**, E703-**D42**	L45-**E192***, V47-**E195**, V47-**L611**, H49-L45, H49-**A624**, N44-**Q36***, L472-**V517***, L616-**D117***, L616-**G485***, E703-**F129***, E703-**W191***
G593D	L45-V47, L45-H49, V47-L45, V47-**Q50**, V47-**D612**, V47-**A614**, L48-**V30***, L48-**L43**, L48-N44, H49-**N51***, H49-L45, H49-**P52***, H49-**A55***, H49-**S630**, G547-**L459**, G548-**V544**, N44-**G33***, L616-R**484***, A621-H49*, A621-**Q50***, E703-**P193***	L45-A621, V47-P29*, V47-N44, V47-**E195**, V47-**L611**, V47-**N615**, H49-H49*, G547-**V544**, L616-**G485***, E703-**L43**, E703-**W191***

Table S 4: Systems destabilizing interface residues as determined by alanine scanning. Common residues between alanine scanning and *BC* are red, *CC*: blue, *DC*: green, *EC*: yellow and *KC*: purple. Residues present in more than one metric are shown in bold

System	Hot spot residues per chain
S140N	**Protomer A**: 28, 30, 31, 36, 38, 39, 42, **44**, 46, 49, 129, 149, 161, **195**, 197, 204, 287, 289, 293, 294, 296, 380, 394, 425, 608, 664, 703, 707, 709, 710
	Protomer B: 27, 28, 30, **35**, 36, 38, 39, 128, 129, 135, 146, 149, 161, 191, 192, **195**, 197, 198, 293, 296
S140R	**Protomer A**: 27, 28, 30, **35**, **36**, 38, 39, 42, **44**, **49**, 56, 129, 133, 146, 149, 153, 157, 161, 191, **192**, 195, 197, 289, 295, 296, 320, 324, 379, 380, **419**, 425, 426, 604, 608, 664, 669, 678, 681, 701, 702, **703**, 705, 709, 710, 711, 715, 719
	Protomer B: 28, 30, 31, 35, 36, 37, 39, 42, **44**, 46, 49, 129, 133, 149, 153, 197, 204, 287, 293, 294, 296, 306
G279D	**Protomer A:** 27, 28, 30, 31, 38, 39, 42, 44, 49, 129, 149, 191, **192**, 195, 197, 198, 289, 293, 320, 324, 380, 417, 418, 427, 604, 664, 678, 702, 707, 709, 715, 719
	Protomer B: 28, 30, 31, 35, 38, 39, 41, 46, 49, 90, 133, 149, 161, 191, 192, 197, 200, 204, 289, 293, 294, 296
G285D	**Protomer A**: 28, 30, 35, 38, 39, 49, 129, 146, 149, 191, **192**, 195, 197, 204, 293, 294, 296, 324, 380, 394, 412, 417, 431, 604, 608, 664, 678, 702, 709, 710, 719
	Protomer B: 26, 28, 30, 31, 35, 38, 39, 41, 44, **46**, **49**, 128, 129, 146, 149, 153, 191, **195**, 197, 198, 289, 293, 294, 296
S315I	**Protomer A**: 28, 30, 35, 36, 38, 39, 41, 42, 49, 56, 128, 129, 133, 146, 149, 152, 161, 191, 192, 195, 197, 204, 289, 293, 294, 296, 320, 324, 380, 418, 425, 426, 431, 435, 439, 608, 664, 702, 703, 705, 707, 709, 710, 715
	Protomer B: 28, 30, 31, 38, 39, 42, 46, 49, 56, 128, 129, 146, 149, 161, 189, 191, 192, 195, 197, 204, 289, 293, 296
S315N	**Protomer A**: 27, 28, 30, 31, 35, 38, 39, 42, 44, **49**, 51, 129, 149, 161, 197, 287, 293, 294, 296, 320, 324, 379, 380, 394, 419, 421, 425, 604, 664, 678, 702, 703, 707, 709, 715, 719
	Protomer B: 26, 28, 30, 35, 36, 39, 42, 44, 128, 133, 135, 149, 155, 195, 197, 204, 289, 293, 296
S315R	**Protomer A**: 26, 28, 30, **35**, 36, 38, 39, 41, **42**, **46**, 49, 133, 134, 149, 191, 201, 203, 204, 289, 293, 296, 324, 336, 380, 394, 409, 418, 431, 435, 608, 664, 678, 702, 703, 707, 709, 715, 719, 723
	Protomer B: 28, 30, **31**, 37, 38, 39, 42, 44, 90, 129, 133, 134, 149, 157, 161, 191, **192**, 195, 197, 204, 289, 293, 294, 296, 306
S315T	**Protomer A**: 27, 28, 30, 35, 36, 38, 39, 46, 49, 51, 56, 129, 133, 149, 161, 191, 192, 197, 204, 289, 293, 294, 296, 324, 380, 394, 418, 425, 426, 427, 431, 604, 664, 678, 707, 709, 710, 711, 715, 719, 723
	Protomer B: 28, 30, 37, 38, 39, 41, **49**, 129, 149, 161, 191, **192**, 197, 200, 204, 293, 294, 296
G316D	**Protomer A:** 28, 30, 35, **36**, 38, 39, 128, 129, 149, 157, 191, 195, 197, 289, 293, 294, 296, 320, 324, 380, 394, 409, 419, 425, 426, 431, 608, 664, 705, 707, 715
	Protomer B: 26, 27, 28, 30, 35, 36, 38, 39, 42, **44**, 46, 90, 146, 149, 153, 191, 195, 197, 200, 204, 217, 293, 294, 296
S457I	**Protomer A**: 28, 30, 37, 38, 39, 42, 44, **49**, 54, 90, 129, 135, 149, 152, 161, **192**, 195, 197, 289, 293, 294, 296, 320, 324, 380, 394, 418, 425, 604, 608, 705, 715, 719
	Protomer B: 27, 28, 30, **35**, 36, 38, 39, 42, 44, 46, **49**, 129, 133, 146, 149, 157, 161, 191, 197, 198, 289, 295, 296
G593D	**Protomer A**: 27, 28, 30, 35, 36, 38, 39, 42, **46**, 49, 51, 128, 129, 133, 149, 195, 197, 204, 287, 289, 293, 294, 296, 320, 380, 394, 412, 417, 418, 419, 425, 427, 431, 435, 608, 620, 664, 678, 709, 710, 715, 719
	Protomer B: 28, 30, 31, 36, 38, 39, 42, 44, **46**, 49, 51, 90, 128, 129, 133, 149, 191, 192, 197, 198, 204, 289, 293, 294, 296

Table S 5: References for the PZA drug resistance mutations

Mutation(s)	Reference
A3E	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2045301/
A3P	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC89359/
A46V	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC89721/
A102V	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC89360/
A134V, T142P and V155A	https://www.ncbi.nlm.nih.gov/pubmed/10645449
V7G, Q10P, G17D, Y34S, H43P, H51P, H51R, D63G, S67P, T114P, V130G, G132D, S104R, A146V and L172P	https://www.ncbi.nlm.nih.gov/pubmed/9692180
D49G, S66P, L116R, T160P and A161P	https://www.ncbi.nlm.nih.gov/pubmed/10390239
V128G and A171E	https://www.ncbi.nlm.nih.gov/pubmed/15616332
A171P	https://www.ncbi.nlm.nih.gov/pubmed/10089282
H51Y, L85R and A171V,	https://www.ncbi.nlm.nih.gov/pubmed/10471589
C14H, L35R, H57Y, L159P and S185T	https://www.ncbi.nlm.nih.gov/pubmed/16848344
D12A, C14R, C72R, L85P, K96N, G132S and T142K	https://www.ncbi.nlm.nih.gov/pubmed/9055989
C138Y and Q141P	https://www.ncbi.nlm.nih.gov/pubmed/8640557
V9A, V9G, D8G, D49A, W68G, W68R L159R and Y103S	https://www.ncbi.nlm.nih.gov/pubmed/11641519
D53A, H82R, G97S, V139M and R140S	https://www.ncbi.nlm.nih.gov/pubmed/10681313
H71R, G78D, G97D, V139G and V139L	https://www.ncbi.nlm.nih.gov/pubmed/11083630
L19R, V21G and G162D	https://www.ncbi.nlm.nih.gov/pubmed/17360809
H57P	https://www.ncbi.nlm.nih.gov/pubmed/10882091
L19P, T47A, P54T, W68L, T76P, R121P, W119R, V139A and H137R	https://www.ncbi.nlm.nih.gov/pubmed/9056006
L35P	https://www.ncbi.nlm.nih.gov/pubmed/19234602
T61P, V155G	https://www.ncbi.nlm.nih.gov/pubmed/10390238
V45G	https://www.ncbi.nlm.nih.gov/pubmed/11641519

Table S 6: Mutational effect on the stability of MtPncA-NAM as per mCSM web server

System	Predicted ΔΔG	Effect	System	Predicted ΔΔG	Effect
A3E	-2.528	Highly Destabilizing	H82R	-0.817	Destabilizing
A3P	-0.516	Destabilizing	L85P	-1.423	Destabilizing
V7G	-2.555	Highly Destabilizing	L85R	-1.387	Destabilizing
D8G	-1.131	Destabilizing	K96N	-1.185	Destabilizing
V9A	-2.267	Highly Destabilizing	G97D	-2.785	Highly Destabilizing
V9G	-2.844	Highly Destabilizing	G97S	-1.615	Destabilizing
Q10P	-0.317	Destabilizing	A102V	-0.603	Destabilizing
D12A	0.477	Stabilizing	Y103S	0.0289	Stabilizing
C14H	-1.314	Destabilizing	S104R	-0.658	Destabilizing
C14R	-0.441	Destabilizing	T114P	-0.567	Destabilizing
G17D	-0.722	Destabilizing	L116R	-1.64	Destabilizing
L19P	-1.451	Destabilizing	W119R	-2.323	Highly Destabilizing
L19R	-1.287	Destabilizing	R121P	-0.145	Destabilizing
V21G	-1.593	Destabilizing	V128G	-2.997	Highly Destabilizing
Y34S	-2.96	Highly Destabilizing	V130G	-2.863	Highly Destabilizing
L35P	-1.489	Destabilizing	G132D	-2.145	Highly Destabilizing
L35R	-1.404	Destabilizing	G132S	-1.735	Destabilizing
H43P	-0.593	Destabilizing	A134V	-0.758	Destabilizing
V45G	-2.956	Highly Destabilizing	H137R	-0.38	Destabilizing
A46V	0.375	Stabilizing	C138Y	-0.274	Destabilizing
T47A	-1.799	Destabilizing	V139A	-1.704	Destabilizing
D49A	-1.022	Destabilizing	V139G	-2.028	Highly Destabilizing
D49G	-1.372	Destabilizing	V139L	-0.867	Destabilizing
H51P	-1.28	Destabilizing	V139M	-0.979	Destabilizing
H51R	-1.688	Destabilizing	R140S	-1.554	Destabilizing
H51Y	0.207	Stabilizing	Q141P	1.0	Stabilizing
D53A	-0.428	Destabilizing	T142K	-0.896	Destabilizing
P54T	-1.795	Destabilizing	T142P	-0.504	Destabilizing
H57P	-0.809	Destabilizing	A146V	-0.557	Destabilizing
H57Y	0.462	Stabilizing	V155A	-2.586	Highly Destabilizing
T61P	-0.276	Destabilizing	V155G	-3.109	Highly Destabilizing
D63G	-1.079	Destabilizing	L159P	-1.018	Destabilizing
S66P	-0.393	Destabilizing	L159R	-0.978	Destabilizing
S67P	-0.14	Destabilizing	T160P	-0.498	Destabilizing
W68G	-3.169	Highly Destabilizing	A161P	-0.579	Destabilizing
W68L	-2.054	Highly Destabilizing	G162D	-1.099	Destabilizing
W68R	-1.953	Destabilizing	A171E	-2.59	Highly Destabilizing
H71R	-1.89	Destabilizing	A171P	-0.67	Destabilizing
C72R	-1.801	Destabilizing	A171V	-0.67	Destabilizing
T76P	-0.325	Destabilizing	L172P	-1.465	Destabilizing
G78D	-2.465	Highly Destabilizing	S185T	-0.362	Destabilizing

Table S 7: Classification of NAM and PZA stablity based on visual inspection in VMD

PZA-stable (21)	*NAM-stable (34)*	*PZA-unstable (33)*	*NAM-unstable (31)*	*PZA-released (29)*	*NAM-released (17)*
Q10P	A3P	A3E	A3E	A3P	V9A
C14R	V9G	V9G	V7G	V7G	Q10P
Y34S	D12A	C14H	D8G	D8G	T47A
A46V	C14H	G17D	C14R	V9A	D49G
H51Y	L19P	L19R	G17D	D12A	H51Y
D63G	L19R	V21G	Y34S	L19P	P54T
S66P	V21G	L35P	L35R	L35R	S66P
W68G	L35P	V45G	H43P	H43P	S67P
L85R	H51R	T47A	V45G	D49A	W68R
S104R	H57P	D53A	A46V	D49G	C72R
R121P	D63G	P54T	D49A	H51P	G97D
G132D	W68L	S67P	H51P	H51R	L116R
V128G	H71R	W68L	D53A	H57P	W119R
H137R	T76P	C72R	H57Y	H57Y	R121P
V139A*	G78D	T76P	T61P	T61P	V139G
V139G*	H82R	H82R	W68G	W68R	V155G
R140S	G97S	G97D	L85P	H71R	L159P
L159R	A102V	G97S	L85R	G78D	
A171E	T114P	A102V	K96N	L85P	
A171V	V130G	T114P	Y103S	K96N	
S185T	G132D	W119R	S104R	L116R	
	G132S	V130G	V128G	Y103S	
	V139A	C138Y	A134V	A134V	
	V139L	G132S	H137R	V139M	
	V139M	V139A*	C138Y	Q141P	
	R140S	V139G*	Q141P	A146V	
	A146V	T142K	T142K	L159P	
	V155A	T142P	T142P	T160P	
	L159R	V155A	T160P	A161P	
	A161P	V155G	G162D		
	A171E	G162D	L172P		
	A171P	A171P			
	A171V	L172P			
	S185T				

Table S 8: Omicron sub-lineage specific GISAID sequence IDs. Adapted from Barozi et al., **[271]**

Omicron sub-lineages	GISAID sequence IDs	
BA.1	EPI_ISL_11765237	EPI_ISL_8097234
	EPI_ISL_11765238	EPI_ISL_8097272
	EPI_ISL_11765240	EPI_ISL_8826405
	EPI_ISL_8826449	EPI_ISL_8826412
	EPI_ISL_9002773	EPI_ISL_8826429
	EPI_ISL_9002788	EPI_ISL_7834433
	EPI_ISL_9002795	EPI_ISL_7834437
	EPI_ISL_9002810	EPI_ISL_7834440
	EPI_ISL_9002815	EPI_ISL_7834441
	EPI_ISL_9002836	EPI_ISL_7834442
	EPI_ISL_7834473	EPI_ISL_7834443
	EPI_ISL_7834476	EPI_ISL_7834444
	EPI_ISL_7834514	EPI_ISL_7834446
	EPI_ISL_7834572	EPI_ISL_7834453
	EPI_ISL_7834591	EPI_ISL_7834457
	EPI_ISL_7566360	EPI_ISL_7552702
	EPI_ISL_7834419	EPI_ISL_7566330
	EPI_ISL_7834432	
BA.2	EPI_ISL_10981812	
	EPI_ISL_8636462	
	EPI_ISL_9149786	
	EPI_ISL_10893972	
	EPI_ISL_10967591	
BA.3	EPI_ISL_7834448	
	EPI_ISL_7852875	
	EPI_ISL_9149817	
	EPI_ISL_9313505	
BA.4	EPI_ISL_11674413	
	EPI_ISL_11674425	
	EPI_ISL_11674426	

	EPI_ISL_11674430
	EPI_ISL_11674431
	EPI_ISL_11674432
	EPI_ISL_11674443
	EPI_ISL_11674447
	EPI_ISL_1167444
	EPI_ISL_12119232
	EPI_ISL_11674410
	EPI_ISL_11674411

Table S 9: Omicron sub-lineage mutated residue physicochemical properties. Adapted from Barozi et al., **[271]**

Mutation	Physicochemical property change
G446S	Hydrophobic - Hydrophilic
S371F	Small/Polar - Hydrophobic
G339D	Non-polar - Polar
Q498R	Polar/uncharged – Polar/+ve charge
F486V	Hydrophobic - Hydrophobic
Y505H	Aromatic - Aromatic
T376A	Hydrophobic - Hydrophobic
S371L	Small - Aliphatic
Q493R	Polar/uncharged - Polar/+ve charge
D405N	Small/charged -Small
N501Y	Small – Aromatic
R408S	+ve charge - Polar
S477N	Polar/tiny - Polar
G496S	Small - Small/Polar
S373P	Small/Polar - Small
T478K	Polar – Polar/+ve charge
E484A	Polar - Nonpolar
S375F	Polar/small - Hydrophobic/Aromatic
L452R	Hydrophobic/aliphatic - Polar /+ve charge
K417N	Polar - Polar
N440K	Polar - Polar

Table S 10: *EC* hub residue flexibility. Adapted from Barozi et al., **[271]**.

	RMSF (nm)						
Residue	**WT**	**BA.1**	**BA.2**	**BA.3_10**	**BA.3_12**	**BA.3_15**	**BA.4**
Gly326	0.192	0.215	0.218	0.141	0.255	0.206	0.184
Asp355	0.096	0.113	0.169	0.097	0.221	0.153	0.127
Phe356	0.096	0.121	0.176	0.099	0.227	0.164	0.134
Arg357	0.103	0.136	0.196	0.117	0.236	0.179	0.147
Glu375	0.107	0.147	0.194	0.140	0.254	0.162	0.145
His378	0.092	0.120	0.154	0.111	0.154	0.153	0.123
Ile379	0.085	0.112	0.145	0.123	0.186	0.192	0.122

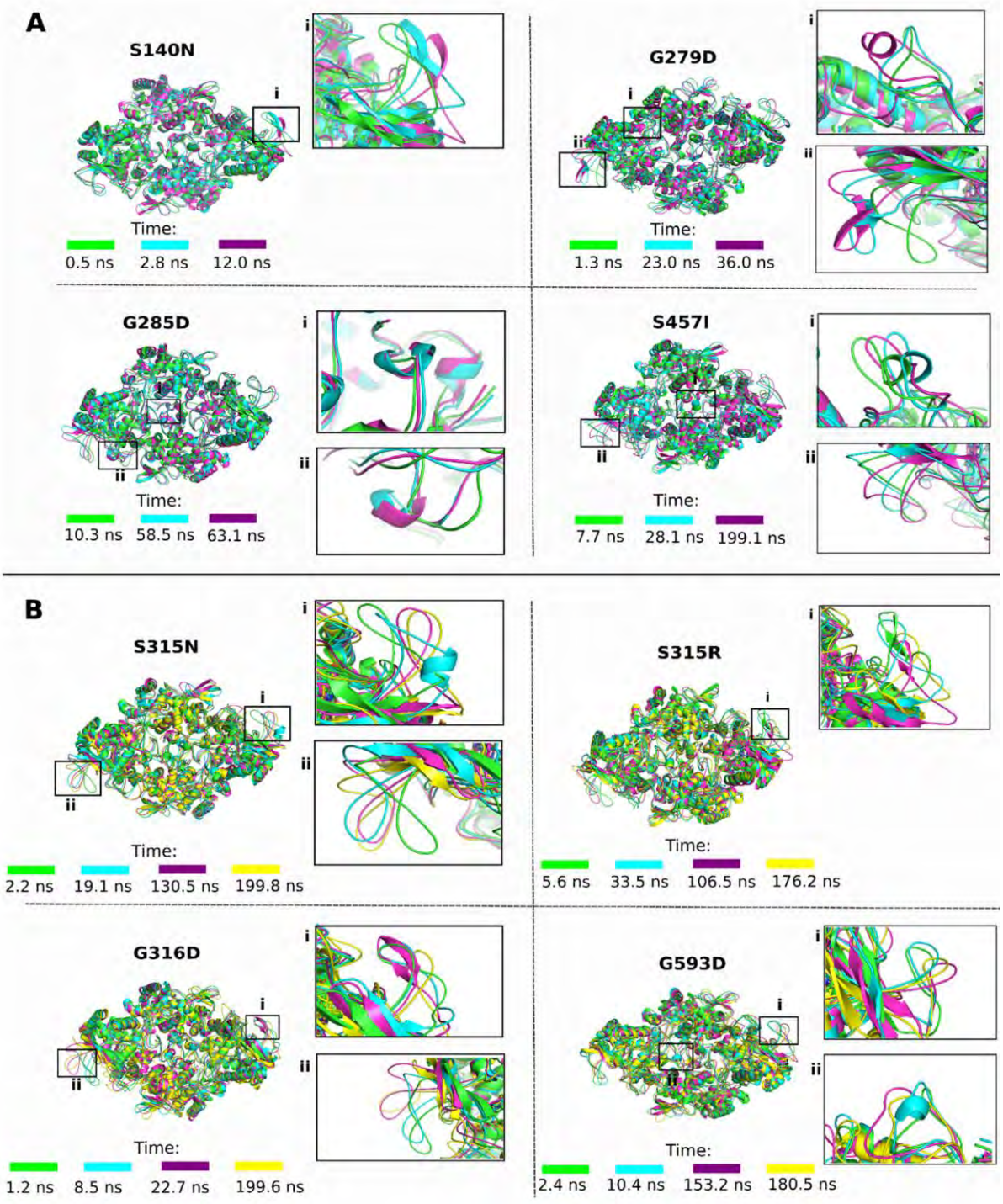

Figure S 1: Overlayed structures from the most sampled clustering in each system highlighting the structural differences (black boxes) between them. The structure colors correspond to the time (ns) when sampled during the 300 ns simulation. The figure is produced with permission from Barozi et al., **[115]**.

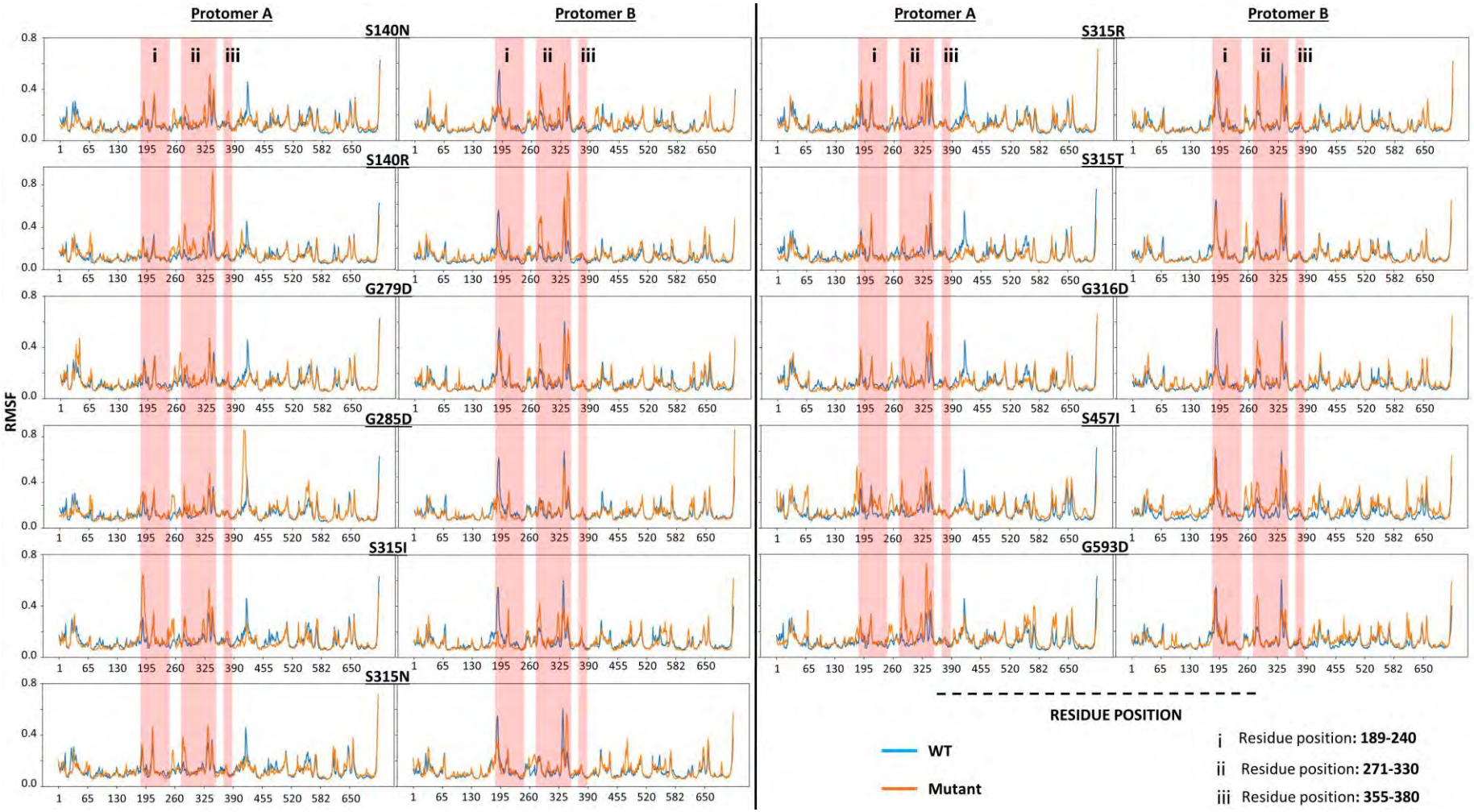

Figure S 2: Superimposed RMSF line plots of the WT (blue) and mutants (orange) for protomers A and B. The x and y-axes show the residue number and RMSF (nm), respectively. The figure is produced with permission from Barozi et al., **[115]**.

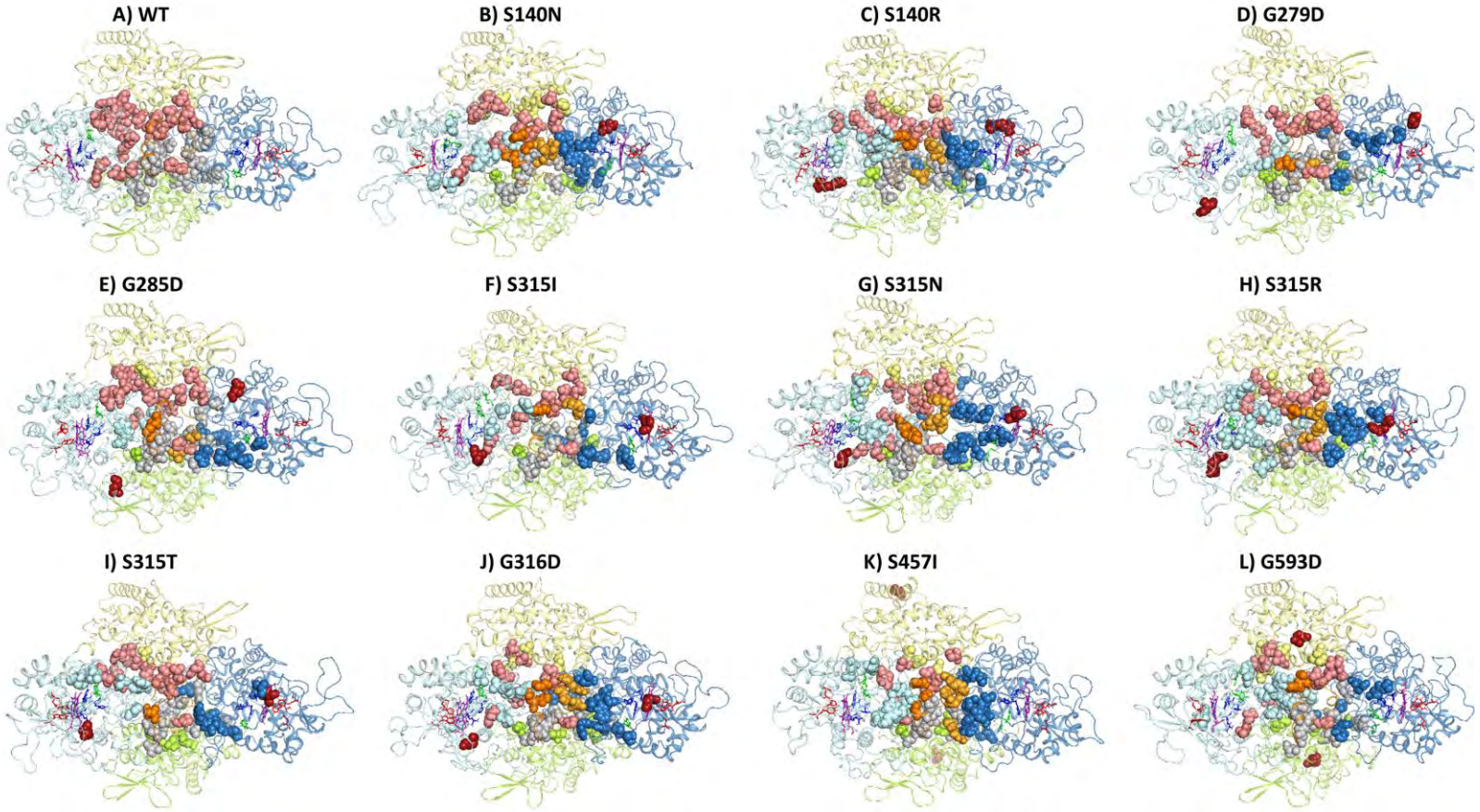

Figure S 3: WT (A) and mutant (B-L) structures showing the distribution of the global top 4% *BC* hubs. N and C-terminal domains of protomer A are shown as pale cyan and yellow, respectively whereas those of protomer B are sky-blue and lime. Dimerization domain is orange and bright yellow for protomer A and B, respectively. WT and mutant common hubs are shown are grey spheres (protomer A) and salmon spheres (protomer B). Mutant unique hubs are colored as the structure domains. Firebrick spheres are mutation positions. The figure is produced with permission from Barozi et al., **[115]**.

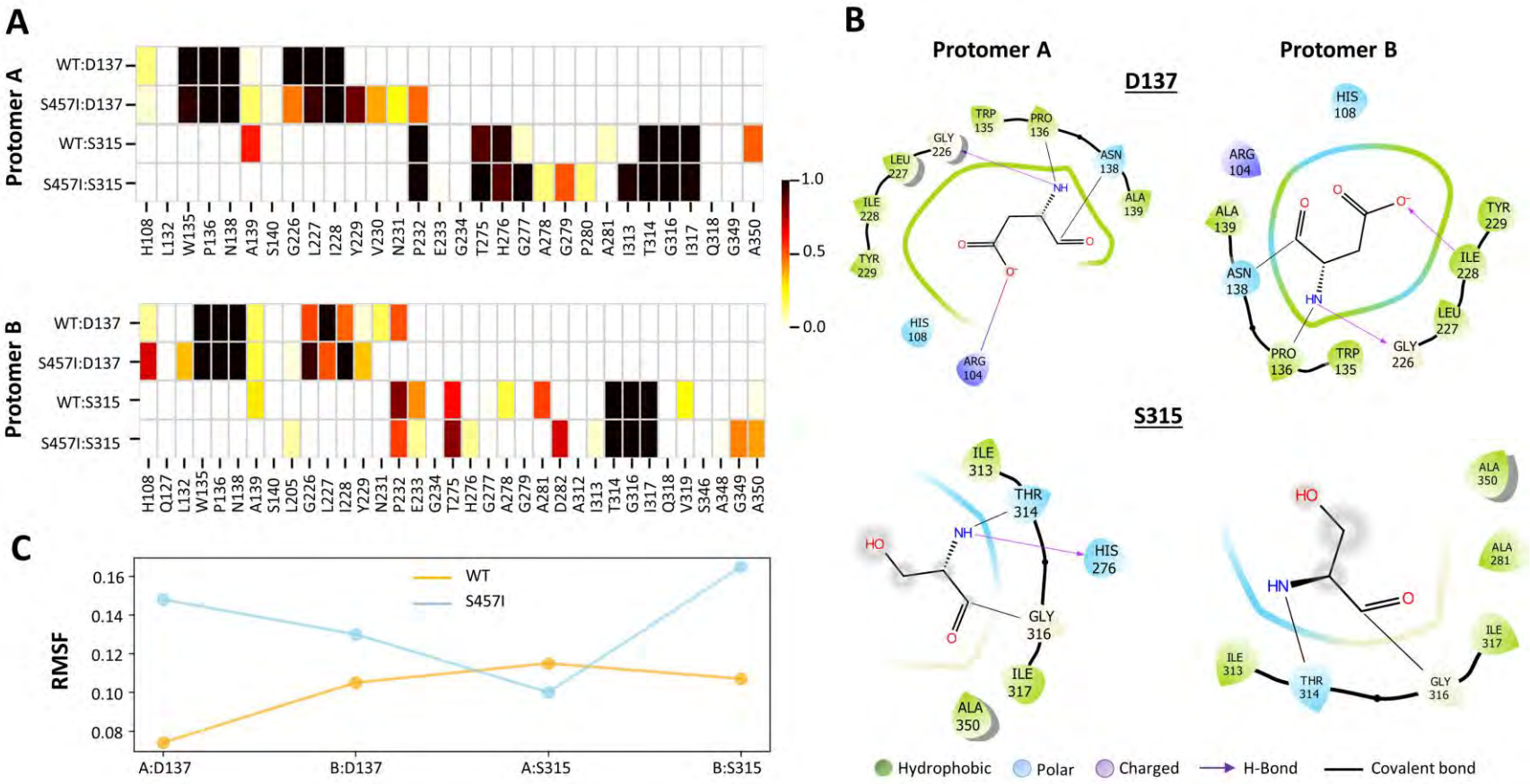

Figure S 4: A shows the heatmap representation of residues Asp137 and Ser315 in protomers A and B, respectively in the WT and S457I. The color scale from yellow to dark red shows the range in residue centrality values. B shows the S457I residue interaction types in in both protomers. C shows residue Asp137 and Ser315 RMSF in the WT (blue) and S457I (orange). The figure is produced with permission from Tastan Bishop et al., **[86]**.

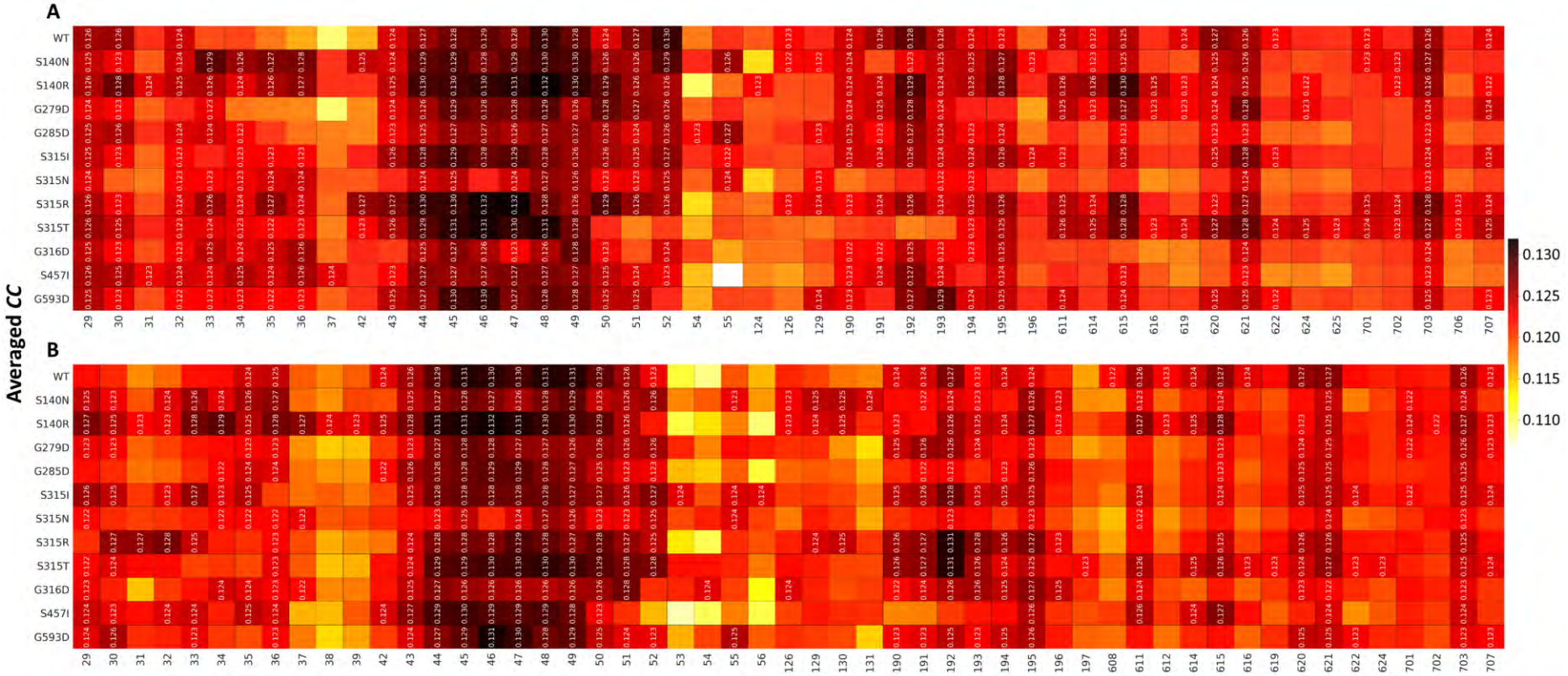

Figure S 5: A and B are heatmap representations of the high centrality valued *CC* residues (hubs) for protomers A and B, respectively as determined by the global 4% cut-off, The hubs per system are annotated with centrality values whereas the homologous residues from the other systems are not. The color scale from yellow to dark red shows the degree of variation of *CC* values across the residues. The figure is produced with permission from Barozi et al., **[115]**.

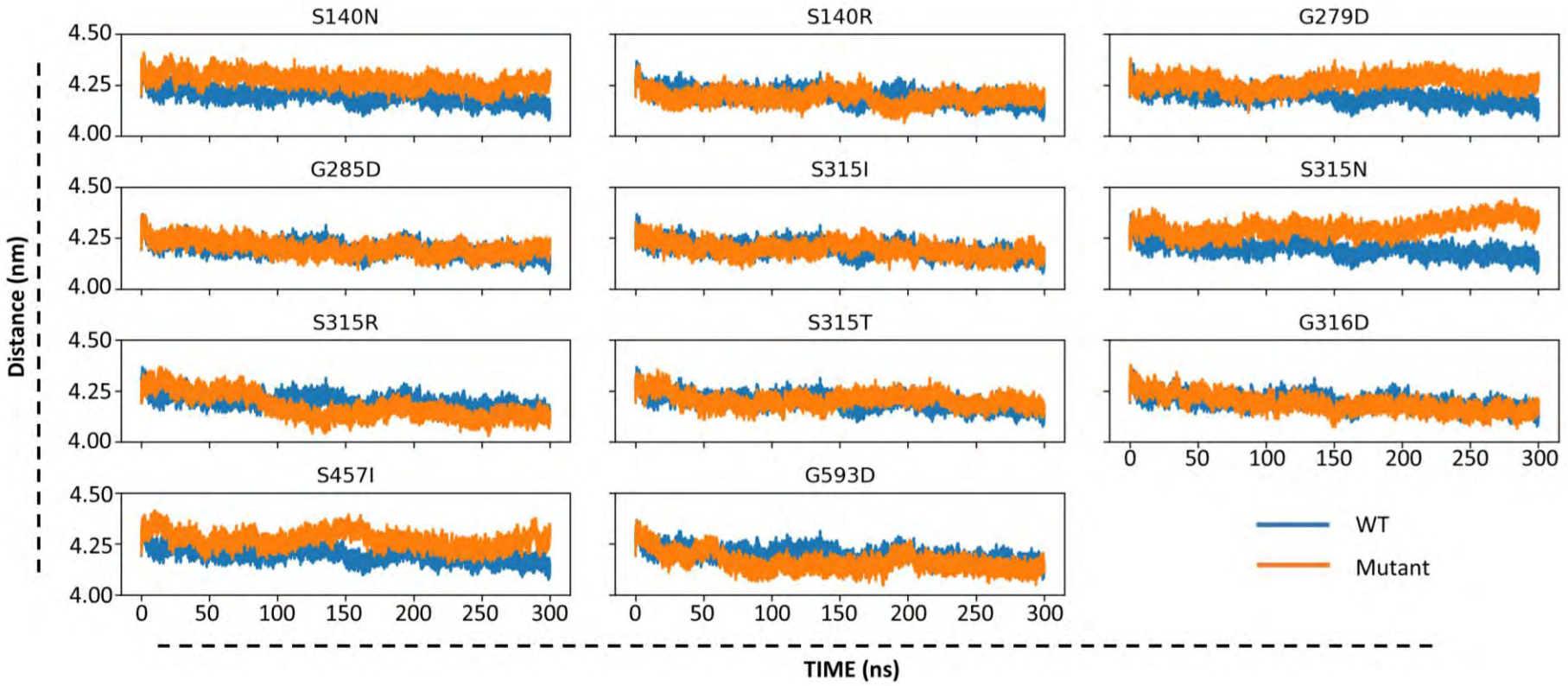

Figure S 6: Comparative line plots of the inter-protomer distance between the WT (blue) and mutant systems (orange) over the 300ns simulation. The x and y-axes show the time (ns) and COM distance (nm). The figure is produced with permission from Barozi et al., **[115]**.

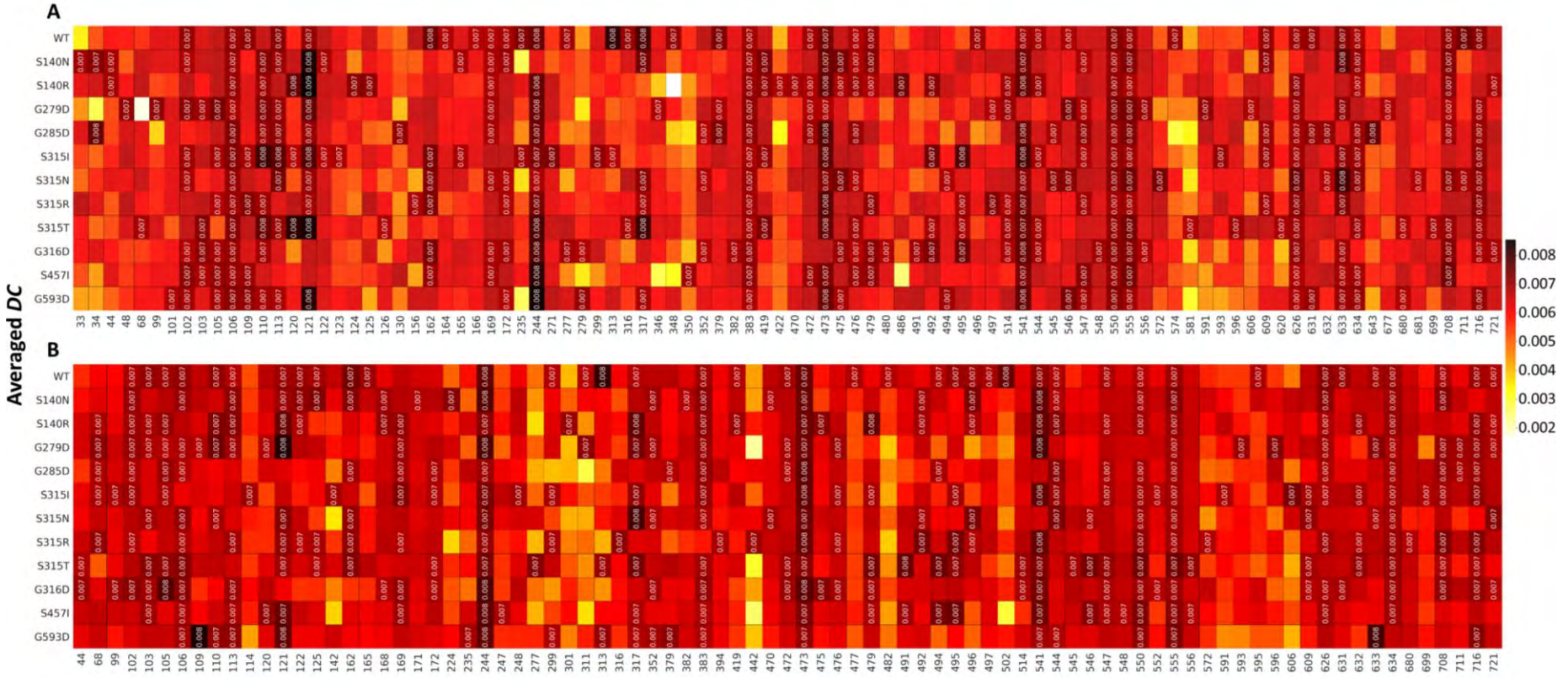

Figure S 7: A and B are heatmap representations of the high centrality valued *DC* residues (hubs) for protomers A and B, respectively as determined by the global 4% cut-off, The hubs per system are annotated with centrality values whereas the homologous residues from the other systems are not. The color scale from yellow to dark red shows the degree of variation of *DC* values across the residues. The figure is produced with permission from Barozi et al., **[115]**.

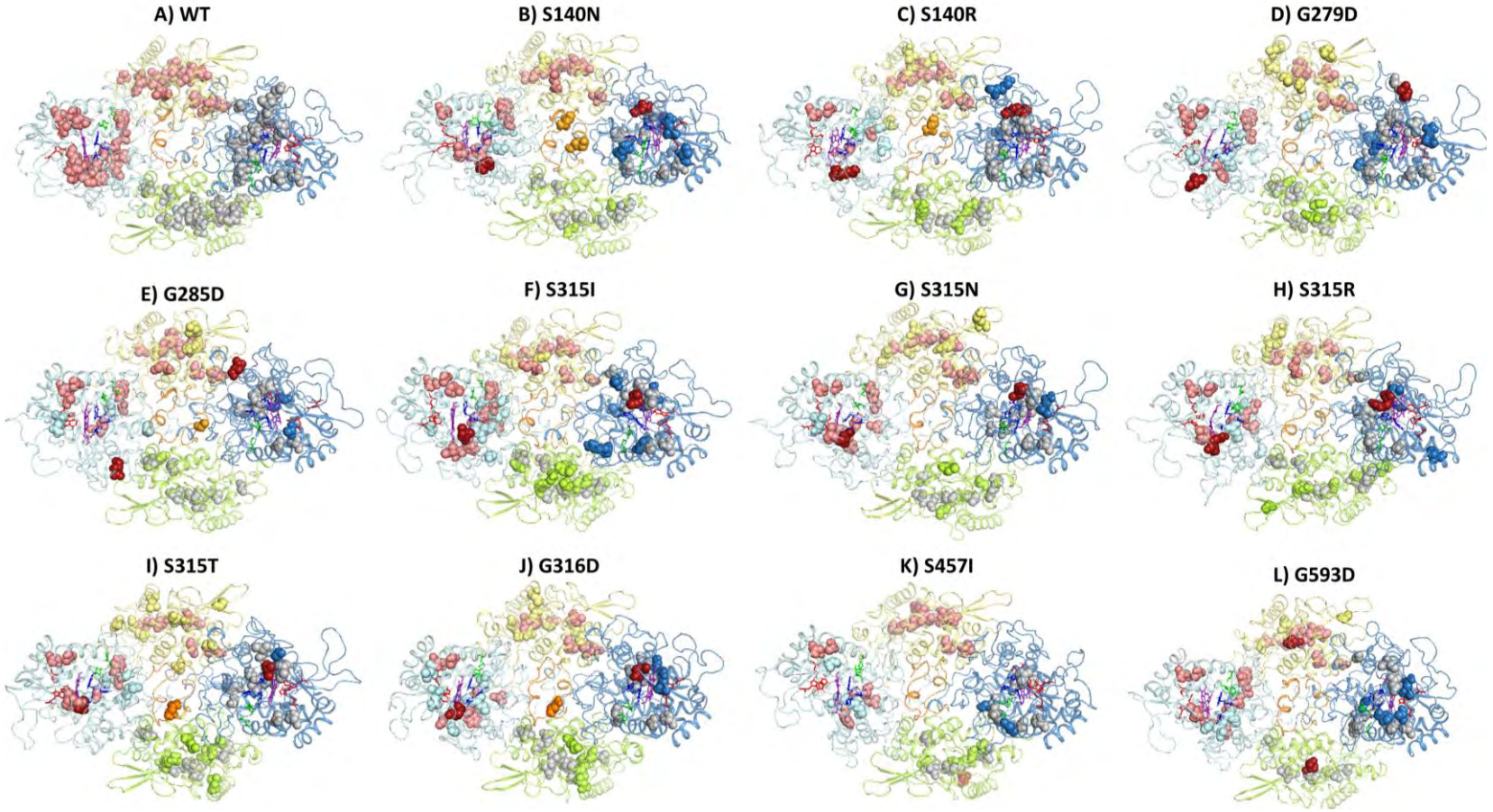

Figure S 8: Cartoon presentation of the WT (A) and mutant (B-L) structures showing the distribution of the global top 4% *DC* hubs. N and C-terminal domains of protomer A are shown as pale cyan and yellow, respectively whereas those of protomer B are sky-blue and lime. Dimerization domain is orange and bright yellow for protomer A and B, respectively. WT and mutant common hubs are shown are grey spheres (protomer A) and salmon spheres (protomer B). Mutant unique hubs are colored as the structure domains. Firebrick spheres are mutation positions. The figure is produced with permission from Barozi et al., **[115]**.

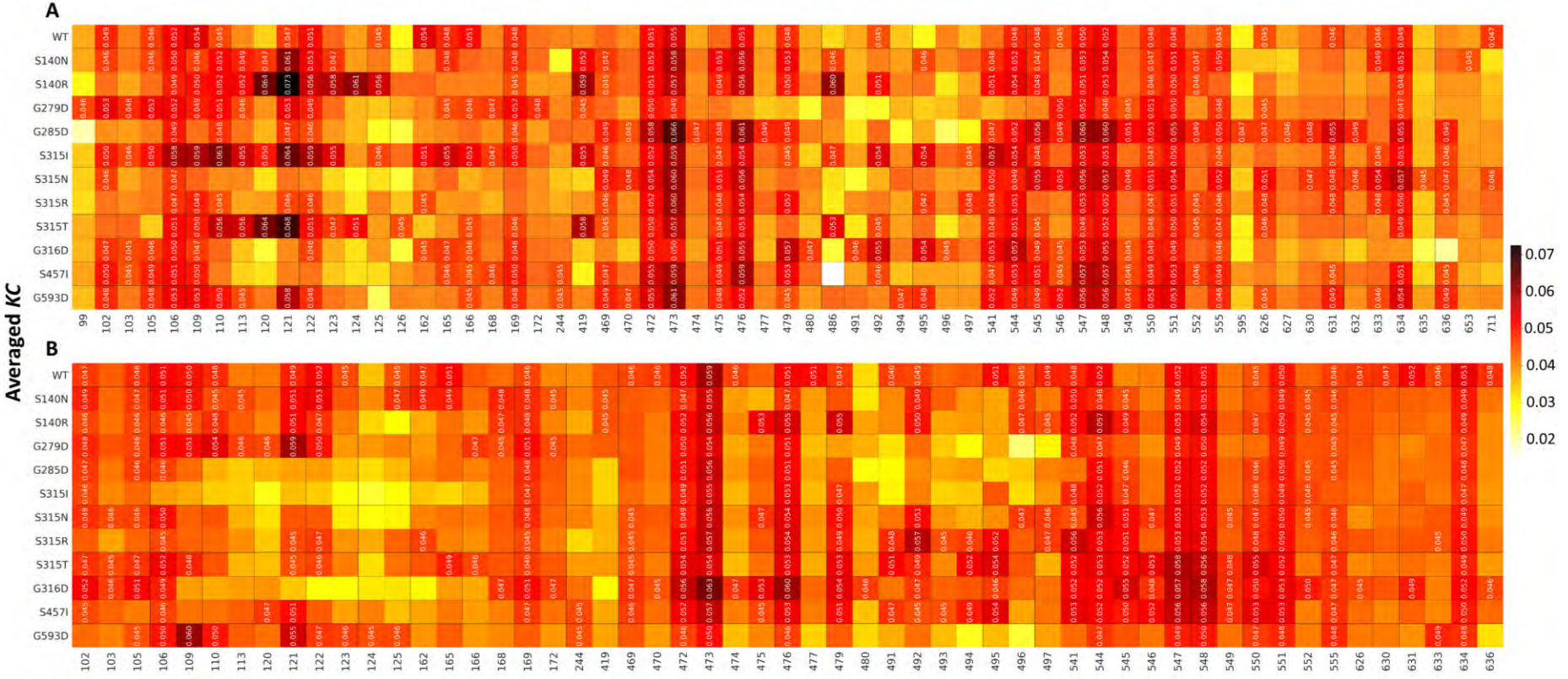

Figure S 9: A and B are heatmap representations of the high centrality valued *KC* residues (hubs) for protomers A and B, respectively as determined by the global 4% cut-off, The hubs per system are annotated with centrality values whereas the homologous residues from the other systems are not. The color scale from yellow to dark red shows the degree of variation of *KC* values across the residues. The figure is produced with permission from Barozi et al., **[115]**.

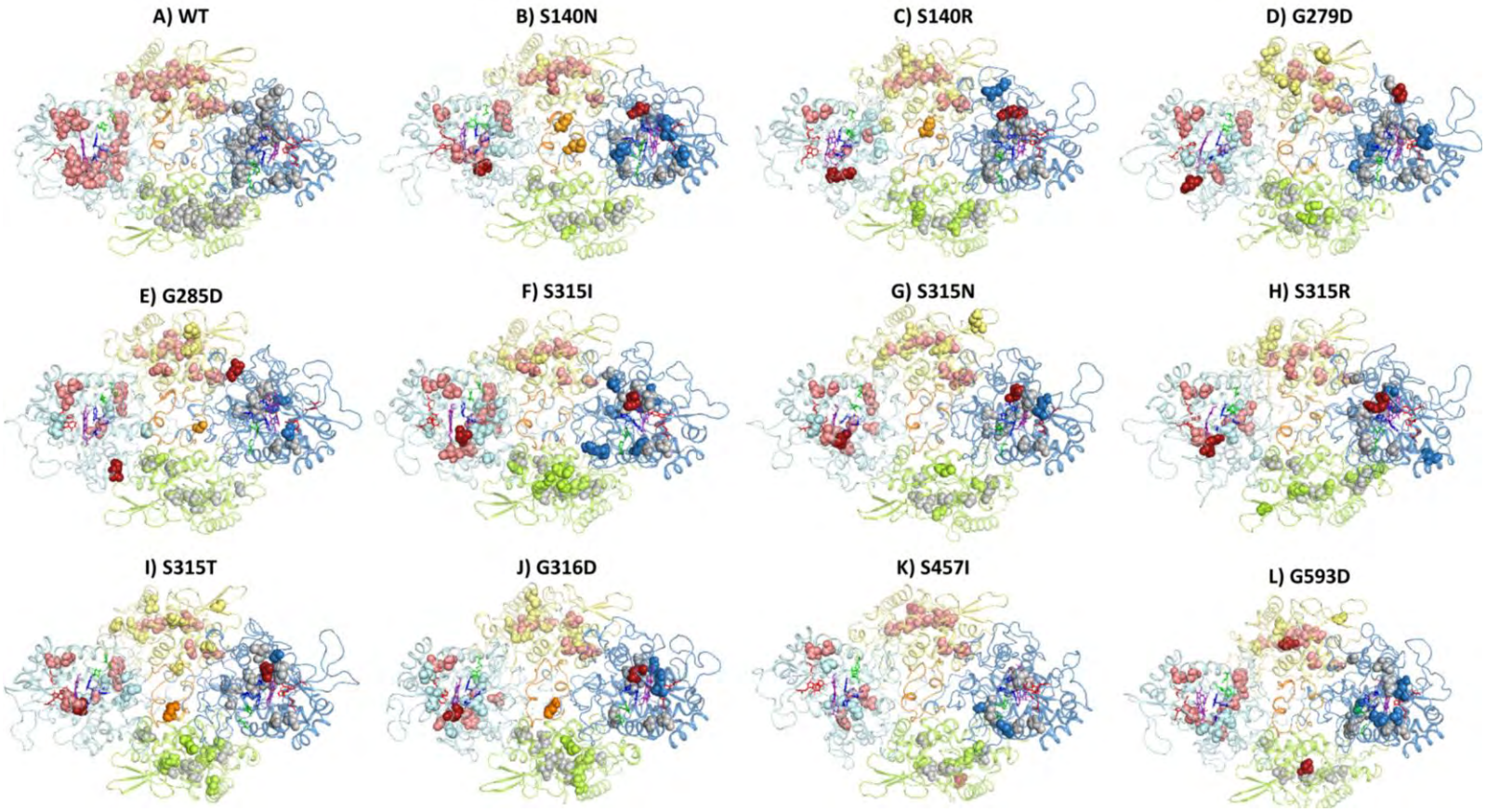

Figure S 10: Cartoon presentation of the WT (A) and mutant (B-L) structures showing the distribution of the global top 4% *KC* hubs. N and C-terminal domains of protomer A are shown as pale cyan and yellow, respectively whereas those of protomer B are sky-blue and lime. Dimerization domain is orange and bright yellow for protomer A and B, respectively. WT and mutant common hubs are shown are grey spheres (protomer A) and salmon spheres (protomer B). Mutant unique hubs are colored as the structure domains. Firebrick spheres are mutation positions. The figure is produced with permission from Barozi et al., **[115]**.

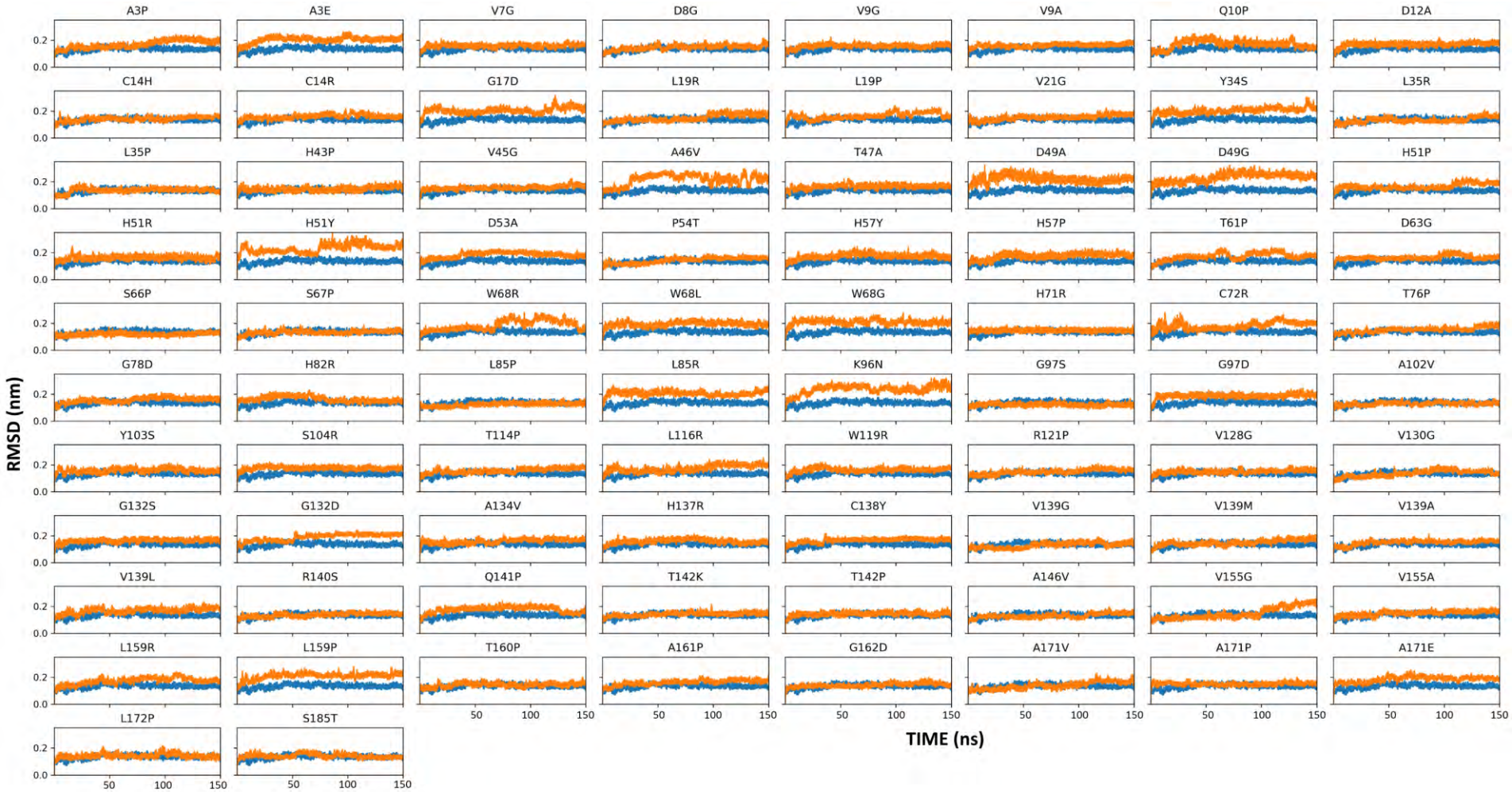

Figure S 11: Line plot presentation of the MtPncA-NAM RMSD progression for the WT (blue) superimposed with each mutant (orange), for the 150 ns simulation time. The x-axis shows time in ns whereas the y-axis shows RMSD in nm.

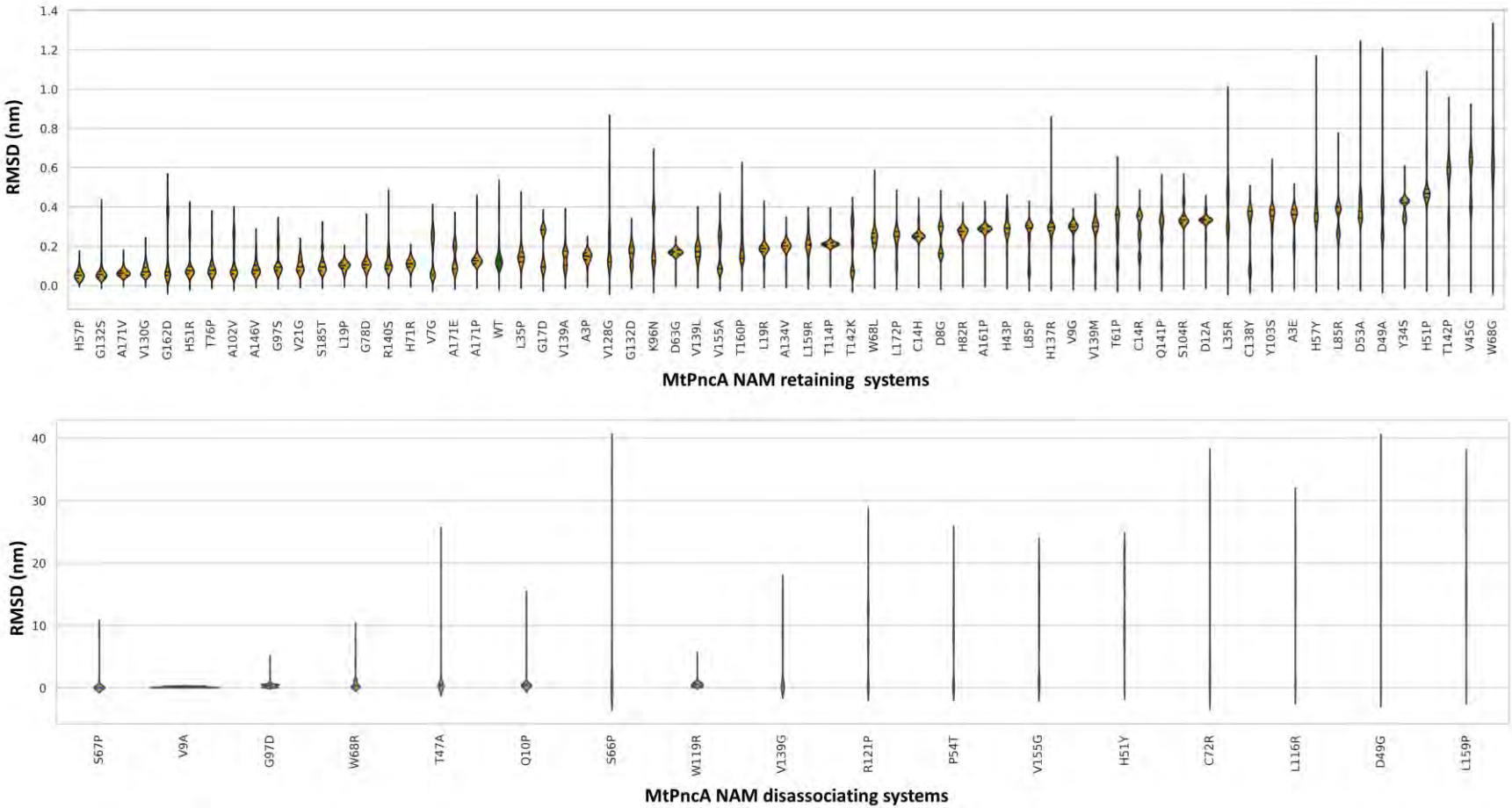

Figure S 12: RMSD violin plot representation of the MtPncA NAM retaining and disassociating systems. The WT is shown in green and the mutant systems in orange. The violin plots are arranged in ascending order of the median RMSD. The x-axis shows the protein systems whereas the y -axis shows the RMSD in nm.

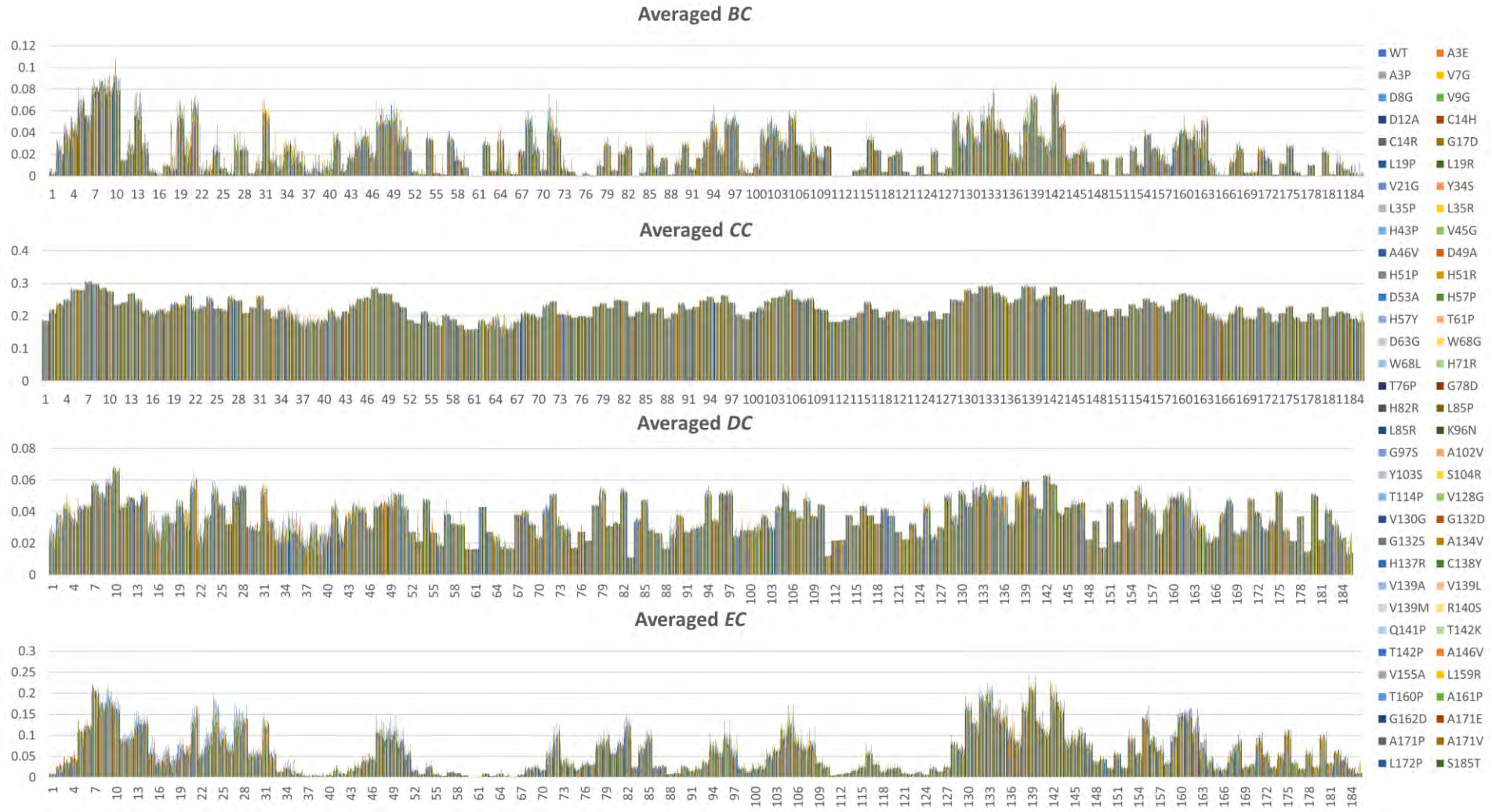

Figure S 13: Histograms showing the distribution of raw centrality values per metric for the WT and all the NAM retaining mutant systems. The x and y-axes show the residue numbers and centrality values, respectively. Each histogram is color coded to a specific system.

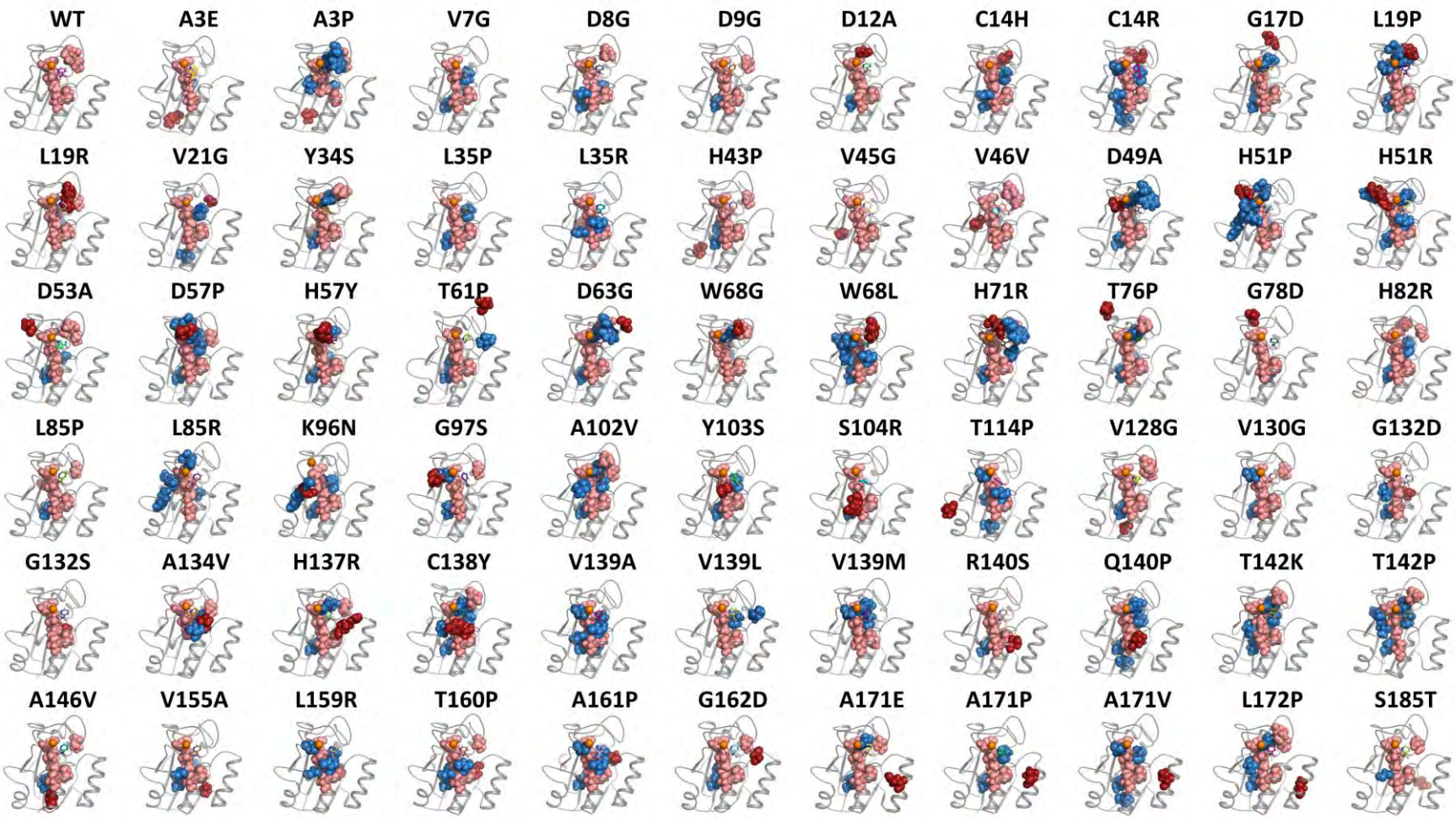

Figure S 14: Cartoon representation of the MtPncA showing the distribution of the global top 5% *BC* hubs in the WT and mutant systems. NAM is shown as sticks while the mutation positions are shown as firebrick spheres. WT hubs and mutant hubs common to the WT are shown as salmon spheres whereas *BC* hubs unique to the mutants (Δ hubs: mutant – WT hubs) are shown as sky-blue spheres.

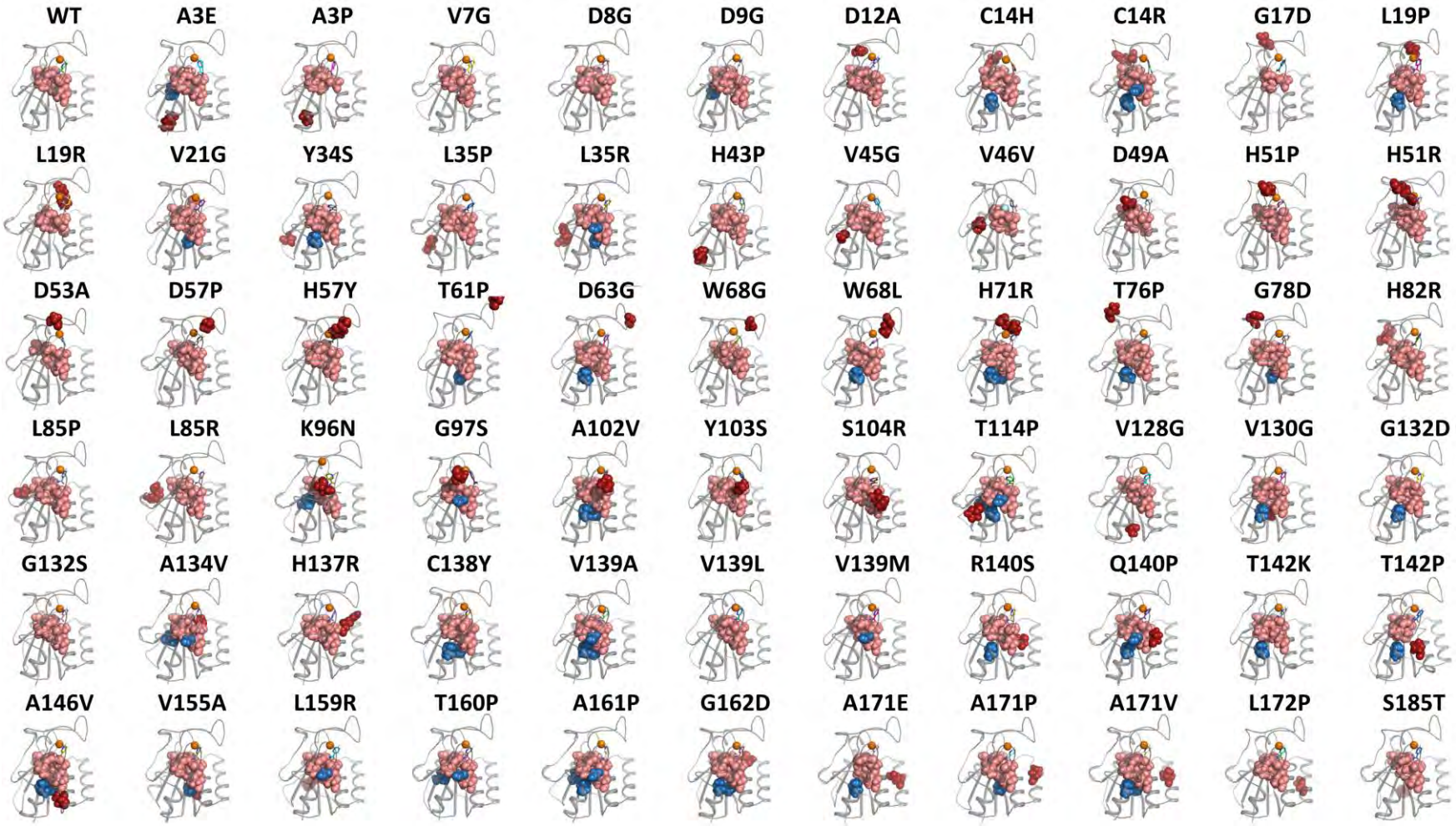

Figure S 15: Cartoon representation of the the global top 5% *CC* hubs in the WT and mutant systems. NAM is shown as sticks while the mutation positions are shown as firebrick spheres. WT hubs and mutant hubs common to the WT are shown as salmon spheres whereas *CC* hubs unique to the mutants (Δ hubs: mutant – WT hubs) are shown as sky-blue spheres.

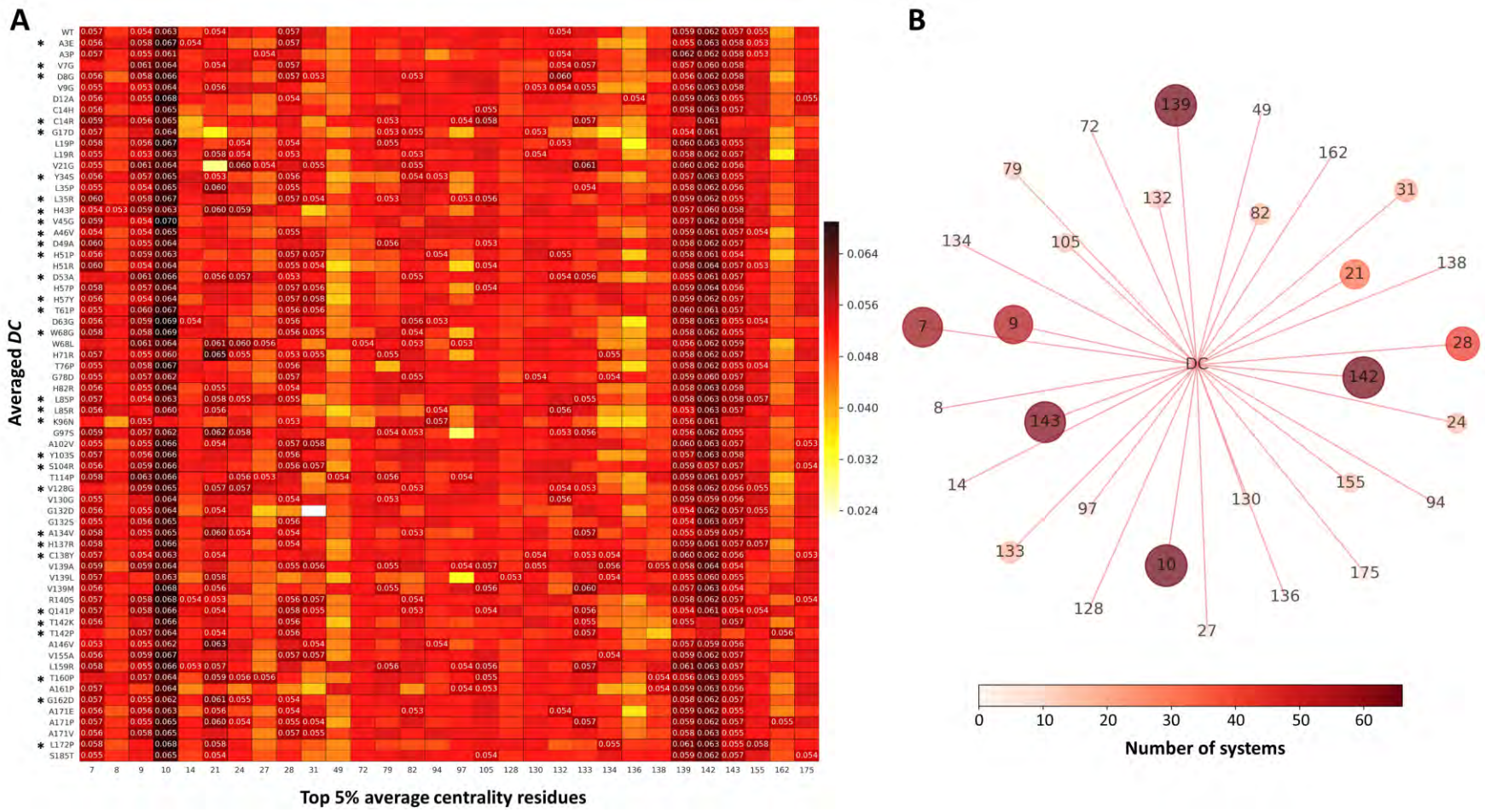

Figure S 16: A is a heat map of the global top 5% *DC* hubs for the WT and NAM retaining mutant systems. System hubs are annotated with their centrality values, whereas homologous residues in alternate systems are not. Color scale from white to dark red shows range of centrality values. Hubs that unstably coordinated NAM have an Asterix. B shows a spring network plot showing the regularity of occurrence of *DC* hub residues across the systems. Larger and dark red nodes correspond to more frequent hubs while small and white nodes to less frequent hubs.

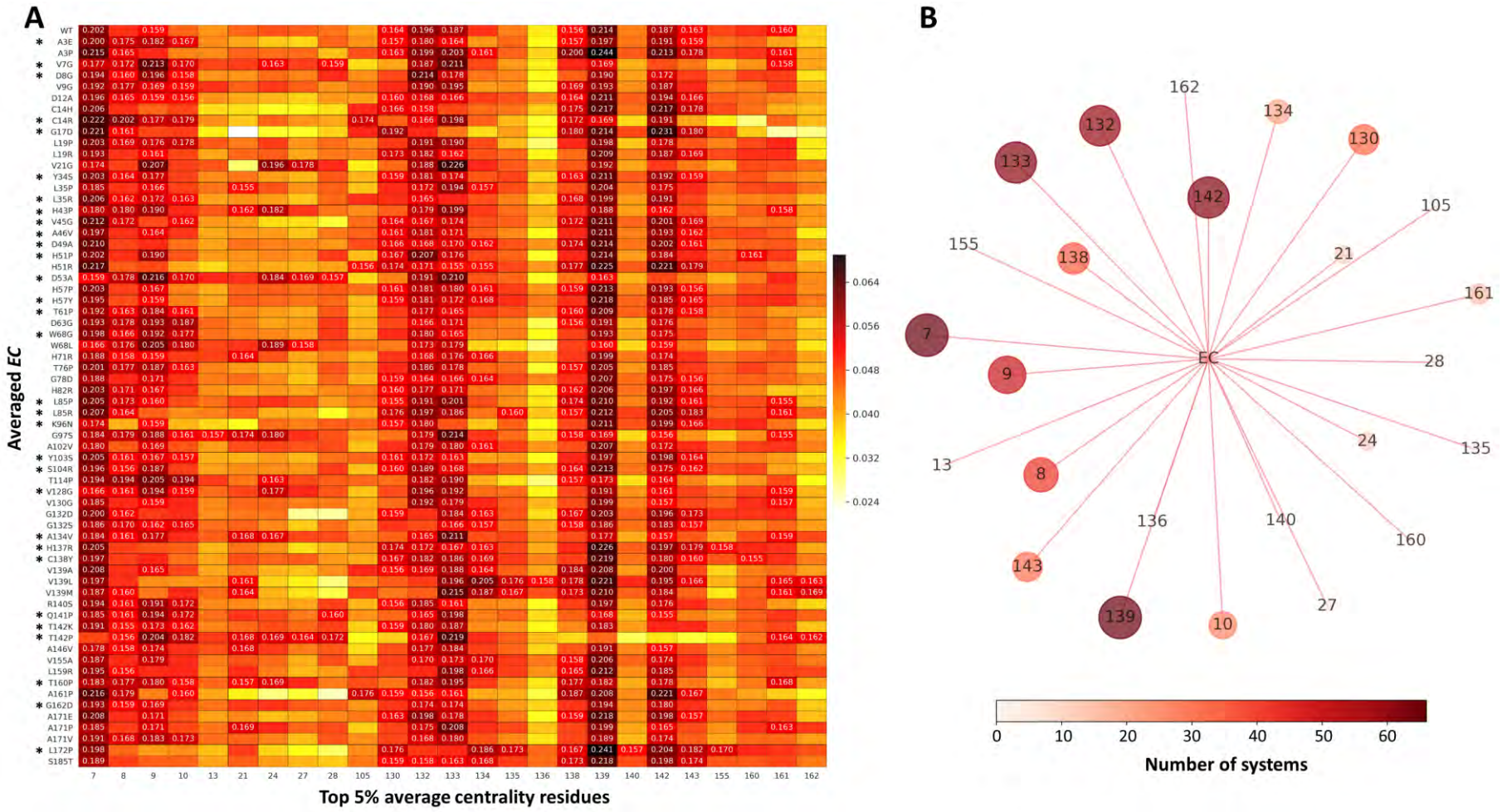

Figure S 17: A is a heat map of the global top 5% *EC* hubs for the WT and NAM retaining mutant systems. Systems hubs are annotated with their centrality values, whereas homologous residues in alternate systems are not. Color scale from white to dark red shows range of centrality values. Hubs that unstably coordinated NAM have an Asterix. B) Spring network plot showing the regularity of occurrence of *EC* hub residues across the systems. Larger and dark red nodes correspond to more frequent hubs while small and white nodes to less frequent hubs.

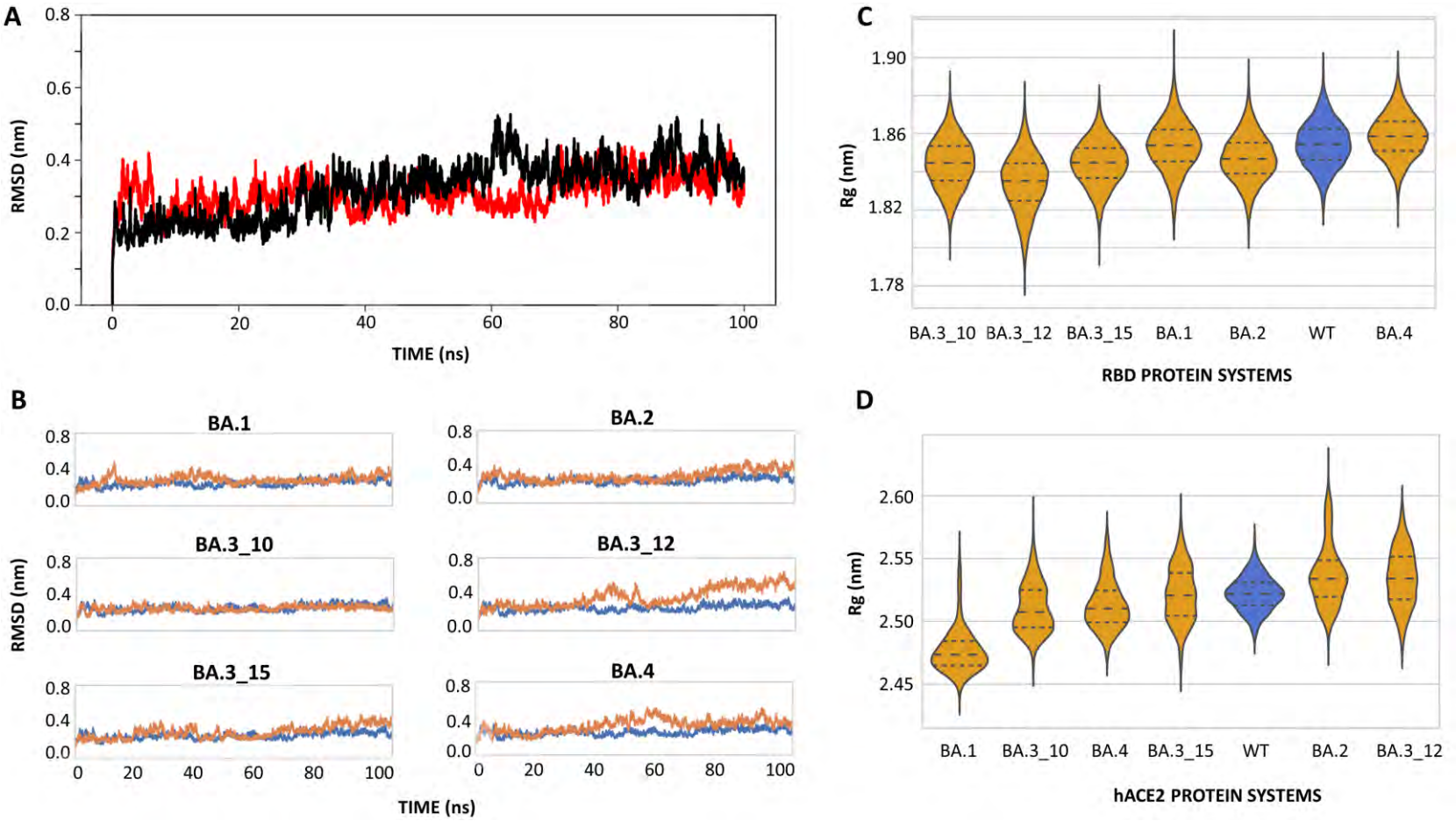

Figure S 18: A shows the duplicate line plots of the WT RBD-hACE2 complex. B Shows the comparative RMSD line plots of the WT (blue) and Omicron sub-lineage (orange) RBD-hACE2 systems. C and D are violin lots of the Rg for the RBD and hACE2 proteins, respectively. The WT is shown in blue and Omicron sub-lineages in orange. Adapted from Barozi et al., **[271]**.

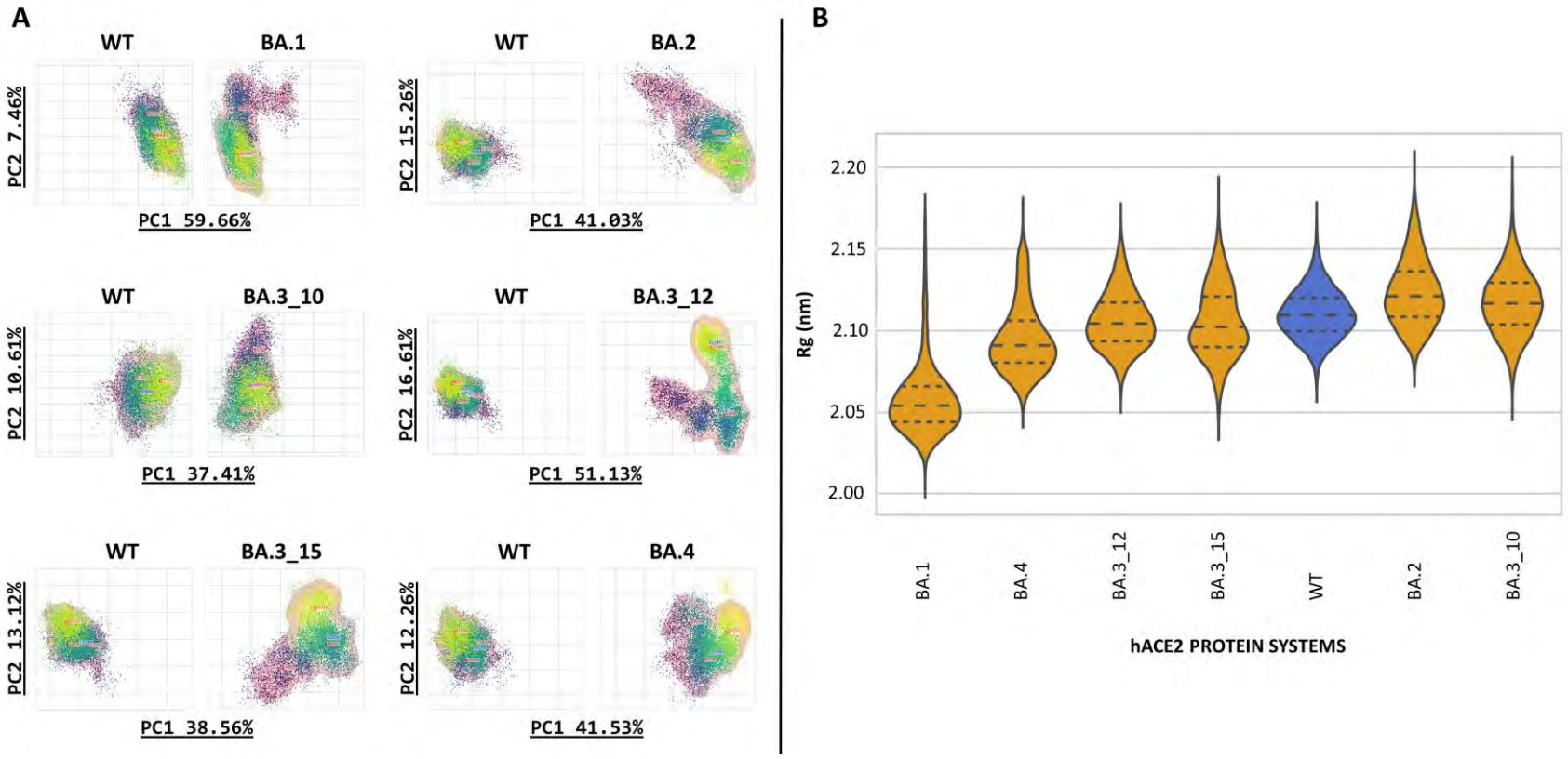

Figure S 19: A shows the ED scatter plots along PC1 (x-axis) and PC2 (y-axis) of hACE2 active pocket groove region for the Omicron sub-lineages in the same eigen space as the WT. B is a violin plot representation of the hACE2 active pocket groove region. The WT is shown in blue and the Omicron sub-lineages in orange. Adapted from Barozi et al., **[271]**.

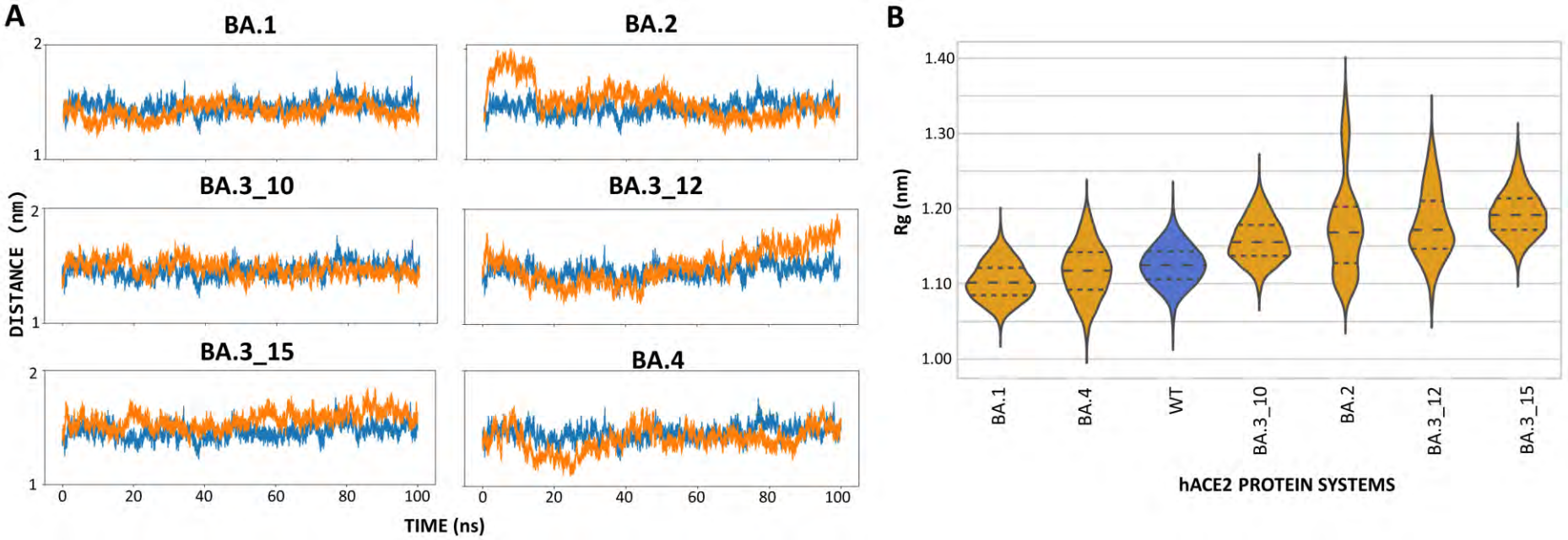

Figure S 20: A Shows the COM distance between hACE2 sub-domain I and sub-domain II for the WT (blue) and Omicron sub-lineages (orange). B is the Rg violin plot representation for the WT (blue) and Omicron sub-lineages (orange). Adapted from Barozi et al., **[271]**.

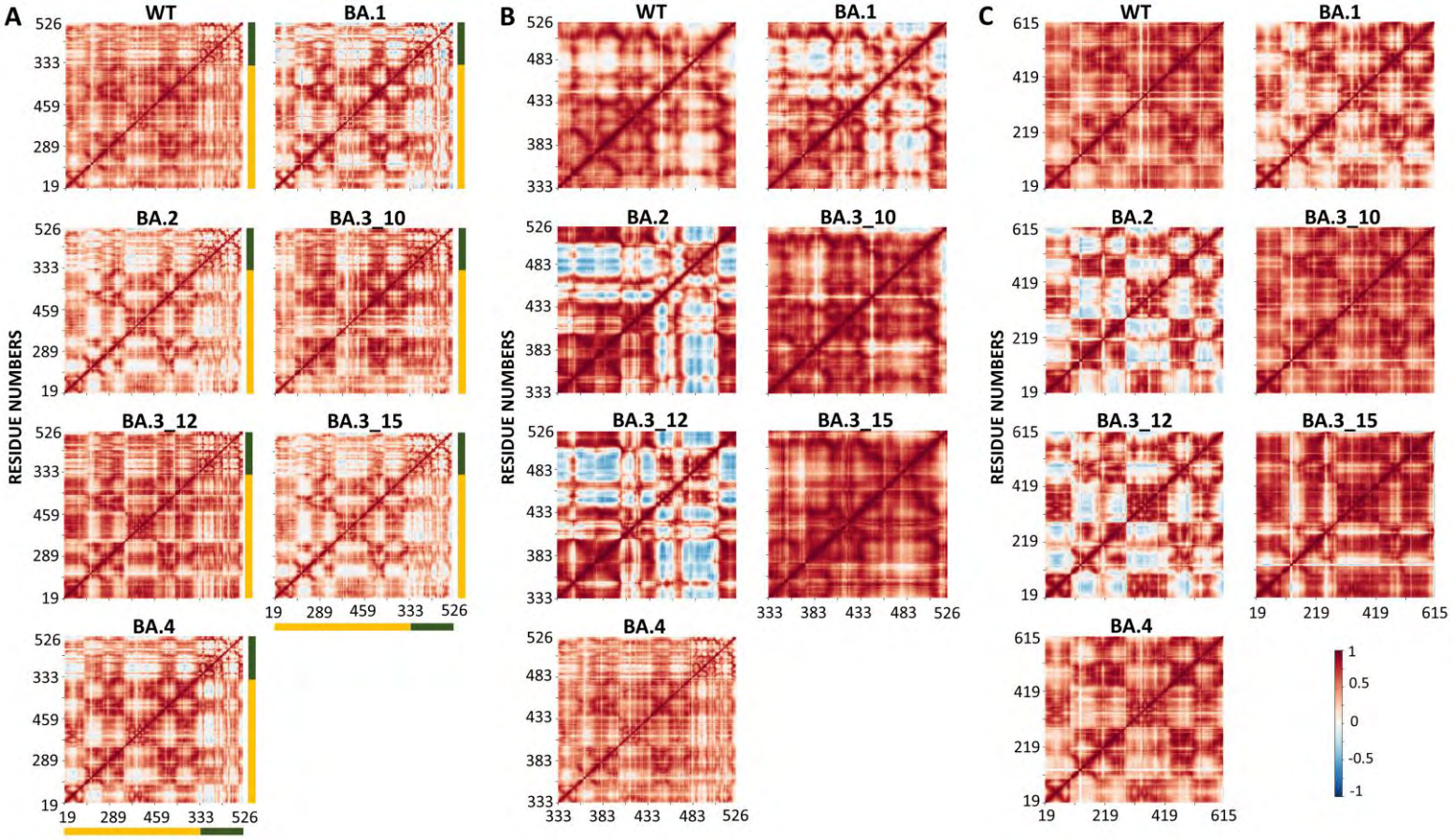

Figure S 21: A, B and C show the heatmap of degree of DCC in the RBD-hACE2 complex, RBD and hACE2 proteins, respectively in the WT and Omicron sub-lineages. The scale from -1 through 0 to 1 signifies the degree of correlation with -1 implying complete anti-correlation, 0 implying no correlation and 1 signifying complete correlation. In A, the orange and green region correspond to the hACE2 and RBD proteins, respectively. Adapted from Barozi et al., **[271]**.

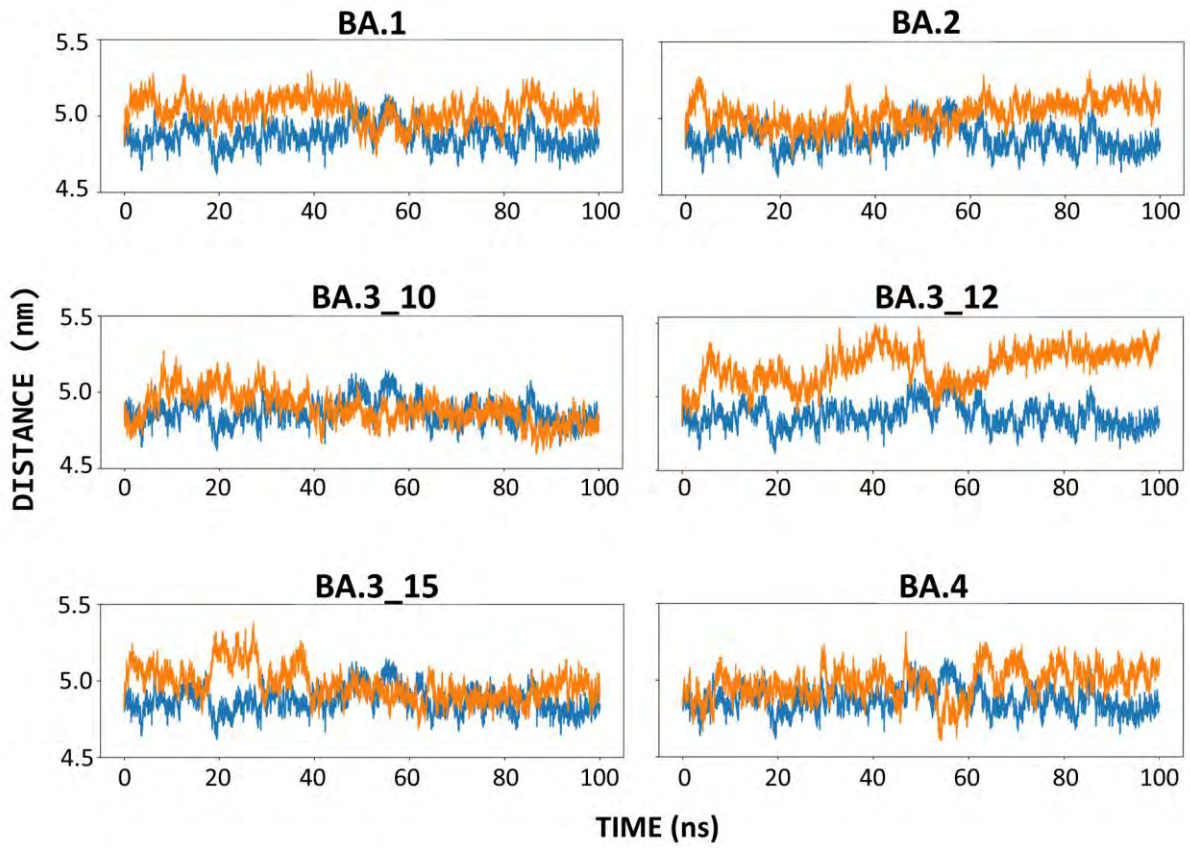

Figure S 22: Comparative line plots of the WT (blue) and Omicron sub-lineage (orange) COM distance between the RBD and hACE2 proteins over the simulation time. Adapted from Barozi et al., **[271]**.

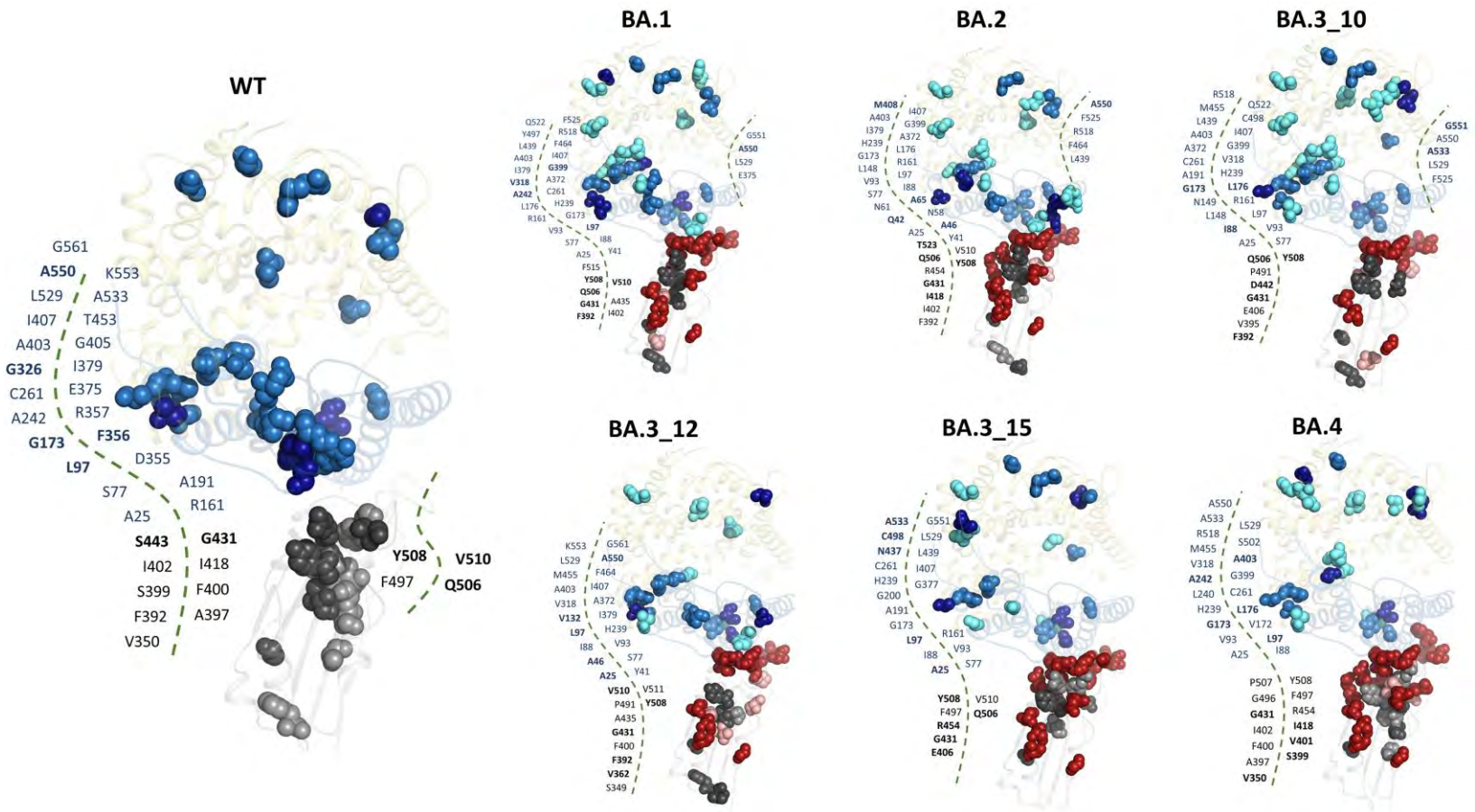

Figure S 23: Cartoon representation of the RBD-hACE2 complex showing the distribution of *DC* hubs in the RBD and hACE2 as identified by the global top 5 and 4%, respectively. The hACE2 sub-domains I and II are shown as sky blue and yellow, respectively while the RBD is grey. *DC* hubs common to the WT are shown grey (RBD) and blue (hACE2) spheres. *DC* hubs unique to the Omicron sub-lineages are shown as salmon (RBD) and teal (hACE2) spheres. The five highest centrality residues are in dark grey (RBD) and dark blue (hACE2) whereas, the mutation positions are shown as firebrick spheres. Adapted from Barozi et al., **[271]**.

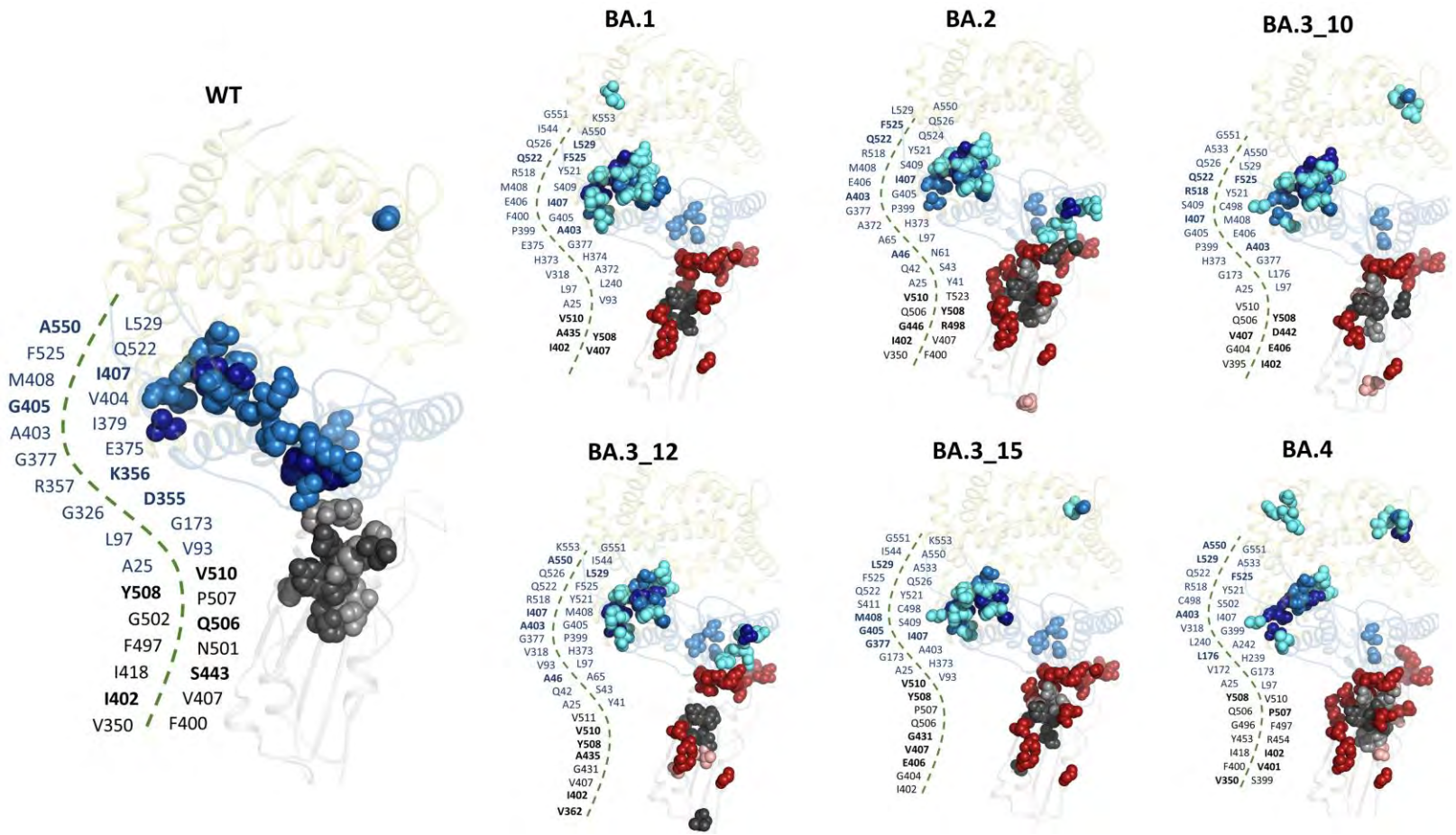

Figure S 24: Cartoon representation of the RBD-hACE2 complex showing the distribution of *KC* hubs in the RBD and hACE2 as identified by the global top 5 and 4%, respectively. The hACE2 sub-domains I and II are shown as sky blue and yellow, respectively while the RBD is grey. *KC* hubs common to the WT are shown grey (RBD) and blue (hACE2) spheres. *KC* hubs unique to the Omicron sub-lineages are shown as teal salmon (RBD) and teal (hACE2) spheres. The five highest centrality residues are in dark grey (RBD) and dark blue (hACE2) whereas, the mutation positions are firebrick spheres. Adapted from Barozi et al., **[271]**.

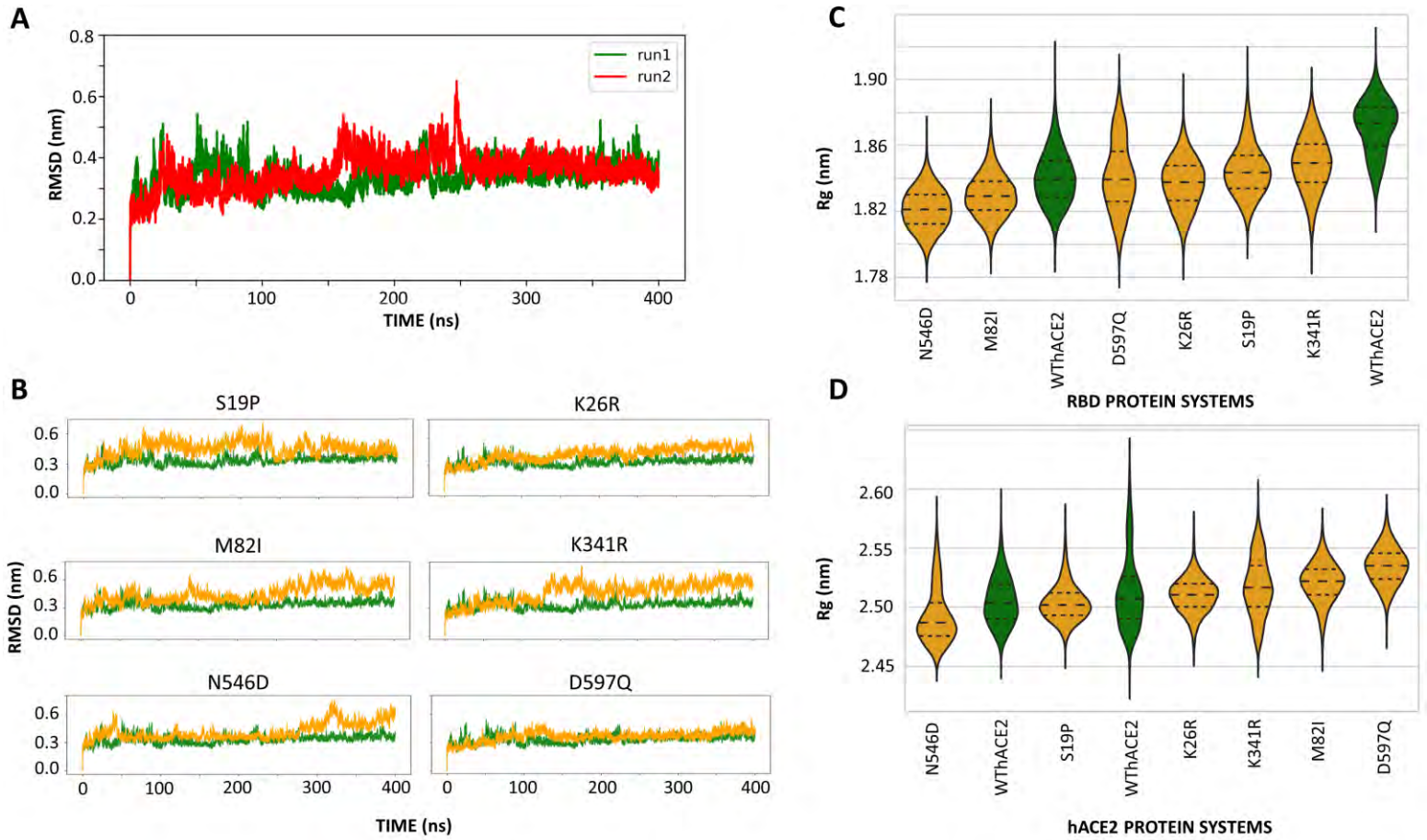

Figure S 25: A shows the WThACE2 RMSD line plots for the duplicate runs, whereas the line plots in B show the RMSD of the WThACE2 run 1 (green) in comparison to the hACE2 mutant systems (orange) over the 400 ns MD simulation. C and D and are Rg violin plot representations for the RBD and hACE2 systems, respectively.

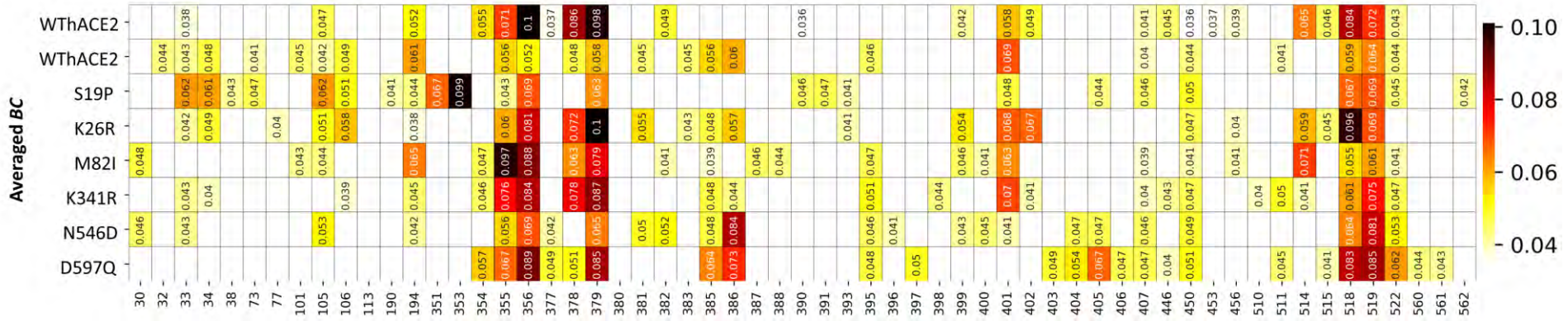

Figure S 26: The hACE2 top 4% high *BC* residues as calculated for each individual system shown as a heatmap. Residues numbers are on the x-axis and the protein systems on the y-axis. The color code from white to dark red shows the degree of centrality.

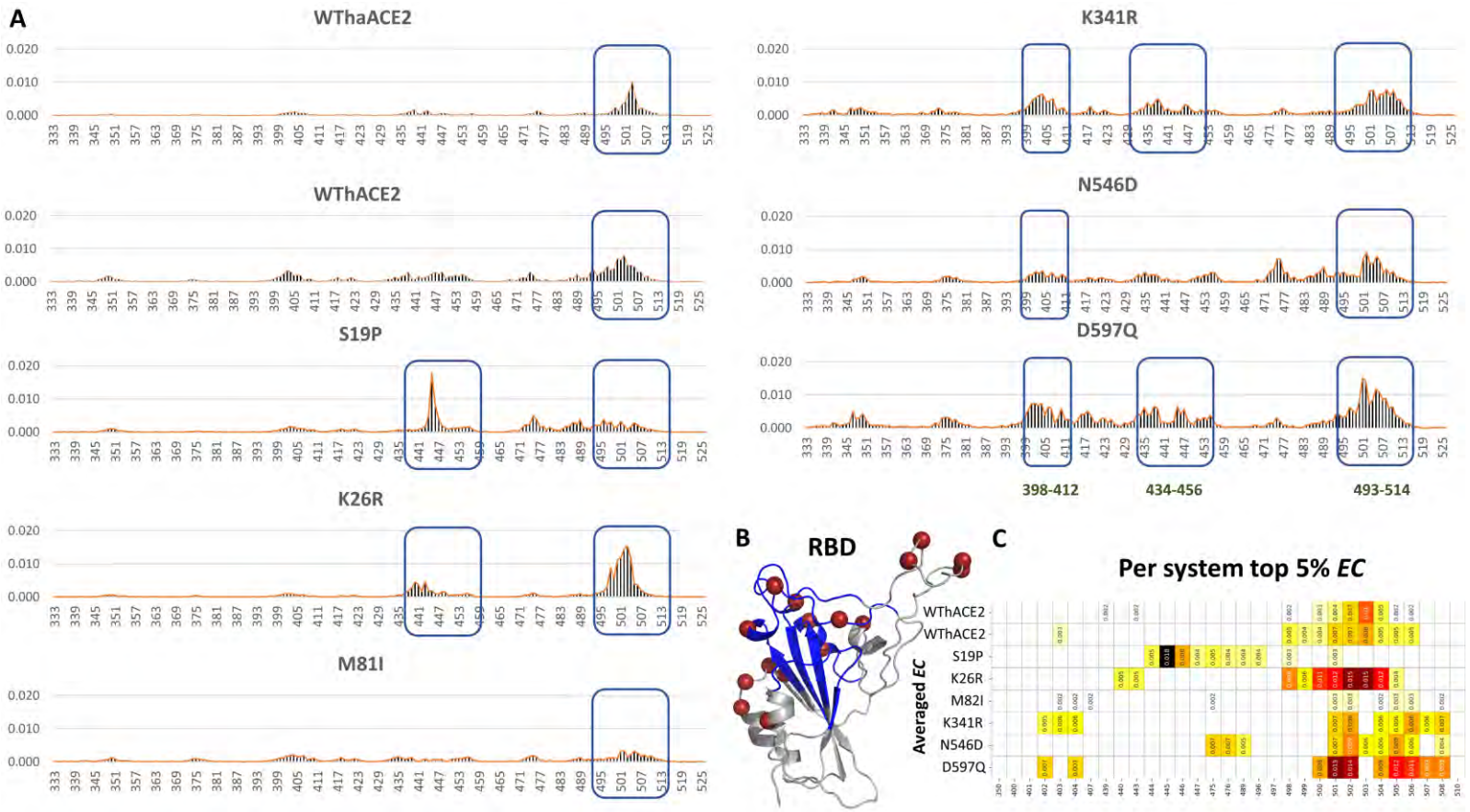

Figure S 27: A shows bar plots of the RBD *EC* distribution per system. Regions with high centrality are marked with blue boxes. The x and y-axes show the residue numbers and *EC* values, respectively. The location of high centrality regions is mapped in blue on cartoon structure of RBD in B and the BA.4 RBD mutations as firebrick spheres. C is a heatmap of the high *EC* residues as calculated per system. The color scale shows the degree of centrality and the high centrality residues are annotated with centrality values.

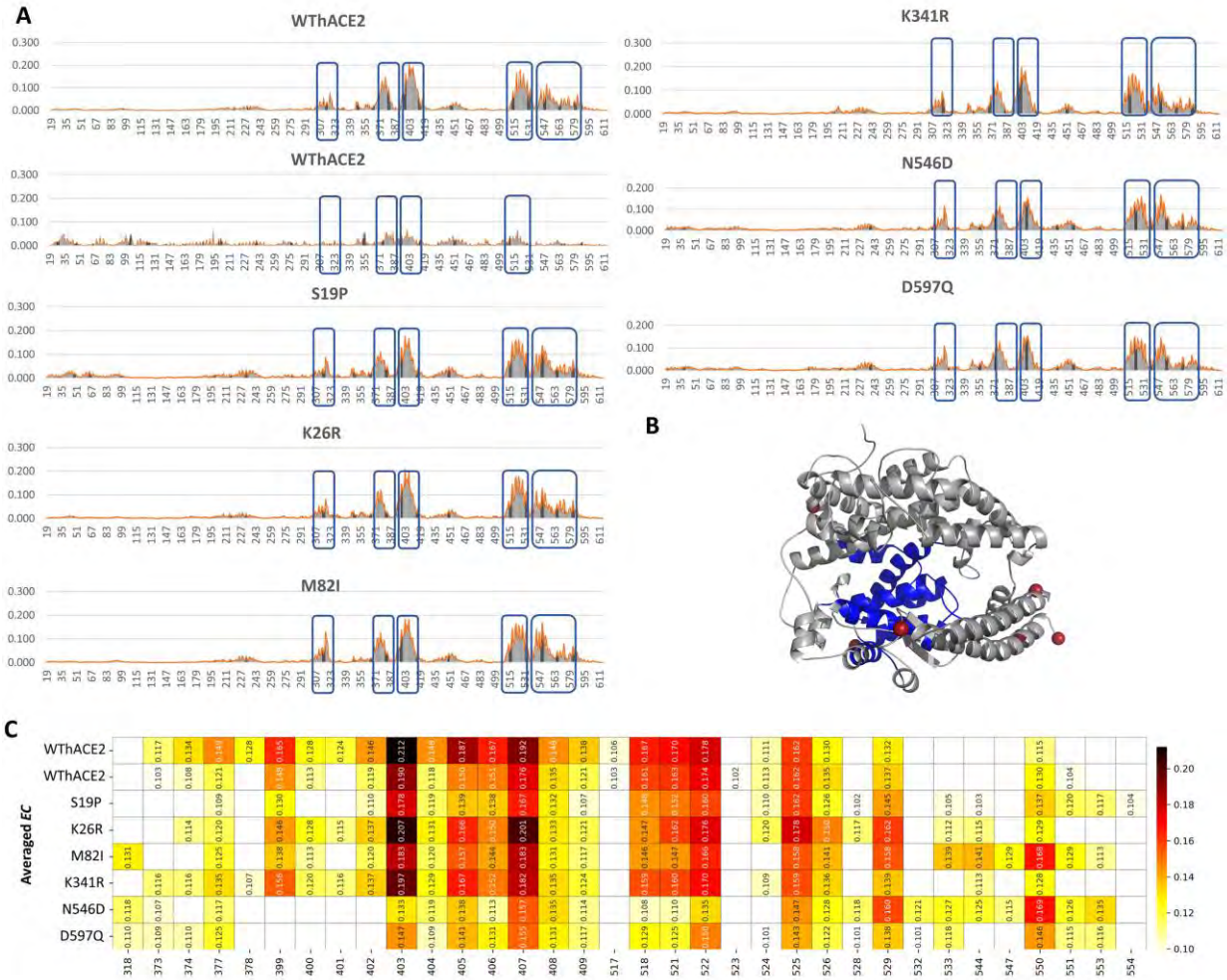

Figure S 28: A shows the distribution of *EC* for the hACE2 protein in each system. The x and y-axes show the residue numbers and the centrality values, respectively. hACE2 regions with high centrality are marked with blue boxes. The high centrality regions were then mapped on the hACE2 structure (B) in blue. Mutation positions on the structure are shown as firebrick spheres. C is a heatmap of the high *EC* residues as calculated per system. The color scale shows the degree of centrality and the high centrality residues are annotated with centrality values.

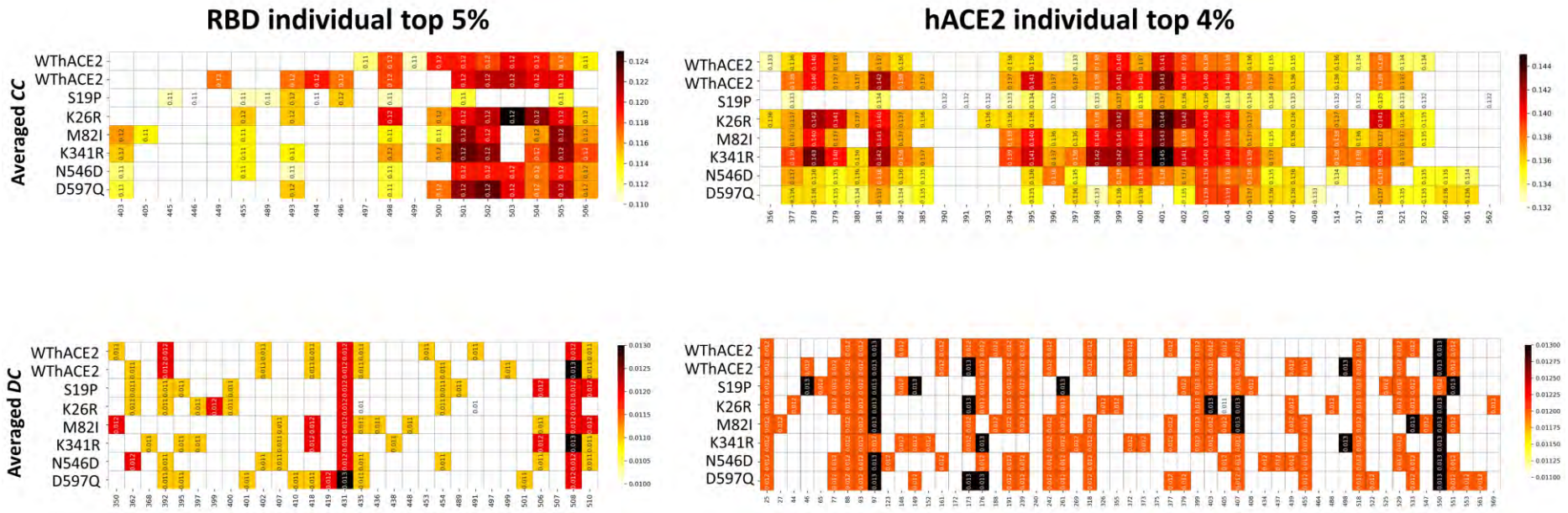

Figure S 29: Heat maps showing the per system high *CC* and *DC* values based on the top 4% and 5% cut-off in the hACE2 and RBD proteins, respectively. The color scale shows the degree of centrality, whereas the high centrality residues are annotated with centrality values.

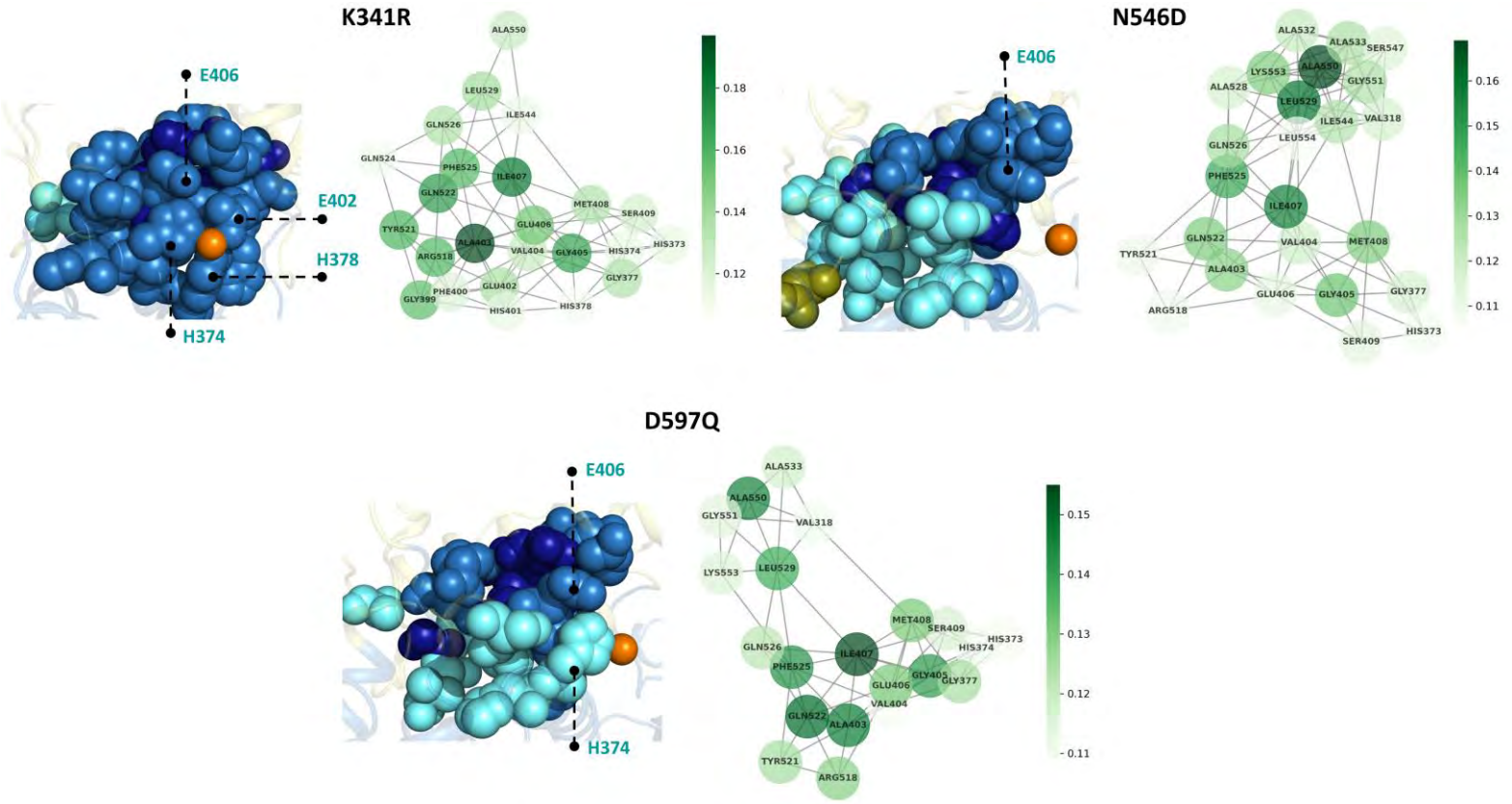

Figure S 30: The network of *EC* hub distribution in the hACE2 proteins. To the left of each sub-plot is the *EC* hub distribution around the zinc ion (orange sphere). WThACE2 hubs are shown as sky-blue spheres and the same colors are used for *EC* hubs common to both the WThACE2 and hACE2 mutant systems. *EC* hubs unique to the complexes with hACE2 mutations are shown as cyan spheres. The zinc coordinating residue hubs are labelled. To the right of each sub-plot is the network between hACE2 *EC* hubs. The node color corresponds to the to the degree of centrality with dark green indicating high *EC* and light green low *EC*.

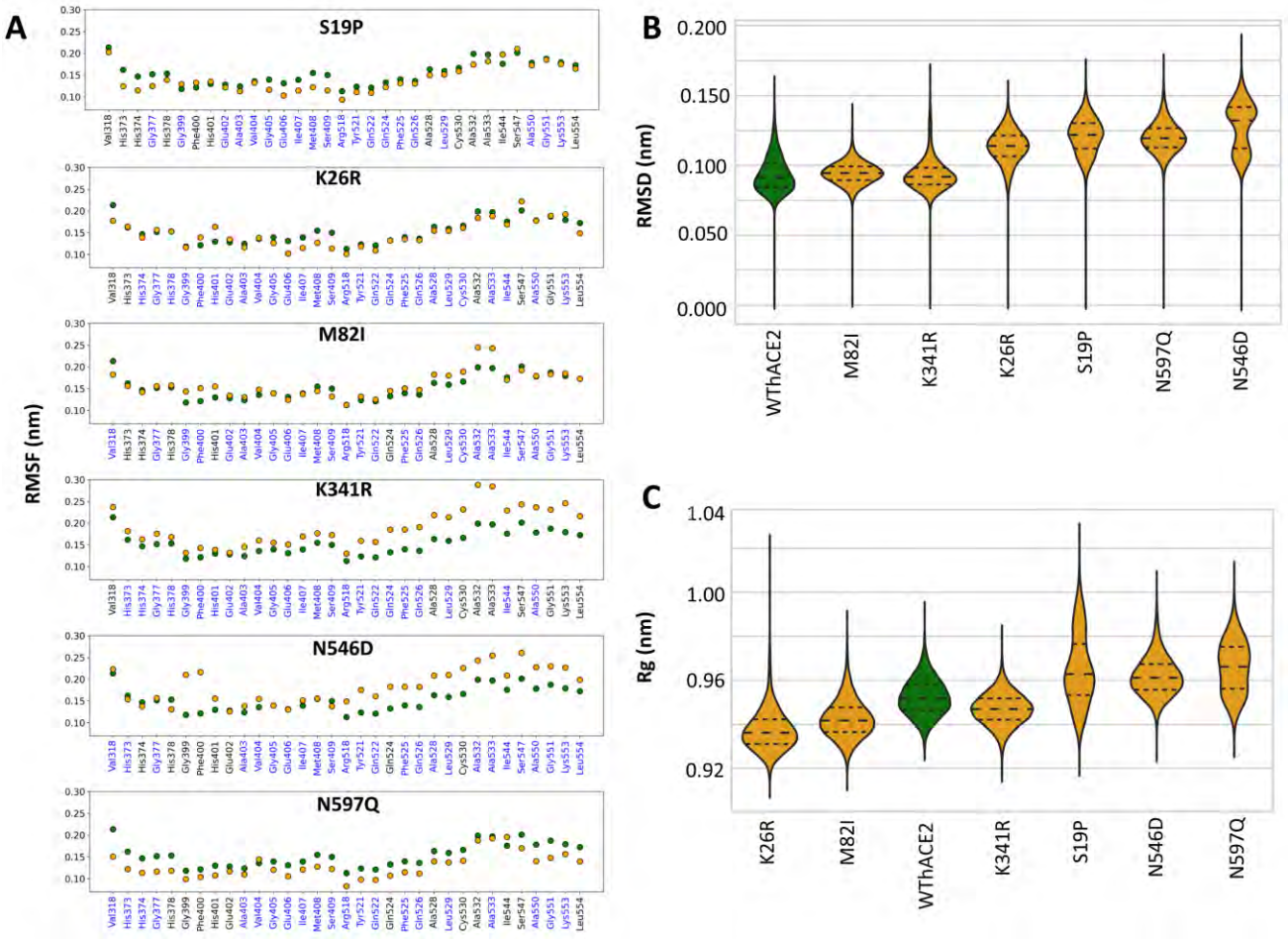

Figure S 31: A shows the RMSF of the WThACE2 *EC* hubs (orange) and hACE2 mutants (orange). *EC* hubs residues unique to the hACE2-mutant containing systems are colored blue. The hACE2 *EC* hub RMSD and Rg is shown in B and C, respectively with the WThACE2 in green and hACE2-mutation bearing systems in orange. The plots are arranged in ascending order of the median.