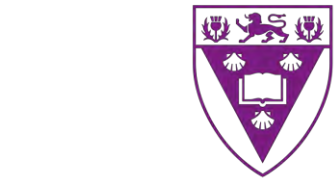


RHODES UNIVERSITY

*Where leaders learn*

# Falcipain 2 and 3 as malarial drug targets: Deciphering the effects of missense mutations and identification of allosteric modulators via computational approaches

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

**DOCTOR OF PHILOSOPHY IN BIOINFORMATICS**


**RESEARCH UNIT IN BIOINFORMATICS (RUBi)**

DEPARTMENT OF BIOCHEMISTRY and MICROBIOLOGY

FACULTY OF SCIENCE

RHODES UNIVERSITY

SOUTH AFRICA

**by**

**CHIAMAKA JESSICA OKEKE**

**ORCID ID: https://orcid.org/0000-0001-5838-5423**


**June 2023**

# Abstract

Malaria, caused by an obligate unicellular protozoan parasite of the genus *Plasmodium*, is a disease of global health importance that remains a major cause of morbidity and mortality in developing countries. The World Health Organization (WHO) reported nearly 247 million malaria cases in 2021, causing 619,000 deaths, the vast majority ascribed to pregnant women and young children in sub-Saharan Africa. A critical component of malaria mitigation and elimination efforts worldwide is antimalarial drugs. However, resistance to available antimalarial drugs jeopardizes the treatment, prevention, and eradication of the disease. The recent emergence and spread of resistance to artemisinin (ART), the currently recommended first-line antimalarial drug, emphasizes the need to understand the resistance mechanism and apply this knowledge in developing new drugs that are effective against malaria. An insight into ART's mechanism of action indicates that ferrous iron ($Fe^{2+}$) or heme, released when hemoglobin is degraded, cleaves the endoperoxide bridge. As a result, free radicals are formed, which alkylate many intracellular targets and result in plasmodial proteopathy. Aside from the existing evidence that mutations in the Kelch 13 protein propeller domain affect ART sensitivity and clearance rate by *Plasmodium falciparum* (*Pf*) parasites, recent investigations raise the possibility that additional target loci may be involved, and these include a nonsense (S69stop) and four missense variants (K255R, N257E, T343P, and D345G) in falcipain 2 (FP-2) protein. FP-2 and falcipain 3 (FP-3) are cysteine proteases responsible for hydrolyzing hemoglobin in the host erythrocytic cycle, a key virulence factor for malaria parasite growth and metabolism. Due to the obligatory nature of the hemoglobin degradation process, both proteases have become potential antimalarial drug targets attracting attention in recent years for the development of blood-stage antimalarial drugs. The alteration of the expression profile of FP-2 and FP-3 through gene manipulation approaches (knockout) or compound inhibition assays, respectively, induced parasites with swollen food vacuoles due to the accumulation of undegraded hemoglobin. Furthermore, missense mutations in FP-2 confer parasites with decreased ART sensitivity, probably due to altered enzyme efficiency and momentary decreased hemoglobin degradation. Hence, understanding how these mutations affect FP-2 (including those implicated in ART resistance) and FP-3 is imperative to finding potentially effective inhibitors.

The first aim of this thesis is to characterize the effects of missense mutations on the partial zymogen complex and the catalytic domain of FP-2 and FP-3 using a range of computational approaches and tools such as homology modeling, molecular dynamics (MD) simulations,

comparative essential dynamics, dynamic residue network (DRN) analysis, weighted residue contact map analysis, amongst others.

The *Pf* genomic resource database (PlasmoDB) identified 41 missense mutations located in the partial zymogen and catalytic domains of FP-2 and FP-3. Using structure-based tools, six putative allosteric pockets were identified in FP-2 and FP-3. The effect of mutations on the whole protein, the central core, binding pocket residues and allosteric pockets was evaluated. The accurate 3D homology models of the WT and mutants were calculated. MD simulations were performed on the various systems as a quick starting point. MD simulations have provided a cornerstone for establishing numerous computational tools for describing changes arising from mutations, ligand binding, and environmental changes such as pH and temperature. Post-MD analysis was performed in two stages viz global and local analysis. Global analysis via radius of gyration (Rg) and comparative essential dynamic analysis revealed the conformational variability associated with all mutations. In the catalytic domain of FP-2, the presence of M245I mutation triggered the formation of a cryptic pocket via an exclusive mechanism involving the fusion of pockets 2 and 6. This striking observation was also detected in the partial zymogen complex of FP-2 and induced by A159V, M245I and E249A mutations. A similar observation was uncovered in the presence of A422T mutation in the catalytic domain of FP-3. Local DRN and contact map analyses identified conserved inter-residue interaction changes on important communication networks. This study brings a novel understanding of the effects of missense mutations in FP-2 and FP-3 and provides important insight which may help discover new anti-hemoglobinase drugs.

The second aim is the identification of potential allosteric ligands against the WT and mutant systems of FP-2 and FP-3 using various computational tools. Of the six potential allosteric pockets identified in FP-2 and FP-3, pocket 1 was evaluated by SiteMap as the most druggable in both proteins. This pipeline was implemented to screen pocket 1 of FP-2 and FP-3 against 2089 repositionable compounds obtained from the DrugBank database. In order to ensure selectivity and specificity to the *Plasmodium* protein, the human homologs (Cat K and Cat L) were screened, and compounds binding to these proteins were exempted from further analysis. Subsequently, eight compounds (DB00128, DB00312, DB00766, DB00951, DB02893, DB03754, DB13972, and DB14159) were identified as potential allosteric hits for FP-2 and five (DB00853, DB00951, DB01613, DB04173 and DB09419) for FP-3. These compounds were subjected to MD simulation and post-MD trajectory analysis to ascertain their stability in their respective protein structures. The effects of the stable compounds on the WT and mutant

systems of FP-2 and FP-3 were then evaluated using DRN analysis. Attention has recently been drawn towards identifying novel allosteric compounds targeting FP-2 and FP-3; hence this study explores the potential allosteric inhibitory mechanisms in the presence and absence of mutations in FP-2 and FP-3.

Overall, the results presented in this thesis provide (i) an understanding of the role mutations in the partial zymogen complex play in the activation of the active enzyme, (ii) an insight into the possible allosteric mechanisms induced by mutations on the active enzymes, and (iii) a computational pipeline for the development of novel allosteric modulators for malaria inhibition studies.

# Declaration

I, **CHIAMAKA JESSICA OKEKE**, declare that this is my unaided work, except where duly acknowledged. It is being submitted to the Faculty of Science at Rhodes University for the Doctor of Philosophy in Bioinformatics degree. It has never been submitted to another university for a degree or examination.

Signature: ……...

Date: ……

# Acknowledgements

I would like to express my deepest appreciation to my supervisor Prof Özlem Tastan Bishop, for her unwavering support and relentless assistance. Her patience, motivation, enthusiasm, and immense knowledge contributed to the successful completion of my PhD.

I sincerely thank Dr Thommas Mutemi Musyoka for his mentorship, valuable comments, suggestion, and encouragement throughout my study.

I extend my gratitude to my colleagues at the Research Unit in Bioinformatics (RUBi), Rhodes University, for the stimulating discussions, the emotional support, and the fun we had especially last year. To my dearest friends (past and present colleagues) Dr Colleen Varaidzo Manyumwa, Dr Lorna Chebon-Bore, Dr Rita Afriyie Boateng-Fosu, Dr Rolland Bantar Tata, Dr Fisayo Olotu, Dr Allan Sanyanga, Chiratidzo Chamboko, Lilian Mbaisi Ang'ang'o, Bilinga Tendwa, Victor Barozi, Zach Mokgerwa Monama, and Rehema, thanks for all the great times we had. To Tendai Muronzi, we started this postgraduate journey together and completing this milestone is a testament to God's grace.

I'm extremely grateful to my friend and sister, Dr Damilola Esther Folley, for her endless encouragement, moral support, and wise counsel. To my friends, Dr T.A. Anjugu, Dr Deborah Kajewole, and Dr Miebaka Jamabo thank you for all your support throughout my study. To my best friend, Yvonne Grace Akande, thank you for keeping me sane and being such a strong support system. To Benigna Anyaeche, my number-one hype woman, thank you for constantly cheering me on and being in my corner. To Cynthia Nwaogaraku, I am blessed to have you as a friend.

Special thanks to the H3ABioNet, who funded my research through the National Human Genome Research Institute of the National Institutes of Health under Award Number U24HG006941.

I greatly appreciate Dr Remy and Dr Chika Nnadozie and family for being my home away from home. My dearest friend and sister in Christ, Chidinma Iheanatu, thank you for your spiritual, physical, and emotional support. Special thanks to my church in Grahamstown; I'm grateful to have served God with you all.

To my mentor, Prof Oyekanmi Nashiru, and my former colleagues from the Department of Genetics, Genomics and Bioinformatics, National Biotechnology Development Agency (NABDA), thank you for your support.

To my mother, Mrs D.A. Okeke, my aunts: Mrs Nkechi Jane Okoye and Mrs Chidiebere Ezeude, my sister and her husband: Chief and Lolo L. Ajuogu, my brothers: Arinze Ignatius Okeke and Chuka Emmanuel Okeke, my cousins and members of my extended family, your faith, hope, understanding, endless patience, encouragement, prayers, was my strength during the storm. Thank you.

# Dedication

Who am I, Lord, that you are mindful of me? You have made me a little lower than the heavenly beings, and have crowned me with glory and honour (**Psalm 8: 4 – 5** "*paraphrased*")

I dedicate this thesis as a tribute to my beloved father, **Mr I.G. OKEKE**, who passed away 25 years ago. To my amazing mother, **Mrs D.A. OKEKE**, who gave me wings to fly, I love you always and forever.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| 3D | Three-dimension |
| ART | Artemisinin |
| ARTR | Artemisinin-resistant |
| *BC* | *Betweenness centrality* |
| CADD | Computer-aided drug discovery |
| Cathepsin K | Cat K |
| Cathepsin L | Cat L |
| *CC* | *Closeness centrality* |
| CED | Comparative essential dynamics |
| CHPC | Center for High Performance Computing |
| *DC* | *Degree centrality* |
| DRN | Dynamic Residue Network |
| *EC* | *Eigenvector centrality* |
| FP-2 | Falcipain 2 |
| FP-3 | Falcipain 3 |
| HTVS | High throughput virtual screening |
| *KC* | *Katz centrality* |
| LBVS | Ligand-based virtual screening |
| MD | Molecular Dynamics |
| NALs | Neutral allosteric modulators |
| NAMs | Negative allosteric modulators |
| Nanosecond | *ns* |
| NMR | Nuclear Magnetic Resonance |

| | |
|---|---|
| PAMs | Positive allosteric modulators |
| PARS | Protein Allosteric and Regulatory Sites |
| PC1 | Principal component 1 |
| PC2 | Principal component 2 |
| PCA | Principal component analysis |
| PDB | Protein Data Bank |
| *Pf* | *Plasmodium falciparum* |
| PIC | Protein interaction calculator |
| PlasmoDB | *Pf* genome resource database |
| RBCs | Red Blood Cells |
| RMSD | Root mean square deviation |
| RMSF | Root mean square fluctuation |
| Rg | Radius of gyration |
| SAMs | Silent allosteric modulators |
| SBDD | Structure-based drug design |
| SBVS | Structure-based virtual screening |
| WHO | World Health Organisation |
| WT | Wildtype |

# Research Outputs

## Primary publication

1. **Okeke, C.J**.[#], Musyoka, T.M.[#], Sheik Amamuddy, O., Barozi, V., 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*, 19, 5647–5666 (2021), DOI:10.1016/j.csbj.2021.10.011. [#] Indicates equal first authorship.

## Secondary publications

1. Hamdi, Y., Zass, L., Othman, H., Radouani, F., Allali, I., Hanachi, M., **Okeke, C.J**., Chaouch, M., Tendwa, M.B., Samtal, C., et al. Human OMICs and Computational Biology Research in Africa: Current Challenges and Prospects. *OMICS A Journal of Integrative Biology*, 25(4), 213–233 (2021). DOI: 10.1089/omi.2021.0004.
2. Othman, H., Zass, L., da Rocha, J. E. B., Radouani, F., Samtal, C., Benamri, I., Kumuthini, J., Fakim, Y. J., Hamdi, Y., Mezzi, N., Boujemaa, M., **Okeke, C. J.**, et al. African Genomic Medicine Portal: A Web Portal for Biomedical Applications. *Journal of Personalized Medicine* 12(2), 265 (2022). DOI: 10.3390/JPM12020265/S1.

## Conference Presentations

1. Oral presentation: P4ND'21 edition, Center for Genomics Research and Innovation (CGRI), National Biotechnology Development Agency, Abuja, Nigeria. 14$^{th}$ – 15$^{th}$ December 2021. Title: Allosteric pockets and dynamic residue network hubs of Falcipain 2 in mutations including those linked to artemisinin resistance.
2. Oral presentation: Chemico and Biomedicinal Research Symposium, Rhodes University. 30$^{th}$ May 2022. Title: Allosteric pockets and dynamic residue network hubs of Falcipain 2 in mutations including those linked to artemisinin resistance.

# Chapter 1

## 1.1 Introduction

Pathogenic organisms develop resistance to currently available drugs due to several factors, including excessive use of drugs (e.g. antibiotics, antimalarials), insufficient controls on drug prescription, inadequate compliance with treatment regimens, and improper dosing [1,2]. Despite the huge advances being made to curb these factors, it is also imperative to understand the resistance mechanism in these pathogenic organisms to curtail the spread of resistance and devise innovative therapeutic measures to combat multidrug-resistant organisms [3].

One of the most important tropical parasitic diseases, malaria, is caused by multidrug-resistant *Plasmodium species* [4]. Among the different *Plasmodium species*, five are known to infect humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi* [5]. *P. falciparum* is the most prevalent species, accounting for approximately 80% of malaria infections and serious complications resulting in death [6]. In most malaria-endemic areas, particularly in sub-Saharan Africa, *P. falciparum* accounted for 99.7% of estimated cases in 2018, as well as 71% in the World Health Organization (WHO) regions of Eastern Mediterranean, 65% in the Western Pacific, and 50% in South-East Asia, [7]. A range of strategies has been implemented across the globe to mitigate and eliminate malaria, including early disease detection, antimalarial drugs, and vector control [8]. Antimalarial drugs are an integral part of this strategy. Among the most extensively used antimalarial drugs are quinolones (amodiaquine, chloroquine, mefloquine, piperaquine, primaquine and quinine), antifolate drugs (chloroguanide proguanil, dapsone, paludrine, pyrimethamine, sulfadoxine and sulfalene), and artemisinin [9,10]. However, the progression of antimalarial drug resistance compromises the prevention, treatment, and eradication of malaria [11]. With the onset and spread of chloroquine resistance in Southeast Asia and South America in the late 1950s, Papua New Guinea in the 1960s, and East Africa in the late 1970s, *Plasmodium* parasites have continuously developed complex mechanisms to resist almost any antimalarial drug, diminishing the effectiveness of all eradication efforts [12–15]. Artemisinin (ART), the current first-line antimalarial drug recommended by the WHO, has recently developed resistance, underscoring the need to investigate and develop novel antimalarial drugs [16].

## 1.2 The erythrocytic cycle of *Plasmodium falciparum*

Malaria eradication strategies can take advantage of numerous potential targets throughout the *Plasmodium* life cycle [17]. A prominent stage in the life cycle of the parasite is the pre-erythrocytic stage, in which the parasite is transmitted as sporozoites which infect the liver of the human host [18,19] (**Figure 1.1**). The sporozoites multiply asexually to form multiple merozoites, which are released into the bloodstream, where they activate the erythrocytic stage [20,21]. During the erythrocytic cycle, the parasite population multiplies substantially through periodic cycles of invasion of merozoites into the red blood cells (RBCs), degradation of hemoglobin, rupture of the schizonts, the release of the merozoites, and re-invasion into non-infected erythrocytes [6,22,23]. The repetitive cycle of asexual multiplication and toxic material emitted from infected RBCs are responsible for the clinical manifestation of malaria [6].

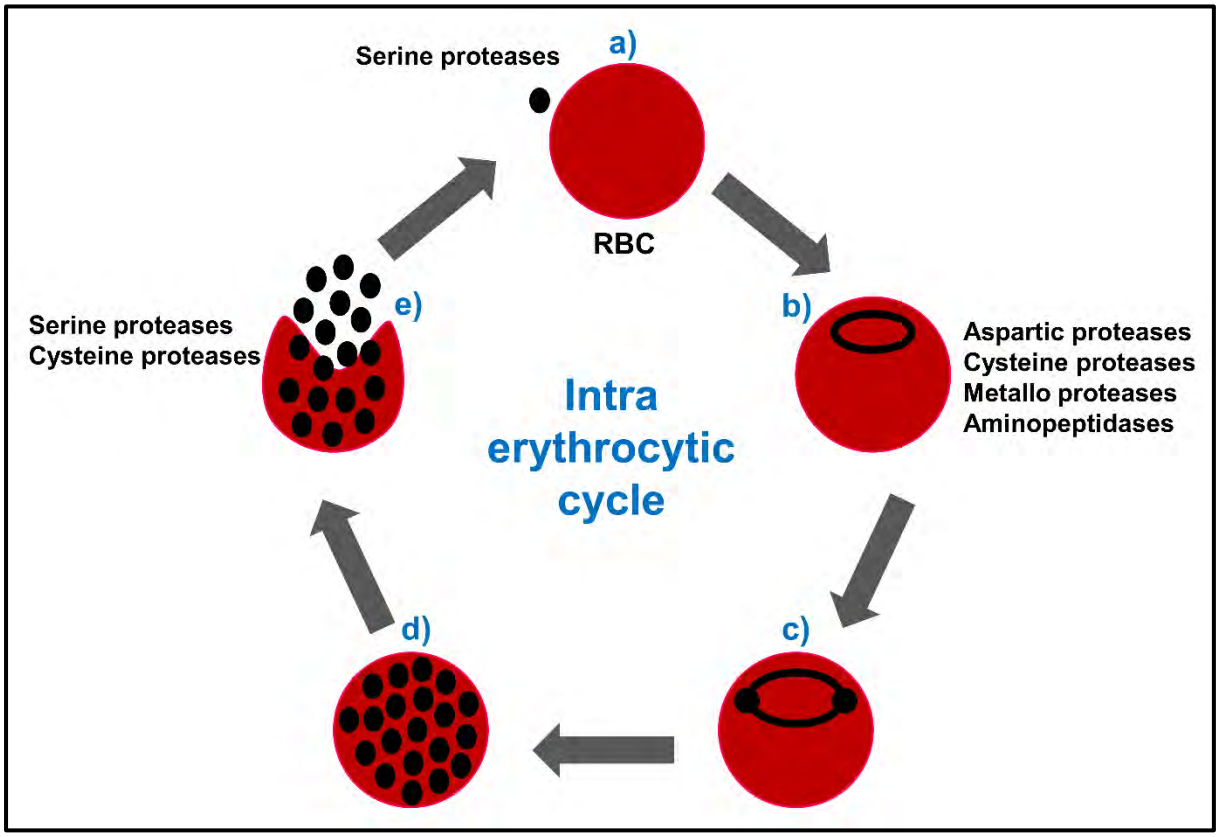



**Figure 1.1: Schematic representation of the role of proteases in the intra-erythrocytic cycle of the malaria parasites**. The asexual life cycle begins with (a) the release of the merozoite into the host red blood cell (RBC), where they mature through a variety of stages (b) ring, (c) trophozoite, (d) schizont, and the rupture of erythrocytes and release of new merozoites which perpetuates the asexual life cycle. (Adapted from [24,25]).

It is imperative to understand the mechanism of action of past and current antimalarials in order to develop new compounds that target the parasite's erythrocytic stage [22,26]. The development of *Plasmodium* parasite relies on the degradation of hemoglobin in the food vacuole, which provides a source of free amino acids for protein synthesis and nutritional requirements for growth and maturation [27]. A toxic heme is released, which is subsequently sequestered as hemozoin [22]. The quinoline group of antimalarials appear to interfere with heme sequestration, which, in spite of widespread resistance to chloroquine, is still a viable target for the development of new antimalarial drugs [22,28,29]. An insight into the mechanism of action of ART suggests the involvement of $Fe^{2+}$ or heme in the cleavage of the endoperoxide bridge of ART, leading to the generation of free radicals, which in turn alkylate a wide range of potential intracellular targets, resulting in the death of the parasite [30,31].

## 1.3 Proteases: An attractive antimalarial drug target

Proteases are a group of enzymes ubiquitously expressed in many living systems and regulate several physiological processes, including enzymatic activation, induction of cellular signalling, and protein degradation [32,33]. Proteases are one of the major virulence factors in diseases ranging from cardiovascular disorders to cancer, as well as various parasitic diseases such as African sleeping sickness, Chagas disease, leishmaniasis, malaria, and schistosomiasis [34–36]. In addition to their implication in parasite pathogenesis, proteases are involved in the invasion of host cells, hemoglobin degradation, immune evasion, and activation of inflammation [34]. Consequently, they are essential for both the survival of pathogenic organisms and the transmission of diseases, and due to their attractive nature, they are crucial for drug discovery [32]. A number of drugs that inhibit the activity of proteases have been successfully developed to combat coagulopathies [37], hepatitis C virus (HCV) [38], and human immunodeficiency virus (HIV) [39]. Proteases are grouped into six distinct classes based on their mechanism of action, and this includes aspartic, serine, cysteine, glutamic, metallo and threonine proteases [40]. Aspartic, glutamic and metallo proteases utilize an activated water molecule which acts as a nucleophile attack on the peptide bond of the substrate [40]. The nucleophiles of serine, cysteine, and threonine proteases are amino acid residues found in the active site of the enzyme (cysteine, serine, or threonine) [40]. According to the MEROPS protease database [41], the *Plasmodium* genome encodes ~170 predicted proteases, but only one-third of them have been identified and analyzed [42]. Considering the malaria parasite's limited ability for *de novo* amino acid biosynthesis, the parasite depends on

hemoglobin degradation as a source of free amino acid for protein biosynthesis and energy metabolism [43,44]. Malaria parasites express a multitude of proteases in the food vacuole for hemoglobin degradation, including a group of aspartic proteases (plasmepsins I, II, III, and IV), a group of cysteine proteases (falcipains 2, 2', and 3), a metalloprotease (falcilysin), dipeptidyl peptidase 1, and aminopeptidase [43]. The hemoglobin degradation occurs sequentially, with plasmepsins cleaving native hemoglobin into smaller peptides, followed by secondary cleavage by falcipains [43]. The sequence of cleavage is debatable as an alternative pathway that documents the involvement of falcipains in the initial cleavage of hemoglobin has been recorded [43]. In turn, falcilysin and dipeptidyl peptidase 1 cleave the small peptides into shorter oligopeptides which are hydrolysed by aminopeptidases to free amino acids [43,45,46]. Due to the obligate nature of the hemoglobin degradation process, proteases have become potential antimalarial drug targets, attracting augmented attention in recent years for the development of blood-stage antimalarial drugs [42,43,46–49]. This study is focused on cysteine proteases.

## 1.4 *Plasmodium falciparum* cysteine proteases

Cysteine proteases are enzymes named due to the function of their catalytic cysteine, which facilitates protein hydrolysis via a nucleophilic attack on the carbonyl carbon of a susceptible peptide bond [50]. Cysteine proteases are subdivided into clans that differ in sequence or structural identity and probably evolved independently but utilize cysteine residue for their enzymatic activity [51]. Clans are grouped into families based on sequence identities and similarities [41,52]. The clan CA, family C1 (papain-family), is the largest clan of cysteine proteases characterized by the presence of Cysteine (Cys), Histidine (His), and Asparagine (Asn) in their catalytic triad [6,51]. The nucleophilicity of cysteine is increased by the catalytic histidine residue, which also serves as a proton donor [53]. These are the most studied cysteine proteases of *Plasmodium* [50]. The genome sequence analysis of *P. falciparum* suggests that clan CA cysteine proteases include four falcipains, three dipeptidyl peptidases, nine serine-rich antigen (SERA) related proteins, and a homolog of calpain [50]. This thesis is focused on falcipains, including falcipain 2 (FP-2) and falcipain 3 (FP-3). Clan CD is the second clan of interest in *Plasmodium* with a caspase-like fold which utilizes Histidine and Cysteine dyad for its catalytic activity [50]. They consist of families C11, C13, C14, C15 and C50, and sequence analysis indicates that members of families C13, C14 and C50 are present in *Plasmodia*

[50,54]. Clan CE is also represented in the *P. falciparum* genome and is distinguished by catalytic residues His, Glu (or Asp), and Cys [50].

### 1.4.1 Functions of *Plasmodium* cysteine proteases

Inhibition studies on cysteine proteases have been used to characterize their functions in malaria parasites [50]. The effects of leupeptin and E-64 against cultured *P. falciparum* parasite were extensively characterized, and both inhibitors were effective against both the trophozoites and schizonts phases of the parasite's life cycle, confirming the role of cysteine proteases FP-2 and FP-3 in hemoglobin degradation [55–57]. The best-characterized functions of *Plasmodium* cysteine proteases are hemoglobin hydrolysis, erythrocyte invasion, and erythrocyte rupture [58]. During the erythrocytic cycle, malaria parasites multiply asexually, resulting in clinical manifestations [50]. As part of this process, erythrocyte cytosol is absorbed by parasites through a specialized organelle known as the cytostome before being transported to an acidic food vacuole where hemoglobin is degraded [28,50]. During the degradation process, the heme component of hemoglobin is broken down into hemozoin pigment while the globin component is hydrolysed to its constituent amino acids [50]. The hydrolysis of hemoglobin provides amino acids necessary for parasite protein synthesis [28,59]. The mechanism of hemoglobin hydrolysis appears to be mediated by a cooperative process involving proteases of many catalytic classes, such as aspartic, cysteine, and metallo proteases [28,50]. Inhibition of cysteine proteases disrupted the hemoglobin degradation process. This led to the accumulation of large quantities of undigested hemoglobin and the enlargement of the food vacuole [57,60]. Additionally, inhibition of cysteine proteases prevents the rupture of erythrocytes associated with the end of the erythrocytic cycle [50]. Consequently, cysteine protease activity appears to be essential to facilitate the release of merozoites, which swiftly invade other erythrocytes to restart the asexual cycle [50]. Earlier research showed that malaria parasite cultures treated with leupeptin accumulated mature schizonts [61–63]. Likewise, in the mature schizont stage parasite, E-64 inhibited the lysis of the parasitophorous vacuole membrane surrounding the intraerythrocytic parasite [64]. Another research using different methods demonstrated that cysteine protease inhibitors, leupeptin and antipain inhibited erythrocyte membrane lysis [65]. These studies suggest the involvement of cysteine proteases in the release of merozoites [50].

## 1.5 Falcipain cysteine proteases

The best characterized cysteine proteases in *P. falciparum* are four papain-family (clan CA, family C1) cysteine proteases known as falcipains [50]. Falcipains are the best biochemical and genetically characterized *Plasmodium* cysteine proteases [24]. These proteases are falcipain 1 (FP-1), falcipain 2 (FP-2), falcipain 2' (FP-2'), and falcipain 3 (FP-3'). The *P. falciparum* genome sequence reveals that FP-1 is encoded in chromosome 14, and the three other proteases (FP-2, FP-2', and FP-3) are encoded within a 12kb stretch of chromosome 11 [50]. The sequence identity (<40%) and function of FP-1 are distantly related to those of the other falcipains [66]. Although the physiological function of FP-1 is not fully understood, a new study raises the possibility that it plays a role in the development of oocysts in the mosquito midgut during the development of parasites [67]. FP-2, FP-2', and FP-3 are closely related and have been identified as the most important hemoglobinases in the acidic food vacuole [68–70]. Among the falcipains, FP-2 has been studied extensively through biochemical characterization, cellular localization, recombinant expression, structural analysis, and biological functions [50,66]. FP-2 and FP-2' share a 93% sequence identity (at the amino acid level) and differ by three amino acid substitutions close to the active site [69]. FP2' has not been previously characterized; however, its physiological role and function as a hemoglobinase have been attributed to its high sequence identity with FP-2 [71]. FP-2 and FP-3 share the following properties: 68% sequence identity (**Figure 1.2A**) and a similarly sized prodomain (**Figure 1.2B**). Additionally, they possess a unique amino acid extension of the catalytic domain, viz., a nose region consisting of ~ 17 amino acids and a $\beta$ arm region of ~ 14 amino acids (**Figure 1.2B**) [50]. These unique features are functionally conserved among *Plasmodium* cysteine proteases but not found in the human homologs cathepsins K (Cat K) and L (Cat L) [72].

### 1.5.1 Expression of falcipains by erythrocytic parasites

Purification of FP-2 by affinity chromatography protocol confirms that FP-2 is responsible for at least 93% of the cysteine protease activity that is identified in trophozoite lysates with standard peptidyl substrates (which is used to study cysteine protease activity) [68]. FP-3 is abundantly expressed by erythrocytic parasites, but it is inactive against peptidyl substrates [70]. FP-2 and FP-3 are expressed sequentially by *Plasmodium falciparum* through the erythrocytic cycle [73]. FP-2 is expressed 12 hours earlier than FP-3, with peak expression in early trophozoites for FP-2 and late trophozoites for FP-3 [70,74]. Immunoblotting analysis revealed the expression of FP-1 throughout the erythrocytic cycle [74], and

immunofluorescence-based localization studies determined that it is active during the invasive merozoite stage [75].

### 1.5.2 Biochemical characterization of FP-2 and FP-3

The biochemical characterization of FP-2 and FP-3 has been aided by the expression of large amounts of the two enzymes in *Escherichia coli,* followed by the refolding of the active enzyme using a systematic microdilution approach [68,70,76]. FP-2 and FP-3 share similar biochemical characteristics but differ in other ways, possibly reflecting differences in function [50]. Both proteases have acidic pH optima, which is consistent with activity in the acidic food vacuole [50]. The specificities of FP-2 and FP-3 for peptide substrates are similar to those of other papain-family proteases, with high selectivity for peptidyl substrates with leucine at the $P_2$ position in the binding pocket [73]. FP-2 is significantly more active against peptide substrates than FP-3 and is more stable at neutral pH [77]. FP-2 is also the only enzyme that can activate and undergo autohydrolysis at this pH [50]. The activities of FP-2 and FP-3 are, however, remarkably similar when the natural substrate hemoglobin is taken into consideration as both proteases cleaved native and denatured hemoglobin under mild reducing conditions [58]. FP-2 and FP-3 have been localized in the acidic food vacuole using cell fractionation [68,70], immunofluorescence, and immunoelectron microscopy techniques [78]. Hydrolysis of FP-3 appears to be delayed until the enzyme enters an acidic environment, most likely the food vacuole [50]. Despite the fact that both FP-2 and FP-3 are synthesized as membrane-bound peptides, FP-2 is processed faster to mature protease, presumably to its susceptibility to autohydrolysis at a neutral pH [73]. Both proteases were inhibited by cysteine protease inhibitors and brefeldin A but not by aspartic or serine protease inhibitors, implying that FP-2 and FP-3 undergo autohydrolysis after leaving the endoplasmic reticulum/Golgi network [78]. These findings suggest that these proteases are essential membrane proteins that reside in a particular cellular compartment before being transported to the food vacuole, where they undergo autohydrolysis to release the active soluble proteases [73].

### 1.5.3 Structural properties of FP-2 and FP-3

As with other cysteine proteases, FP-2 and FP-3 are synthesized as inactive zymogen complexes with a prodomain and a mature (active) domain linked by a salt bridge and hydrophobic interactions (**Figure 1.2B**) [79,80]. The zymogen complex is composed of the prodomain, which is an N-terminal extension of the mature domain. **Figure 1.2B** depicts the

zymogen complex structure, with the prodomain region (coloured in wheat) spanning over the mature domain (coloured in grey). The prodomain of cysteine proteases serves two distinct functions: to maintain the enzyme in an inactive state (zymogen complex) until it reaches the appropriate site where protease activity occurs and to function as a structural template to ensure folding during translation [81]. During transport to the food vacuole via the endoplasmic reticulum/Golgi system, the prodomain is proteolytically removed, releasing the activated mature enzyme [79]. Expression of recombinant papain-like cysteine proteases in *E. coli* by dilution in alkaline buffer revealed that the mature enzyme of FP-2 and FP-3 is the only papain-family proteases capable of refolding to active enzymes and does not require their prodomains [76,82]. Nevertheless, the prodomain serves as a signal for intracellular localization and inhibits the premature cleavage of the mature domain [77]. The 3D protein structures of the mature domain of FP-2 in complex with cystatin [83], iodoacetamide [84], and E-64 [56] and FP-3 in complex with leupeptin [56] and a vinyl sulfone inhibitor [85] have been resolved [86]. The mature domain of FP-2 contains residues 244 – 484 and 250 – 492 for FP-3. The mature domain has two main sections [left (L) and right (R)] consisting of $\alpha$-helices and $\beta$-sheets, respectively, which are linked together by highly dynamic loops (**Figure 1.2C**) [83]. The centrally located trench-like active site, including the catalytic triad (Q285, H417, N447 for FP-2, and C293, H425, and N455 for FP-3), are organised into four subsites, namely, S1, S2, S3, and S1' [56]. Residues in the active sites of FP-2 and FP-3 are detailed in **Table 1.1**.

**Table 1.1: Position of the different subsite residues in FP-2 and FP-3 proteins** (whole protein numbering) [87].

| | **Protein** | |
|---|---|---|
| **Subsite** | **FP-2** | **FP-3** |
| **S1** | Q279, C282, G283, C323, N324 | Q287, C290, G291, C331, Y332 |
| **S2** | L327, I328, S392, L415, N416, A418, D477 | Y335, I336, S400, P423, N424, A426, E485 |
| **S3** | K319, N320, Y321, G325, G326 | K327, N328, N329, G333, G334 |
| **S1'** | V393, A394, V395, S396, A400, H417, N447, W449 | I401, A402, A403, S404, A408, H425, N455, W457 |

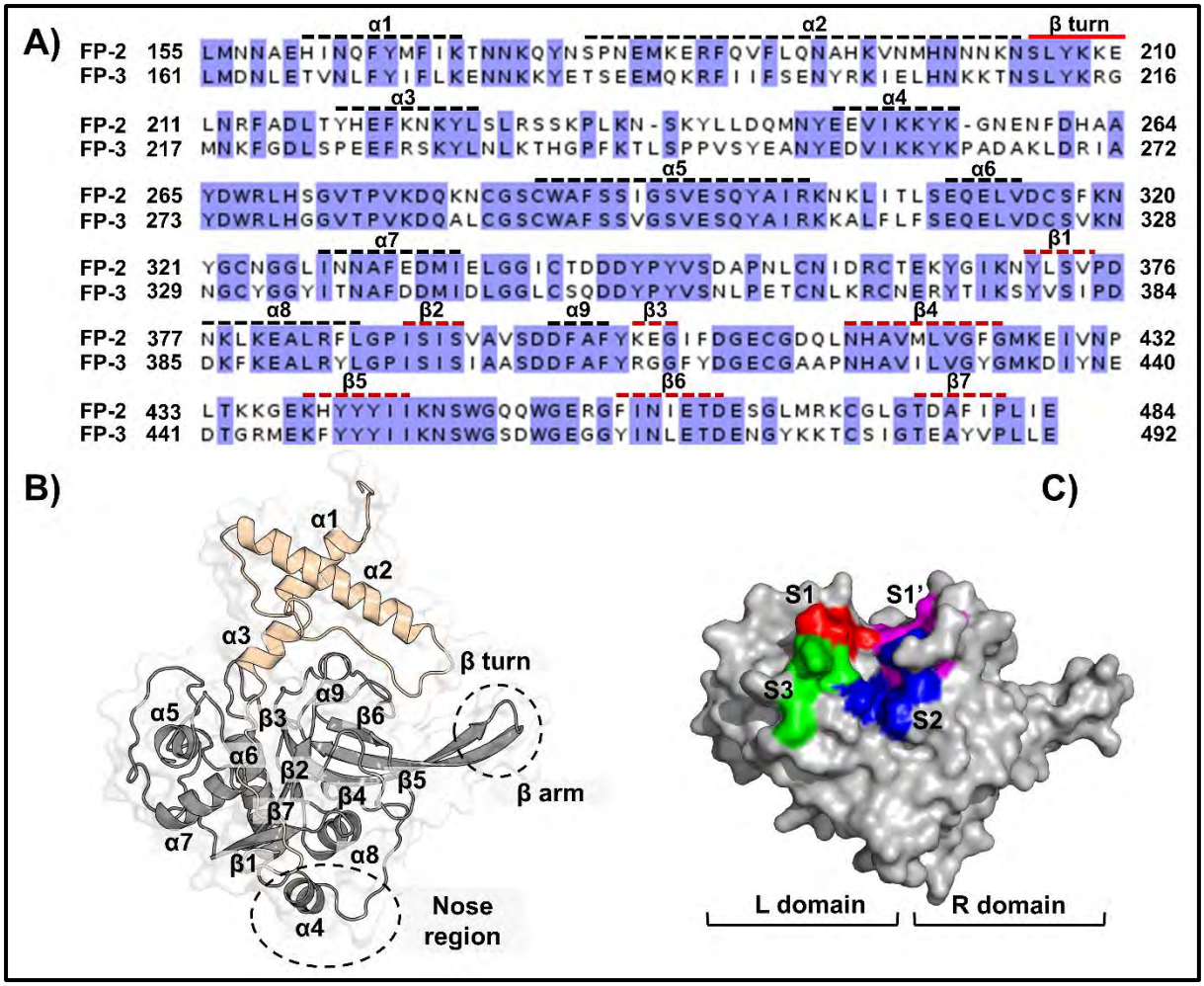



**Figure 1.2: Sequence and structural information of FP-2 and FP-3. A) Sequence information and residue alignment showing the level of conservation between FP-2 and FP-3.** Conserved residues are highlighted in blue. B) Structural fold, as well as the position of various secondary structural elements. C) Position of the subsites S1, S1', S2 and S3. The prodomain (residues 155 – 243 for FP-2 and 161 – 249 for FP-2) is coloured in wheat, and the catalytic domain is grey. Residue positions are numbered based on the whole protein sequence. The different subsite residues are listed in **Table 1.1** [87].

## 1.6 Understanding the relationship between hemoglobin degradation and artemisinin resistance

The emergence and spread of resistance to artemisinin and its derivatives in the Greater Mekong sub-region (Cambodia, Lao People's Democratic Republic, Myanmar, Thailand, and Vietnam) and several areas in Africa (specifically Eritrea, Rwanda, and Uganda) poses a serious threat to global eradication malaria [88–93]. Artemisinin and its semi-synthetic derivatives, including arteether, artemether, artesunate, and dihydroartemisinin, are sesquiterpene lactones that inhibit nearly all erythrocytic stages of *Plasmodium falciparum* parasites life cycle [29,94]. These consist of asexual stages, which are responsible for the

clinical manifestation of malaria, and sexual stages, in which the Anopheles mosquitoes transmit malaria via their mature gametocytes to other human hosts [94]. The exact mechanism of action of artemisinin is uncertain; however, some research suggests that ART-based antimalarials are highly reactive in the presence of hemoglobin-derived heme or ferrous iron, suggesting that artemisinin activity depends on parasite uptake and hemoglobin hydrolysis [95]. Falcipain cysteine proteases play a critical role in hemoglobin degradation, and research by Klonis et al. [95] demonstrated that inhibiting cysteine proteases effectively counteracted artemisinin action, confirming the importance of hemoglobin degradation in the mechanism of action of ART [95]. Due to the emergence of ART-resistant (ARTR) parasites, there has been considerable discussion about the possible correlation between the loss of ART potency and the alteration of hemoglobin in ARTR parasites. Mutations in hemoglobinase genes may disrupt hemoglobin digestion and heme detoxification, leading to ART resistance [96]. Alteration of the expression profile or catalytic activity of FP-2 via gene manipulation (knockout) or compound inhibition assays resulted in parasites with swollen food vacuoles due to the accumulation of undigested hemoglobin [74]. Intriguingly, these parasites showed decreased sensitivity to ART, establishing a link between hemoglobin digestion and ART activity [95,97]. A knockdown of FP-3 in an FP-2 knockout background generated an ART-resistant phenotype that manifested at the early ring stage, suggesting these proteases are involved in ART resistance [98,99]. As a result of the identification of mutations in FP-2, decreased hemoglobin degradation, and enlargement of the food vacuole, the hypothesis that altered FP-2 activity contributes to artemisinin resistance has been substantiated [95,97].

## 1.7 Mutations in Falcipains

Considering the attractive nature of the hemoglobin degradation pathway in antimalarial drug discovery, mutations in genes encoding hemoglobinases have emerged as a major contributor to drug resistance [97]. A nonsense mutation at codon 69 (S69stop) and missense mutations in FP-2 has been reported to confer parasites with reduced ART sensitivity, possibly stemming from altered enzyme efficiency and reduction in hemoglobin digestibility [97,100]. These mutations are K255R, N257E, T343P, and D345G (numbered according to the whole protein sequence). Additional mutations occurring in FP-2 and FP-3 have been reported in the *Pf* genomic resource database (PlasmoDB) [101,102]. Currently, the structural and dynamic impacts of these mutations have not been determined. Therefore, this study presents a novel

insight towards understanding these mutations and their probable effect on the catalytic function of the respective protein structures.

## 1.8 Background information on computational approaches used in this thesis

Besides affecting the structure and function of a protein, mutations also affect the binding affinity of small molecules [103]. In order to combat the emergence of drug resistance in diseases of global health importance, a better understanding of mutations is crucial [104]. Mutational analysis can also be used to design possible inhibitors that are effective against mutation-induced resistance [105]. The first part of this thesis analyses the effects of mutations on FP-2 and FP-3. Several computational approaches have been developed over the years for predicting the effect of missense mutations on protein stability and function [103,106]. These approaches include but are not limited to molecular dynamics (MD) simulation, comparative essential dynamics (CED), dynamic residue network (DRN) analysis, and residue contact map analysis. Various mutant structures associated with global health diseases have been studied using these approaches [107–119]. The second part of this thesis involves the identification of potential inhibitors against the mutant systems of FP-2 and FP-3. The ability to target disease-causing mutations and elucidate their mutational effects is vital when designing drugs [120]. Traditional drug discovery is time-consuming, energy-intensive, and financially burdensome, hence the introduction of computer-aided drug discovery (CADD) [121]. Structure-based drug design (SBDD) is a computational approach to identify or design potential inhibitors suitable for clinical testing based on the knowledge of three-dimensional (3D) structures (obtained either experimentally or by computational homology modeling) of biological targets [120,122]. The SBDD approach is widely used for virtual screening of large libraries of small molecules and includes methods such as molecular docking, molecular dynamics, and structure-based pharmacophore design [120,122].

### 1.8.1 Homology modeling

A protein is composed of one or more amino acid chains (known as polypeptides) linked by covalent peptide bonds, and the amino acid sequence and function of a protein are determined by its gene [103,123]. When the amino acids are assembled sequentially, their interactions form the secondary structure. This structure minimizes the energy within the protein and folds into a stable conformation, producing a 3D structure [103]. It is essential for a protein to maintain

a 3D structure to remain functionally active; thus, generating a 3D structure is critical to characterizing gene mutation-induced changes in function and stability. A useful and accurate computational technique for predicting protein structure is homology modeling [124]. This method involves the prediction of the 3D structure of a target protein from its amino acid sequence and a homologous (template) protein whose structure has been experimentally determined by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy [125,126]. The steps utilized in homology modeling include i) identification and selection of template, ii) sequence alignment and alignment correction, ii) model building, iv) loop modeling, v) side-chain modeling, vi) model optimization, and vii) model validation [124].

### 1.8.2 High throughput virtual screening

Drug discovery centers around the identification of lead compounds that show pharmacological activity against biological targets and the subsequent optimization of their pharmacological properties and potencies [127]. Consequently, the pharmaceutical industry has adopted high throughput virtual screening (HTVS), the experimental screening of huge libraries of chemicals, to identify new compounds against therapeutically relevant targets [127]. HTVS is a prominent strategy for drug discovery that employs computational algorithms to identify novel bioactive molecules [128]. HTVS is widely classified into two types: ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS) [128]. LBVS methods focus on identifying molecules sharing similar chemical and physical properties based on the premise that similar compounds can exert similar effects on a drug target [129]. These techniques typically ignore any data pertaining to the drug target and concentrate solely on the ligand [129]. SBVS methods, also known as receptor-based methods, require the availability of a 3D structure of the drug target obtained through X-ray crystallography, NMR spectroscopy, or homology modeling [128,129]. These methods entail the explicit molecular docking of each ligand into the target's binding site, generating a predicted binding mode and a measurement of the quality of the fit between the compound and the binding site of the target [129]. Based on this information, ligands that strongly bind to the target protein are separated from those that do not [129]. With the availability of more and more 3D structures of target proteins, in addition to the fact that receptor-based approaches tend to produce more reliable and accurate results, receptor-based approaches have gained prominence over ligand-based techniques [129]. The following are the major steps of SBVS: (i) preparation of the compound library and target protein, (ii) determination of the most favourable binding position, and (iii) ranking of

the docked complexes [128]. SBVS method will be used in this thesis to identify potential antimalarial compounds. Molecular docking is one of the most commonly used SBVS methods that mimic the interactions between a ligand and a protein at the atomic level [130]. This allows the characterization of ligand conformation at the binding site of the appropriate protein [122]. Molecular docking algorithms perform quantitative predictions binding energies and ranks the docked compounds based on the binding affinity of ligand-protein complexes [122].

### 1.8.3 Molecular dynamics simulation

MD simulation is a computational tool used extensively to study the conformational dynamics of protein structures, which are key to understanding their properties and functions [131]. MD simulation is based on a general model of the physics governing interatomic interactions that describes the movement of every atom in a protein system or other molecular system moves over time [132,133]. The use of MD simulation has proven useful in predicting the response of biomolecules to perturbations such as mutations and ligand binding [133]. The availability of 3D protein structures determined experimentally or obtained via homology modeling is a starting point for MD simulations [134]. In MD simulation, the protein motions and forces acting on the atom are based on the classical Newton's laws of motion (equation 1.1), where $F_i$ is the force acting on atom $i$, $m_i$ is the mass of atom $i$, $r_i$ position vector of atom $i$, and t represents the time [135]. The forces between interacting atoms (bonded and non-bonded) and their potential energies are calculated using a model known as a molecular mechanical force field [133,136]. Force fields are comprised of a mathematical formula that describes the energy of a protein in relation to its atomic coordinates [137]. The widely used biomolecular force fields include Assisted Model Building and Energy Refinement (AMBER96 [138] and AMBER99SB-ILDN [139]), Chemistry at HARvard Macromolecular Mechanics (CHARMM) [140], GROningen Molecular Simulation (GROMOS) [141], and Optimized Potential for Liquid Simulations (OPLS) [142]. After obtaining the forces acting on individual atoms, Newton's law of motion is applied to calculate accelerations and velocities as well as update atom positions [143].

$$\boldsymbol{F}_i = m_i \frac{d^2\,\boldsymbol{r}_i(t)}{dt^2} \quad (1.1)$$

**Equation 1.1**: Newton's Law of Motion

Molecular dynamics simulations require many calculations, which are typically performed on supercomputers or computer clusters with hundreds of processors. In this study, all MD calculations were calculated on the Center for High Performance Computing (CHPC) cluster in Cape Town, South Africa. There are many MD simulation software packages with the same names as their default force fields, including AMBER [144], CHARMM [145], GROMACS [146] and NAMD [147,148].

MD simulations utilize a series of analyses to measure global and local protein conformations [149]. These include the root mean square deviation (RMSD), the root mean square fluctuation (RMSF), and the radius of gyration (Rg). RMSD calculates the variation in the backbone atoms of a protein structure from its initial structural conformational to its final position [150]. The deviations produced during a simulation can be used to determine how stable a protein is compared to its conformation, so the lower the deviation, the more stable the protein [150]. RMSF is a measure of individual residue flexibility. By measuring how much a particular residue fluctuates during the simulation, RMSF can reveal the contribution of each amino acid to the molecular motion of a protein structure [151]. Rg measures the compactness of a protein structure [152].

### 1.8.4 Comparative essential dynamics

Protein structures undergo conformational changes induced by perturbation [153]. These changes vary from the conformational shifting of loops and side chains in the catalytic or allosteric binding pocket to partial or complete folding of the part or whole protein structure [154,155]. The overall motion of protein systems obtained from MD simulations exists as a large data set. This includes atomic coordinates, dihedral angles, and the dynamic properties of the protein system [156]. Comparative essential dynamics analysis (CED) employs principal component analysis (PCA) to reduce the dimensionality of the large data obtained during molecular dynamics simulation and to identify the dominant motions in protein systems [157]. This method is useful in comparing conformational changes of different MD trajectories. In this study, the conformational changes associated with mutations in our protein are of interest. Using covariance-matrix-based mathematical techniques, PCA extracts the principal element in data (the dominant motion in protein systems in this case) from the atomic coordinates describing the proteins' accessible degrees of freedom (DOF). [157,158]. A set of eigenvectors and eigenvalues are obtained by diagonalizing the original Cartesian covariance matrix that contains positional atomic fluctuations about the average structure in all three coordinate axes

(x,y,z) [159]. The eigenvalues represent the total amount of variance, and the eigenvectors represent the direction of the overall motions [159,160]. The first principal component (PC1) is the eigenvector with the greatest variability in the dataset, and each subsequent eigenvectors or PC*n* (n = 2, 3, 4, …., n), in turn, has the highest variance if it is perpendicular to the previous PC [161].

### 1.8.5 Dynamic Residue Network Analysis

MD simulations offer multi-layered data that can be analyzed in more depth than post-MD analysis [162]. Hence, we proposed the use of DRN analysis, which is a computation tool developed by the Research Unit in Bioinformatics (RUBi) [163]. DRN analysis has emerged as an integrated computational tool for understanding the mutation-induced effects on protein structure and function [108,115,163]. In this approach, each residue C$\beta$ (C$\alpha$ for Gly) is represented as a node, and the inter-node juxtaposition depicts the edges that compose a network graph. In this study, we applied five DRN metrics, namely: averaged *betweenness centrality* (*BC*), averaged *closeness centrality* (*CC*), averaged *degree centrality* (*DC*), averaged *eigenvector centrality* (*EC*), and averaged *Katz centrality* (*KC*).

#### 1.8.5.1 Averaged *betweenness centrality*

Averaged *betweenness centrality* metrics measure the amount of information flowing through a node [164]. It is thus defined as how often a node lies on the shortest path between other nodes in a protein network [165]. *BC* metric is computed using equation 1.2.

$$\overline{BC}(v) = \frac{1}{m}\sum_{i=1}^{m}\sum_{u=1}^{n-1}\frac{\delta(s_i, t_i|v_i)}{\delta(s_i,t_i)} \quad (1.2)$$

**Equation 1.2**: Calculation of averaged *BC*

*v* represents the entire set of nodes; *m* represents the number of frames. $\delta(s,t)$ represents the number of shortest paths linking nodes *s* and *t*. $\delta(s,t|v)$ represents the number of these paths passing through another node, *v*; and *i* represents the frame number [163].

#### 1.8.5.2 Averaged *closeness centrality*

According to averaged *CC* metric, each node is ranked by its proximity to all other nodes in the protein network [166]. The closeness centrality of a node is inversely proportional to the total distance from other nodes. *CC* metric is computed using equation 1.3.

$$\overline{CC}(v) = \frac{n-1}{m}\sum_{i=1}^{m}\sum_{u=1}^{n-1} d(v,u) \quad (1.3)$$

**Equation 1.3**: Calculation of averaged *CC*

where *d(v, u)* is the shortest path between nodes *v* and *u,* and *n* is the number of nodes in the graph [163].

#### 1.8.5.3Averaged *degree centrality*

Averaged *DC* metric assigns significance based on the number of connections a node has in the protein network, where a highly connected node engages in many interactions within the network [164]. *DC* metric is computed using equation 1.4.

$$\overline{DC}(k) = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1, j\neq i}^{n} A_{ijk} \quad (1.4)$$

**Equation 1.4**: Calculation of averaged *DC*

where *n* is the number of nodes and $A_{ijk}$ is the $jk^{th}$ adjacency for the $i^{th}$ frame [163].

#### 1.8.5.4 Averaged *eigenvector centrality*

Averaged EC metric measures the number of connections a node has as well as the degree of centrality of its connections [167]. *EC* metric is computed using equation 1.5.

$$A \cdot \overrightarrow{EC} = \lambda \cdot \overrightarrow{EC} \quad \text{(a)}$$

$$\overline{EC}(i) = \frac{1}{m}\sum_{k=1}^{m} EC_{ik} \quad \text{(b)} \qquad (1.5)$$

**Equation 1.5**: Calculation of averaged *EC*

Where (a) *EC* is the eigenvector, and lambda ($\lambda$) is the eigenvalue for the eigen decomposition of adjacency matrix A. In NetworkX, this is obtained by power iteration. (b) Averaged *EC* is computed for $i^{th}$ residue by computing the vector for each MD frame and averaging [163].

#### 1.8.5.5 Averaged *Katz centrality*

Averaged *KC* metric calculates the relative influence of individual nodes with a protein network by considering the centrality of its direct and indirect connections [168]. *KC* metric is computed using equation 1.6.

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

$$\overline{KC}(i) = \frac{1}{m} \sum_{k=1}^{m} KC_{ik} \quad \text{(b)} \qquad (1.6)$$

**Equation 1.6**: Calculation of averaged *KC*

*KC* is a modification of *EC* that uses a dampening coefficient and a constant in order to influence adjacency values [163].

### 1.8.6 Weighted residue contact map

In the protein 3D structure, residue-residue contacts are pairs of residues that are spatially close to one another [169]. A pair of residues are said to be in contact when the distance between their C$\beta$ atoms (C$\alpha$ for Glycine) is less than the cut-off distance [170]. In this study, we defined the cut-off distance as 6.7 Å. Protein structures are shaped by intra-molecular interactions between residues; thus, the prediction of the changes in residue contact frequencies can be useful in mutation analysis, drug designs, and other areas of research that study the dynamic properties of proteins [171,172]. Weighted residue contact maps implemented by MDM-TASK-web are useful tools in investigating local residue contact frequencies throughout MD simulations [163]. In our study, we implemented this analysis in comparing the change in residue contact frequencies associated with mutations.

## 1.9 Project Motivation and Aims

In the fight against malaria, the emergence of resistance to ART and its partner drugs compromises the successful treatment, prevention, and elimination of the disease [11]. Despite the ongoing research for new antimalarial drugs to counter drug-resistant *Plasmodium falciparum*, it is crucial to understand the parasite's resistance mechanism to existing antimalarial drugs. FP-2 and FP-3 are cysteine proteases responsible for degrading the host hemoglobin in the parasite's food vacuole to provide amino acids required for intraerythrocytic parasite development [50,97]. Alteration of the catalytic activities of FP-2 and FP-3 via gene manipulation studies has been reported to confer parasites with reduced ART sensitivity [97]. Hence, a priori understanding of the structural and dynamic impacts of mutations in FP-2 and FP-3 is paramount. The first aim of this thesis is to characterize the effects of missense mutations occurring in FP-2 and FP-3 using a variety of computational approaches and tools. From the *Pf* genome resource database (PlasmoDB) [102], we obtained a collection of missense

mutations occurring in FP-2 and FP-3. The effects of these mutations have not been established; hence, this study presents a novel understanding of the impact of these missense mutations on catalytic efficiency and possible allosteric mechanisms through which hemoglobin degradation function may be affected. FP-2 and FP-3 have proven to be attractive targets for the development of antimalarial drugs. The presence of mutations in both proteins makes it imperative to identify potential hits that can bypass resistance. The second aim of this thesis is the identification of potential modulators against the mutant systems of FP-2 and FP-3 via *in silico* approaches.

## 1.10 Research Objectives

The specific objectives include

PART 1: Mutation analysis

i. Generation of accurate 3D protein structures of the wildtype (WT) and mutant systems
ii. Analyses of the conformational changes associated with mutations using MD simulations and comparative essential dynamic analysis
iii. Investigation of the effects of the mutations on the communication networks on the WT and mutant systems of FP-2 and FP-3

PART 2: Potential hit identification

i. Identification of potential allosteric modulators against the WT and mutant systems of FP-2 and FP-3 via high throughput virtual screening
ii. Investigation of the stability of selected protein-ligand systems using MD simulation
iii. Characterization of the changes in the communication network of the selected protein-ligand complexes in WT and mutants.

# Chapter 2

## Structural, dynamic and network communication effects of missense mutations on the partial zymogen and the catalytic domain of Falcipain 2

Part of the work presented in this chapter has been published as highlighted below:

**Chiamaka Jessica Okeke**, Thommas Mutemi Musyoka, Olivier Sheik Amamuddy, Victor Barozi, Özlem Tastan Bishop. Allosteric pockets and dynamic residue network hubs of Falcipain 2 in mutations including those linked to artemisinin resistance. *Computational and Structural Biotechnology Journal*, 19, 5647–5666 (2021), DOI:10.1016/j.csbj.2021.10.011.

Authorship contribution: **C.J.O.** retrieved all data, performed homology modeling, molecular dynamics simulations, comparative essential dynamics, DRN analysis, residue contact map, writing, and editing the manuscript. T.M.M. identified potential allosteric pockets and evaluated the distal effects associated with mutations, formal analysis, writing, and manuscript editing. O.S.A. performed formal analysis on comparative essential dynamics, cryptic pocket identification, and manuscript editing. V.B. performed global top 5% DRN calculations. Ö.T.B.: Conceptualization, formal analysis, methodology, DRN analysis and writing related PyMOL scripts, writing, and manuscript editing.

Additional work in this chapter, which is not included in the published article, was solely performed by **C.J.O.** These include DRN analysis on the catalytic domain and all analyses on the partial zymogen complexes.

## 2.1 Introduction

The introduction of missense mutations on drug target proteins is a common strategy utilized by pathogens to evade therapeutic pressure [173–175]. These mutations may lead to structural and functional changes in proteins with drug-binding efficacy consequences [176–178]. Additionally, changes in genetic information may also induce gene expression mechanisms, especially those coding for carrier and transporter proteins, ultimately promoting physiological conditions that reduce drug susceptibility [179]. Historically, most previous research on drug resistance mechanisms has examined mutations occurring within the binding site environments

[180–185]. A direct consequence of such mutational events is the alteration of pocket physicochemical properties, including the size/volume, active site residue interactions and conformation dynamics crucial to drug binding. However, studies evaluating drug resistance mechanisms of pathogens due to missense mutations that are distally located from the active sites have increased significantly in the past few years [108,115,116,186–189]. These mutations have been found to confer resistance through alteration of the WT native conformational ensembles via subtle allosteric mechanisms [108,115,190].

Allostery is an intrinsic protein feature by which perturbations at a distal site arising from ligand binding or mutations induce changes in protein function, especially at the active site [191]. These distant signals rarely alter the overall structural conformation of a protein but cause changes in its functional sites' dynamics and thermodynamic properties. Similarly, mutations occurring in key drug targets have been found to regulate protein function leading to the loss of native properties and function [192]. The rising importance of the allosteric effects of mutations points to another uncharted land of opportunities in drug design and other biomedical applications.

### 2.1.1 Prodomain inhibitory effect of falcipains

FP-2 and other proteases are in charge of degrading hemoglobin during the host erythrocytic cycle, which is a crucial component of the malaria parasite's pathogenicity. [6,56,85]. There are two main domains of falcipains: the prodomain and the mature (catalytic) domain [79], and together they make the zymogen (inactive protein precursor). The prodomain is further subdivided into two domains: the N-terminus, which directs the prodomain to the food vacuole [178] and the C-terminus, which is necessary for inhibiting falcipains [193]. Our study is focused on the inhibitory part of the prodomain (C-terminal), and together with the mature (catalytic) domain, is herein termed a "partial zymogen complex". Two inhibitory motifs in the prodomain, ERFNIN and GNFD, cover the active site cleft of the catalytic domain and prevent falcipain from activating [83,193,194]. The ~50 kDa zymogens are transported through the endoplasmic reticulum/Golgi network to the food vacuole, where the acidic environment stimulates the dissociation of the prodomain and the release of the ~27 kDa active enzymes [68,84] (**Figure 2.1**). According to previous mutagenesis studies, residues located at the interface between pro- and mature domains contribute to the formation of salt bridges (R185-E221, E210-K403 in FP2 and R202-E238 in FP3) and hydrophobic interactions (F214, W449, W453 in FP2 and F231, W457, W461 in FP3) [80]. Due to the importance of these interactions

in dissociating the prodomain and their responsibility in auto-processing, they are considered "hotspot" interactions [195].

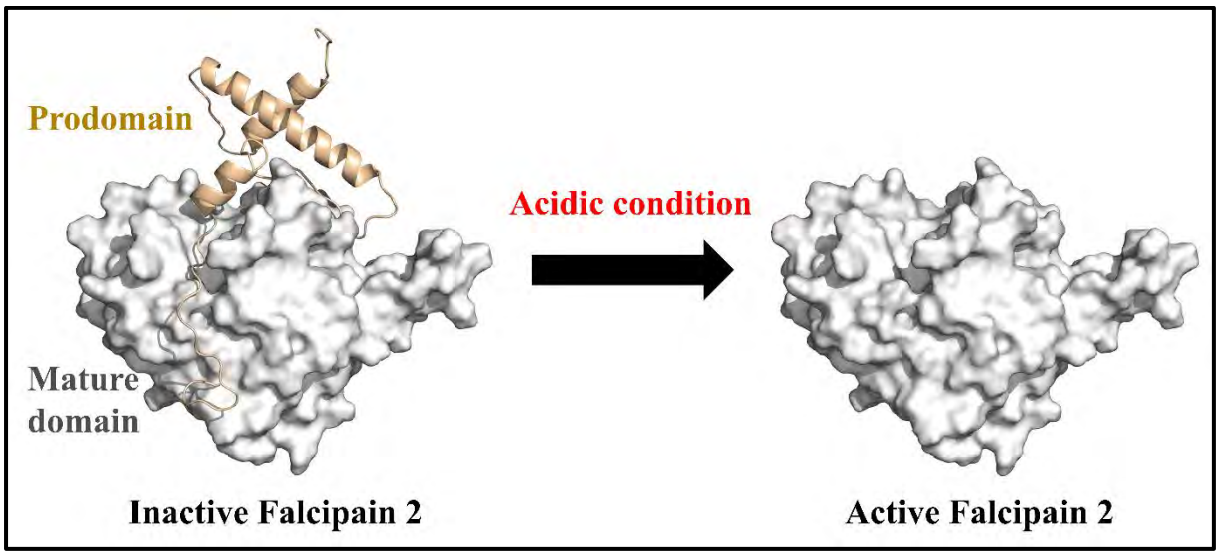



**Figure 2.1: The activation of partial zymogen complex of FP-2**. Prodomain is coloured wheat (cartoon), and the mature (catalytic) domain is grey (surface).

## 2.2 Proposed Work

From PlasmoDB version 9.3 [101,102], we identified missense mutations occurring in the partial zymogen complex and the catalytic domain of FP-2. Prior to reaching the acidic food vacuole, the inhibitory domain in the partial zymogen complex blocks the activity of the mature domain. We hypothesize that mutations in the partial zymogen complex could hinder the auto-processing and proteolytic maturation of the mature domain, hence the objectives of this study. In this study, 29 missense mutations located in the partial zymogen complex or the catalytic domain of FP-2, were evaluated to understand the effect on the structure and function of the protein. A key requirement for this structure-based study is the availability of reliable protein 3D structural information. Thus, accurate 3D homology models of WT and mutant systems were calculated using the homology modeling technique. The WT and mutant systems of the partial zymogen and catalytic domain of FP-2 were subjected to 500 ns all-atom MD simulation. Post-MD simulation, weighted contact map analysis, DRN calculations, and alanine scanning were performed to understand the dynamic properties of the mutant proteins and communication changes within the protein structures (**Figure 2.2**).

## 2.3 Methodology

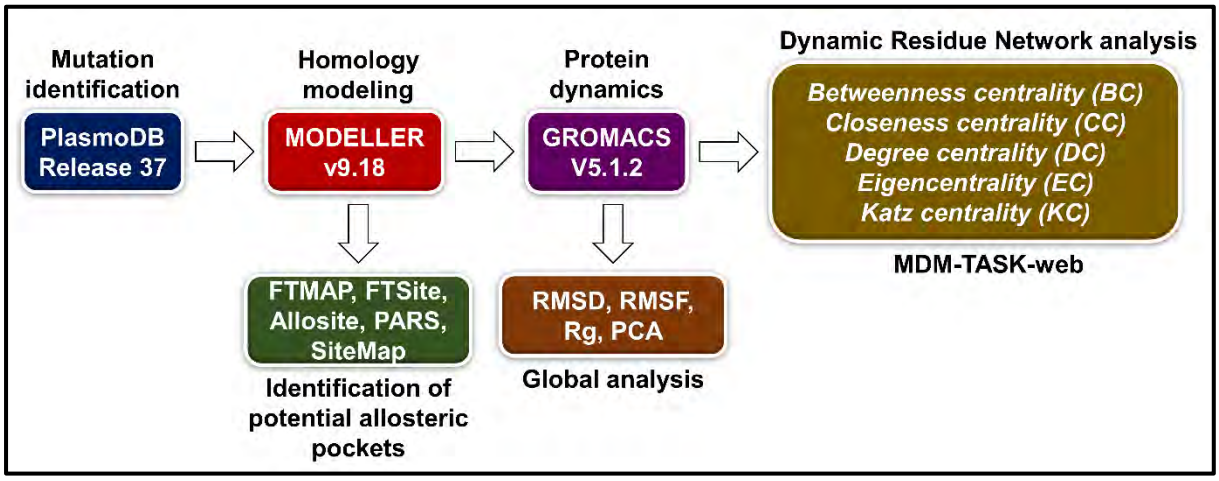



**Figure 2.2: A workflow of the methods and tools utilised in this chapter.**

### 2.3.1 Data retrieval and identification of missense mutations

The three-dimensional (3D) structure of the catalytic domain of FP-2 [PDB ID: 2OUL] [196] was retrieved from the RCSB Protein Data Bank (PDB) [197,198]. Using PyMOL [199], all crystallographic waters and bound ligand (chagasin) were removed. Although structural annotation of FP-2 indicated the absence of mutations, UniProt data (UniProt ID: Q8I6U4) [200] revealed the presence of four missense mutations (K255R, N257E, T343P, D345G) in the crystal structure. Thus, the protein sequence of FP-2 (accession number: PF3D7_1115700) was retrieved from PlasmoDB version 9.3 [101,102], and the 3D structure of FP-2 was accurately generated via homology modeling. The WT structure of the partial zymogen complex was modeled by Thommas Mutemi Musyoka. PlasmoDB version 9.3 [101,102] (release 37, accessed 21-Aug-2018), FP-2 missense mutations (including those linked to ART mutations) were identified (**Table 2.1**), and the corresponding 3D mutant structures (partial zymogen and catalytic domain) were generated.

**Table 2.1: List of mutations occurring in the partial zymogen and catalytic domain of FP-2**. In bold are the four ART conferring missense mutations. Mutation numbering is based on the entire protein sequence.

| Domain | Missense mutations |
|---|---|
| Partial zymogen | A159V, F165I, M167T, N197K, N204K, N224S, S228T |
| Catalytic domain | M245I, E248D, E249A, **K255R**, **N257E**, **T343P**, **D345G**, A353T, V393I, A400P, Q414E |

### 2.3.2 Homology modeling of FP-2 WT and mutant 3D structures

Homology modeling is an important computational tool to predict protein structure [124]. Due to the presence of mutations in the retrieved crystal structure, the 3D structure of the WT catalytic domain was re-modeled. PDB ID: 2OUL was identified as the best template as described by Musyoka et al. 2016 [87]. Homology modeling was done using MODELLER version 9.19 [201] using a very slow refinement. One hundred models were generated, and the normalised discrete protein energy (z-DOPE) score [202] was used to assess the quality of the models. The best model was selected based on the lowest z-DOPE score. Additional evaluation of the best model was performed using Verify3D [203], ProSA [204] and PROCHECK [205]. Verify3D evaluates the quality of a model by determining the compatibility of a protein model (3D) with its amino acid sequence (1D) [203]. It then generates an averaged 3D-1D score based on empirical and energetic methods. ProSA utilises a statistical-based approach where the quality of a model is evaluated by relegating a score and comparing it to the scores of relatable experimentally determined protein structures available in the PDB [204]. PROCHECK evaluates the model's stereochemical quality by assessing each residue's geometry and comparing it to the stereochemical parameters derived from well-refined, high-resolution, experimentally determined protein structure [205]. Overall, the best WT model was selected based on the consensus of these evaluation tools. The WT partial zymogen complex was modeled by Thommas Mutemi Musyoka [206]. Using the modeled WT structure as a template, mutant structures were predicted. The quality of the mutant structures was evaluated as described above. The protein interaction calculator (PIC) webserver was used to assess the presence of four disulphide bonds involved in the maintenance of the structural integrity in the modeled structures [207]. The protein structures (WT and mutants) were protonated to a pH of 5.5 to match the acidic environment of the food vacuole using the PROPKA tool from PDB2PQR (version 2.1.1) [208,209].

### 2.3.3 Identification of potential allosteric pockets

The identification of potential allosteric pockets was performed by Thommas M. Musyoka using a combination of structure-based approaches, including Allosite [210], FTMap [211], FTSite [211], Protein Allosteric and Regulatory Sites (PARS) [212], SiteMap [213], and Auto-Ligand [214]. Allosite predicts the location of allosteric sites in protein structures via algorithms such as pocket-based analysis and support vector machine (SVM) classifiers [210]. FTMAP predicts the location of allosteric sites by distributing small organic probe molecules

of different shapes, sizes, and polarity on the surface of a protein structure. This server determines the most suitable position for each probe, clusters the probes, and ranks the clusters according to their average energy. Consensus sites are termed regions that bind various probe clusters, and the site with the greatest quantity of clusters is referred to as the main hotspot. Secondary hotspots are all other consensus sites [211]. FTSite identifies and ranks binding sites based on consensus sites detected by the FTMap server. It ranks the consensus sites according to the number of non-bonded contacts between the protein structure and all probe clusters [211]. PARS employs the LIGSITE algorithm [215] and Normal Mode analysis to predict allosteric sites [212]. SiteMap predicts the location of allosteric sites using an algorithm similar to Goodford's GRID algorithm [216] while employing its definition of hydrophobicity. It selects site points based on energetic and geometric properties, which are compiled into sets to define the sites [213]. Auto-Ligand utilises a grid-based representation to predict potential binding affinity. It computes the potential affinity grid around a protein structure and uses a flood-fill and site-optimisation process to identify the best contiguous region within the energy grid [214].

### 2.3.4 Molecular dynamics simulation

To understand the effect of mutations on the partial zymogen and catalytic domain of FP-2, 500 nanoseconds (ns) all-atom MD simulations were conducted using the GROMACS 5.1.2 package [217] (WT and 11 mutations in the catalytic domain; WT and 18 mutations in the partial zymogen complex). We initially determined the most suitable force field parameters by performing triplicate test MD runs on the FP-2 WT using three different force fields; AMBER96 [138], AMBER99SB-ILDN [139], and CHARMM36 [140]. Subsequently, AMBER99SB-ILDN [139], the force field of choice, was employed to generate WT and corresponding mutant topology files. MD simulations were performed under periodic boundary conditions and solvated using the TIP3P (Transferable Intermolecular Potential with 3 Points) water model [218] in a triclinic box with a 1.75 nm distance between the protein structures and the box edges. Consequently, 0.15M of NaCl was added to neutralise the system charges. Energy minimisation was achieved using the steepest descent algorithm (emstep) with a preliminary energy step (nsteps) of 0.01nm without constraints until a tolerance limit of < 1000 kj $mol^{-1}$ $nm^{-1}$ was obtained. Once the system had converged, a two-step canonical ensemble equilibration (each lasting 100 ps) was used to ensure that the solvated system had reached the proper temperature and pressure. The Berendsen thermostat [219] was used to set the

temperature to 300 K (NVT - constant number of particles, volume, and temperature). The Parrinello-Rahman barostat [220] was then used to perform pressure equilibration at 1 atm (NPT - constant number of particles, pressure, and temperature). The equilibrated systems were subjected to production MD for 500 ns with a two fs integration step (femtoseconds). All bonds were constrained using the LINCS algorithm [221] during the production and equilibration phases. The Particle-mesh Ewald (PME) algorithm [222] was used to calculate long-range electrostatics with a Fourier grid spacing of 0.16 nm. A non-bonded cut-off distance of 1.4 nm was applied for the Coulomb and van der Waals interactions. After the production phase, the GROMACS built-in *gmx trjconv* command was used to remove periodic boundary conditions from each system's trajectory. Each system's global and local structural changes were calculated using the following packages: *gmx rms* to estimate the root mean square deviation (RMSD), *gmx rmsf* to estimate the root mean square fluctuation (RMSF), and *gmx gyrate* to estimate the radius of gyration (Rg). The results were analysed in RStudio [223] and Python libraries like Matplotlib [224], Numpy [225], Pandas [226], and Seaborn [227]. The conformational variability of each system throughout the simulation was analysed by calculating their corresponding all *versus* all RMSD (pairwise RMSD) distribution using *pytraj* [228]. Pairwise RMSD distributions highlight the degree of structural diversity of a particular ensemble, in this case, a protein system [229]. This may be due to inherent flexibility, mutation, or ligand binding.

### 2.3.5 Comparative essential dynamics

Essential dynamics aims to identify atoms' collective and dominant motions, thereby revealing the mechanisms underlying atomic fluctuations [230]. To assess the distribution of pocket conformational sampling across MD simulations for the WT and mutant systems of the partial zymogen and catalytic domain, we performed comparative essential dynamics using the *compare_essential_dynamics.py* script in MDM-TASK-web [163]. Essentially, the comparative essential dynamics approach filters the data and aligns all MD frames (using C$\alpha$ atoms) to a common reference before proceeding to a single decomposition. In this case, we aligned the WT and 11 mutant systems of the catalytic domain of FP-2 to a single frame from the WT simulation before specifying the region of interest, which is used to compute and decompose the covariance matrix. This was also applied to the WT and 18 mutant systems of the partial zymogen complexes. The regions of interest were the active site residues and the predicted pocket residues, and this method was independently applied to these regions. By

default, three C-terminal residues from each trajectory are excluded before structural alignment. The distribution of active sites or predicted pocket conformation was then depicted as scatterplots from the first principal component (PC1) and second principal component (PC2) axes. The percentages of total variance explained by each PC are displayed along the axes. In addition, we visualised the lowest-energy conformation of the total protein obtained from the maxima of the 2D nuclear density estimates generated from the two principal components to detect significant conformational changes.

### 2.3.6 Dynamic residue network analysis

DRN analysis was performed to investigate the influence of missense mutations on the flow of information within the protein residue network utilising the newly established MDM-TASK-web server [163]. The concept of DRN is based on the mathematical graph theory where residues in a protein structure are represented as nodes (C$\beta$ and C$\alpha$ for Glycine), and interconnecting residues are represented as edges based on a cut-off distance of 6.7 Å and a step size of 50 frames. In this chapter, five DRN metrics were calculated over the equilibrated 30 ns portion for each trajectory. These include *betweenness centrality* (*BC*), *closeness centrality* (*CC*), *degree centrality* (*DC*), *eigenvector centrality* (*EC*) and *Katz centrality* (*KC*). The equations used to calculate these metrics are described in Chapter 1, section 1.8.4.

### 2.3.7 Determination of centrality hubs

The concept of "*centrality hub*" was defined in our previous study as any node included in the set of nodes with the highest centrality for any given metric [119]. These *hubs* are defined as the global top 5% centrality nodes calculated across all metrics. The global 5% centrality hub was calculated by concatenating the centrality values of each protein and sorting them in ascending order to determine the residues with the highest centrality values. This approach was applied for each centrality metric. The centrality hubs were analysed and represented as heat maps using RStudio [223] and Python libraries such as Matplotlib [224] and Seaborn [227]. In order to understand the functionality of the identified hubs, we have introduced the following terminology: (1) *Persistent hub*: a hub that remains throughout the system. (2) *Distinctive hub*: Hub absent in WT but is present in mutant systems. The five centrality metrics represent different measures of importance within a protein network; thus, *persistent hubs* discussed here will be specific to a single centrality metric and will not be used across metrics.

## 2.4 Results and Discussion

In this study, our objective was to elucidate the structural and dynamic effects of mutations in the partial zymogen and catalytic domains of FP-2 using a combination of computational approaches.

### 2.4.1 Identification of missense mutations and generation of three-dimensional (3D) WT and mutant structures of FP-2

From PlasmoDB (release 37, accessed 21-Aug-2018), a total of 29 mutations (including those linked to ART resistance) distributed across the catalytic domain (**Figure 2.3**) and partial zymogen (**Figure 2.4**) were identified (**Table 2.1**). Seven of these mutations were located in the promoter region, whereas 11 were in the catalytic domain. A previous residue-residue interaction analysis established the importance of interactions between the partial promoter region α-helices and the catalytic domain in maintaining the integrity of the prodomain section, which is crucial for its zymogen regulatory role [206]. These interactions include *pi-pi* (F165-F168) and hydrogen bonds (S228 – G326, and S228 – N416). Additionally, hydrogen bonds between A400 and N212 (located within the ERFNIN motif) and M245 – L242 are important in anchoring the promoter region. Thus, mutations at positions F165, S228, M245, and A400, may disrupt these native contacts and may lead to a compromised structural fold and stability of the promoter region. All mutations except F165I, M167T, and N197K (prodomain promoter region) are located in the loop regions. An interesting observation unique to FP-2 is the presence of a cluster of five mutations (M245I, E248D, E249A, K255R, and N257E) within the nose region. The implications of these mutations on the catalytic efficiency and functional role of FP-2 are yet to be reported. However, a recent study by Siddiqui and colleagues identified four mutations, K255R, N257E (located in the nose region), T343P and D345G, linked to K13 protein mutations (F446I and P574L) which confer parasites with reduced artemisinin sensitivity. Although V393I and A400P were the only mutations located within the active pocket environment (S1' subsite) that could disrupt important native contacts with the hemoglobin substrate, there is no evidence that they are associated with ART resistance.

#### 2.4.1.1 Catalytic domain

Deciphering mutational effects is critical in biomedical research. The first step is the accurate prediction of 3D protein structures. Although FP-2 has several crystallised structures in the PDB, the 3D structure was generated via homology modeling due to the presence of four

missense mutations in the crystal structure. Using MODELLER version 9.19 and a very slow refinement, 100 models for the WT and 11 mutant systems, respectively, were created. The best model with the lowest normalised Discrete Optimized Protein Energy (z-DOPE) score for each system was selected and subjected to further validation using VERIFY3D, ProSA, and PROCHECK. The results from each validation tool are outlined in **Table S1**. The selected structures were then protonated as described in the methodology section and used for further catalytic domain analysis.

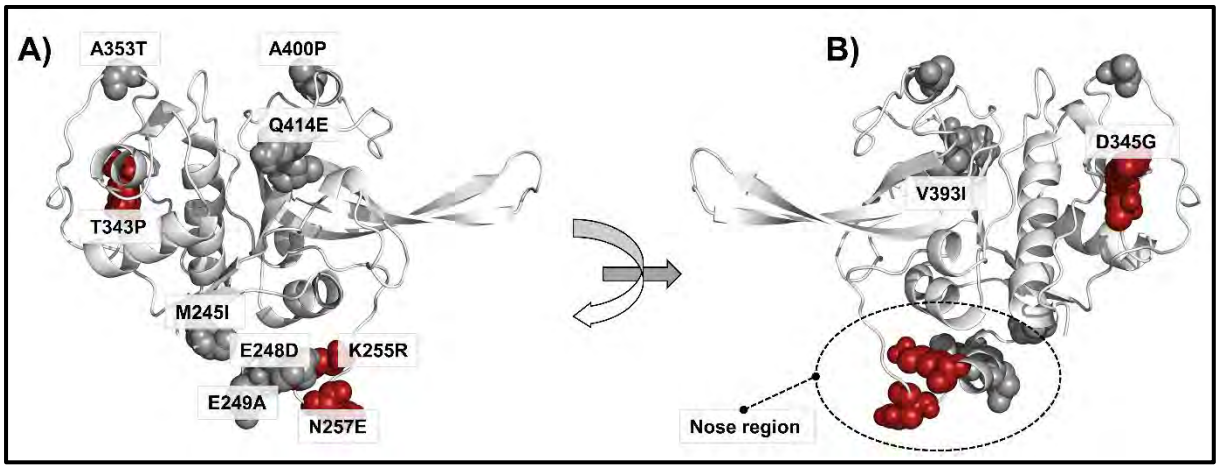



**Figure 2.3: Structural representation of the mutations occurring in the catalytic domain of FP-2**. Firebrick spheres and the remaining mutations by grey spheres represent the ART-linked mutations.

### 2.4.1.2 Partial zymogen complex

The WT of the partial zymogen complex of FP-2 was modeled by Thommas Mutemi Musyoka [206]. A total of 18 mutant structures were built using MODELLER version 9.19. The z-DOPE score was used to validate the quality of the mutant structures. Quality assessment of the top models was performed using VERIFY3D, ProSA, and PROCHECK. The results from each validation tool are outlined in **Table S2**. The selected structures were then protonated as described in the methodology section and used for further partial zymogen complex analysis.

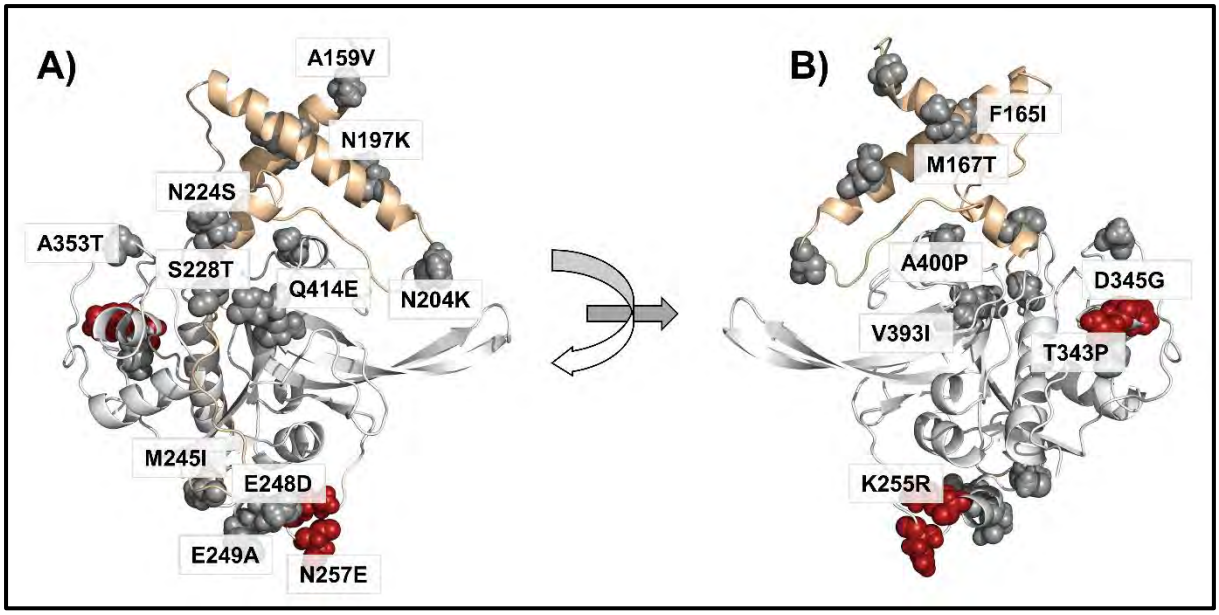



**Figure 2.4: Structural representation of the mutations occurring in the partial zymogen complex of FP-2**. The promoter region is colored in wheat, and the catalytic domain is white. Firebrick spheres and the remaining mutations by grey spheres represent the ART-linked mutations.

### 2.4.2 Identification of potential allosteric pocket

Integration of results from different tools revealed six potential allosteric pockets (named Pockets 1 – 6) (**Figure 2.5**). Pocket 1 is located around the nose region loop, a structural characteristic unique to FP-2 and related homologs from other Plasmodium parasite species. Intriguingly, this pocket contains two ART-resistant linked missense mutations, K255R and N257E, and three other missense mutations (M245I, E248D, E249A) near the pocket. Pockets 2 and 4 constitute residues in highly dynamic loops adjacent to the S1 and S1′ subsites. Pocket 2 contained the missense mutation A535T, whereas A400P bordered pocket 4. Pocket 3 is located at the base of the L-domain. AutoLigand only identified Pockets 5 and 6 as one pocket, which we separated into two. They were made up of two small grooves. The interface of Pocket 3 and 6 shared two other mutations (T343P and D345G) associated with ART resistance (**Table 2.2**). Based on the druggability analysis performed with SiteMap and AutoLigand (**Table S3**), all pockets, with the exception of Pocket 1, were determined to be less likely to bind small compounds (D-score < 0.6).

**Table 2.2: Residues in potential allosteric pockets** (pockets 1 – 6) in FP-2 based on the whole protein sequence.

| Pockets | Residues |
|---|---|
| 1 | Y247, E248, I251, K252, K255, G256, N257, E258, N259, F260, H262, D376, N377, K378, L379, K380, E381, A382, Y443, E465, S466, G467, L468 |
| 2 | Q311, L313, V314, D315, C316, S317, F318, K319, N320, Y321, GLY322, C323, N324, D334, V350, S351, D352, A353, P354, N355, L356, C357, N358, I359 |
| 3 | Y298, K302, N303, K304, I306, D334, M335, I336, E337, LEU338, G339, G340, I341, T343, T363, E364, K365, Y366, G367, I368, K369, N370, L482, I483, E484 |
| 4 | D398, F401, Y402, K403, E404, G405, I406, F407, D408, G409, E410 |
| 5 | S271, G272, V273, T274, A299, N303, L305 |
| 6 | P275, V276, K277, D278, K280, T307, S309, E310, D344, D345, P348, Y349, G450, Q451 |

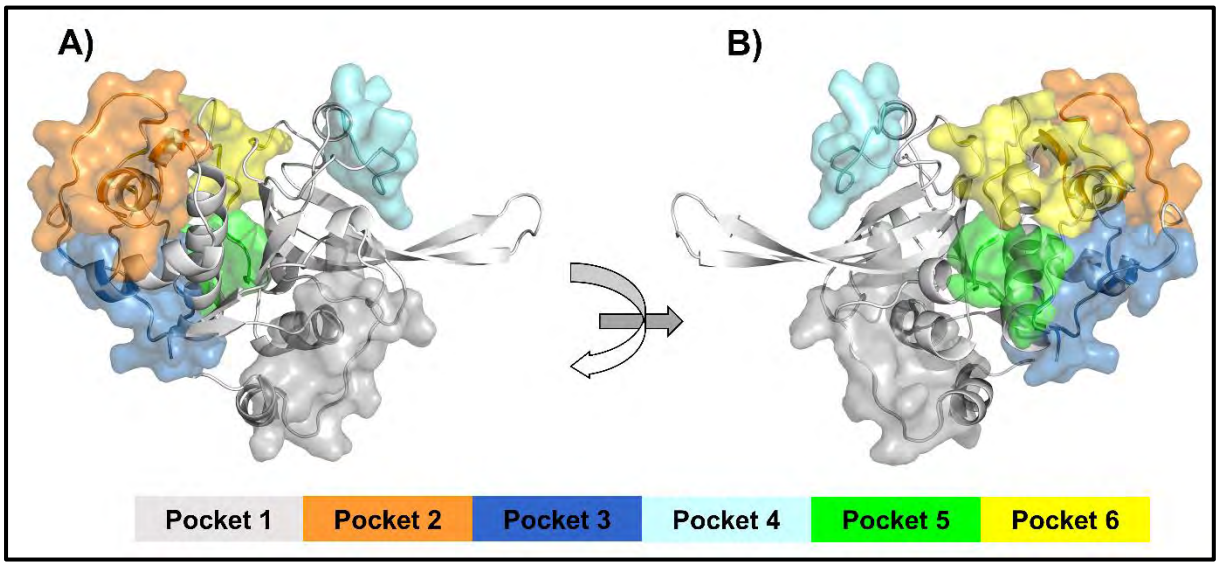



**Figure 2.5: Structural representation of the potential allosteric pockets in FP-2**. Identified pockets are shown in the surface presentation and colored accordingly.

### 2.4.3 Mutation-induced structural dynamic changes identified using MD simulations

To effectively analyse the properties of a biomolecular system using MD simulations, it is crucial to determine the most suitable force field parameters. As a preliminary step, we evaluated three force fields *viz*. AMBER96 [138], AMBER99SB-ILDN [139], and CHARMM36 [140] in a triplicate run per force field for the partial zymogen and catalytic

domain WT system. According to the backbone RMSD analysis, AMBER99SB-ILDN was the most appropriate force filed in describing the properties of FP-2, which are composed of ~ 50% loop regions (**Figure 2.6**). Consequently, the dynamic changes of WT FP-2 and mutant proteins (partial zymogen complex and catalytic domain) were assessed over 500 ns MD simulations. The cleaned MD trajectories for all systems (devoid of PBC, solvent and ions) were visualised using VMD [231] to evaluate any significant structural conformational changes. Overall post-MD analysis was performed using RMSD (whole protein, central invariant core, and the active site residues) and Rg (whole protein, central invariant core, and the various potential allosteric pocket residues) to determine the protein stability and conformational changes of the mutant proteins compared to the respective WT system. The calculation results are discussed below.

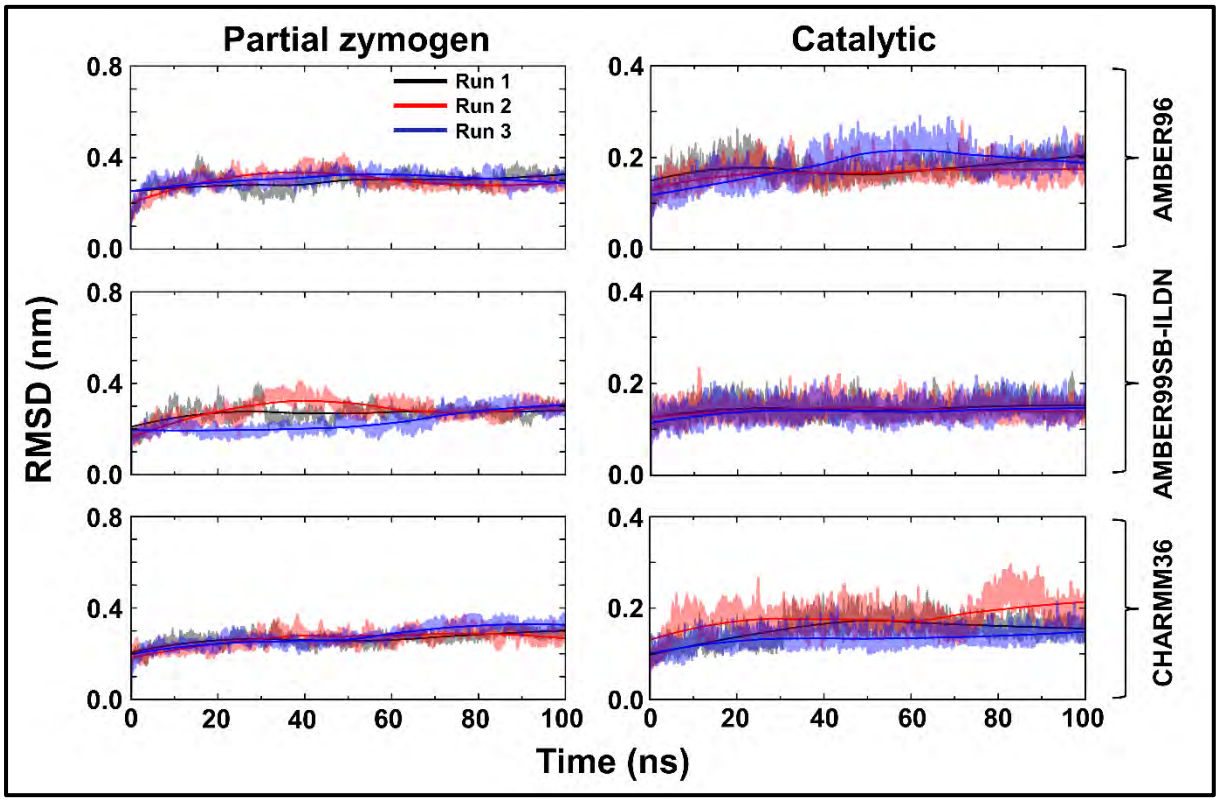



**Figure 2.6: Identification of the most accurate force field for the partial zymogen complex and the catalytic domain of FP-2.** The backbone RMSD analyses of the three force fields over the triplicate MD simulations are shown in line plots. Run 1 is coloured in black, run 2 in red and run 3 in blue.

### 2.4.3.1 Catalytic domain

Understanding the dynamic changes in a protein structure using MD simulation has become a crucial approach to deciphering atomistic details for molecular recognition as well as protein function under various conditions [232]. The stability of a system relative to the initial structure

(in this case, the WT system) can be evaluated by the deviation patterns generated during MD simulations where higher RMSD values indicate destabilisation [233]. Using backbone RMSD, the global conformational variability of the mutant systems was evaluated, and a comparison with the WT structure was made. For that, we used various RMSD analysis techniques.

Initially, the most commonly used RMSD *versus* timeline plots for evaluating the positional divergence of a structure to the starting configuration over the course of the simulation were determined (**Figure S2.1**). The majority of mutant systems displayed convergence after the first few nanoseconds. All RMSD values converge within ~0.4 nm, indicating that the structural integrity of the various protein structures was maintained. To determine the conformational variability throughout the simulations, we calculated the RMSD distribution violin plots for the whole protein, central invariant core, and the active site pockets, excluding the first 20 ns of the equilibration time (**Figure 2.7 A-C**). Inspection of the whole protein revealed that the majority of the protein systems displayed unimodal distribution, suggesting that the systems sampled a single conformational equilibrium during the simulation. The exceptions were mutant systems M245I, E249A, V393I, and A400P, which showed bimodal distribution. In the central invariant core, which consists of $\alpha$-helices and $\beta$-sheets, all mutant systems exhibited a unimodal distribution except for A400P. To determine if mutations caused conformational changes within the trench-like binding pocket, the RMSD of residues forming the different subsites (S1, S2, S3, and S1'), including the catalytic triad residues, were determined. The active site RMSD results revealed that both E248D and V393I demonstrated bimodal distribution, whereas the remaining mutant systems exhibited a unimodal distribution. V393I is located within the active site, which could explain the different conformational sampling.

Further analysis to compare how the starting conformation differs from that of each frame, all *versus* all frame (pairwise RMSD) plots were computed using *pytraj* [228]. Pairwise RMSD is an effective method for comparing and contrasting the conformations along a trajectory for various protein structures. Similar to the classical RMSD *versus* timeline plots, the starting structure of most systems showed significant variation compared to other frames along the simulation (**Figure 2.7D**). Once more, M245I and A400P showed prominent bimodal behaviour. We attributed the bimodal behaviour of M245I to the presence of a cryptic pocket explained in **chapters 2.4.5 and 2.4.6.**

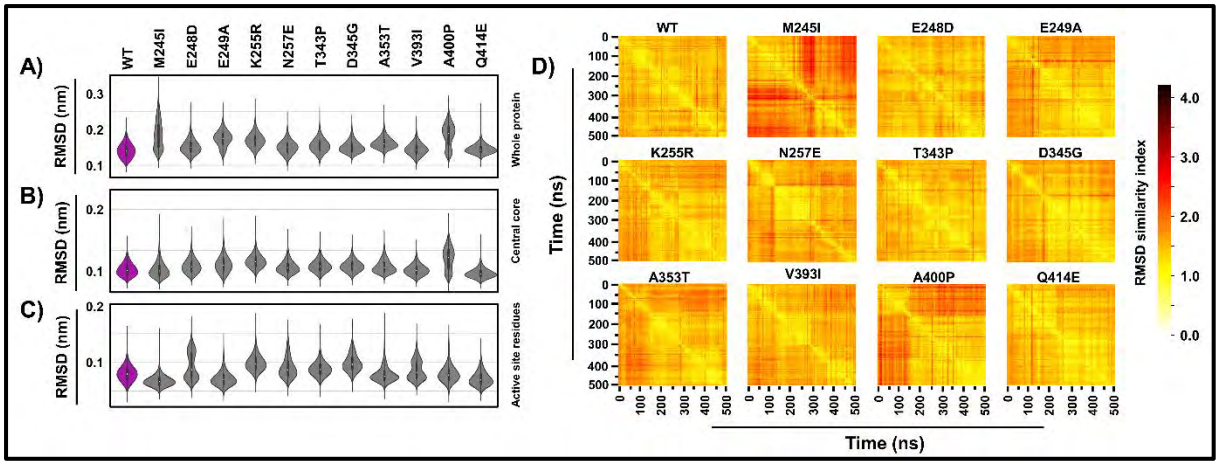



**Figure 2.7: Backbone RMSD analysis of the WT and mutant systems of the catalytic domain of FP-2.** RMSD violin distribution plots of A) whole protein, B) central invariant core, and C) binding active site residues of the WT and mutant protein systems. WT ensemble plots are shown in purple. D) All vs all C$\alpha$ RMSD of the WT and mutant protein systems. The x and y-axes represent time (ns). The color scale indicates the degree of conformational variation between frames (most similar = white, different = dark). This figure is produced with permission from Okeke et al. 2021 [118].

### 2.4.3.2 Partial zymogen complex

To determine the dynamical properties of the partial zymogen complex systems, violin distribution plots of the whole systems, central invariant core, and the binding pocket residues were determined (**Figure 2.8A-C**). In contrast to results from the catalytic domain, a distinct change in the conformational variability was observed in the whole protein of the partial zymogen complexes (**Figure 2.8A**). The WT and some mutant systems (F165I, M167T, N197K, E249A, K255R, A353T, V393I, A400P, and Q414E) displayed unimodal violin distribution, an indication that the systems visited a single conformational space during the MD simulation. The other mutant systems exhibited bimodal distribution. An explanation for this is the presence of the highly dynamic prodomain segment, which impedes the active site in both proteins while mediating its inhibitory function of the partial zymogen. The violin distribution RMSD profiles of the invariant central core were computed to ascertain the dynamical characteristics of the systems in the absence of loops (**Figure 2.8B**). This core region, composed primarily of secondary structural elements, exhibited a unimodal distribution profile in all systems. This indicates that the high dynamics of partial zymogen complexes strongly influenced the resulting overall conformational properties. The active site residues in WT and mutant systems exhibited unimodal distribution, with F165I, N257E and V393I exhibiting slight differences (**Figure 2.8C**). F165I is located in the $\alpha$1 helix, which is known to form important interactions that stabilize and maintain the structural integrity of the

prodomain. N257E is located in the nose region and is part of the artemisinin-associated mutations. V393I is a part of subsite 1' (S1') residues, forming the active sites in FP-2. Classical RMSD *vs* timeline plots for the WT and mutant systems are represented in **Figure S2.2**.

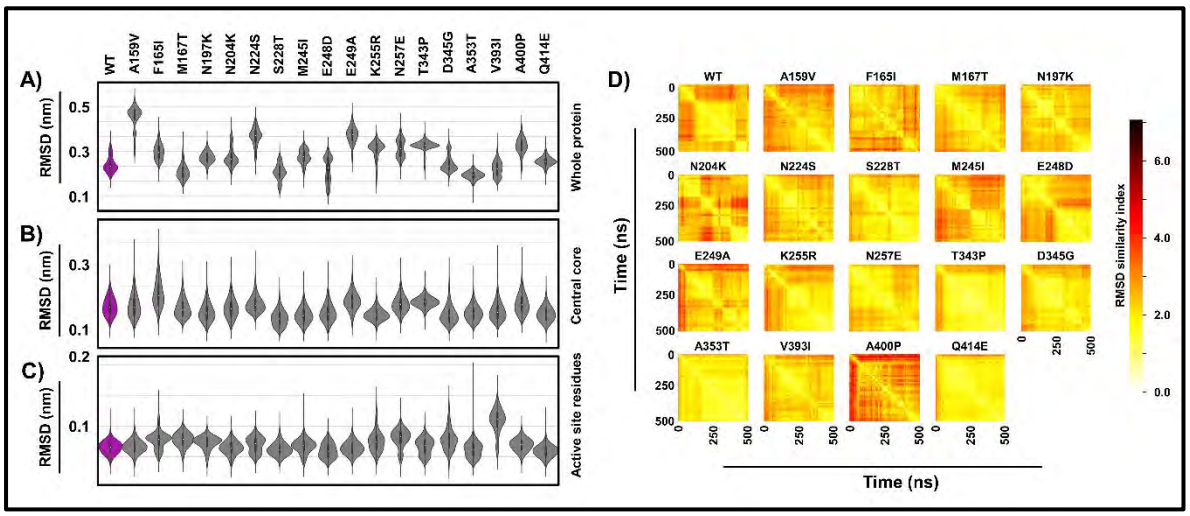



**Figure 2.8: Backbone RMSD analysis of the WT and mutant systems of the partial zymogen complex of FP-2.** RMSD violin distribution plots of the A) whole protein, B) central invariant core, and C) binding active site residues of the WT and mutant protein systems. WT ensemble plots are shown in purple. D) All vs all C$\alpha$ RMSD of the WT and mutant protein systems. The × and y-axes represent time (ns). The color scale indicates the degree of conformational variation between frames (most similar = white, different = dark).

These findings suggest that subtle conformational changes caused by mutations in the partial zymogen and catalytic domains of FP-2 exist. Interestingly, the central invariant core of the WT and mutant systems in the partial zymogen and catalytic domain of FP-2 remained intact throughout the stimulation, suggesting that the effect of mutations on the flexible regions of FP-2 was responsible for the different conformational samplings.

### 2.4.4 Local per-residue flexibility analysis using RMSF

Using the C$\alpha$ RMSF, we calculated per-residue flexibility profiles for different systems and identified significant mutation-related changes. At first, regions of high flexibility in both the partial zymogen complex and the catalytic domain of FP-2 were identified. As expected, the highly flexible loop regions in the partial zymogen complex and catalytic domain exhibited the highest RMSF scores. In the partial zymogen complex, the prodomain loop segment occluding the binding pocket (residues 232 – 246), the nose region (residues 256 – 260), and the $\beta$-arm hairpin (residues 430 – 436) exhibited high flexibility across all systems (**Figure 2.9A**).

Additionally, residues 348 – 359 exhibited high fluctuations in only mutant systems A159V, M245I, and E249A. Interestingly, these residues form residues in pockets 2 and 6. Besides the nose and $\beta$-arm hairpin regions of the catalytic domain, loops around the S2 subsite also exhibited high fluctuations in the absence of the prodomain segment (**Figure 2.9B**). This was probably caused by the lack of stabilising interactions between the prodomain and catalytic domain, previously identified by Musyoka et al. [206]. In M245I, residues 348 – 359 displayed high fluctuations, and these residues form pockets 2 and 6. In the subsequent section of this chapter, we establish the relationship between the high fluctuations in these regions and a cryptic pocket found in these mutant systems.

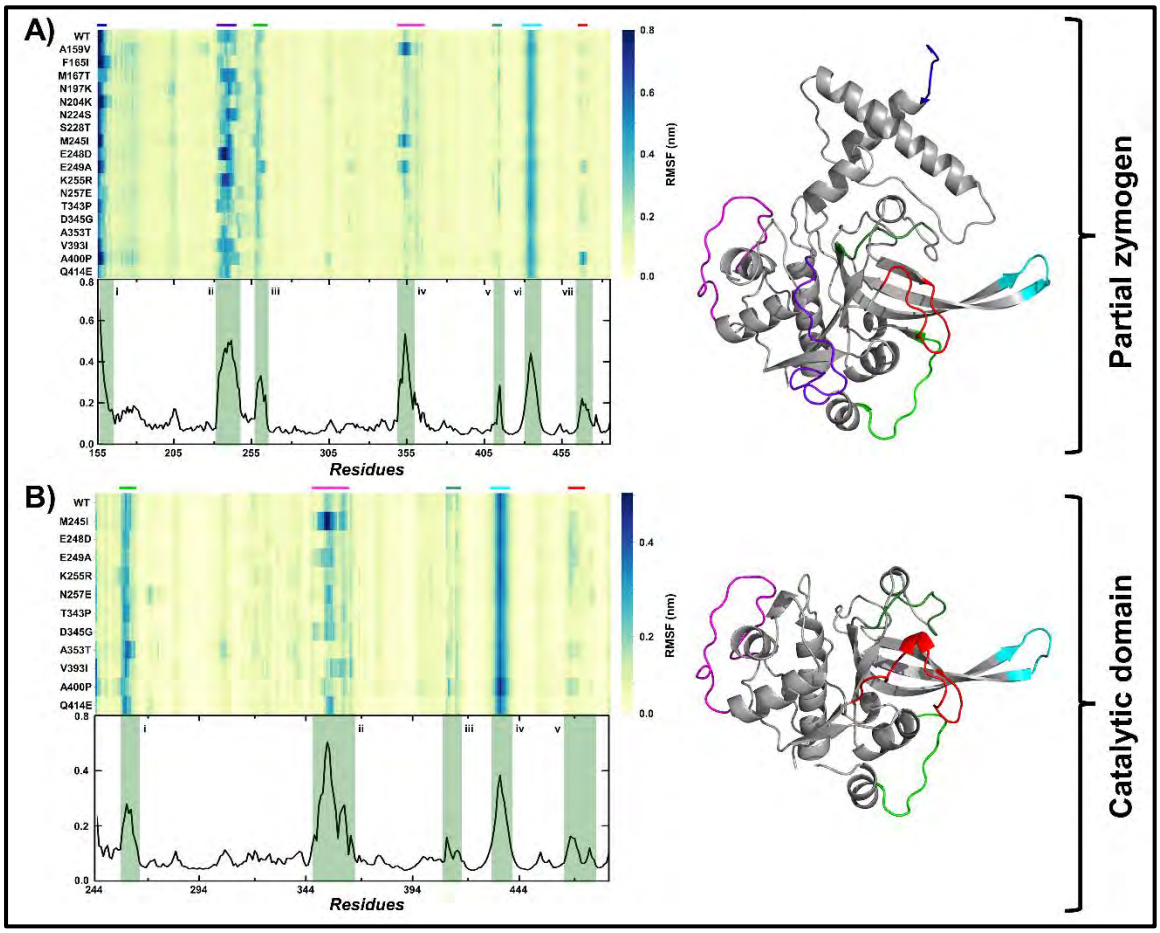



**Figure 2.9: Heatmaps showing the per-residue raw RMSF values across all systems.** RMSF values of the WT and mutant systems of A) the partial zymogen complex and B) the catalytic domain of FP-2. Protein loops displaying increased RMSF, as shown using different color bars, and their positions are shown in the 3D structures.

### 2.4.5 Mutation-linked changes in the putative allosteric pockets were identified via Rg and RMSF

Using Rg calculations, we investigated the effects of mutations on the overall compactness of the entire protein and the potential binding pockets (pockets 1 – 6). Distribution of Rg values for the whole protein of the WT and mutant systems of the catalytic domain of FP-2 indicated that all systems had nearly similar Rg values throughout the simulations (**Figure S2.3B**), signifying the absence of unfolding conformational events resulting from the destabilisation effect of mutations. In the partial zymogen complex, all systems maintained a unimodal profile from the violin plots (**Figure S2.3A**). Regardless of the loopy regions within the protein, the globular nature and presence of two pairs of disulphide bonds may account for the perceived compactness. Mutant A159V in the partial zymogen complex displayed higher Rg values while maintaining a unimodal distribution. Additionally, we analysed the Rg and RMSF of the identified pockets to determine whether there is an association between mutations and changes in the pockets.

#### 2.4.5.1 Catalytic domain

As seen in **Figure 2.10**, putative allosteric pockets Rg exhibited more variability than the Rg of the whole protein. In pocket 1, WT, M245I, and N257E displayed bimodal behaviour (**Figure 2.10A**). A353T appeared to sample a wider conformational range while maintaining a flatter distribution. The heatmap of pocket 1 RMSF values revealed high fluctuations in residues G256 – F260, with mutant systems A353T and Q414E having the highest fluctuations in this region. Among all the pockets, pocket 2 presented the most varied Rg distributions. M245I, E249A, and N257E sampled two major conformations linked to increased loop flexibility from residue RMSF calculations (**Figure 2.10B**). In the M245I system, residues S351 – N355 demonstrated the highest flexibility compared to WT and the other mutant systems. Interestingly, according to RMSF data, in the presence of A353T mutation located in pocket 2, the pocket became the most rigid compared to the WT and other mutant systems.

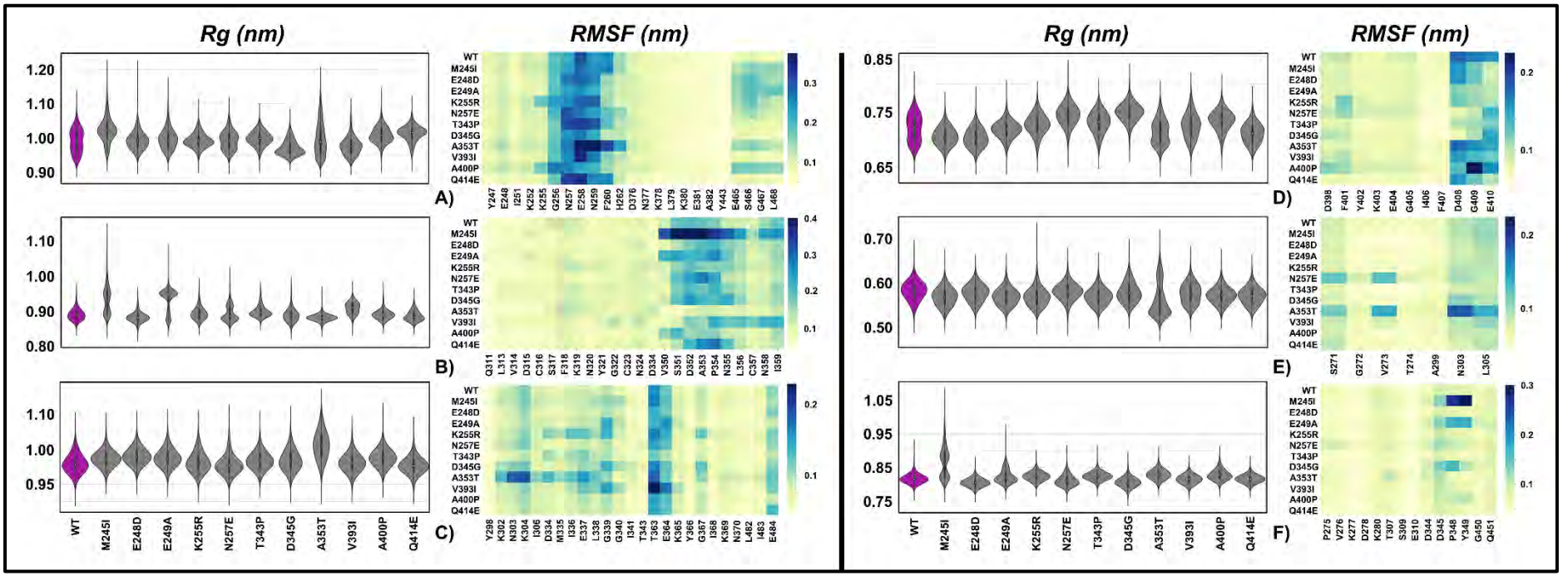



**Figure 2.10: Rg and RMSF values for Pockets 1 – 6 (A – F) of the WT and mutant systems of the catalytic domain of FP-2.** Rg and RMSF are represented as violin distribution plots and heatmaps, respectively. The purple ensemble in the violin plots represents the WT system. The heatmaps show pocket residues on the x-axis and WT and mutant systems on the y-axis. This figure is produced with permission from Okeke et al. 2021 [118].

In pocket 3, all the mutant systems showed Rg distributions similar to WT, except A353T, where a significantly higher mode was observed (**Figure 2.10C**). This observation may be linked to the relatively high flexibility of K302 – K304 loop residues, which caused it to be located further from the central core helix. For pocket 4, in contrast to the WT and A353T mutant system, the majority of mutant systems adopted a single conformation, and overall, the pocket residues showed similar flexibility (**Figure 2.10D**). In pocket 5, only A353T showed significant differences in conformational distribution (**Figure 2.10E**). Finally, in pocket 6, only M245I demonstrated a distinct conformational distribution compared to other systems (**Figure 2.10F**). These results will be linked to section 2.4.6.1.

#### 2.4.5.2 Partial zymogen complex

In pocket 1, mutant systems E249A, K255R, A353T, and A400P demonstrated bimodal behaviour compared to the WT and other mutant systems (**Figure 2.11A**). A400P displayed the highest conformational variability, which could be attributed to the high flexibility of residues E465 – L468. High RMSF flexibility in residues N259 – H262 and E465 – L468 accounts for the conformational diversity of E249A. Pocket 2 demonstrated the most variation among the other pockets. M245I and E249A sampled two major conformational distributions (**Figure 2.11B**). Interestingly, these mutations showed similar results in the catalytic domain, suggesting that these mutations exerted the same effect on the partial zymogen complex and the catalytic domain. RMSF results revealed high fluctuations in residues V350 – C357, which could be attributed to the conformational sampling observed in the presence of these mutations. We also observed similar high fluctuations in these regions in A159A. Although A159V maintained a unimodal conformation, a notable higher mode was observed. All mutant systems in pocket 3 exhibited a similar Rg distribution to the WT, with the exception of F165I, N204K, and A400P, which showed a marginally higher mode (**Figure 2.11C**). These results were supported by the high flexibility in residues K302 – K304. In Pocket 4, all mutants exhibited a similar Rg distribution to the WT, with A159V, N197K, K255R, and A400P displaying a higher Rg distribution (**Figure 2.11D**). Only F165I exhibited a slightly Rg distribution compared to other systems in pocket 5 (**Figure 2.11E**). Lastly, in pocket 6, A159V and M245I exhibited bimodal distributions, while the other systems exhibited unimodal distributions. RMSF data revealed high flexibility in residues D345, P348, and Y349. We also observed similar flexibility in these regions in mutant A159V and noted a higher unimodal Rg distribution (**Figure 2.11F**).

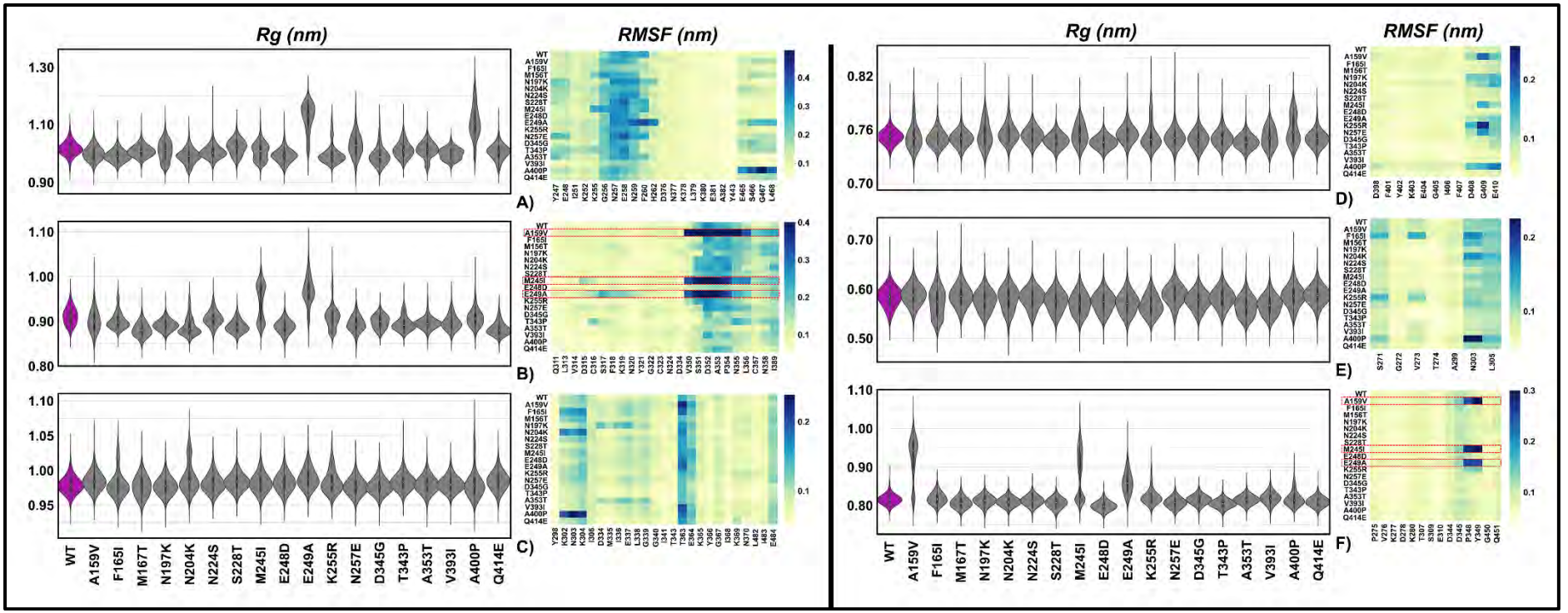



**Figure 2.11: Rg and RMSF values for Pockets 1 – 6 (A – F) of the WT and mutant systems of the partial zymogen complex of FP-2.** Rg and RMSF are represented as violin distribution plots and heat maps. The purple ensemble in the violin plots represents the WT system. The heatmaps show pocket residues on the x-axis and WT and mutant systems on the y-axis.

### 2.4.6 Essential dynamics revealed diverse conformational changes in the active site region and identified putative allosteric pockets

Comparative essential dynamics of WT and mutant trajectories were performed to investigate the effect of mutations on the conformational sampling of the binding sites and the identified putative allosteric pockets obtained during MD simulation. This study analysed the extreme projections of the first and second principal components (PC1 and PC2) responsible for the major protein motion contributions.

#### 2.4.6.1 Catalytic domain

**Figure S2.4** shows the dominant motions and the nature of conformational differences of the active site residues of the first and second principal modes of the WT and mutant systems. Comparative essential dynamics revealed that the position of the binding site residues of the WT and some mutations varied less over the MD simulation. The findings revealed that binding site residues were remarkably similar along the axis that explained the highest percentage of variance (71.09% along PC1), while minuscule differences found along PC2 (4.06%) explained the lower percentage of the total variance. This implies potential stability within the active site residues of these systems. E248D, A400P, and artemisinin-associated mutations K255R, N257E, T343P, and D345G moved significantly, thus accounting for most of the change in the respective protein systems. These results correlated with the RMSD results of the active site residues, as determined in **Figure 2.7C**.

Analysis of the comparative essential dynamics of the potential allosteric pockets showed greater variability compared to active site residues. In pocket 1, PC1 and PC2 accounted for 53.42% and 9.60%, respectively. In pocket 2, PC1 and PC2 accounted for 38.29% and 16.56%, respectively. In Pocket 3, PC1 and PC2 accounted for 75.32% and 4.19%, respectively. In pocket 4, PC1 and PC2 accounted for 30.53% and 25.87%, respectively. In pocket 5, PC1 and PC2 accounted for 83.12% and 3.57%, respectively. Lastly, in pocket 6, PC1 and PC2 accounted for 51.14% and 23.00%, respectively. The comparative essential dynamics results correlated with the Rg and RMSF analysis of the potential allosteric pockets revealing subtle changes in pockets 1, 3, 4, and 5 of the WT and mutant systems (**Figure S2.5**). Pockets 2 and 6 revealed more conformational diversity in the mutant systems compared to the WT (**Figure 2.12A and B**). In pocket 2, E248D, K255R, D345G, A400P, and Q414E varied at the beginning of the simulation but remained stable throughout the simulation. Visual inspection of these

trajectories revealed minimal conformational changes for this pocket. Interestingly, M245I sampled different conformations throughout the simulation leading us to visualise the trajectories.

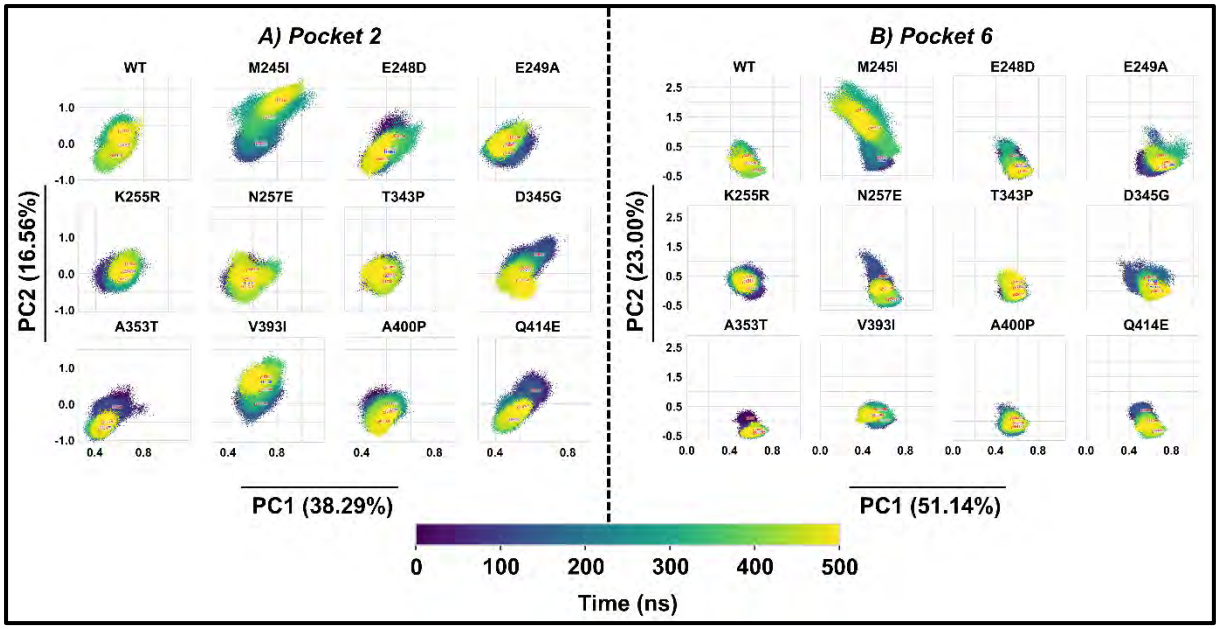



**Figure 2.12: Comparative essential dynamic analysis of pockets 2 and 6 of the WT and mutant systems of the catalytic domain of FP-2 along PC1 and PC2.** The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue. This figure is produced with permission from Okeke et al. 2021 [118].

Visualisation of the M245I trajectory revealed the conformational diversity of the pockets such that residues in pocket 2 moved into pocket 6. In Pocket 6, M245I, E249A, N257E, and D345G varied during the simulation. Visual inspection revealed that N257E and D345G displayed minimal conformational changes for this pocket. Upon inspection, Olivier Sheik Amamuddy discovered an unusual opening of pocket 6 in M245I where residues in pocket 2 moved into pocket 6 (**Video S1**) [118]. The absence of this opening in WT and other mutant systems suggests that this is a cryptic pocket. While visual inspection of E249A did not reveal the presence of the cryptic pocket observed in M245I, it revealed the opening of pocket 6 and the slight movement of residues in pocket 2 into pocket 6. WT and M245I pose extracted by a k-means algorithm at 448,180 ps showing the absence and presence of cryptic pocket in WT and M245I, respectively, are represented in **Figure 2.13A and B**. WT, and E249A poses extracted by a k-means algorithm at 447,196ps showing the opening of pocket 6 E249A is represented in **Figure 2.13C and D**.

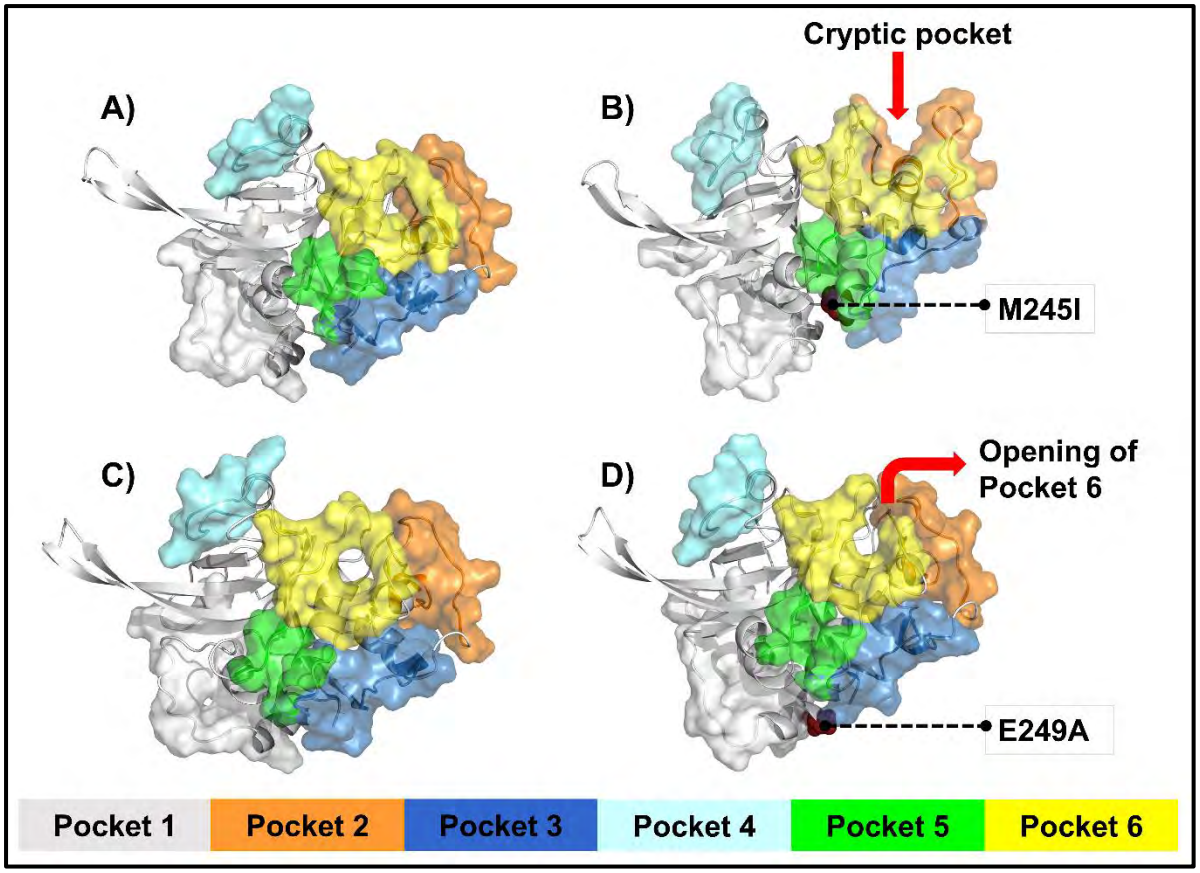



**Figure 2.13: Cartoon representation of the WT and mutant systems M245I and E249A extracted at various timestamps.** A) WT and B) M245I extracted at 448,180 ps**,** revealing the presence of the cryptic pocket in M245I and absence in the WT. C) WT and D) E249A extracted at 447,196 ps revealing the opening of pocket 6 and absence in the WT.

### 2.4.6.2 Partial zymogen complex

Comparative essential dynamics examines relationships between different protein structures to provide insight into the nature of mutation-induced conformational differences. Firstly, we analysed the conformational differences of the active site residues of the WT and mutant systems. The results showed that binding site residues were strikingly similar along the axis that explained the highest percentage of variance (82.36% along PC1), whereas negligible differences were found along PC2 (8.24%), which explained the lower percentage of the total variance (**Figure S2.6**). These results correlated with the RMSD results of the active site residues, as determined in **Figure 2.8C**, suggesting potential stability within these systems' active site residues. Next, we analysed the conformational differences of the different potential allosteric pocket residues of the WT and mutant systems.

In pocket 1, mutants E249A and A159V exhibited the highest divergence in conformational sampling along PC1, accounting for 58.57%, and PC2 accounting for 10.52% of the total

variance (**Figure S2.6A**). The differences between these mutants were consistent with the observed differences in compactness derived from Rg distribution and RMSF values in **Figure 2.11**. Visual inspection of these mutant's pockets revealed contrasting surface topologies compared to the WT.

In pocket 2, PC1 and PC2 accounted for 58.70% and 16.64% of the total variance, respectively. A159V, M245I, and E249A showed a more dispersed conformational distribution, while the WT and the other mutant systems displayed a more compact distribution (**Figure 2.14A**). Interestingly, in the catalytic domain of FP-2, M245I and E249A also displayed a more dispersed conformational distribution, suggesting that irrespective of the presence and absence of the prodomain, these mutations exerted the same effect on FP-2.

In pocket 3, PC1 and PC2 accounted for 92.30% and 2.06% of the total variance, respectively (**Figure S2.6B**). The distribution of the pocket conformations was consistently similar along both PC1 and PC2. A159V and A400P exhibited the highest divergence compared to the WT. Visual inspection of A159V and A400P trajectories revealed some topographical differences in the pockets compared to the WT.

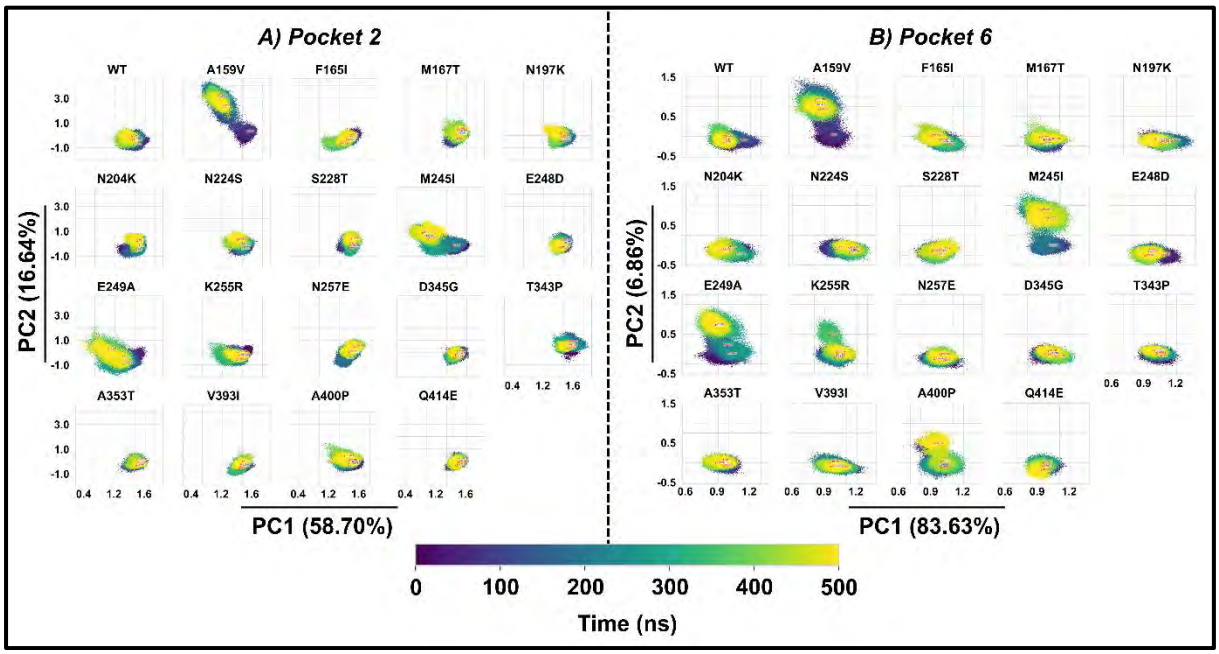



**Figure 2.14: Comparative essential dynamic analysis of pockets 2 and 6 of the WT and mutant systems of the partial zymogen complex of FP-2 along PC1 and PC2.** The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

In pocket 4, PC1 and PC2 accounted for 85.32% and 4.40% of the total variance, respectively. WT and all mutant systems displayed a compact distribution with the exception of A159V and A400P, which varied at the beginning of the simulation but remained stable throughout the simulation (**Figure S2.6C**).

In pocket 5, PC1 and PC2 accounted for 90.49% and 2.49% of the total variance, respectively. All systems maintained a compact distribution throughout the simulation (**Figure S2.6D**).

In pocket 6, PC1 and PC2 accounted for 83.63% and 6.86% of the total variance, respectively (**Figure 2.14B**). Mutant systems A159V, M245I, E249A, K255R, and A400P exhibited the highest range of conformational sampling. Visual inspection of these trajectories revealed that K255R and A400P displayed a slight opening of pocket 6 towards the end of the simulation. However, A159V, M245I, and E249A revealed the presence of the cryptic pocket where pocket 6 opened up, and residues in pocket 2 moved in, forming a single groove. WT and A159V poses extracted by a k-means algorithm at 497,370 ps showing the absence and presence of cryptic pocket in WT and A159V, respectively, represented in **Figure 2.15A and B**. WT, and M245I poses extracted by a k-means algorithm at 406,840 ps showing the absence and presence of cryptic pocket in WT and M245I respectively is represented in **Figure 2.15C and D**. WT, and E249A poses extracted by a k-means algorithm at 435,390 ps showing the absence and presence of cryptic pocket in WT and E249A respectively is represented in **Figure 2.15E and F**. Interestingly, in the catalytic domain of FP-2, the presence of a cryptic pocket involving pockets 2 and 6 was observed M245I, and a slight opening of pocket 6 in E249A suggesting that irrespective of the presence and absence of the prodomain, these mutations exerted the same effect on FP-2.

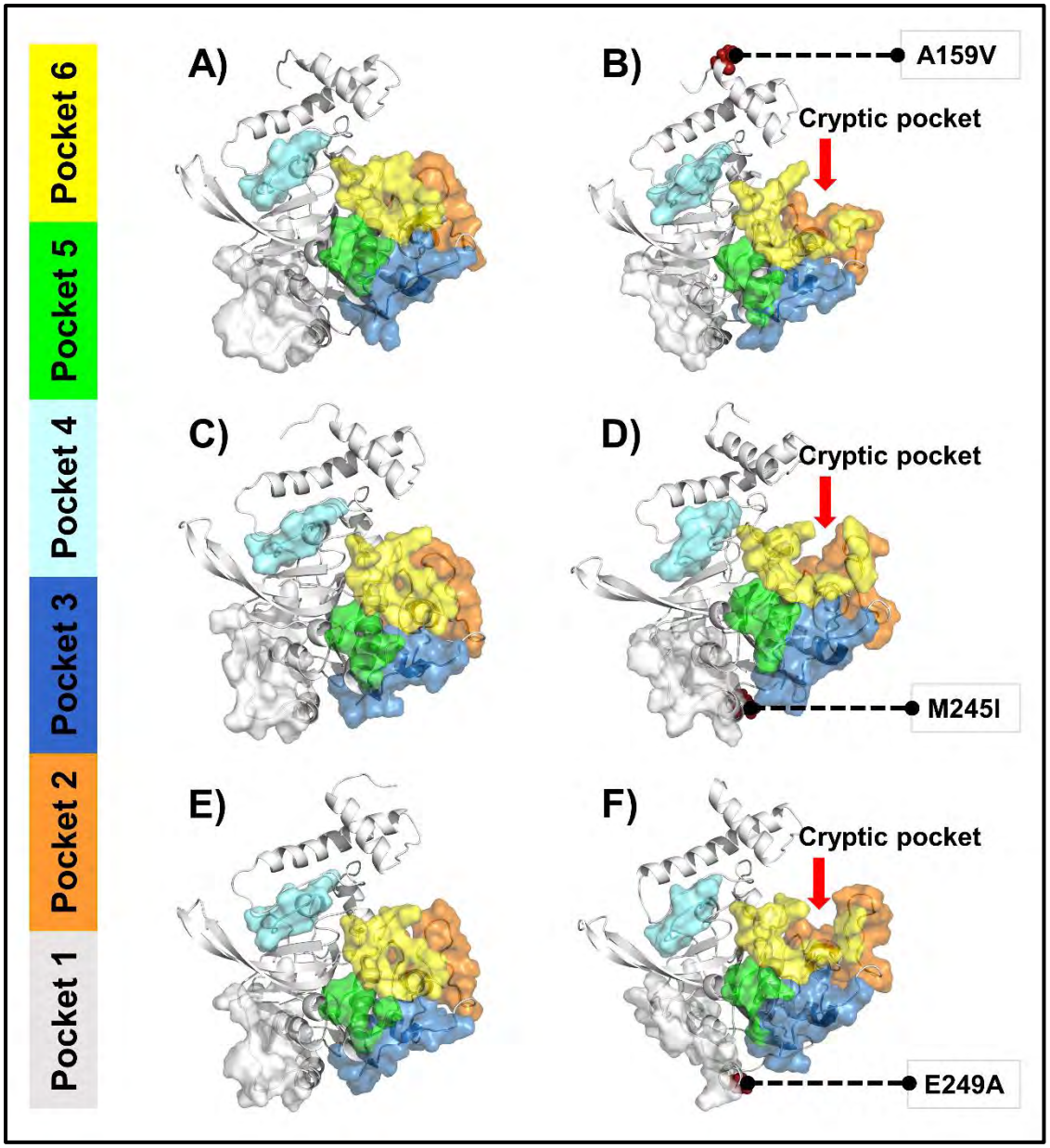



**Figure 2.15: Cartoon representation of the WT and mutant systems A159V, M245I, and E249A extracted at various time stamps.** A) WT and B) A159V extracted at 497,370 ps revealing the presence of the cryptic pocket in A159V and its absence in the WT. C) WT and D) M245I extracted at 406,840 ps, revealing the presence of the cryptic pocket in M245I and its absence in the WT. E) WT and F) E249A were extracted at 435,390 ps, revealing the presence of the cryptic pocket in E249A and its absence in the WT.

### 2.4.7 Dynamic residue network analysis reveals the communication pattern exerted by mutations

The function of protein systems is often related to the structural organisation of their amino acid residues and how the dynamics of these amino acid residues change upon perturbation [234]. Elucidating the interaction networks of residues in proteins using graph techniques has

become a key strategy in predicting signalling effects associated with perturbations on their backbone connectivity arising from mutations and ligand binding [107,112,113]. This approach allows a global analysis of all residue interactions within a protein network, allowing for the identification of key residues involved in the flow of information. Utilising this approach, subtle long-range communication effects critical for protein function that are not detected by conventional MD methods. The MD analysis revealed structural differences between the WT and mutant ensembles.

Consequently, additional analysis was performed to find alterations in residue communication brought on by mutations. In order to identify conserved dynamic residue network changes arising from mutations, key residues crucial for protein communication were calculated using five different network centrality metrics in both WT and mutant systems. The metrics included *BC*, *CC*, *DC*, *EC*, and *KC* centrality as implemented on the MDM-TASK-web [163].

#### 2.4.7.1 Catalytic domain

Heatmap representations of the averaged *BC*, *CC*, *DC*, *EC*, and *KC* of the WT and mutant systems of FP-2 are in **Figure 2.16**. From all the metrics used, the majority of the residues within the loop regions, including those in the nose region and $\beta$-hairpin in the WT and mutant systems, exhibited the lowest centrality values. Whereas residues within the $\alpha$-helices ($\alpha$5, $\alpha$6, $\alpha$7, and $\alpha$8) and $\beta$-sheets ($\beta$2, $\beta$3, $\beta$4, $\beta$5, and $\beta$6) which form the central invariant core and the trench-like binding pockets, exhibited high centrality values, an indication of the presence of increased interactions with other residues in the network (**Figure 2.16**). We examined different centrality distributions for each system to determine whether mutations caused changes in the overall distribution. Upon visual inspection, we observed some differences in the *BC* distribution between the WT and mutant systems, suggesting that the communication pattern were perturbed in the presence of mutations (**Figure 2.17A**). *CC* distributions were very similar in all systems, suggesting that the presence of mutations did not perturb *CC* distributions (**Figure 2.17B**). The *DC* distributions showed the most significant differences between WT and mutant systems among the centrality measures (**Figure 2.17C**). This indicates that the communication pattern was perturbed in the presence of mutations. *EC* distributions displayed differences between the WT and mutant systems (**Figure 2.17D**). Lastly, *KC* distribution showed differences between the WT and mutant systems (**Figure 2.17E**). In order to identify conserved dynamic residue network changes arising from mutations, key residues crucial for protein communication were ranked, and the top 5% of residues were marked as important

residues for protein communication (**Figure 2.18**). From the various metrics, several residues were persistent in the WT and mutant systems. This include *BC*: F288, S293, I389, M420, Y441, and I480. *CC*: S289, S290, G292, S293, P388, I389, S390, I391, V419, M420, and L421.

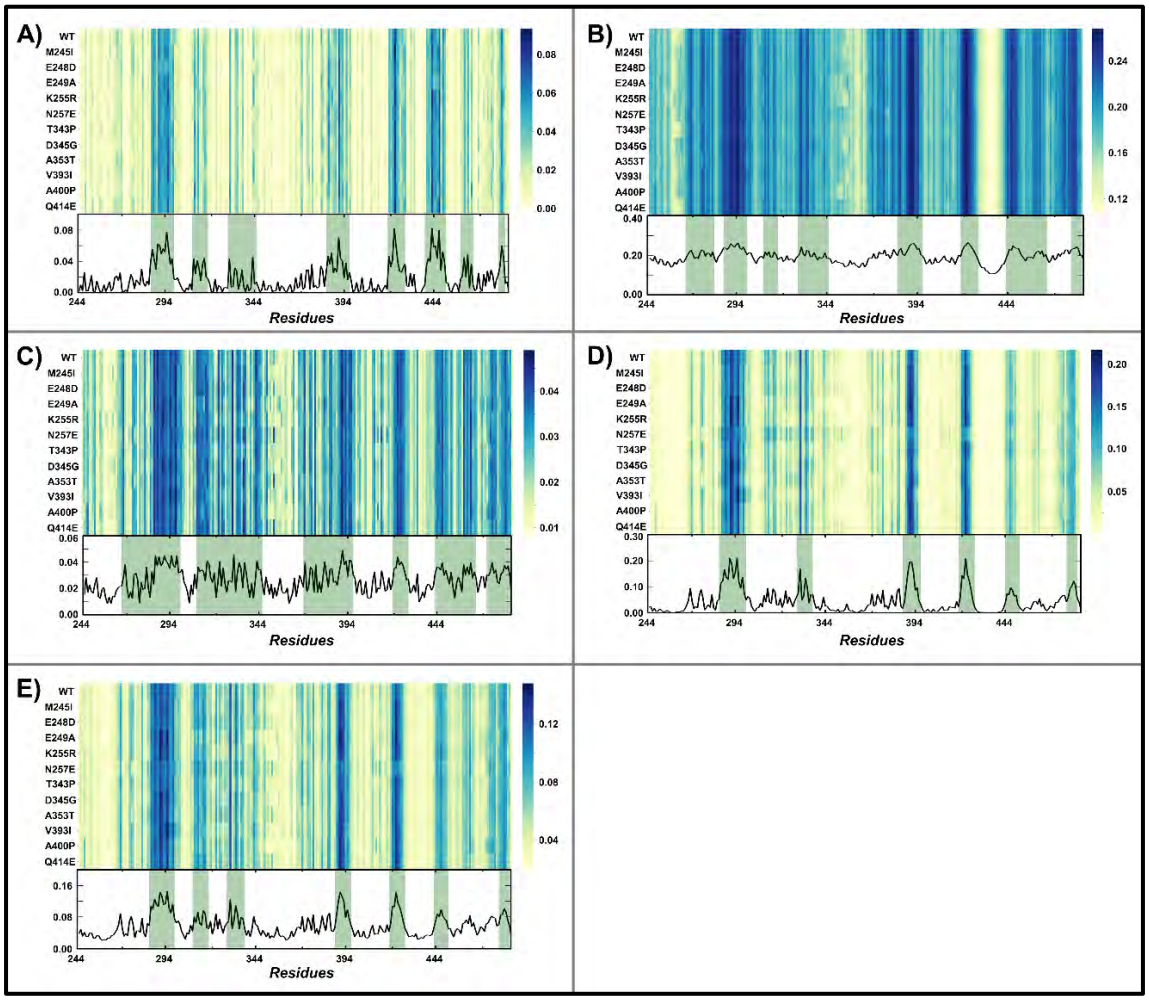



**Figure 2.16: Heatmaps showing per-residue centrality scores of the WT and mutant systems of the catalytic domain of FP-2.** A) Averaged *Betweenness centrality* (*BC*), B) Averaged *Closeness centrality* (*CC*), C) Averaged *Degree centrality* (*DC*), D) Averaged *Eigenvector centrality* (*EC*), and E) Averaged *Katz centrality* (*KC*).

*DC*: I328, and I389. *EC*: W286, S289, S290, S293, I328, I389, and S390. *KC*: W286, S289, S290, S293, I328, and I389. From *BC* analysis, the usage of the catalytic thiol (C285) increased in the majority of the mutants with the exception of K255R, N257E and the WT (**Figure 2.18A**). This was observed with S289, which is located at the central $\alpha$-helix. The usage of residues L421 (located at the antiparallel $\beta$-sheet) and W447 (located in the S1' subsite) in communication was lost in the majority of the mutant systems. From *CC* analysis, the majority of the top 5% residues retained their usage in communication as with the WT (**Figure 2.18B**). Using the *DC* metric, the usage of residue N320 (located in the S3 subsite) and A394 (located

in the S1' subsite) was lost in nearly all mutant systems (**Figure 2.18C**). However, the majority of the mutant systems had increased usage of residues D334, M420, I445, and I461. From the *EC* results, the high centrality of residues A287 and I291 in WT was lost in nearly all mutant systems (**Figure 2.18D**). From *KC* analysis, the usage of residues A287 and I291 was lost in the majority of the mutant system (**Figure 2.18E**). This analysis highlights the value of combining different centrality metrics when evaluating the potential consequences of mutations. Özlem Tastan Bishop identified the allosteric residue communication path in M245I, K255R, T343P, A353T, and V393I, originating from distinct average *BC* hubs [118].

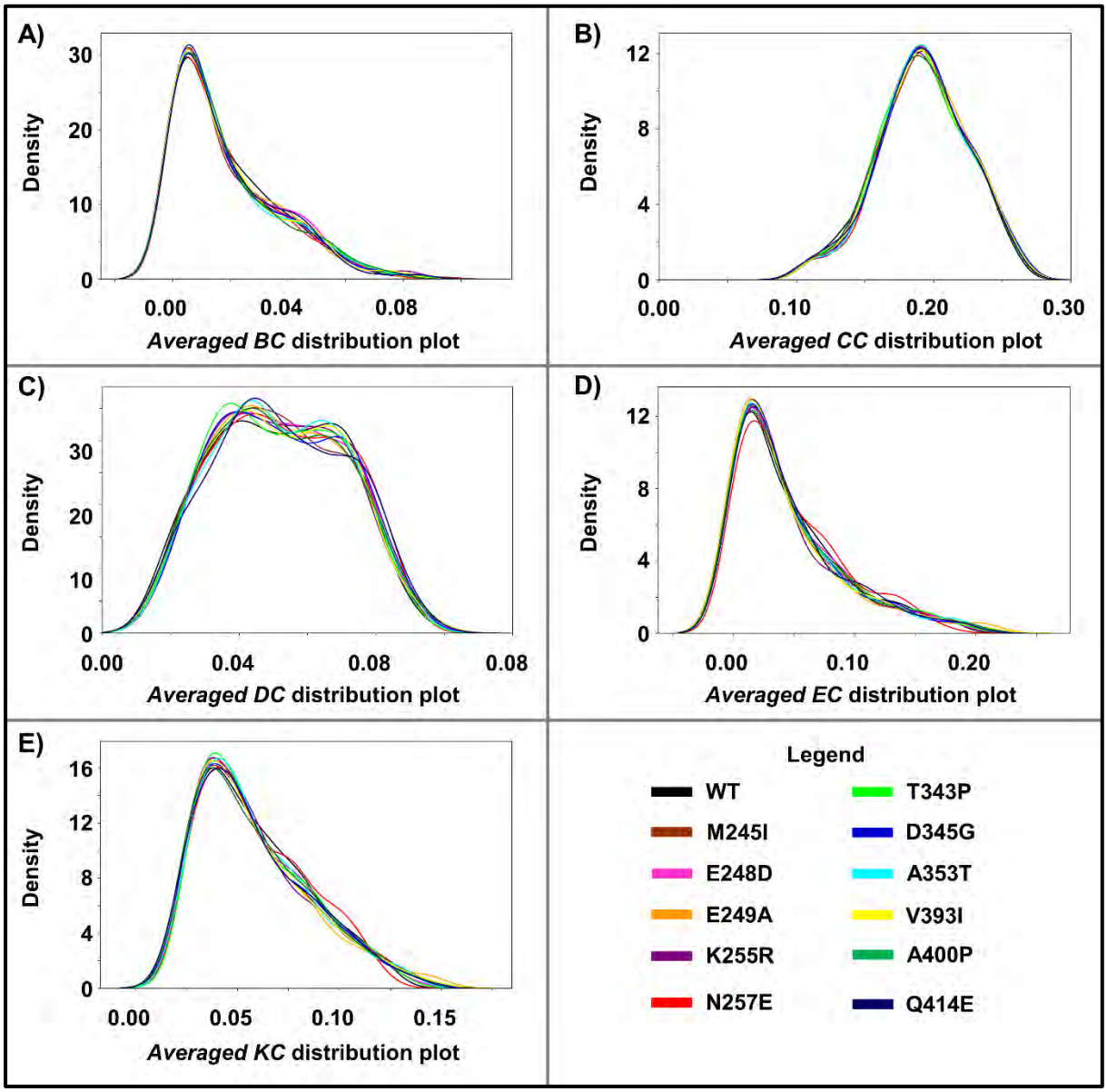



**Figure 2.17: Density distribution of all centrality metrics of the WT and mutant systems of the catalytic domain of FP-2.** A) Averaged *Betweenness centrality* (*BC*), B) Averaged *Closeness centrality* (*CC*), C) Averaged *Degree centrality* (*DC*), D) Averaged *Eigenvector centrality* (*EC*), and E) Averaged *Katz centrality* (*KC*).

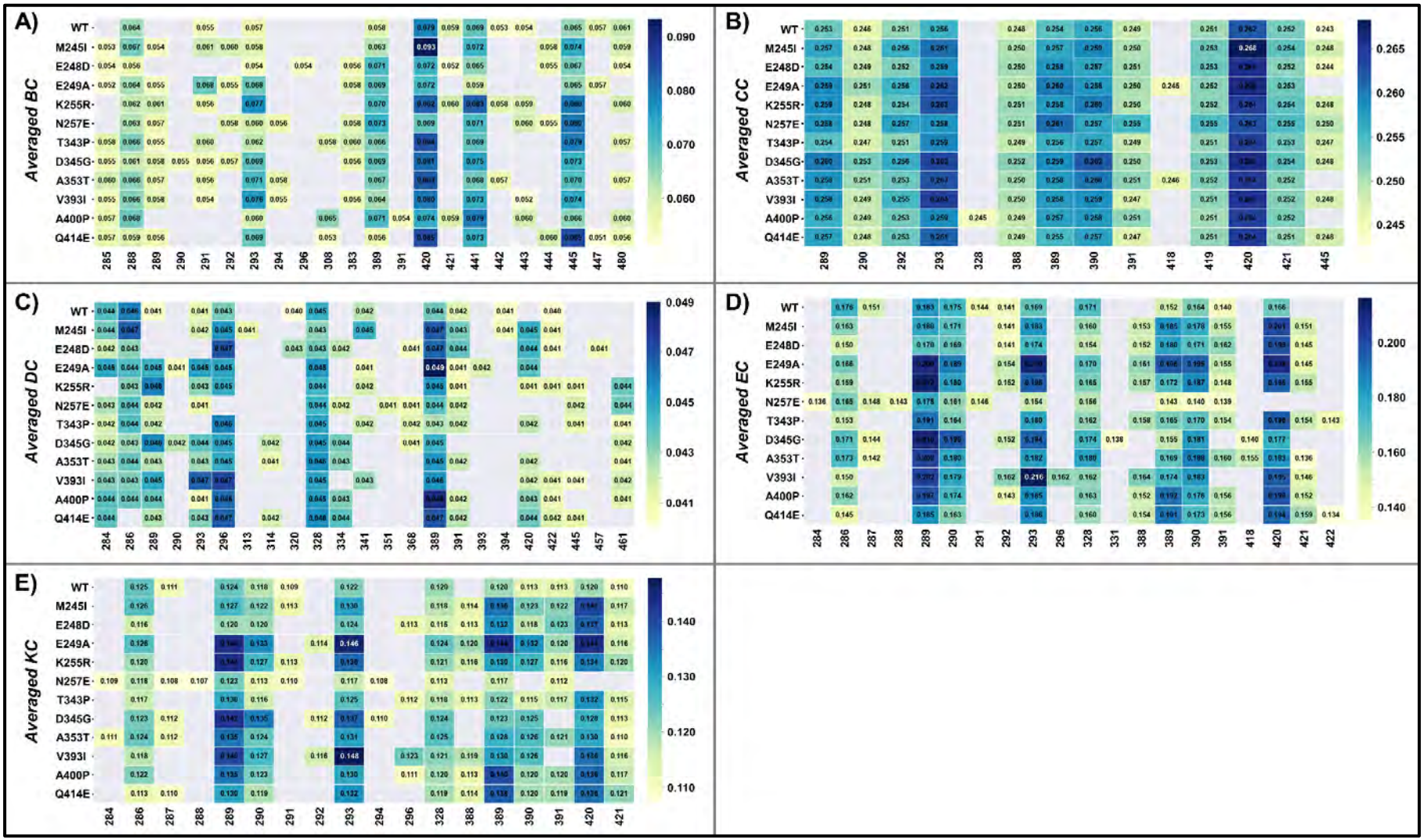



**Figure 2.18: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems of the catalytic domain of FP-2.** Hubs are shown along the x-axis, and protein systems along the y-axis. Detected hubs are annotated with a centrality value.

### 2.4.7.2 Partial zymogen complex

In all protein systems (WT and mutants), residues in the $\alpha$-helical structural elements, which form the central invariant core, exhibited high centrality values in all metrics (**Figure 2.19**). All systems exhibited high *BC* values in $\alpha$2 and $\alpha$3 helix regions, forming two highly conserved motifs (ERFNIN and GNFD) characteristic of the papain family. The functional importance of residues forming these two has been determined previously [206]. Using interaction analysis, several inter-residue contacts between the two motifs that are critical in maintaining the structural integrity of the prodomain segment were identified [206]. Additional interactions between $\alpha$2 and $\alpha$3 residues and the catalytic domain (H199-D398, Y207-D408, K208-E404) were also identified as crucial in anchoring the prodomain segment for it to mediate its natural regulatory function.

Furthermore, residues known to form two salt bridges (R185-E221 and E210-K403) required to activate the partial zymogen complex had high *BC* values. These results were similar to *CC* and *DC* metrics, respectively. *EC* and *KC* metrics have similar results to the *BC* metrics. However, residues in the prodomain ( $\alpha$2 and $\alpha$3) were devoid of high centrality values. We examined the distributions of the centrality metrics to determine whether mutations triggered adjustments within the normal distribution (**Figure 2.20**). As seen with results in the catalytic domain, notable differences were observed in the distribution of the centrality metrics. The WT and mutant systems showed differences in the *BC* distributions, which were skewed to the right (**Figure 2.20A**). The *CC* distribution showed a normal distribution, with minor changes observed in the WT and mutant systems (**Figure 2.20B**). The *DC* distribution mimicked the normal distribution and showed the greatest difference between the WT and mutant systems on centrality measures (**Figure 2.20C**). *EC* (**Figure 2.20D**) and *KC* (**Figure 2.20E**) distributions were highly similar to the BC distribution and showed differences between the WT and mutant systems. Further analysis was conducted to gain insight into the changes in the residues with the highest centrality (5%).

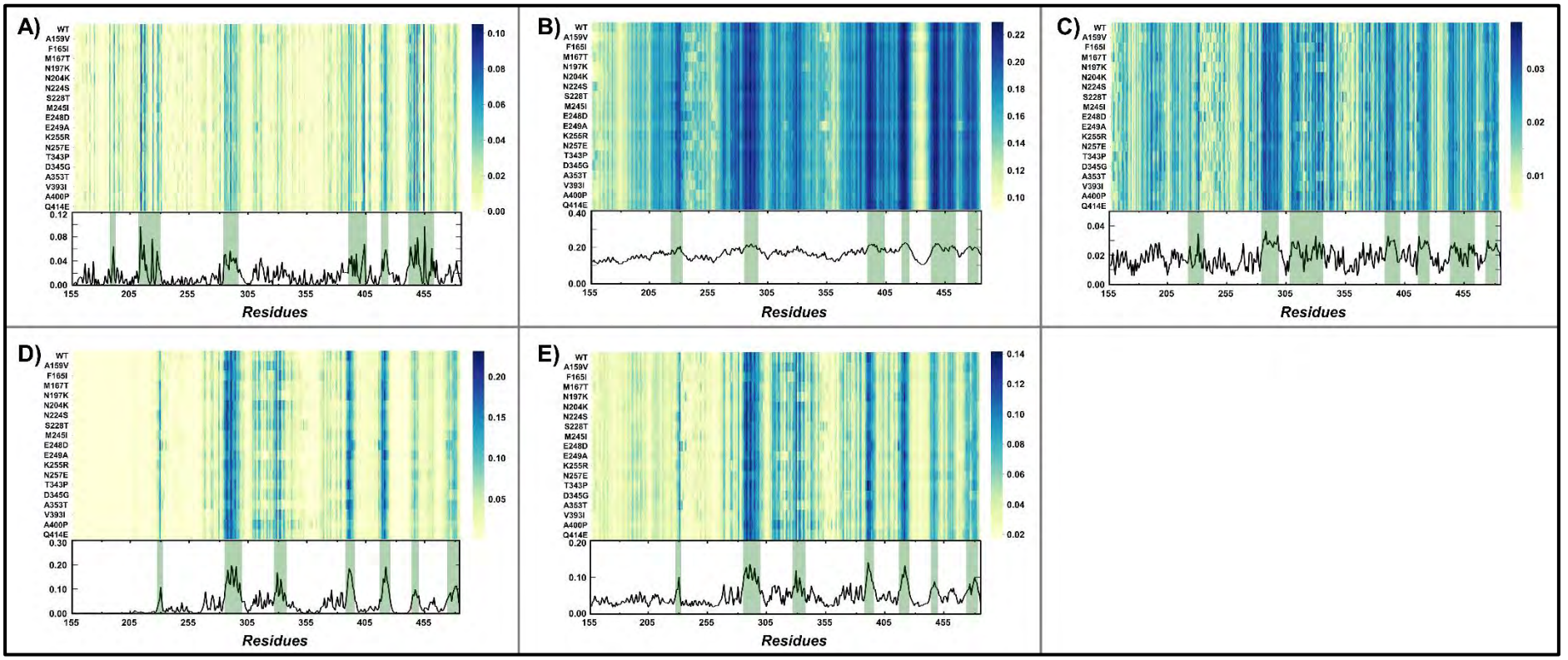



**Figure 2.19: Heatmaps showing per-residue centrality scores of the WT and mutant systems of the partial zymogen complex of FP-2.** A) *Betweenness centrality* (*BC*), B) *Closeness centrality* (*CC*), C) *Degree centrality* (*DC*), D) *Eigenvector centrality* (*EC*), and E) *Katz centrality* (*KC*).

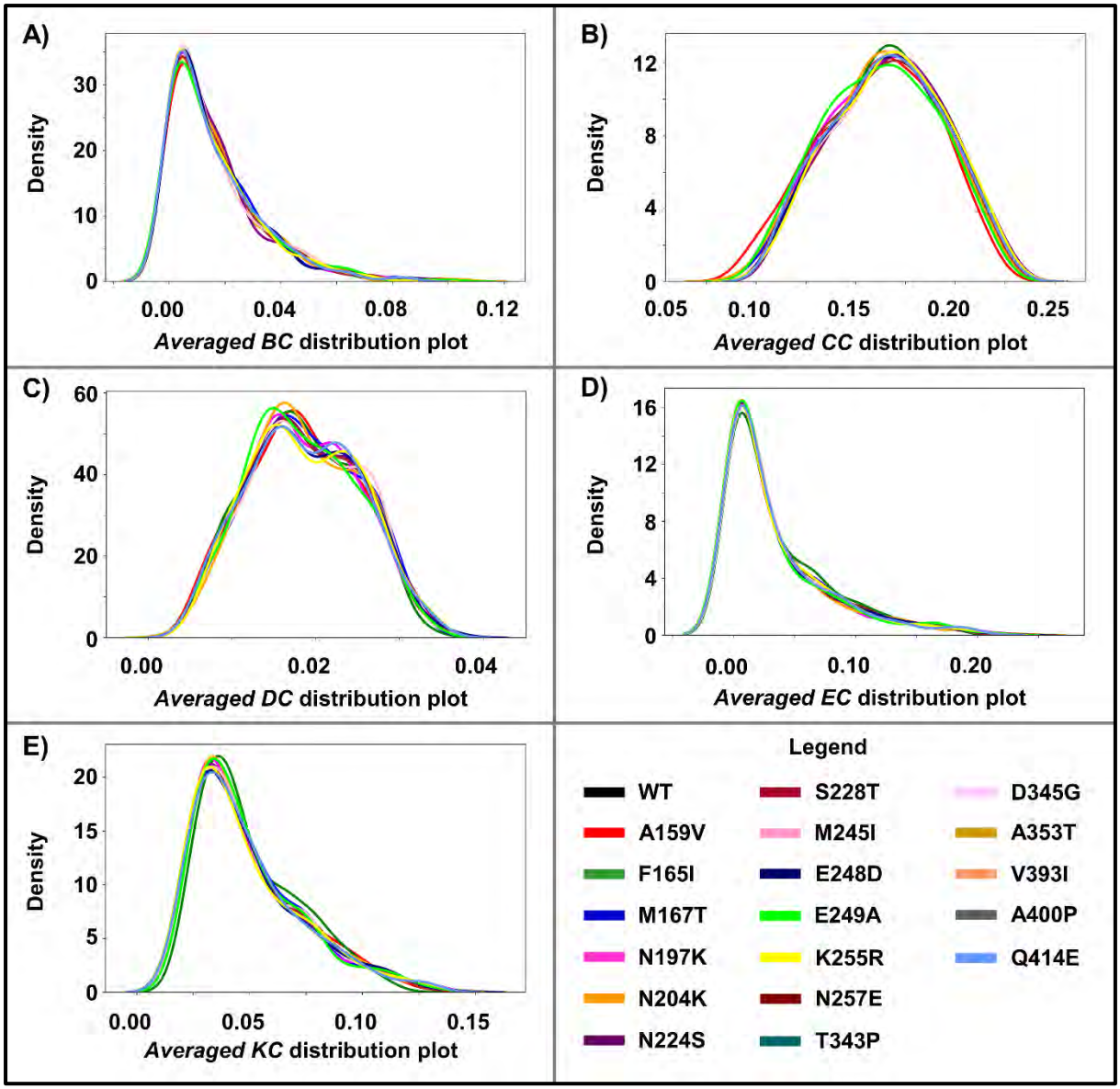



**Figure 2.20: Density distribution of all centrality metrics of the WT and mutant systems of the partial zymogen complex of FP-2.** A) Averaged *Betweenness centrality* (*BC*), B) Averaged *Closeness centrality* (*CC*), C) Averaged *Degree centrality* (*DC*), D) Averaged *Eigenvector centrality* (*EC*), and E) Averaged *Katz centrality* (*KC*).

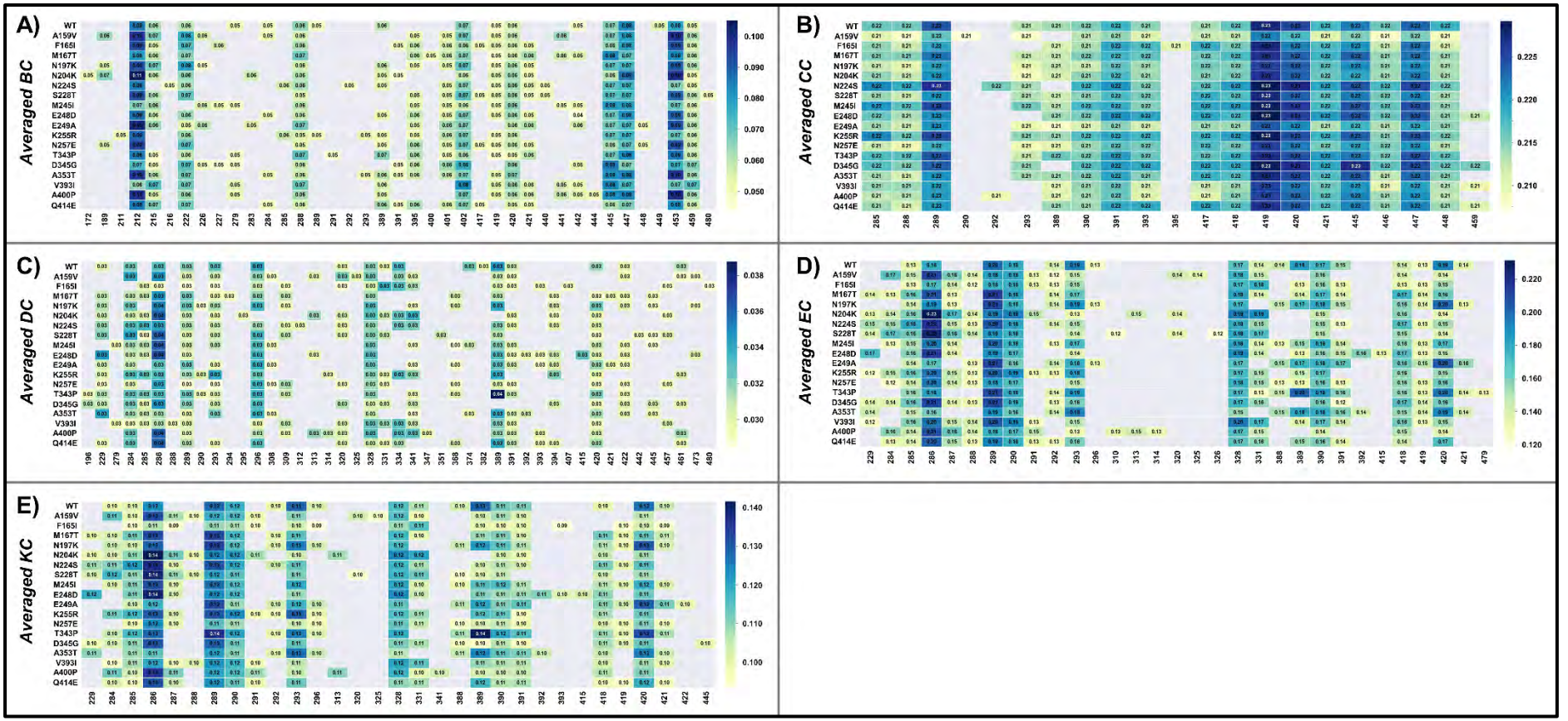



**Figure 2.21: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems of FP-2.** Hubs are shown along the x-axis, and protein systems along the y-axis. Detected hubs are annotated with centrality values

According to **Figure 2.21**, *persistent hubs* for the averaged *BC* were N212, F222, F288, Y402, I445, N447, W453, and I459. *Persistent hubs* for the averaged *CC* were C285, F288, S289, S390, I391, V393, H417, A418, V419, M420, L421, I445, K446, N447, S448. *Persistent hubs* for the averaged *DC* were W286, S289, and I328. *Persistent hubs* for the averaged *EC* were C285, W286, S289, S290, S293, I328, S390, A418, and M420. *Persistent hubs* for the averaged *KC* were C285, W286, S289, S290, S293, I328, S390, and M420.

***Betweenness centrality* (*BC*)**

*BC* determines how often residuals are used when navigating the shortest path in the network. Hydrophobic residue W453 appeared as a *persistent hub* in all systems. W453 forms a hydrophobic interaction with F214 and W449, which is crucial for auto-processing [80]. In this interaction, the prodomain segment of the inactive zymogen complex dissociates into the active enzyme and hydrolyses hemoglobin [80]. Therefore, we hypothesise that the loss of any *persistent hubs* may influence the communication network of the protein structure. The global 5% *distinctive hubs* of the WT and mutant systems were further examined.

Interestingly, active site residue W449 appeared as a *distinctive hub* in the WT and was absent in all mutant systems (**Figure 2.21A**). F214 forms hydrophobic interactions with W449 and W453, and this interaction is crucial for auto-processing [80]. Active site residue Q279 appeared in the WT and M245I, T343P, D345G, V393I, and A400P but was absent in the other mutations. Additionally, active site residues G283, C285 (catalytic thiol), V395, A400, and H417 were distinctive in some mutant systems but absent in the WT. Interface residues Y226 and L227 were distinctive in some mutant systems but absent in the WT. Overall, we hypothesise that the presence of mutations on the partial zymogen complex of FP-2 disrupted the communication network, thereby altering the activation of its active enzyme.

***Closeness centrality* (*CC*)**

*CC* quantifies the shortest path between node *i* and every other node in a protein network. When comparing all the centrality metrics, *CC* analysis had the most *persistent hubs*. Some mutations had *distinctive hubs*: A159V (S290), F165I (V395), N224S (G292), E248D (I459), D345G (I459), V393I (I459), A400P (G292), and Q414E (I459) (**Figure 2.21B**).

***Degree centrality* (*DC*)**

*DC* is the simplest centrality metric, accounting for the total number of connections linked to a particular node in a protein network. In *DC* calculations, a significant increase in the number of hubs was observed in both WT and mutant systems (**Figure 2.21C**). Also, *DC* calculations exhibited the diverse distribution of *distinctive hubs* in the mutant systems. N197K, N204K, N224S, S228T, E248D, E249A, K255R, N257E, D345G, A353T, V393I, A400P, and Q414E gained a number of distinctive hubs, which were found around the active site pocket (**Figure 2.21C**). *Distinctive hubs* L308, L313, G325, I341, and G473 were acquired in A159V. Residues L308, L313, and I341 are found around the interface of pockets 2 and 6. Recall that we found a cryptic pocket in section **2.4.6.2**. We, therefore, postulate that this communication path caused the opening of pocket 6 and the subsequent formation of the cryptic pocket. *Distinctive hubs* C285, S309, I341, V393, F407, G457, and I480 were acquired in F165I. Interestingly, we observed a communication path in the distribution of the *distinctive hubs* in F165I (**Figure 2.22B**). This path occurs in the catalytic domain, which is triggered by F165I that occurs in the prodomain. The path originates from F407-S309-C285-I341-V393 -G457-I480. M165T gained *distinctive hubs* C285, V294, S309, I368, A394, L421, I445. While this did not form a communication path, it occurred around the active site pocket. M245I gained the following distinctive hubs: Q279, C285, L308, I341, I368, and I445. Despite the lack of a clear path, we noticed a pattern in the distribution of the hubs (**Figure 2.22C**). Residues L308 and I341 border around pockets 2 and 6, while I368 is found in pocket 3. Residues Q279, C285 (catalytic thiol), and I445 occur in the active site. We postulate that this communication path is responsible for the presence of the cryptic pocket identified in section **2.4.6.2**. We observed a communication path in T343P which consists of distinctive hubs V196, C285, S290, S309, I341, and I368.

Interestingly, this path originates from I368-I341 and then goes through the mutated residues 343 to connect with S309-S290-C285-V196 (**Figure 2.22D**). In our previous study on the catalytic domain of FP-2, we identified an allosteric communication path in the mutant system T343P (L308-I291-S289-C285) [118]. Since T343P is one of the mutations associated with ART resistance, we speculate that this allosteric communication path formed by six *distinct* averaged *DC* hubs translocating from I368-I341through the mutant residue (343) to residues around the active site is likely responsible for resistance. Overall, the *DC* metric provided insight into the communication patterns initiated by the presence of mutations.

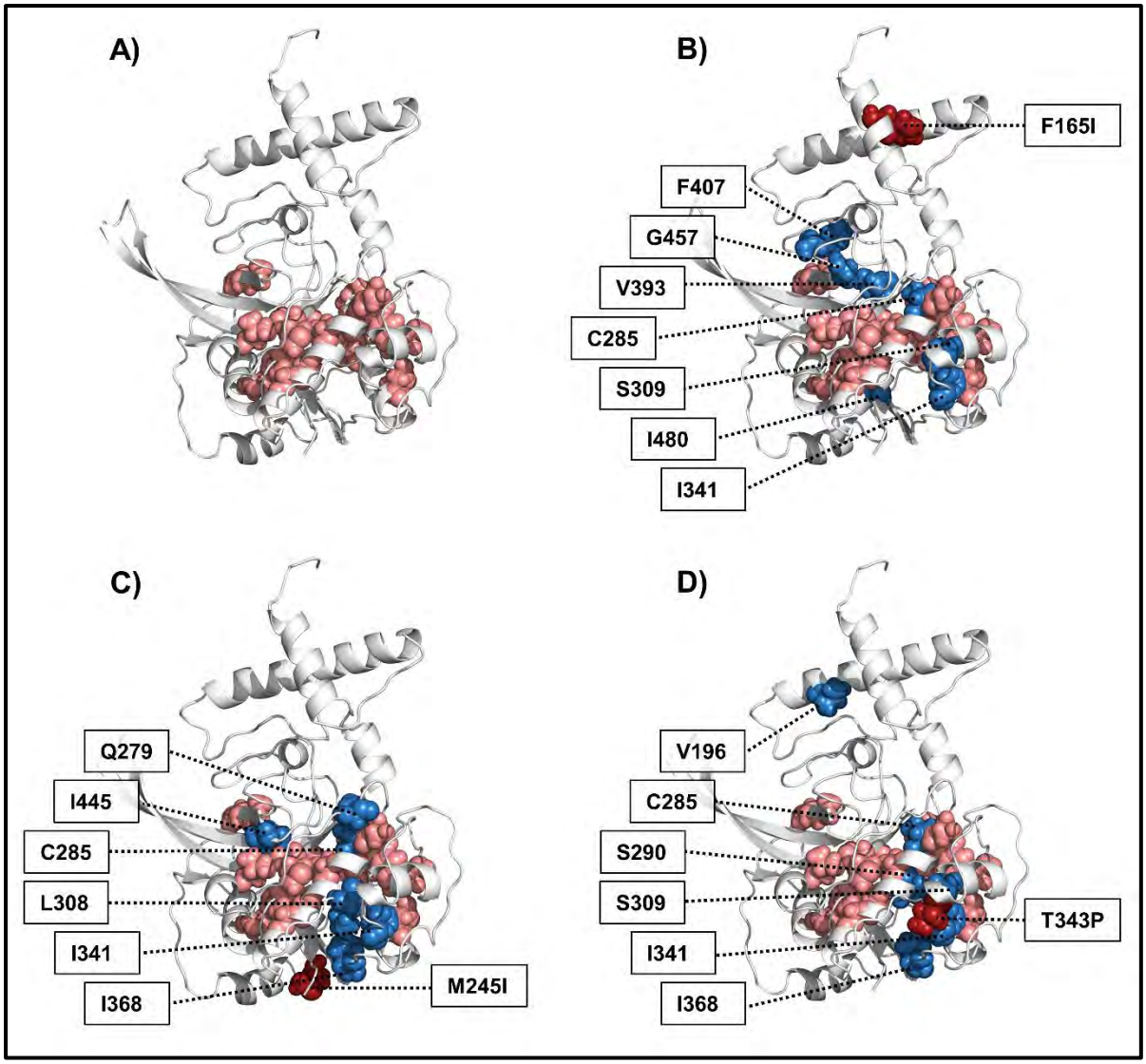



**Figure 2.22: Cartoon representation of the distribution of the averaged *BC* hubs of the WT and mutant systems F165I, M245I, and T343P**. A) WT with averaged *BC* hubs depicted as salmon-colored spheres. B) F165I with its distinct hubs (skyblue) and the WT hubs in salmon. C) M245I with its distinct hubs (skyblue) and the WT hubs in salmon. D) T343P with its distinct hubs (skyblue) and the WT hubs in salmon.

### *Eigenvector centrality* (*EC*)

We observed the loss of hubs in the mutant systems from the *EC* calculations compared to the WT (**Figure 2.21D**). We also observed some *distinctive hubs* in the mutant systems. In A159V, distinctive hubs include S284, A287, F288, I291, N320, and G325. These residues occur around the binding pocket where N320 and G325 are part of the S3 subsite. L229, S284, A287, F288, I291, E310, L313, V314, N320, G325, G326, S392, N415, and T479 appeared as *distinctive hubs* in some mutant systems. While this centrality metric did not observe a definite path, the distinctive hubs were identified as significant in the protein structure. N320, G325,

and G326 are part of residues forming subsite 3 (S3). S392 and N415 are part of residues forming subsite 2 (S2). L313 and V314 are pocket 2 residues. E310 is found in pocket 6.

***Katz centrality* (*KC*)**

As seen in *EC* calculations, loss of hubs from the WT system was observed in some mutant systems (**Figure 2.21E**). We observed some *distinctive hubs* in the mutant systems. This include L229, A287, F288, I291, N320, G325, S392, V393, N415, V419, V422, and I445. It's interesting to note that the majority of these *distinct hubs* were *distinct* in the mutant system according to *EC* calculations, indicating a correction between these two metrics. Additionally, the distribution of both *EC* and *KC* mimicked each other and was skewed to the right (**Figure 2.20D** and **E**).

Overall, we examined the effect of mutations on the partial zymogen complex of FP-2 using the five different centrality metrics and identified *persistent* and *distinctive* hubs that contribute to intra and inter-communication within the various protein systems.

## 2.5 Conclusion

In this chapter, we utilised various computational tools to systematically examine the structural and dynamic effects of 29 missense mutations distributed within the partial zymogen complex and the catalytic domain of FP-2. FP-2 plays a significant role in the development cycle of plasmodial proteases during the blood stage in the host erythrocytes, making this protein an important antimalarial drug target. Even though an approved antimalarial drug targeting FP-2 is yet to be found, numerous compounds of diverse chemical classes with up to nanomolar inhibitory potencies have been identified [235–238]. With the ongoing research surrounding this protease, there is a huge possibility of antimalarial drugs targeting them soon. However, the much-anticipated breakthrough could be short-lived as parasite isolates have already demonstrated the development of polymorphisms that result in missense mutations in the gene coding for FP-2. Recently, the possible association between the observed mutations in FP-2 and reduced ART sensitivity was identified, necessitating an urgent need for studies to evaluate their effects [97].

Structural visualisation of the distribution of the mutations in FP-2 revealed the position of the mutations. With the exception of V393I and A400P located within the binding pocket, all other mutations were distally located from the binding site. We identified six potential allosteric sites

using a combination of structure-based approaches and performed various computational analyses to investigate the overall effect of mutations on the partial zymogen complex and the catalytic domain of FP-2. In the catalytic domain, we observed a cryptic pocket in the presence of mutant M245I and a semi-cryptic pocket in the presence of E249A. We observed a cryptic pocket in the partial zymogen complex in the presence of A159V, M245I, and E249A.

Interestingly, we found that M245I and E249A had the same effect on pockets 2 and 6 in the presence and absence of the promoter region. Using the five centrality metrics (*BC*, *CC*, *DC*, *EC*, and *KC*), we identified conserved communication network changes between the WT and mutant systems. These changes occurred on high centrality residues perceived to be critical for the hemoglobinase activity in FP-2.

# Chapter 3

# Structural, dynamic and network communication effects of missense mutations on the partial zymogen and the catalytic domain of Falcipain 3

## 3.1 Introduction

FP-3 shares a 68% sequence identity with FP-2, consisting of $\alpha$-helices, $\beta$-sheets, and loops [50]. Both proteases require a reducing environment and an acidic pH to function properly [239]. A key difference, however, is that FP-3 transforms into an enzyme only at acidic pH levels [239]. This pH is ideal for its activity and stability, especially against native hemoglobin [239]. FP-3 is the second *P. falciparum* hemoglobinase that is suitable for hydrolyzing native hemoglobin in the food vacuole. While the concentration of FP-2 in trophozoites exceeds that of FP-3, the latter is thought to cleave native hemoglobin approximately twice as fast as the former [239]. Thus, confirming FP-3 as an important target for inhibiting hemoglobin degradation [70].

## 3.2 Proposed study

This chapter determined the structural, dynamic and residue communication effects of 12 missense mutations in the partial zymogen complex and the catalytic domain of FP-3 using a range of computational approaches. As a quick starting point, the mutant structures were predicted using homology modeling. MD simulations were performed to describe the dynamic nature of the protein. The global dynamic attributes of each ensemble were performed, followed by local analysis based on DRN calculations and weighted residue contact maps (**Figure 3.1**).

## 3.3 Methodology

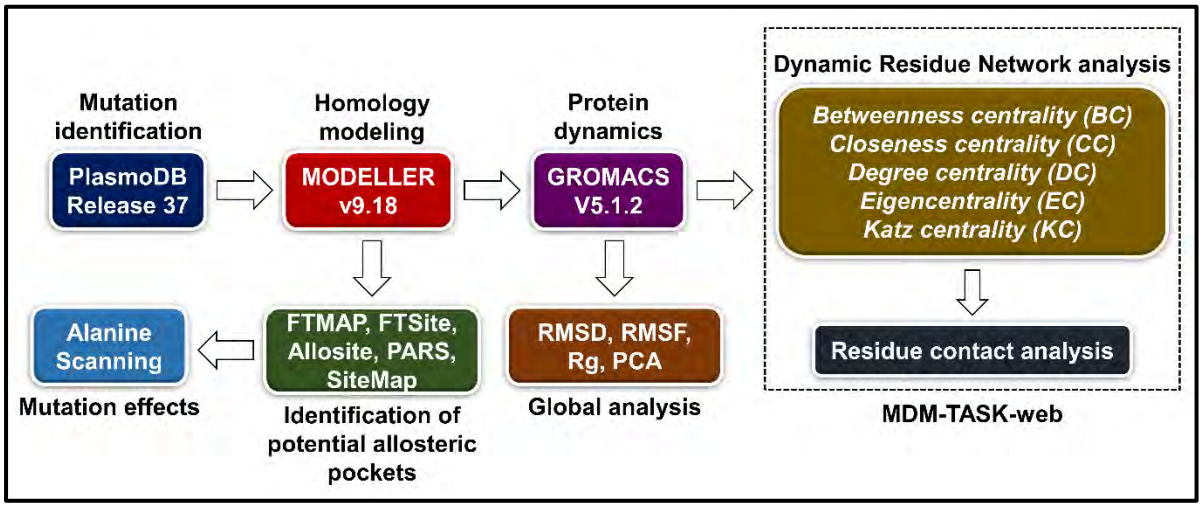



**Figure 3.1: A workflow of the methods and tools utilised in this chapter**.

### 3.3.1 identification of missense mutations and preparation of mutant 3D structures

The 3D structure of the catalytic domain of FP-3 [PDB ID: 3BWK] [85] was retrieved from the RCSB PDB [197,198] and modeled by Thommas Mutemi Musyoka [87] using MODELLER version 9.10 [240]. The WT structure of the partial zymogen complex of FP-3 was modeled using MODELLER version 9.18 [241] by Thommas Mutemi Musyoka [206]. PlasmoDB [102] (release 37, accessed 21-Aug-2018) was queried to identify missense mutations linked with the partial zymogen complex and the catalytic domain of FP-3. These include E177Q and I193L (occurring in the partial zymogen complex) and N371K, R411K, A422T, E446G, and N468Y (occurring in the catalytic domain). Subsequently, MODELLER version 9.19 [201] was used to predict mutant structures using the corresponding WT templates. The quality of the mutant structures was evaluated as described in Chapter 2. The protein structures (WT and mutants) were protonated to a pH of 5.5 to mimic the acidic environment of the food vacuole using the PROPKA tool from PDB2PQR (version 2.1.1) [208,209].

### 3.3.2 Identification of potential allosteric pockets

A combination of structure-based tools such as Allosite [210], FTMap [211], FTSite [211], PARS [212], SiteMap [213] and Auto-Ligand [214] was used to identify potential allosteric pockets as outlined in chapter 2.

### 3.3.3 Molecular dynamics simulations

The WT and mutant systems of the partial zymogen complex and the catalytic domain of FP-2 were subjected to 500 ns of all-atom MD simulation using the GROMACS 5.1.2 package [217] (WT and 5 mutations in the catalytic domain; WT and 7 mutations in the partial zymogen complex). First, using three different force fields viz. AMBER96 [138], AMBER99SB-ILDN [139], and CHARMM36 [140], we performed triple test MD on FP-3 to identify the most suitable force field parameters. The optimal force field, AMBER99SB-ILDN [139], generated WT and mutant topology files. MD simulation was carried out using all parameters outlined in Chapter 2.

### 3.3.4 Comparative essential dynamics

MD simulations provide the positional motions of each atom, and these motions are dissected into principal components (PC), where each component represents the functional state of the protein structure [242]. To investigate the conformational dynamics of the predicted allosteric pockets of the WT and mutant systems of the partial zymogen complex and catalytic domain, we performed comparative essential dynamics using the *compare_essential_dynamics.py* script in MDM-TASK-web [163] as described in chapter 2.

### 3.3.5 Dynamic residue network (DRN) analysis

To understand the impact of missense mutations on the residue interaction network and highlight residues involved in the structure and function of FP-3, we performed DRN analysis using MDM-TASK-web [163]. In this chapter, we considered five DRN metrics: *betweenness centrality* (*BC*), *closeness centrality* (*CC*), *degree centrality* (*DC*), *eigenvector centrality* (*EC*) and *Katz centrality* (*KC*).

### 3.3.6 Residue contact map analysis

Residue contact map analysis was performed to determine how frequently a residue of interest interacts with its neighbouring residues throughout the entire MD trajectory using the *contact_map.py* and *contact_heatmap.py* scripts from MDM-TASK-web [163,165]. The weighted contact maps were created with a cut-off distance of 6.7 Å and a step size of 50 frames. The weighted contact maps were combined into a heat map using the *contact_heatmap.py* script. In this case, "*weighted*" refers to contact maps over MD simulations, indicating the frequency.

### 3.3.7 Assessment of the prodomain interface residues via Alanine Scanning

To predict energetically important amino acid residues within the partial zymogen interface, we performed alanine scanning using the ROBETTA web server [243]. Alanine scanning calculates the contribution of each residue to the binding free energy of a protein-protein interface by mutating interface residues to alanine. Residues with $\Delta G \geq 1$ kcal/mol denote destabilisation, -0.8 to 0.99 kcal/mol denote neutralisation, and < -0.8 denote stabilisation. The pdb structure of the partial zymogen mutant models was edited using PyMOL [199]. Residues in the prodomain were designated chain A and the catalytic domain as chain B.

## 3.4 Results and Discussion

In this chapter, we utilized various computational techniques to elucidate the structural and dynamic effects of mutations in the partial zymogen and catalytic domain of FP-3.

### 3.4.1 Identification of missense mutations and generation of three-dimensional (3D) mutant structures of FP-3

A total of seven missense mutations distributed across the partial zymogen and five across the catalytic domain of FP-3 were identified from PlasmoDB [102] (release 37, accessed 21-Aug-2018). Two of these mutations were located in the partial zymogen complex, while five were in the catalytic domain (**Figure 3.2**). The two mutations located in the partial zymogen complex (E177Q and I193L) were situated in the $\alpha$1 and $\alpha$2 helices. In the catalytic domain, most mutations occur in the loop regions, with the exception of E446G and N468Y, which occur in the antiparallel $\beta$-sheet. Furthermore, E446G is located in the arm region, which is a highly dynamic structural element proposed to capture hemoglobin for processing [83]. Thus, mutant

models were built using MODELLER version 9.19. For each system, the best model with the lowest z-DOPE score was selected and validated using VERIFY3D, ProSA, and PROCHECK. **Table S4** summarizes the results of each validation tool.

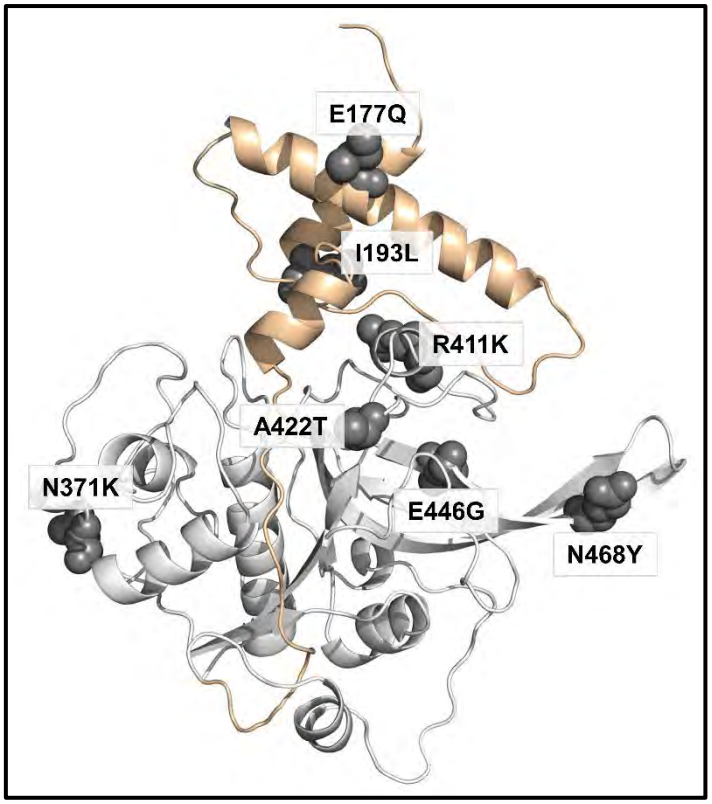



**Figure 3.2: Structural representation and location of missense mutations occurring in FP-3 protein**. The prodomain is colored in wheat, and the catalytic domain is coloured in white.

### 3.4.2 Identification of potential allosteric pocket

Six potential allosteric pockets (named Pockets 1 – 6) were identified by a combination of structure-based tools that utilized different algorithms (**Figure 3.3**). Residues forming the potential allosteric pockets are highlighted in **Table 3.1.** FP-2 and FP-3 are homologous proteins that share a 68% sequence similarity; hence we calculated the sequence identity of the potential allosteric residues between FP-2 and FP-3 using MAFFT [244] and Jalview [245]. From the calculations, FP-2 and FP-3 exhibited the following allosteric pocket sequence identity: pocket 1 (48%), pocket 2 (67%), pocket 3 (44%), pocket 4 (58%), pocket 5 (63%), and pocket 6 (63%). Pocket 1 is located around the nose region loop, which is an intrinsic structural attribute of FP-3 and its homologs. The residues in pockets 2 and 4 are located in the highly dynamic loops adjacent to S1 and S1'. Pocket 4 contained missense mutation R411K. Pocket 3 is formed by residues residing at the bottom of the L-domain. Pockets 5 and 6 are characterized by two small grooves. According to the druggability analysis performed with

SiteMap and Auto-Ligand, pocket 1 was determined to be druggable, pockets 2 and 4 were moderately druggable, while pockets 3, 5, and 6 were less likely to bind small molecules (**Table S5**).

**Table 3.1: Residues in potential allosteric pockets** (pockets 1 – 6) in FP-3 based on the whole protein sequence.

| Pockets | Residues |
|---|---|
| 1 | Y254, E255, I258, K259, K262, P263, A264, D265, A266, K267, L268, D269, R270, D385, K386, F387, K388, E389, A390, E473, N474, G475, Y476 |
| 2 | Q319, L321, V322, D323, C324, S325, V326, K327, N328, N329, G330, C331, Y332, D342, V358, S359, N360, L361, P362, E363, T364, C365, N366 |
| 3 | Y306, K310, A312, F314, M343, I344, D345, L346, G347, G348, L349, E372, R373, Y374, T375, I376, K377, S378, Y379, P489, L490, L491 |
| 4 | D405, D406, F409, Y410, R411, G412, G413, F414, Y415, D416, G417, E418 |
| 5 | G279, G280, V281, T282, A307, K311, A312, L313 |
| 6 | P283, V284, K285, D286, A288, L315, S317, E318, Q352, D353, D354, Y355, P356, Y357, G458, S459 |

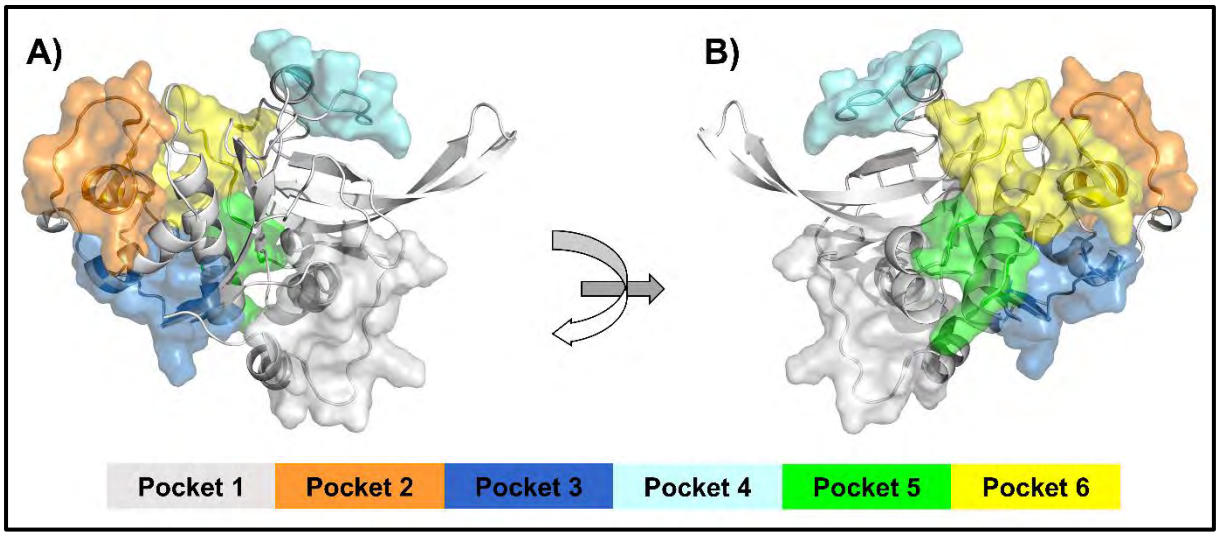



**Figure 3.3: Structural representation of the potential allosteric pockets in FP-3**. Identified pockets are shown in the surface presentation and colored accordingly.

### 3.4.3 Mutations-induced structural changes identified using MD simulations

The use of molecular dynamics simulations has greatly increased our understanding of complex biological systems; however, the selection of an appropriate force field that best describes the structural and dynamic properties of a macromolecule is of vital importance [246]. Consequently, we evaluated three force fields: AMBER96 [138], AMBER99SB-ILDN [139], and CHARMM96 [140] in a triplicate run per force field for the WT system of the partial zymogen and the catalytic domain. According to the backbone RMSD analysis, AMBER99SB-ILDN described the properties of FP-3 most accurately (**Figure 3.4**). Subsequently, the dynamic changes of the WT and mutant systems of the partial zymogen complex and the catalytic domain of FP-3 were assessed over 500 ns MD simulation. Post-MD simulations analysis involved the pre-processing of each trajectory where the periodic boundary conditions (PBC) were removed using the trjconv tool in GROMACS [217]. Next, the protein stability and conformational changes induced by mutations were assessed using RMSD (whole protein, central invariant core, and the active site residues) and Rg (whole protein, central invariant core, and the various potential pocket residues) calculations.

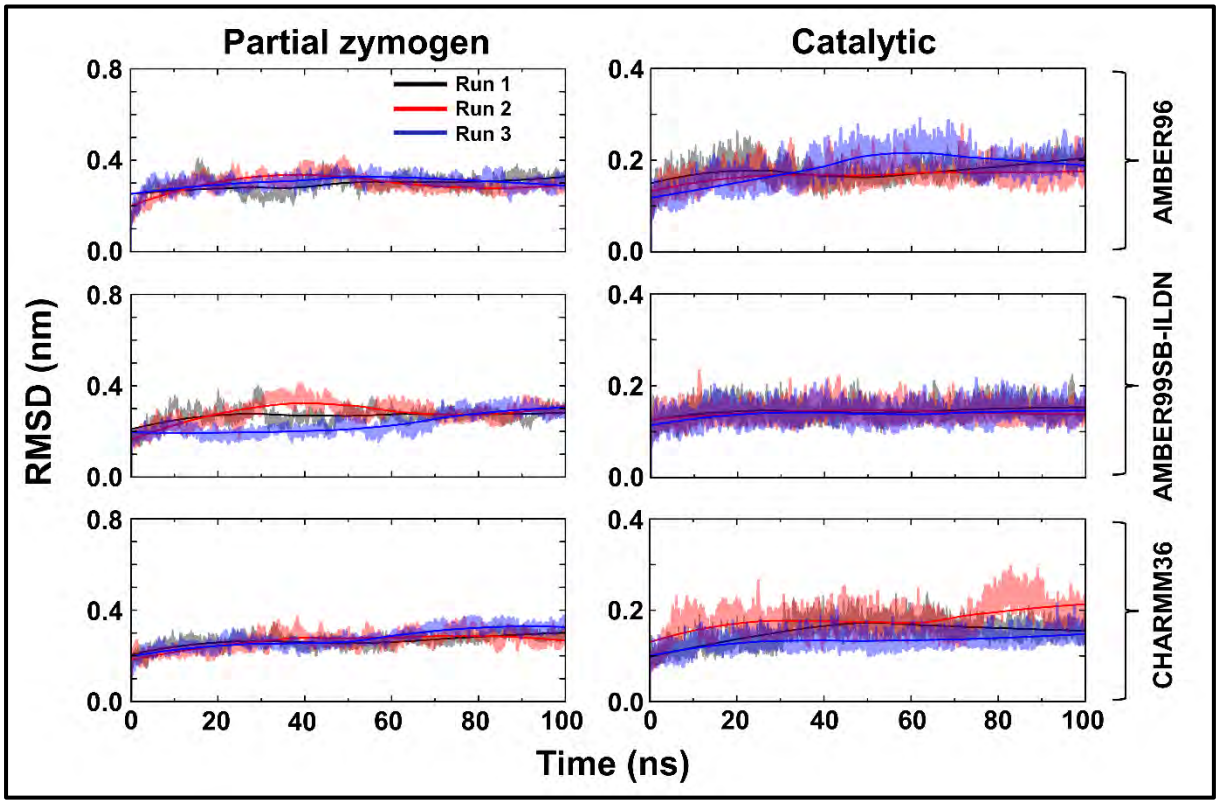



**Figure 3.4: Identification of the most accurate force field for the partial zymogen complex and the catalytic domain of FP-3.** The backbone RMSD analysis of the three force fields over the triplicate MD simulations is shown in line plots. Run 1 is coloured in black, run 2 in red, and run 3 in blue.

### 3.4.3.1 Catalytic domain

To observe the global conformational dynamics over the simulation time, various analyses of atom-positional RSMDs were performed for the WT and mutant systems. These dynamics were evaluated using backbone RMSD, and the mutant systems were compared to the WT structure. Initially, the stability of the WT and mutant systems throughout the entire simulation was evaluated using the classical RMSD *versus* timeline plots. A consistent convergence was observed in the WT and mutant systems (RMSD ranged from 0.1 to 0.3 nm) (**Figure S3.1**). As a next step, we calculated RMSD distribution violin plots for the active site residues and the central invariant core to determine the conformational space sampled by each system. An all-encompassing unimodal RMSD distribution was observed for all systems (whole protein, central invariant core, and active site residues), indicating that all systems visited a single conformational space (**Figure 3.5A, B, and C**).

Additionally, RMSD plots from all *versus* all frames were generated using *pytraj* [228] to examine the differences between the starting conformation and the whole frame conformation. It is important to note that each system's starting structure varied significantly compared to other frames during the simulation, similar to the classical RMSD *versus* timeline plot (**Figure 3.5D**). A422T, located around the binding pocket, exhibited the most divergent results.

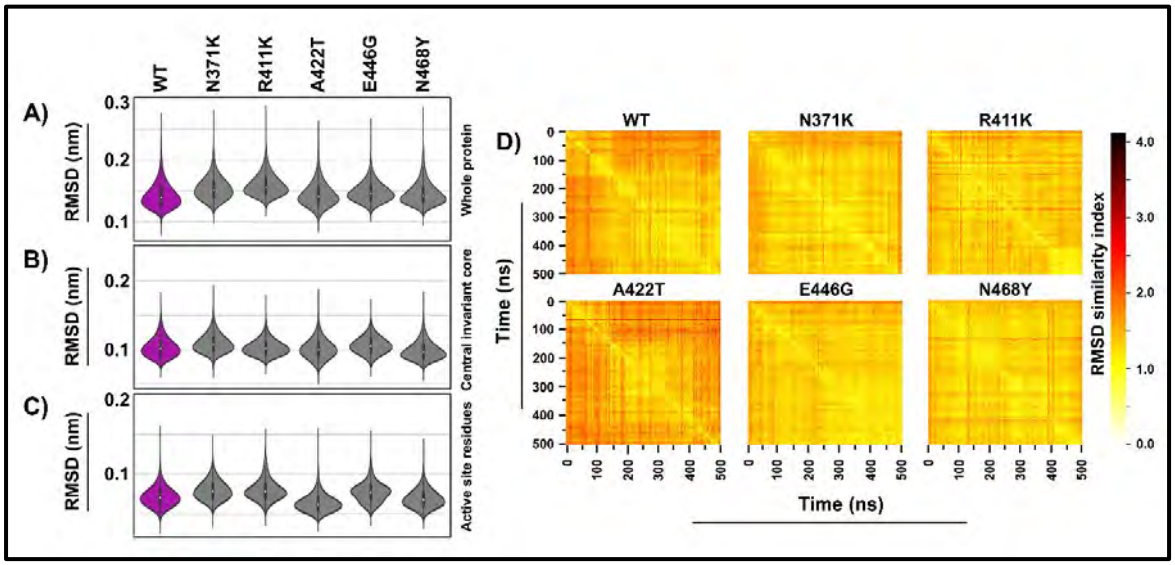



**Figure 3.5: Backbone RMSD analysis of the WT and mutant systems of the catalytic domain of FP-3.** RMSD violin distribution plots of the A) whole protein, B) central invariant core, and C) binding active site residues of the WT and mutant protein systems. WT ensemble plots are shown in purple. D) All vs all C$\alpha$ RMSD of the WT and mutant protein systems. The x and y-axes represent time (ns). The color scale indicates the degree of conformational variation between frames (most similar = white, different = dark).

### 3.4.3.2 Partial zymogen complex

To characterize the conformational changes induced by the presence of mutations, we analyzed the global dynamics of the WT and mutant systems. Based on the RMSD versus timeline plots, both WT and mutant systems showed high values, indicating they sampled different energetic landscapes (**Figure S3.2**). As seen in **Figure 3.6A,** great variability in conformational distribution was observed in the partial zymogen complexes of FP-3. Most systems exhibited bimodal distribution, indicating that the systems evolved through different conformations while undergoing MD simulation. Using VMD [231], the WT and mutant systems were visualised. The central core (consisting of $\alpha$-helices and $\beta$-sheets), where the trench-like binding pocket is located, exhibits vibrational-like movements. However, major conformational diversity was observed within the loop regions linking the central core structural elements. These observations corroborated the RMSD results of the central invariant core, which exhibited unimodal distribution (**Figure 3.6B**). The effect of the mutations on the active site residues is described in **Figure 3.6C**. While most systems exhibited unimodal distributions, notable differences in the RMSD values were observed.

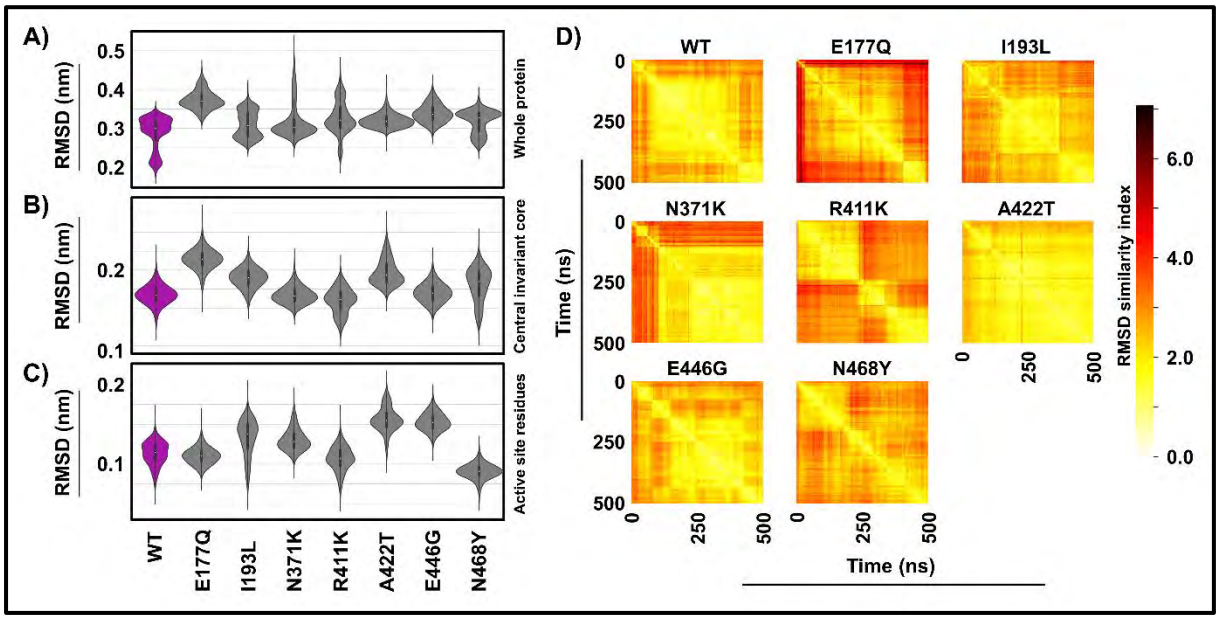



**Figure 3.6: Backbone RMSD analysis of the WT and mutant systems of the partial zymogen complex of FP-2.** RMSD violin distribution plots of the A) whole protein, B) central invariant core, and C) binding active site residues of the WT and mutant protein systems. WT ensemble plots are shown in purple. D) All vs all C$\alpha$ RMSD of the WT and mutant FP-2 proteins. The × and y-axes represent time (ns). The color scale indicates the degree of conformational variation between frames (most similar = white, different = dark).

### 3.4.4 Local per-residue flexibility analysis using RMSF

We determined per-residue flexibility profiles for all systems using the C$\alpha$ RMSF and discovered significant mutation-related changes. The highest RMSF values were observed in the loop regions in the partial zymogen complex and the catalytic domain. In the partial zymogen complex, notable fluctuations were observed in the nose region (residues 243 – 264) and the $\beta$-arm hairpin (residues 438 – 444) (**Figure 3.7A**). Several mutations exhibited higher fluctuations than the WT system. These include E177Q, N371K, and R411K. In the catalytic domain, the $\beta$-arm hairpin and loops around the subsite residues exhibited high fluctuations (**Figure 3.7B**).

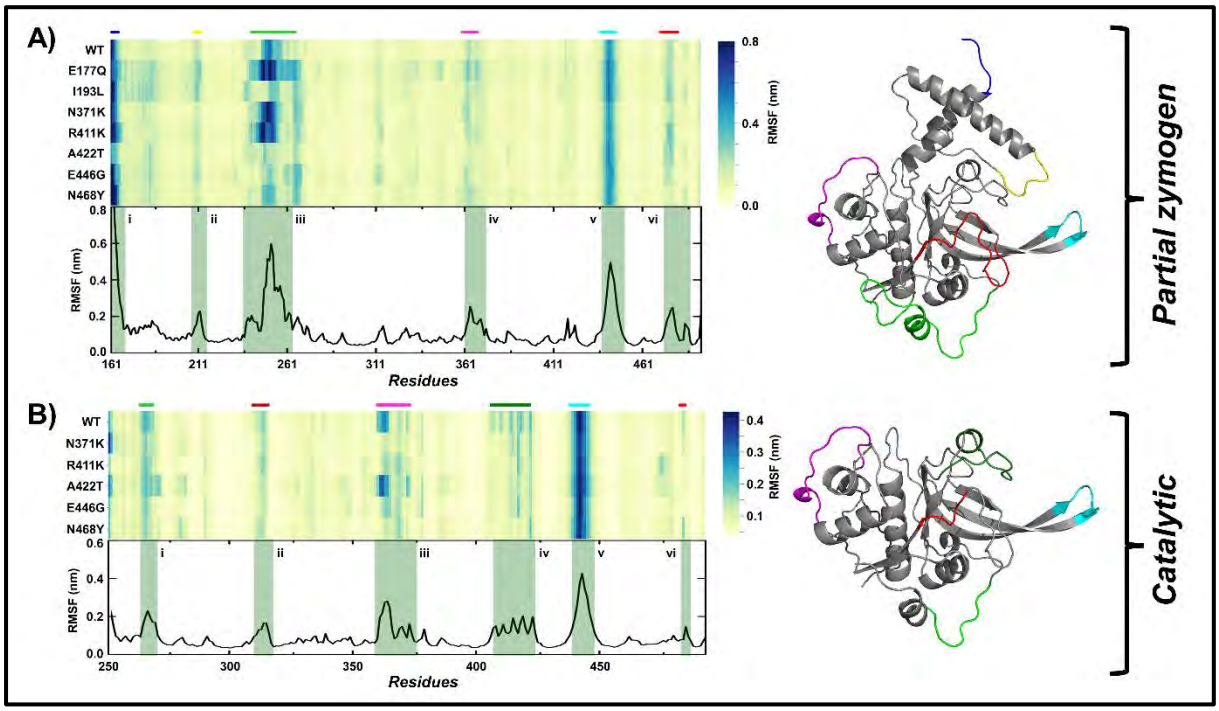



**Figure 3.7: Heatmaps showing the per-residue raw RMSF values across all systems.** RMSF values of the WT and mutant systems of A) the partial zymogen complex and B) the catalytic domain of FP-3. Protein loops displaying increased RMSF, as shown using different color bars, and their positions are shown in the 3D structures.

### 3.4.5 Mutation-linked changes in the putative allosteric pockets were identified via Rg

The overall compactness of the WT and mutant systems of the partial zymogen complex and the catalytic domain of FP-3 was monitored by measuring the corresponding radius of gyration. As can be seen from the Rg violin distribution plots, all systems maintained a unimodal distribution, which indicates that irrespective of the presence of the mutations, the systems

remained compact throughout the simulation (**Figure S3.3**). However, the partial zymogen complex observed a notable increase in the Rg values of E177Q, I193L, R411K, and N468Y. Putative allosteric pockets were further analysed using Rg and C$\alpha$ RMSF to determine the relationship associated with each mutation and the different pockets.

#### 3.4.5.1 Catalytic domain

As seen in **Figure 3.8A,** in pocket 1, all systems displayed unimodal behaviours with observed increases in Rg values. The heatmap of pocket 1 RMSF values revealed high fluctuations in P263 – L268. A422T exhibited higher fluctuations in residues D269 – R270. In addition, R411K and A422T exhibited high fluctuation in residues E473 – Y476. In pocket 2, all systems, with the exception of A422T, exhibited unimodal distribution (**Figure 3.8B**). A422T displayed a bimodal distribution with increased fluctuations in residues V358 – E363. In pocket 3, all systems exhibited unimodal distribution with minimal differences in the RMSF plot (**Figure 3.8C**). In pocket 4, all systems exhibited unimodal distribution. Interestingly, an increase in RMSF values was observed in the WT, suggesting that the mutation increased the pocket's compactness (**Figure 3.8D**). In pocket 5, all systems exhibited unimodal distribution with minimal differences with increased fluctuation in residues G279 – V281 in mutants N371K, A422T, and N468Y (**Figure 3.8E**). Lastly, in pocket 6, all systems exhibited unimodal distribution with minimal differences with increased fluctuation in residues D353 – Y357 in mutants N371K and A422T (**Figure 3.8F**).

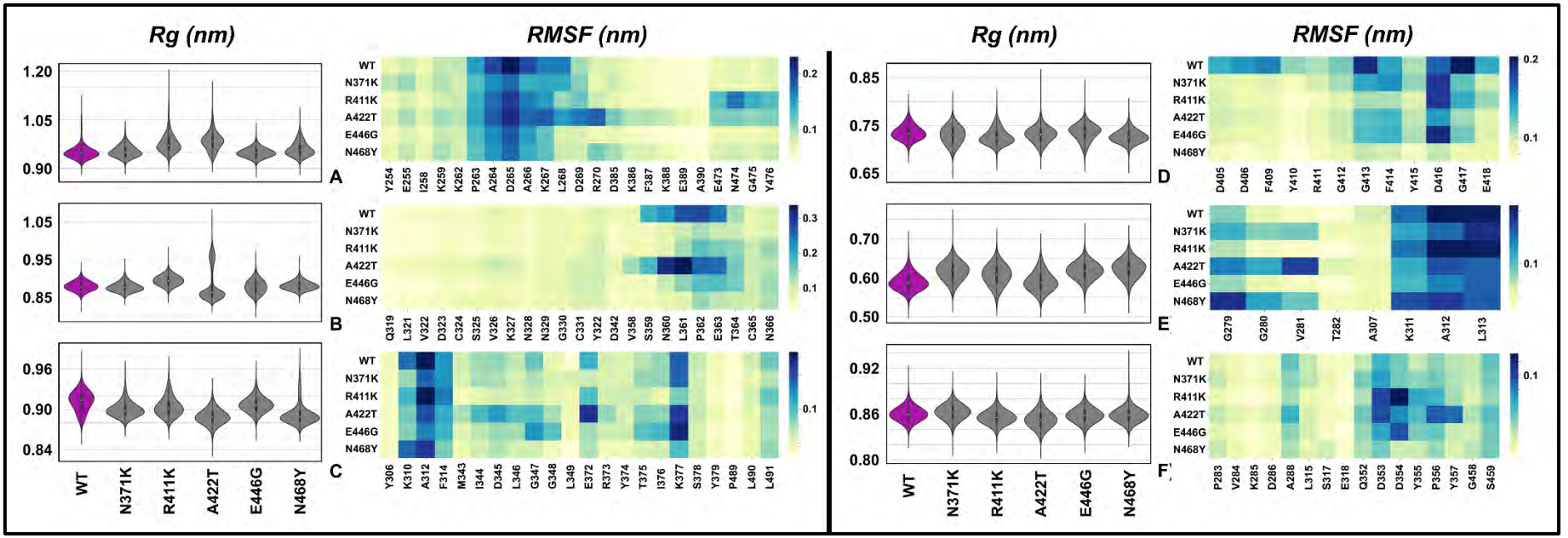



**Figure 3.8: Rg and RMSF values for Pockets 1 – 6 (A – F) of the WT and mutant systems of the catalytic domain of FP-3.** Rg and RMSF are represented as violin distribution plots and heatmaps, respectively. The purple ensemble in the violin plots represents the WT system. The heatmaps show pocket residues on the x-axis and WT and mutant systems on the y-axis.

### 3.4.5.2 Partial zymogen complex

In Pocket 1, WT, I193L, R411K, A422T, and N468Y exhibited unimodal distribution, while E177Q, N371K, and E446G displayed bimodal distribution (**Figure 3.9A**). In comparison to the WT, all mutant systems exhibited higher fluctuations in residues P263 – L268, with mutant E177Q exhibiting extended fluctuations in residues I258 – K262. In Pocket 2, the presence of mutations I193L, N371K, R411K, A422T, and N468Y maintained a unimodal distribution (**Figure 3.9B).** The WT, E177Q, and E446G exhibited a somewhat bimodal distribution. According to the RMSF heatmap plot, A422T displayed no fluctuations in pocket 2 residues, which caused an increase in the compactness of the pocket (**Figure 3.9B**). E177Q had increased fluctuations in residues V326 – Y332. In pocket 3, WT, E177Q, R411K, A422T, and N468Y exhibited a unimodal distribution, while I193L, N371K, and E446G sampled a wider conformation with a flatter distribution (**Figure 3.9C**). E177Q displayed higher fluctuations in residues in pocket 3. In pocket 4, all systems exhibited a unimodal distribution except I193L, A422T, and N468Y, which sampled a wider conformational range (**Figure 3.9D**). In pocket 5, WT, I193L, A422T, E446G, and N468Y displayed a unimodal distribution. E177Q and N371K sampled a wider conformational range while maintaining a flatter distribution (**Figure 3.9E**). Only R411K exhibited a distinct bimodal distribution which correlated to the higher fluctuations in residues K311 – L313. Finally, in pocket 6, all systems displayed a unimodal distribution, with E177Q exhibiting the highest Rg values and higher residue fluctuations (**Figure 3.9F**).

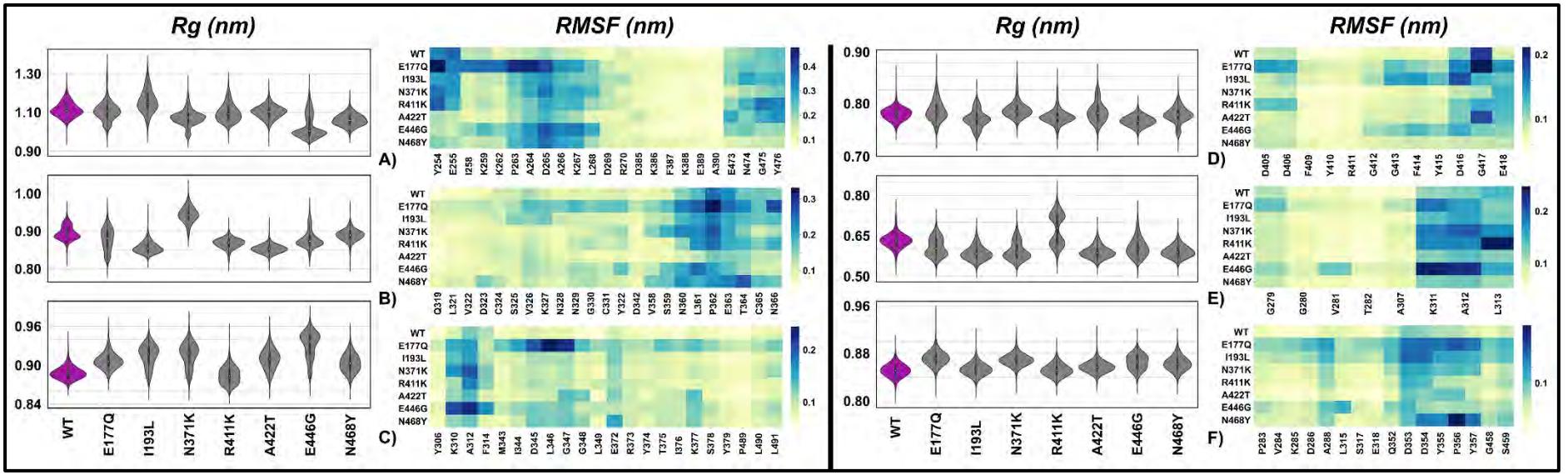



**Figure 3.9: Rg and RMSF values for Pockets 1 – 6 (A – F) of the WT and mutant systems of the partial zymogen complex of FP-3.** Rg and RMSF are represented as violin distribution plots and heatmaps, respectively. The purple ensemble in the violin plots represents the WT system. The heatmaps show pocket residues on the x-axis and WT and mutant systems on the y-axis.

### 3.4.6 Essential dynamics revealed diverse conformational changes in the active site region and identified putative allosteric pockets

In order to examine the biologically important motions and the possible effects of mutations on the active sites and putative pockets, we performed comparative essential using a new tool we developed from MDM-TASK-web [163]. Our analysis was based on the extreme projections of the first and second principal components (PC1 and PC2), which reflect the major protein motion contributions.

#### 3.4.6.1 Catalytic domain

The dominant motions of the active site residues during MD simulation are shown and labelled along PC1 and PC2, and each colour corresponds to a specific trajectory time (**Figure S3.4**). Regardless of the mutations or lack thereof, the lowest energy equilibrium conformation of the catalytic residues indicated a common stable conformation for both states. The percentage variance along PC1 and PC2 (39.73% and 9.48%) explains the minute differences in catalytic residues of the WT and mutant systems.

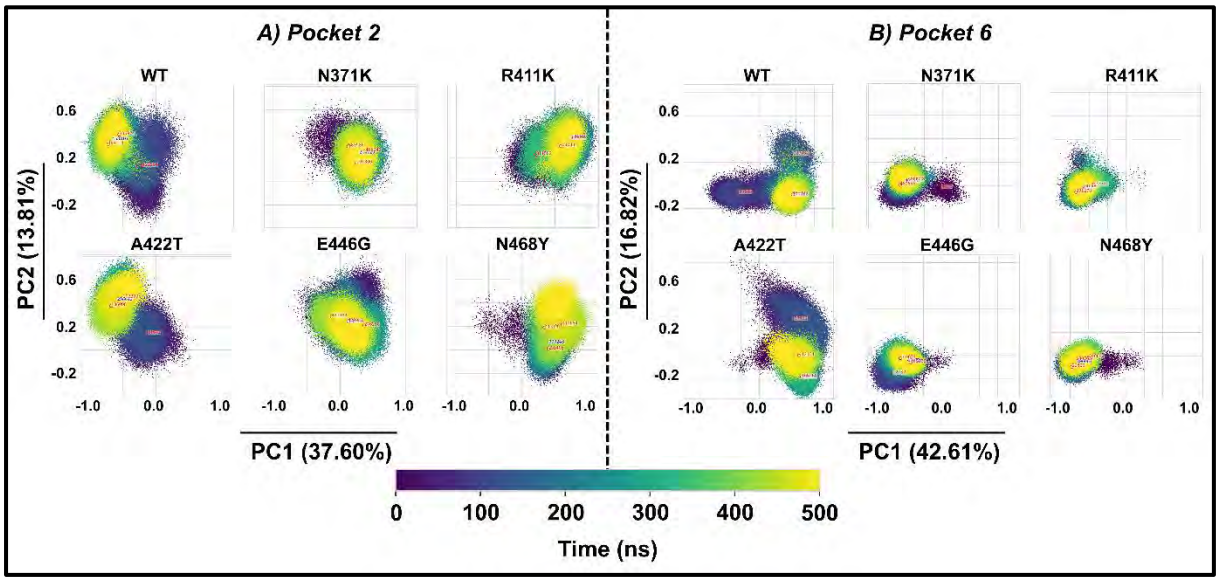



**Figure 3.10: Comparative essential dynamic analysis of pockets 2 and 6 of the WT and mutant systems of the catalytic domain of FP-3 along PC1 and PC2.** The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

As seen along PC1, which accounted for 32.92% of the total variance in Pocket 1, mutant A422T showed the highest conformational divergence, while mutant E446G showed the least

(**Figure S3.5A**). The level of divergence between the WT and mutant systems corresponds with the observed differences in compactness, as seen in **Figure 3.8**.

In Pocket 2, PC1 and PC2 accounted for 37.60% and 13.81% of the total variance, respectively (**Figure 3.10A**). Both axes exhibited a larger conformational distribution in the WT and mutant A422T. An examination of the WT trajectory revealed no conformational changes in this pocket. Observing the A422T trajectory revealed the conformational changes associated with pocket 2. This correlated with the Rg results, which revealed the bimodal conformational sampling of pocket 2 in mutant A422T.

In Pocket 3, PC1 and PC2 accounted for 41.17% and 12.17% of the total variance, respectively. The conformational distribution of WT and mutant systems in Pocket 3 was similar (**Figure S3.5B)**

In Pocket 4, PC1 and PC2 had nearly identical variances, averaging 21.83% and 21.07% of the total variance, unlike other pockets. A closer look at the trajectory revealed that pocket 4 became more rigid due to the mutation (**Figure S3.5C**).

In Pocket 5, PC1 and PC2 accounted for 43.44% and 13.22% of the total variance, respectively. Both WT and mutant systems in Pocket 5 showed similar conformational distributions (**Figure S3.5D**). Visualization of R411K MD simulation revealed another unusual opening of Pocket 5. While residues in Pocket 5 did not move into their neighbouring pocket (3), they maintained an opening and closing motion.

Finally, in Pocket 6, PC1 and PC2 accounted for 42.61% and 16.82% of the total variance, respectively (**Figure 3.10B**). Despite appearing to evolve during the simulation, visual inspection of the WT trajectory revealed no conformational changes as the pocket remained intact. Among the mutant systems measured along PC2, A422T had the highest range of conformational sampling. Upon visualizing the trajectory, we found evidence of a cryptic pocket where Pocket 6 opened, with residues from Pocket 2 migrating to Pocket 6, forming a groove linking both pockets (**Movie S1**). Additionally, analysis of the pose obtained by applying the k-means algorithm of mutant A422T MD simulation at 258,642 ps revealed an unusual opening of pocket 6. The absence of this opening in WT and other mutant systems suggests this pocket is cryptic (**Figure 3.11**). Interestingly, this cryptic pocket was discovered to be present in mutant M245I in FP-2, as described by Okeke et al. [118].

In general, the comparative essential dynamics results of the putative allosteric pocket correlates with the analysis of Rg and RMSF in section **3.4.5.1**.

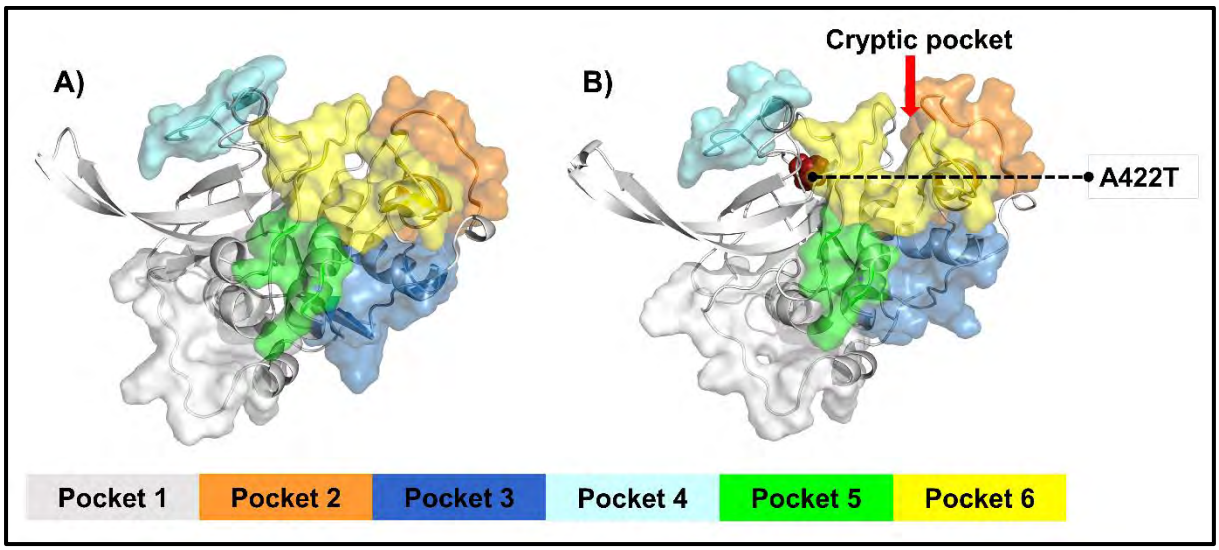



**Figure 3.11: Cartoon representation of A) WT and B) A422T** extracted at 203,390 ps, revealing the presence of the cryptic pocket in M245I and absence in the WT.

### 3.4.6.2 Partial zymogen complex

**Figure S3.6** illustrates and labels the motion of the catalytic residues during MD simulation. The percentage variance along PC1 and PC2 were 37.74% and 15.94%, respectively. E177Q varied significantly along PC1, thus accounting for the changes in the binding site residues. N371K varied along PC1. From this analysis, the presence of mutations affected the binding site. We subsequently analysed the effects of the mutants on the different allosteric pockets.

Pocket 1 analysis revealed the percentage variance along PC1 and PC2 as 34.84% and 17.09%, respectively. Unlike the WT, mutant system E177Q displayed the highest conformational variance along PC1 and PC2. According to the structural analysis of the lowest energy conformation of pocket 1 in E177Q, the surface topology of pocket 1 differs from that of pocket 1 in the WT. In FP-3, several structure-based approaches determined pocket 1 as druggable; therefore, this mutation may interfere with potential allosteric ligand binding (**Figure S3.7A**).

In pocket 2, PC1 and PC2 accounted for 23.96% and 16.31% of the total variance, respectively. A notable variance was observed along PC1 and PC2 in mutant system E177Q (**Figure 3.12A**).

In pocket 3, PC1 and PC2 accounted for 41.58% and 20.07% of the total variance, respectively. **Figure S3.7B** revealed a conformational similarity between the WT and mutant systems, with

the exception of E177Q, I193L, E446G, and N468Y. It was evident from the trajectory visualization that with mutation E446G, pocket 3 disembarked and fused with pockets 5 and 6. One end was fused with pocket 5, while the other was fused with pocket 6 (**Video SX**).

In pocket 4, PC1 and PC2 accounted for 29.57% and 18.50% of the total variance, respectively. Notable differences between the WT and mutant systems were observed, with E177Q displaying the most significant differences (**Figure S3.7C**).

In pocket 5, PC1 and PC2 accounted for 46.25% and 19.09% of the total variance, respectively (**Figure S3.7D**). Significant differences were observed in mutant systems E177Q, I193L, N371K, E446G, and N468Y. Analysis of the trajectory of E177Q revealed an opening and closing mechanism in pocket 5, where residues in pocket 5 moved into its neighbouring pocket (3).

Lastly, pocket 6 analysis revealed that PC1 and PC2 accounted for 35.77% and 21.06% of the total variance, respectively. Mutants E177Q, N371K, E446G, and N468Y varied significantly along PC2 compared to the WT. Visual inspection of these mutant systems revealed the topological changes associated with this pocket (**Figure 3.12B**). However, no cryptic pockets were revealed.

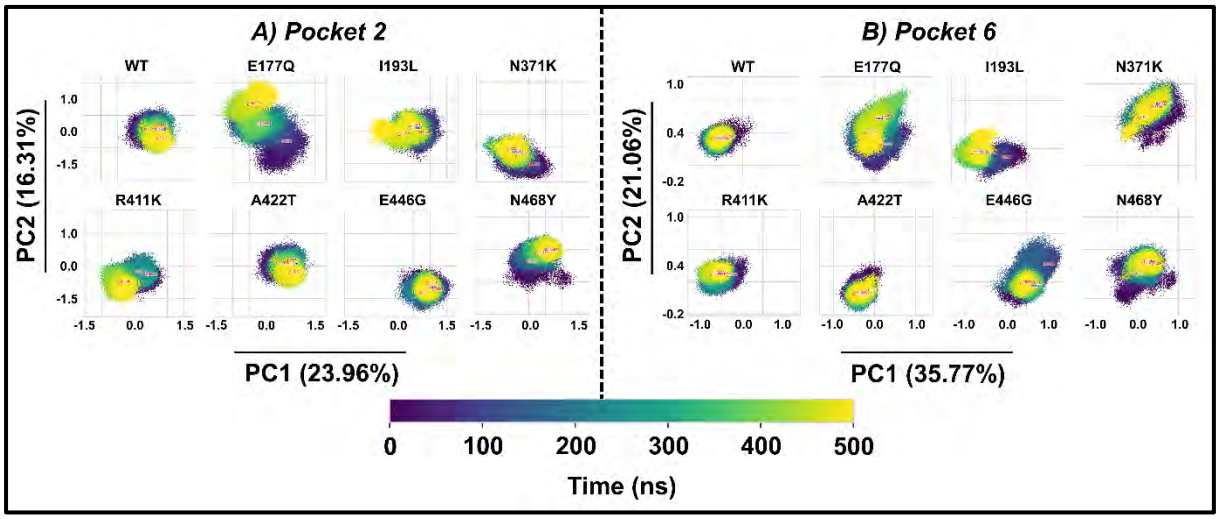



**Figure 3.12: Comparative essential dynamic analysis of pockets 2 and 6 of the WT and mutant systems of the partial zymogen complex of FP-3 along PC1 and PC2**. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

### 3.4.7 Dynamic residue network analysis

Interactions between residues in a protein structure are critical to its functional and structural stability [247]. In order to identify conserved dynamic residue network changes arising from mutations, key residues crucial for protein communication were calculated using five different network centrality metrics in both WT and mutant systems. The metrics used included *BC*, *CC*, *DC*, *EC*, and *KC* centrality implemented on the MDM-TASK-web [163]. We defined the term "*hub*" as any node whose centrality is among the top 5% of all metrics, defined as the top 5% of centrality nodes. For the purpose of understanding the function of identified hubs, we introduced the terms "*persistent hub*" and "*distinctive hub*". The former refers to hubs in both the WT and mutant systems, while the latter refers to hubs absent from the WT but present in the mutant system.

#### 3.4.7.1 Catalytic domain

The raw values of the averaged *BC*, *CC*, *DC*, *EC*, and *KC* of the WT and mutant systems are represented as heatmaps (**Figure 3.13**). Similar to FP-2, most residues within the loop region, including the nose region and the $\beta$-hairpin, exhibited the lowest centrality values, whereas residues within the $\alpha$-helices and $\beta$-sheets that make up the central invariant core showed the lowest centrality values. Group and trench-like binding pockets showed high centrality values (**Figure 3.13**). As a way of increasing the robustness in detecting residues critical in the communication network, we calculated the global 5% of WT and mutant systems for each metric. From the various metrics, several residues were consistently identified as key in communication in the WT and mutant systems.

According to **Figure 3.14**, *persistent hubs* for averaged *BC* were F296, I397, I428, I453, and V488. *Persistent hubs* identified for *CC* include S297, S298, G300, S301, P396, I397, S398, I399, V427, I428, and L429. *DC persistent hubs* were W294, S297, I336, and I397. *EC persistent hubs* were S297, S298, G300, S301, I336, S398, and I428. *KC persistent hubs* were S297, S298, S301, I336, I397, S398, and I428. Upon mapping these hubs to the protein structure, all the identified Persistent hubs are located within the central invariant core of FP-3, where the main binding pocket resides (**Figure 3.15**).

In summary, the heat maps representation of the hubs identified according to the global top 5% in each of the five centrality metrics enabled us to identify key residues unaffected mutations, i.e., persistent hubs.

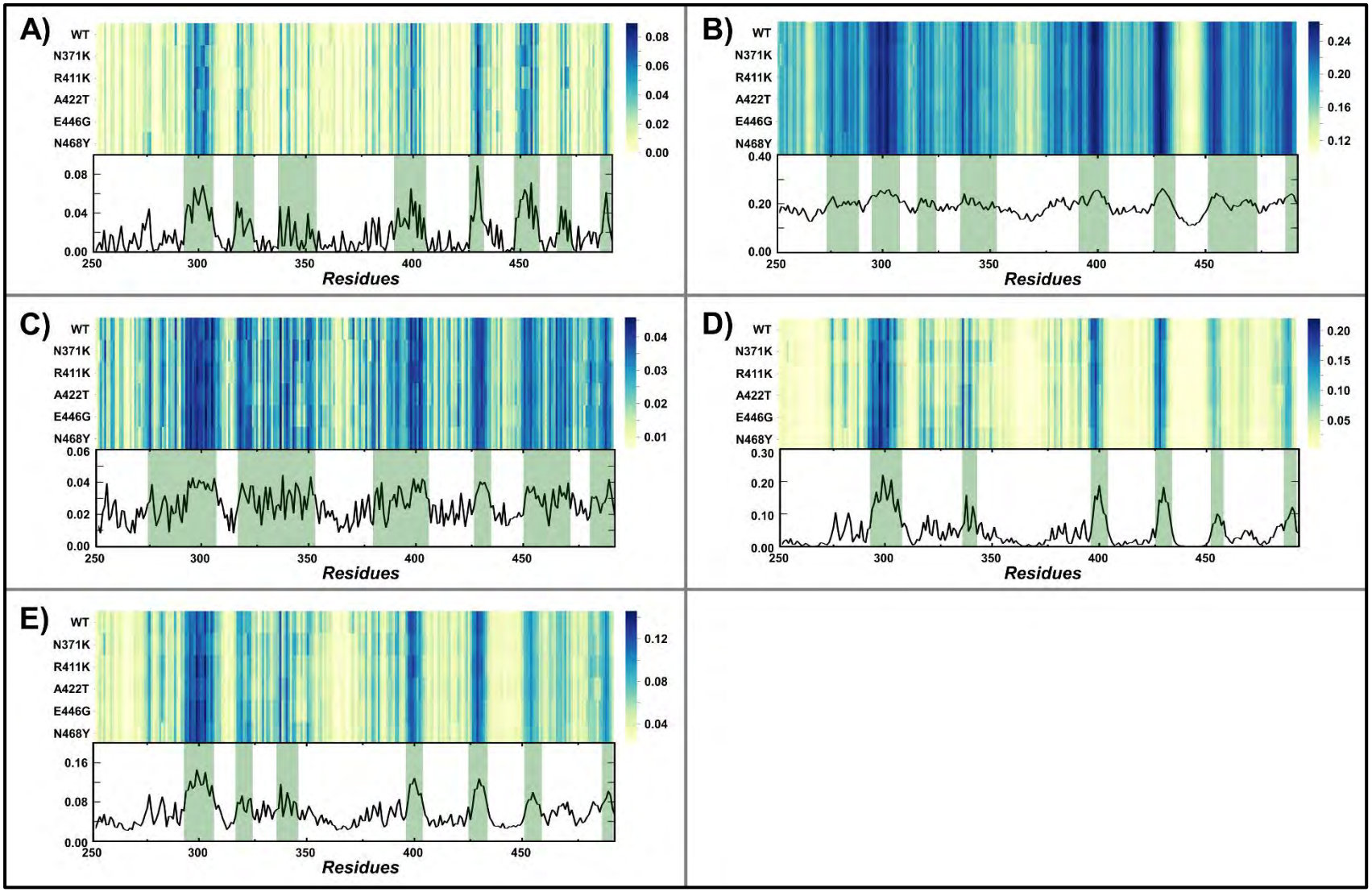



**Figure 3.13: Heatmaps showing per-residue centrality scores of the WT and mutant systems of the catalytic domain of FP-3.** A) *Betweenness centrality*, B) *Closeness centrality*, C) *Degree centrality*, D) *Eigenvector centrality*, and E) *Katz centrality*.

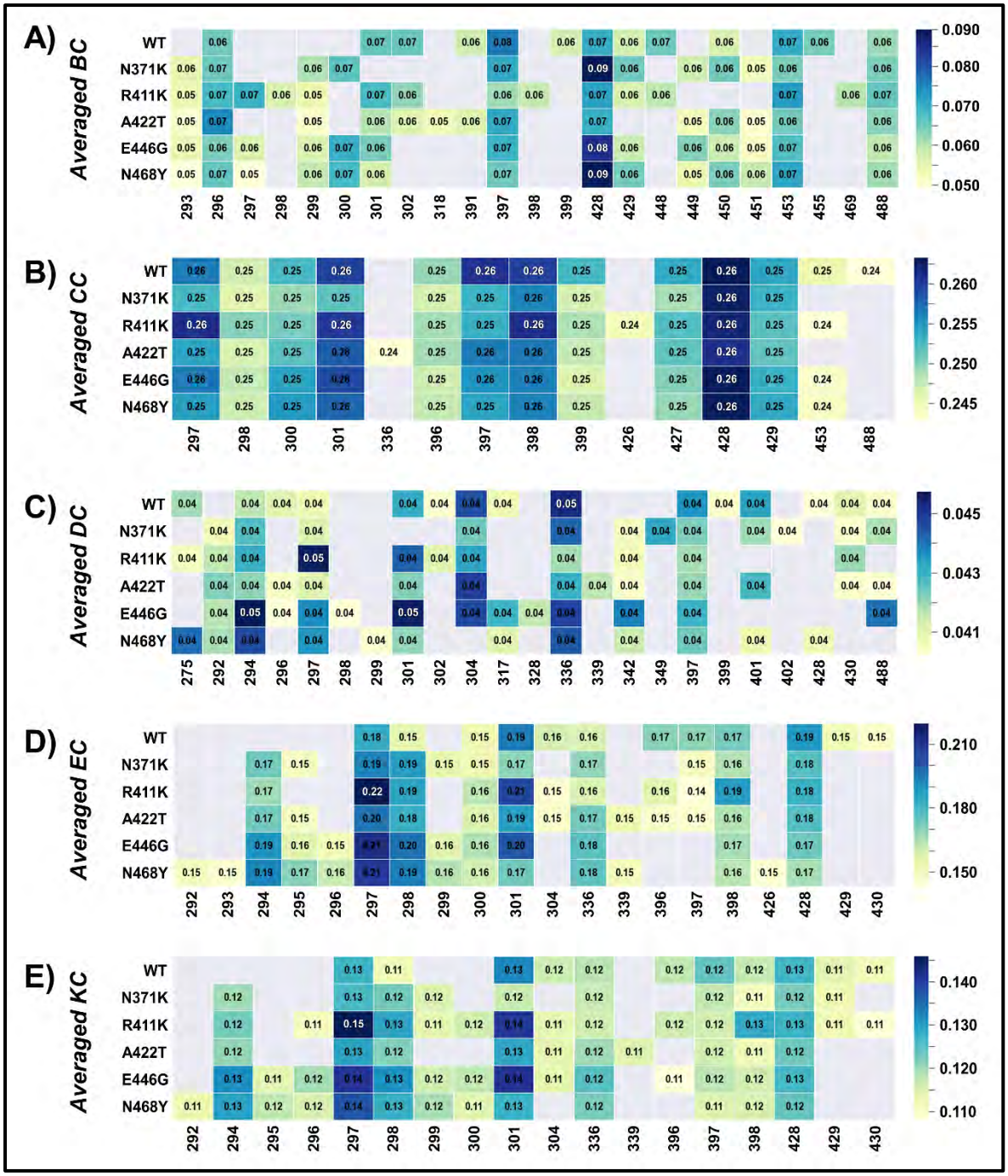



**Figure 3.14: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems of the catalytic domain of FP-3.** Hubs are shown along the x-axis protein systems along the y-axis. Detected hubs are annotated with centrality values.

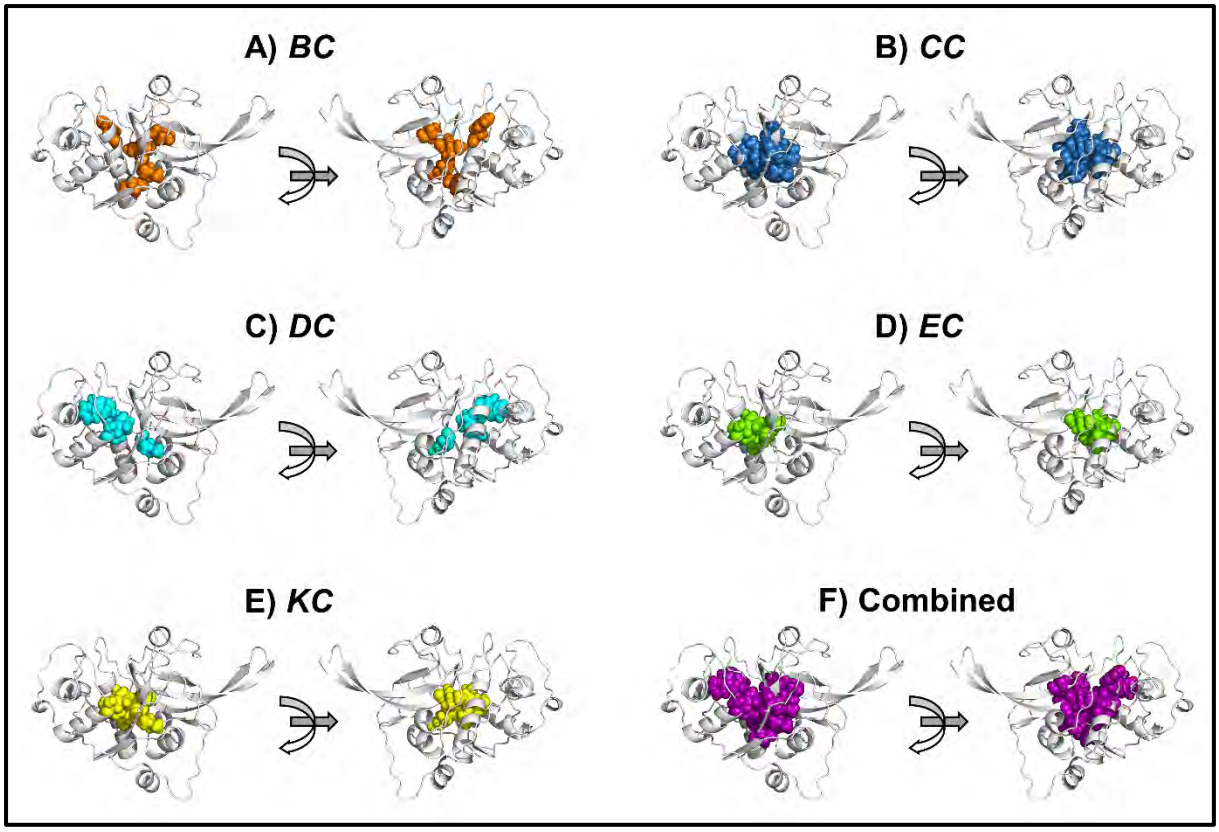



**Figure 3.15: Cartoon representation of the distribution of the persistent hubs in the catalytic domain of FP-3.** Hubs are represented in spheres. (A) Averaged *BC* (orange), (B) Averaged *CC* (sky-blue), (C) Averaged *DC* (cyan), (D) Averaged *EC* (chartreuse), (E) Averaged *KC* (yellow), (F) a collective representation of *persistent hubs* in purple spheres.

***Betweenness centrality (BC)***

According to **Figure 3.14A**, five residues were identified as *persistent hubs,* which implies that the mutation did not alter the communication network of these hubs in the catalytic domain of FP-3. We, therefore, hypothesize that the loss of any of these *hubs* may influence the system's communication. We further analyzed the residues with significantly high *BC* values by calculating the global top 5% *distinctive hubs* of the mutant systems coloured in sky-blue spheres (**Figure 3.16**). Notably, the catalytic thiol (C293) appears as a *distinctive hub* in all mutant systems but is absent in the WT system. V299, located at the central $\alpha$-helix, also appeared as a *distinctive hub* in all mutant systems but was absent in the WT system. Y449 and Y451 were *distinctive hubs* in all mutant systems except mutant R411K. Y449 and Y451 are located in the $\beta$4, part of the $\beta$-hairpin. Notably, I399 and 455 were *distinct* in the WT and absent in all mutant systems. I399 is located in the $\beta$2 around the active site, and N455 is a subsite residue forming part of the active site. S298, S398, and L469 were *distinctive hubs* in

only mutant system R411K. S298 is located in the $\alpha$6, S398 in the $\beta$2, and L469 in the $\beta$5. Interestingly, residue E318 is the only *distinctive hub* specific to mutant A422T. Visual inspection of the location of the *distinctive hub* specific to mutant A422T showed that these hubs (C293, V299, and E318) formed a short communication pathway (**Figure S3.8**). The path originates from the catalytic thiol C293 adjacent to the mutant residue A422T and ends at residue E318. Visual inspection of the relationship between the distinctive hubs and the putative allosteric pockets shows that C293 and V299 are embedded between Pockets 2, 3, and 6, while E318 is located in Pocket 6 (**Figure S3.8**). We hypothesize that this allosteric communication is responsible for the cryptic pocket, as explained in section **3.4.6.1**.

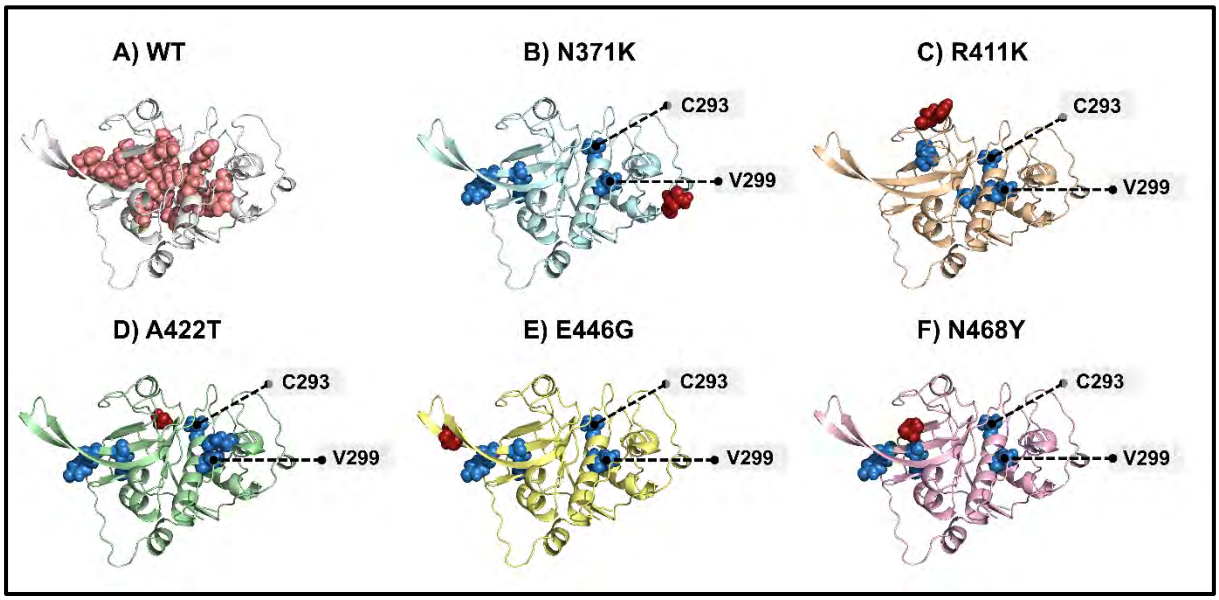



**Figure 3.16: Cartoon representation of the distribution of averaged *BC* hubs in the WT and mutant systems of FP-3.** A) WT containing all identified hubs (salmon spheres) and B-F) distinctive hubs of each mutant system with respect to WT (sky-blue spheres). Catalytic thiol (C293) is labelled with a black stick if identified as a *distinctive hub*. Mutant positions are shown in spheres and colored in firebrick.

### *Closeness centrality (CC)*

*CC* measures the shortest path between node *i* and all other nodes in a protein network. The *CC* hubs were persistent through the WT and mutant systems and were mainly located at the centre of the protein structure (**Figure 3.14B**). We observed no distinctive hubs in the mutant systems. The following *distinctive hubs* were observed in these systems: WT (V488), R411K (A426) A422T (I336). I336 and A426 are subsite 2 residues that form part of the active site. V488 is located in $\beta$6, which is near the active site.

***Degree centrality (DC)***

A notable increase in the number of *hubs* in both WT and mutant systems was observed in *DC* calculations except A422T (**Figure 3.14C**). An analysis of the mutant systems' top 5% distinctive hubs was performed compared to the WT systems. Residues S292 and D342 appeared as *distinctive hubs* in all mutant systems but were absent in the WT system. Additional *distinctive hubs* were observed in several mutant systems. In N371K, L349 and A402 (S1' residue) appeared as *distinctive hubs*. In A422T, A339 (located around the active site) appeared as a *distinctive hub*. In E446G, S298 and N328 (S3 residue) were *distinctive hubs*. Lastly, in N468Y, V299 was the only additional hub observed. Overall, the number of neighbouring nodes around a given node was evident due to the mutants in FP-3.

***Eigenvector centrality (EC)***

The *EC* metric is similar to *DC*, but it takes a step further by considering how connected node *i* is and how many connections it has throughout the communication network. W294 appeared as a *distinctive hub* in all mutant systems (**Figure 3.14D**). Residues L429 and V430 appeared as *distinctive hubs* WT system but were absent in all mutant systems. Additionally, some *distinctive hubs* were discovered for some mutant systems. S295 and V299 appeared as *distinctive hubs* in N371K. In the mutant system, A422T, S295 and A339 appeared as *distinctive hubs*. S296 and V299 appeared as *distinctive hubs* in E446G. In N468Y, S292, C293 (catalytic thiol), S296, V299, A339, and A426 (S2 residue) appeared as *distinctive hubs*.

***Katz centrality (KC)***

*KC* measures the relative influence of a residue within a network. W294 appeared as a *distinctive hub* in all mutant systems. Additional *distinctive hubs* were observed in some mutant systems: V299 in N371K; S296, V299, and G300 in R411K; A339 in A422T; A295, S296, and G300 in E446G; S292, A295, S296, and G300 in N468Y (**Figure 3.14E**).

### 3.4.7.2 Partial zymogen complex

The raw values of the averaged *BC*, *CC*, *DC*, *EC*, and *KC* of the WT and mutant systems are represented as heatmaps (**Figure 3.17**). All systems exhibited high centrality values in $\alpha$-helices ($\alpha$2, $\alpha$3, $\alpha$5 $\alpha$6, $\alpha$7, and $\alpha$9) and $\beta$-sheets ($\beta$2, $\beta$3, $\beta$4, and $\beta$5).

According to **Figure 3.17**, *persistent hubs* for the centrality metrics in the partial zymogen are as follows. *Averaged BC*: F296, I397, I453, N455, W461, and I467. *Averaged CC*: S297, S398, I399, A426, V427, I428, L429, I453, and N455. *Averaged DC*: I336, I397, I401. *Averaged EC*: W294, S297, S298, G300, S301, I336, S398, and I428. *Averaged KC*: W294, A295, S297, S298, S301, I336, A339, I397, S398, and I428.

***Betweenness centrality (BC)***

**Figure 3.18A** shows that six residues remained *persistent hubs* in the partial zymogen complex, indicating that the mutation did not disrupt communication. Interestingly, in the partial zymogen complex, hydrophobic residue W461 was *persistent* in all systems. W461 forms a hydrophobic interaction with W457, which is crucial for auto-processing [80]. As explained in Chapter 2, these hydrophobic interactions are responsible for cleaving the prodomain and subsequent hemoglobin hydrolysis. We further analysed the global top 5% *distinctive hubs* of the mutant systems coloured in sky-blue spheres (**Figure 3.19**). WT hubs are represented as salmon spheres; *distinctive hubs* are represented as skyblue spheres, and mutant positions are represented as firebrick spheres. F228 and Y410 appeared as *distinctive hubs* in most mutant systems in the partial zymogen complex but were absent in the WT system (**Figure 3.19**). At the interface between the inactive zymogen and active enzyme, these residues form hydrophobic clusters with other residues (Y213, F220, Y415, W457, and W461), which are crucial in the auto-processing of the enzyme [195]. In mutant systems of A422T and E446G, hydrophobic residue W457 appeared to be a distinctive hub. This residue forms a hydrophobic interaction with W461, which is crucial for auto-processing [80,195]. In mutants R411K and A422T, the catalytic thiol (C293) appeared as a distinctive hub. Overall, mutations in the partial zymogen complex altered the communication network, affecting the activation of active enzymes and ultimately leading to the accumulation of undegraded hemoglobin.

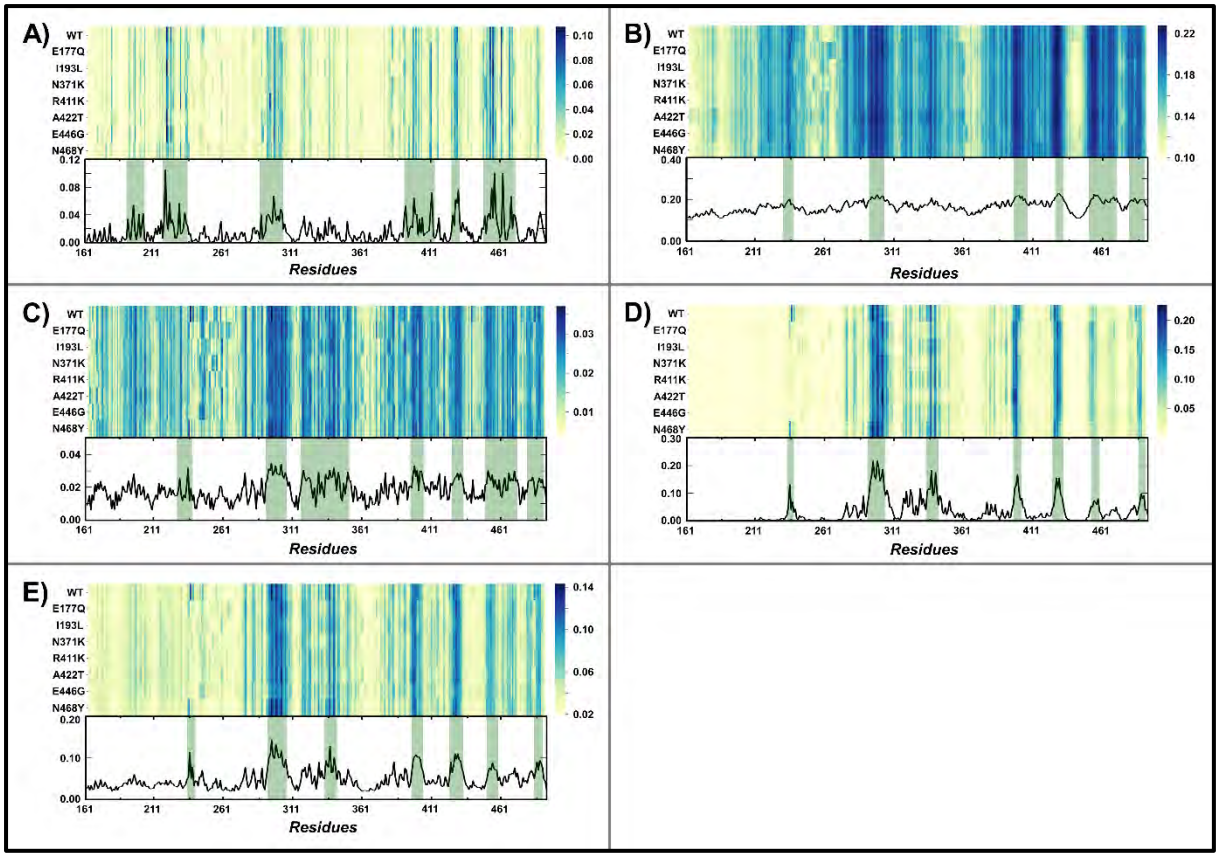



**Figure 3.17: Heatmaps showing per-residue centrality scores of the WT and mutant systems of the partial zymogen complex of FP-3.** A) *Betweenness centrality* (*BC*), B) *Closeness centrality* (*CC*), C) *Degree centrality* (*DC*), D) *Eigenvector centrality* (*EC*), and E) *Katz centrality* (*KC*).

### *Closeness centrality (CC)*

In protein networks, *CC* measures the shortest path between node *i* and all other nodes. CC hubs were mostly located at the center of the protein structure and were persistent through WT and mutant systems. Subsite 1' residue and catalytic thiol residue, H425, was *distinctive* in most mutant systems but absent in the WT (**Figure 3.18B**). Additional *distinctive hubs* were only acquired by mutant systems E177Q and A422T. E177Q has the following *distinctive hubs*: G300, P396, V430, and I467. In A422T, Q287 (Subsite 1 residue and catalytic thiol residue), W294, S298, G300, and I336 (Subsite 2 residue) appeared as *distinctive*.

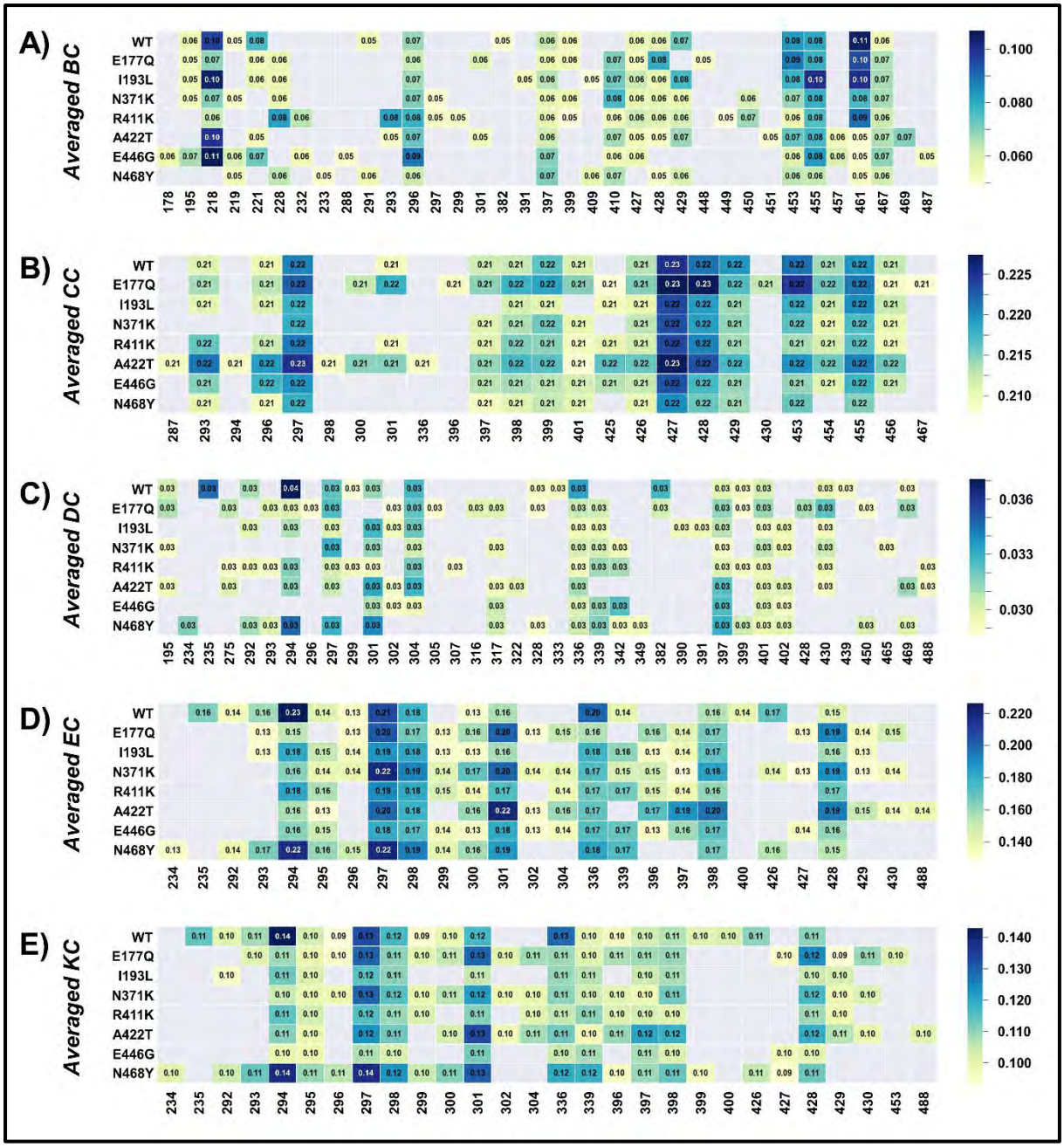



**Figure 3.18: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems of the partial zymogen complex of FP-3.** Hubs are shown along the x-axis, and protein systems along the y-axis. Detected hubs are annotated with centrality values.

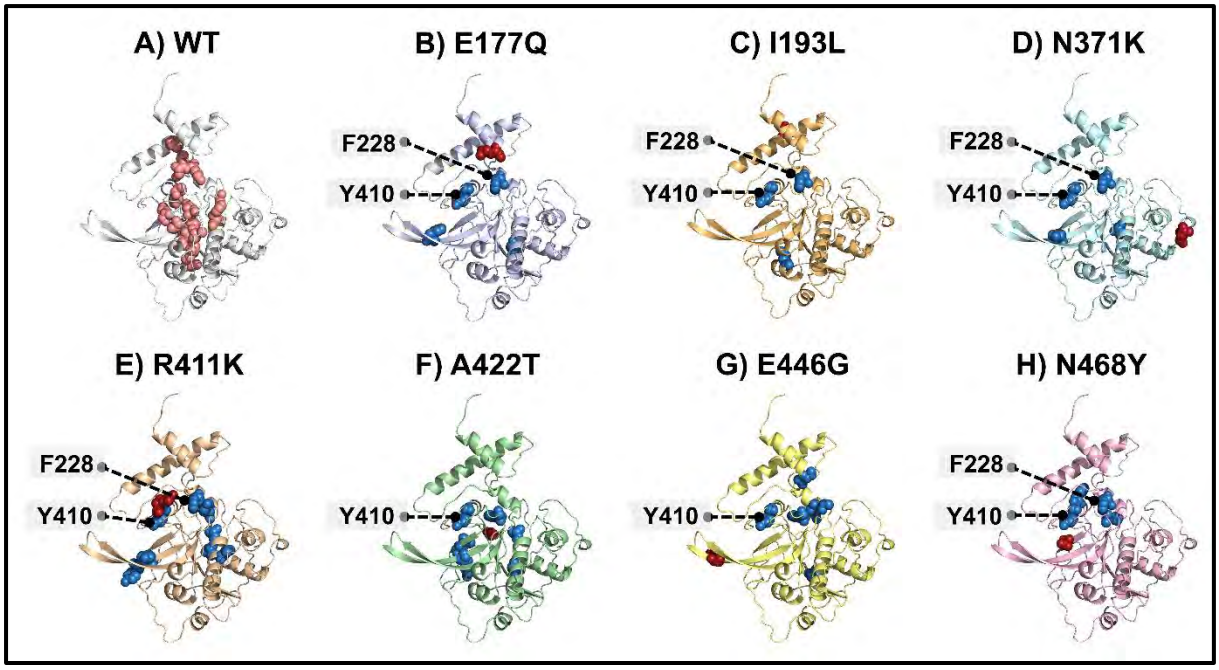



**Figure 3.19: Cartoon representation of the distribution of averaged *BC* hubs in the WT and mutant systems of the partial zymogen complex of FP-3.** A) WT containing all identified *hubs* (salmon spheres) and B – F) *distinctive hubs* of each mutant system with respect to WT (sky-blue spheres). Mutant positions are shown in spheres and colored in firebrick.

***Degree centrality (DC)***

According to the centrality metrics calculated, *DC* had the lowest number of persistent hubs; however, a notable increase in the number of *distinctive hubs* was observed in both WT and some mutant systems (**Figure 3.18C**). In the WT system, L235, G333 (Subsite 3 residue), and N439 appeared as *distinctive hubs*, absent in all mutant systems. The catalytic thiol C293 was identified as a *distinctive hub* in mutants E177Q, R411K, and N468Y. Active site residue A402 appeared as a *distinctive hub* in mutants I193L, N371K, A422T, E446G, and N468Y. A339 also appeared as a *distinctive hub* in all mutant systems, with the exception of A422T. In addition to the highlighted *distinctive hubs*, additional *distinctive hubs* were observed in the various mutant systems. Mutant E177Q had A275, F296, V302, Q305, F316, S317, I428, and Y450. Mutant I193L had V302, A390, and L391. Mutant N371K had S317, D342, and G465. Mutant R411K had W275, A307, D342, and V488. Mutant A422T had W275, V302, S317, V322, and V488. Mutants E446G had V302, S317, and D342. Mutant N468Y had N234, S317, D342, L349, and Y450.

***Eigencentrality (EC)***

The presence of mutations caused a shift in the communication network of the partial zymogen complex of FP-3 as determined by *EC* calculations. In the WT system, L235 and S400 (Subsite 2 residue) appeared as *distinctive hubs* but were absent in all mutant systems (**Figure 3.18D**). Residues V299, V302, S304, P396, I397, V427, L429, and V430 appeared as *distinctive hubs* in most mutant systems and were absent in the WT. **Figure 3.20** highlights the persistent, distinctive, and extinct hub distributions. We defined extinct hubs as hubs present in the WT but lost in the mutant systems. The persistent hubs are represented as salmon spheres, the distinctive hubs are represented as skyblue spheres, and the extinct hubs are represented as green spheres. The distribution of distinct hubs within each mutant system and their communication paths were observed. As mentioned in chapter **3.4.6.2**, visual inspection of mutants E177Q and E446G trajectories revealed an interaction between Pockets 3, 5, and 6. To understand the relationship between the communication network of residues and its effect on the potential allosteric pockets listed above, we mapped the *distinctive hubs* of mutants E177Q, N371K, and E446G on their respective proteins (**Figure 3.20**). In the presence of mutation E177Q, a communication path was observed. This path originates from V299 - V302 - S304 - P396 - I397 - L429 - V430 - V427 (**Figure 3.20B**). Visual inspection of the relationship between the *distinctive hubs* and the putative allosteric pockets shows that V299, V302, S304, P396, and I397 are embedded between Pockets 3, 5 and 6, while L429, V430, and V427 are found around the binding pockets. In the presence of N371K, a communication path was established. This path originates from V299 - V302 - S304 - P396 - I397 - L429 - V430 - V427 (**Figure 3.20C**). Next, we looked at mutant E446G and its allosteric communication path. Recall that in Chapter **3.4.6.2**, we established a relationship between pockets 3, 5 and 6, where a section of Pocket 3 disembarked and fused with Pocket 5 and Pocket 6, where one end fused into Pocket 5 and the other end to Pocket 6. We hypothesize that the *distinctive hubs* identified by *eigencentrality* could be associated with the relationship between pockets 3, 5 and 6. The communication path that occurs in the presence of E446G is shown in **Figure 3.20D**. This path originates from V299 - V302 - S304 - P396 - I397 - V427. Visual inspection of the relationship between the distinctive hubs and the putative allosteric pockets shows that V299, V302, S304, P396, and I397 are embedded between Pockets 3, 5 and 6, while V427 are found around the binding pockets. *Distinctive hubs* were observed in the remaining mutant systems. I193L: V299, P396, I397, and L429. R411K: V299, S304, P396, and I397. A422T: V302, S304, P396, I397, L429, V430 and V488. N468Y: N234 and V299.

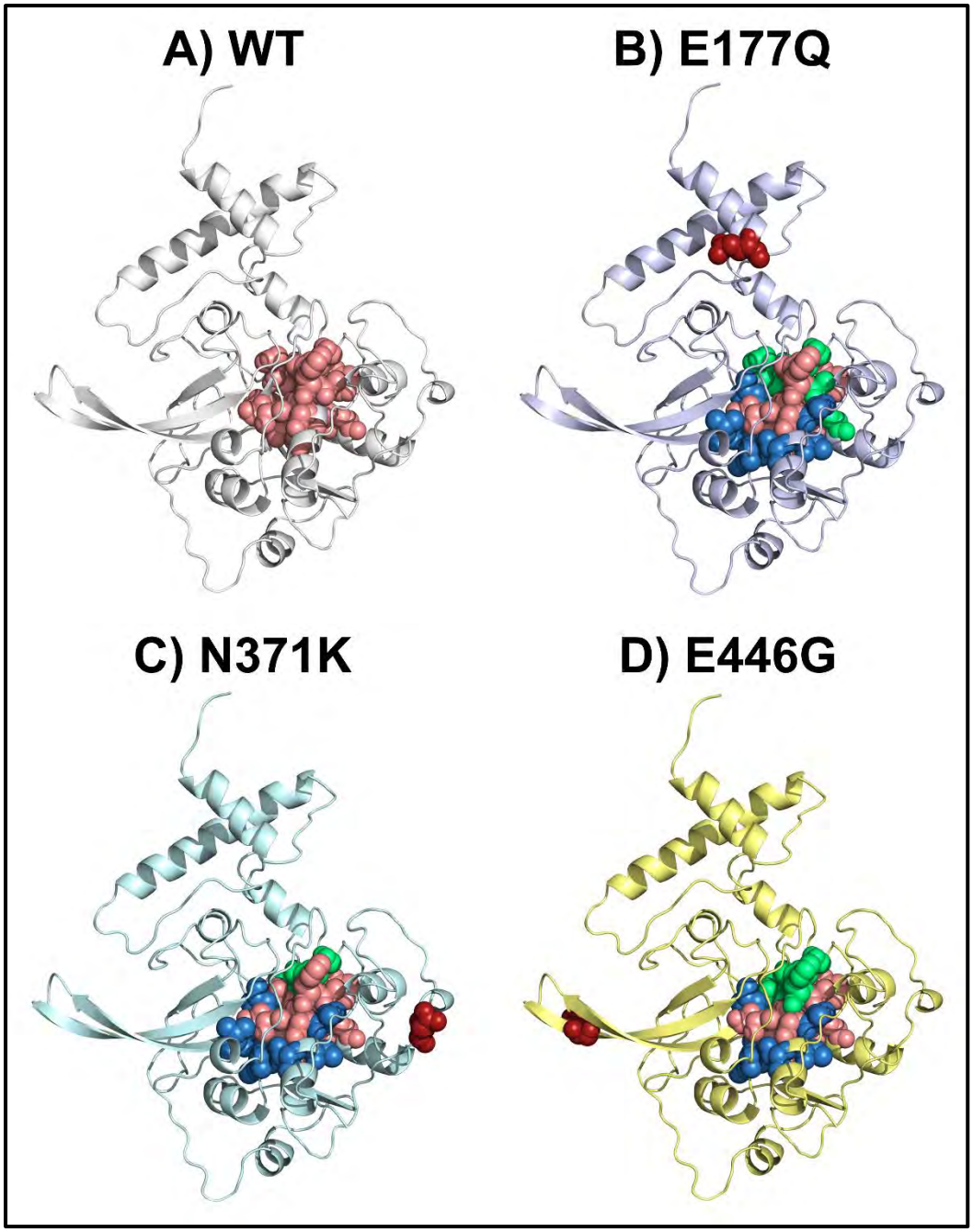



**Figure 3.20: Cartoon representation of the distribution of the averaged *EC* hubs of the WT and mutant systems E177Q, N371K, and E446G.** A) The distribution of the averaged *EC hubs* of WT (salmon spheres). B) E177Q, C) N371K, and D) E446G with their *distinctive*, *persistent*, and *extinct hubs*. *Distinctive hubs* are represented as sky-blue spheres, persistent hubs common to the WT hubs are represented as salmon spheres, and extinct hubs present in the WT and absent in the mutant system are represented as green spheres. Mutant positions are represented as firebrick spheres.

***Katz centrality (KC)***

A similarity between the *EC* and *KC* metrics was observed. Similar *persistent hubs* were identified by both metrics, with the *KC* metrics identifying A295, I397, and A339 as additional *persistent hubs* (**Figure 3.18E**). In mutant system E177Q, *distinctive hubs* V302, S304, V427, L429, V430, and I453 were observed. In mutant systems I193L, E446G, and N468Y, V427 was the only observed *distinctive hub*. In the mutant system, N371K, V302, S304, L429, and V430 were observed as *distinctive hubs*. In R411K, S304 and L429 were observed as *distinctive hubs*. In A422T, V302, S304, ;429, V430, and V488 were observed as *distinctive hubs*.

Using the five centrality metrics, we analyzed the effect of mutations on the catalytic domain of FP-3 structure by identifying *persistent* and *distinctive hubs*—these *hubs* contribute to the intra and inter-communication within the WT and mutant systems.

## 3.4.8 Residue interaction changes due to mutations identified using contact map analysis

An analysis of weighted contact maps was conducted to investigate the interactions between residues surrounding the mutated residues. Contact maps which consist of interactions such as van der Waals, hydrogen bonds and electrostatic bonds, were calculated using MD-TASK [165]. A residue can be identified with loss, gain, or reduced interactions if this analysis assigns a value ranging from 0 to 1, where 0 signifies the absence of interaction and 1 signifies the presence of interaction. The resulting contact map of the mutant systems was compared to the WT, and a heat map was prepared based on the contact frequency.

### 3.4.8.1 Catalytic domain

In the catalytic domain of FP-3, we observed notable changes around mutant systems A422T, E446G, and N468Y (**Figure 3.21B**). Compared to the WT, the amide-amide interaction between residues A422 and G420 decreased by two-fold in the A422T mutant system. Additionally, active site residue N424 interaction increased in A422T mutant system. A slight difference was observed in the contact frequency between A422 and L469. In comparison to the WT, the interaction between E446 and D436 increased four-fold in the E446G mutant system. Additionally, interactions between N439 and F448 decreased respectively in E446G mutant system. Finally, subtle changes in the interaction of E463 in both WT and mutant system N468Y were observed (**Figure 3.21B**).

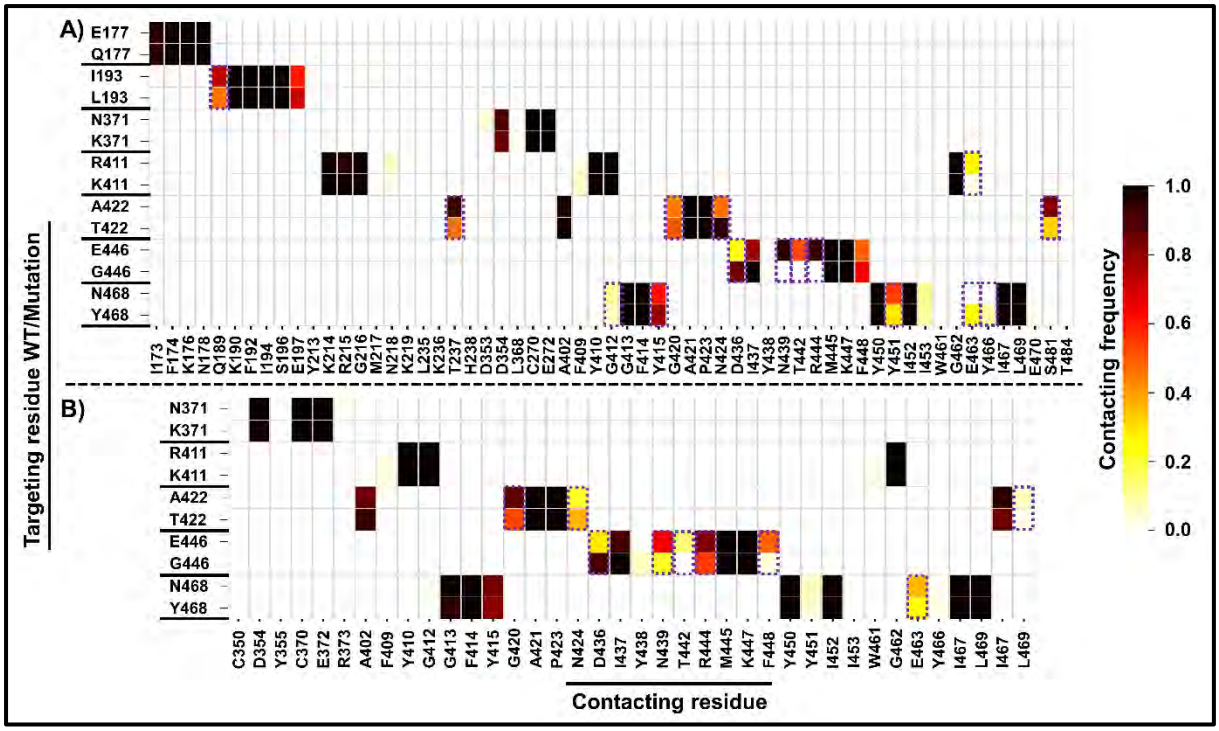



**Figure 3.21: Residue contact analysis heatmaps of the WT and mutant systems of the partial zymogen complex of FP-3.** Heatmaps show the interaction changes around the mutations in (A) partial zymogen complex and (B) catalytic domain of FP-3. Dashed boxes show mutations causing significant changes in the interaction between the mutated residues and their neighbouring residues compared to the WT. The colour code represents the contact frequency, with 0 representing no contact and 1 representing strong contact between the target and the contacting residue.

### 3.4.8.2 Partial zymogen complex

In the partial zymogen complex of FP-3, as compared to the WT, only mutants I193L, A422T, E446G, and N468Y exhibited significant changes in the interaction network (**Figure 3.21A**). Compared to the WT, the interaction between I193 and Q189 decreased by two-fold in the I193L mutant system. Compared to the WT, some residues gain and loss interactions were observed in A422T. The interaction between A422 and T237 decreased three-fold in the A422T mutant system. The amide-amide interaction between A422 and N424 increased three-fold in the A422T mutant system. The interaction between A422 and S481 decreased four-fold in the A422T mutant system. The following interactions occurred in mutant E446T; E446 and D436 increased by four-fold, E446 and N439 decreased by ten-fold, E446 and T442 decreased by six-fold, and E446 and M445 decreased by ten-fold. Finally, in mutant N468, the interaction between N468 and Y451 decreased two-fold.

Besides evaluating the immediate mutation effects on the mutated residues, contact maps around the identified *BC distinctive hubs* (F228 and Y410) were determined. According to **Figure 3.22**, the presence of mutations reduced the contact frequency between F228 and F220. Another reduction in contact frequency was observed in F228 and K231. An increase in contact frequency was observed in F228 and L233. An increase in contact frequency was observed between F213 and Y213 and F410 and R215. In mutant systems, R411K, A422T, and E446G, contact frequency were increased in Y410 and W461. In these mutant systems, a reduction in contact frequency was observed in Y410 and Y415 and Y410 and I467. Additionally, in mutant system E177Q, a reduction in contact frequency between F228 and A408 and an increase in contact frequency between Y410 and A408 were observed. Overall, contact map analysis establishes the mutations' effect on the interactions between the interface residues.

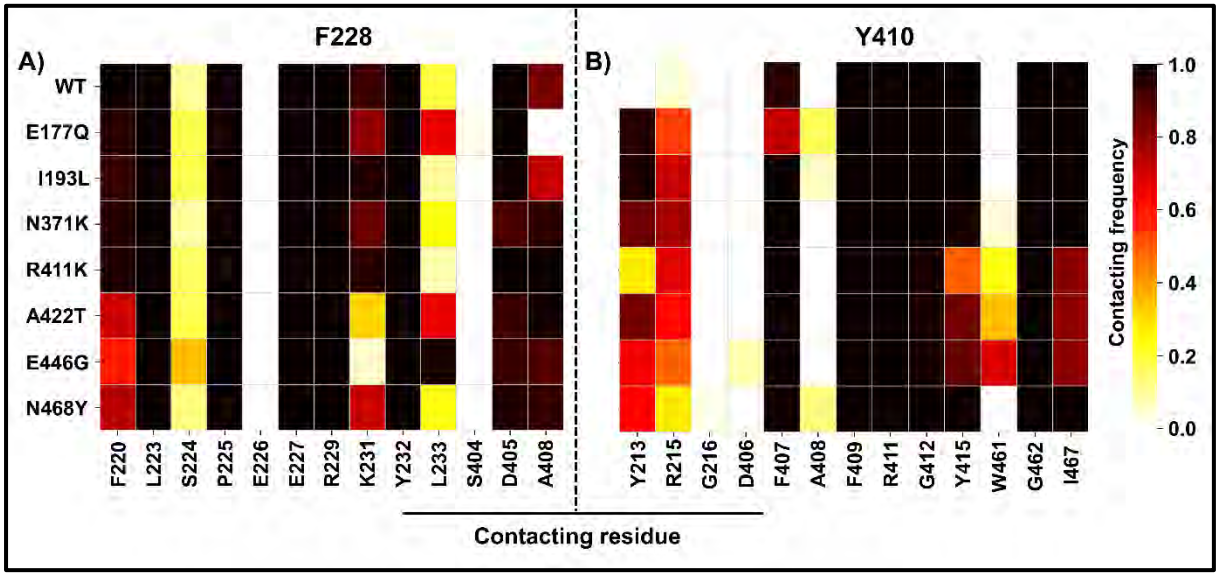



**Figure 3.22: Residue contact analysis heatmaps of the distinctive hubs of the WT and mutant systems of the partial zymogen complex of FP-3.** Heatmaps show the interaction changes between *distinctive hubs* A) F228 and B) Y410. The colour code represents the contact frequency, with 0 representing no contact and 1 representing strong contact between the target and the contacting residue.

### 3.4.9 Destabilization of prodomain interface residues by computational alanine scanning

We applied alanine scanning using the ROBETTA web server [243] to determine the effect of mutations on the interface residues on the partial zymogen complex of FP-3. The *distinctive hubs* (F228 and Y410) identified by the *BC* metric were in the interface region; hence we performed alanine scanning to assess their overall binding energy on the mutant systems. **Table 3.2** summarizes the destabilizing interface residues in mutant systems. The mutations in the

partial zymogen complex exhibited a destabilization effect on the interface residues (**Table 3.2**). Hydrophobic residues Y213, F220, F228, Y410, Y415, and W457 associated with activating the mature enzyme were identified as destabilizing residues. Interestingly, destabilization F228 and Y410 were identified as distinctive *BC* hubs. Overall, the effect of mutation on the protein structure via computational alanine scanning.

**Table 3.2: Destabilizing residues in the mutant systems** identified by Alanine scanning. Hydrophobic clusters are bold, while common residues between alanine scanning and *BC* metric are bold red.

| System | Destabilizing residues |
|---|---|
| E177Q | 212, **213**, 215, **220**, **228**, 232, 233, 234, 236, 244, 255, 332, 335, 337, 400, 406, 409, 425, **457**, 485, 487 |
| I193L | 212, **213**, 215, 218, 219, **220**, **228**, 231, 232, 233, 235, 238, 243, 244, 254, 335, 337, 379, 406, 409, **410**, **415**, **457** |
| N371K | 212, **213**, 215, 218, 219, **220**, **228**, 232, 233, 234, 237, 241, 250, 254, 335, 379, 400, 409, **410**, 411, 425, **457**, 460, 485, 486 |
| R411K | **213**, 215, **218**, 219, **220**, **228**, 231, **232**, 233, 234, 242, 243, 244, 250, 254, 294, 332, 335, 381, 409, 411, 416, **457**, 486 |
| A422T | **213**, 215, **218**, **220**, **228**, 232, 233, 234, 238, 241, 243, 244, 249, 289, 335, 409, 411, 425, **457**, 485 |
| E446G | 201, 211, **213**, 215, 218, **220**, **228**, 229, **232**, 233, 236, 242, 244, 257, 335, 405, 406, 409, 425, **457**, 486 |
| N468Y | 212, **213**, 218, 219, **220**, **228**, 229, 232, **233**, 234, 241, 243, 244, 248, 250, 253, 294, 335, 338, 341, 379, 405, **409**, 416, 425, **457** |

## 3.5 Conclusion

This chapter systematically investigated the structural and dynamic effects of twelve missense mutations associated with the partial zymogen complex and the catalytic domain of FP-3 using a range of computational tools we developed recently [119,163,165] in conjunction with existing methods. Structural visualization of the distribution of the mutations in FP-3 revealed the position of the mutations. All mutations were distally located from the active sites with the exception of A422T, which occurs around the binding site. From the identified putative pockets, only mutant R411K occurred in pocket 4. As a subsequent step, we carried out MD

simulations observed through post-MD evaluation approaches. We analyzed the global dynamic effects of the whole protein via RMSD, Rg, and RMSF, the binding pocket, and potential allosteric pockets of the WT and mutant proteins. Furthermore, we combined the results of the pocket analysis with one of our newly developed tools, the comparative essential dynamics tool. Observations from the combined global analysis in the partial zymogen complex of FP-3 include 1) an opening and closing mechanism in pocket 5 in mutant E177Q. 2) The interaction of pockets 4, 6, and 7 in mutant E446G. Observations from the combined global analysis in the catalytic domain of FP-3 include 1) discovering a cryptic pocket via a unique mechanism where some residues in Pocket 3 moved into Pocket 7 in mutant A422T. 2) in the presence of R411K mutation, an unusual opening of Pocket 6 was observed. We evaluated key communication residues of FP-3 and its allosteric behaviour with mutations using the five centrality metrics (*BC*, *CC*, *DC*, *EC*, *KC*). These approaches identified conserved communication network changes between WT and mutant systems. Most importantly, we identified several persistent hubs across the WT and mutant systems. We previously defined a hub as any residue comprising residues with the highest centrality across all metrics (global top 5%). These hubs are considered functionally important as they are unaffected by mutations. Some *persistent hubs* identified were specific to each centrality metric, while others were common among the centrality metrics. We also identified *distinctive hubs* specifically present in the mutant systems and absent in the WT. A summary of the DRN observations includes 1) The identification of F228 and Y410 as *distinctive hubs* in mutant systems in the partial zymogen complex. These residues form a hydrophobic cluster with other residues (Y213, F220, Y415, W457, and W461) at the interface between the inactive zymogen and active enzyme [195]. 2) Identifying hydrophobic residue W457, a *distinctive hub* mutant systems of A422T and E446G. This residue forms a hydrophobic interaction with W461, which is crucial for auto-processing the inactive zymogen complex to the active form [80]. 3) Identification of an allosteric path in mutant A422T in the catalytic domain. From this observation, we hypothesize that mutations in the partial zymogen complex affect the communication network.

# CHAPTER 4

# *In silico* identification and analysis of novel allosteric modulators of Falcipain 2 and 3 in conjunction with its missense mutations.

## 4.1 Introduction

Despite the combined efforts toward the successful global elimination of malaria, the spontaneous acquisition of mutations by enzymes in plasmodial parasites, which act as pharmacological targets, remains the greatest threat [248]. Thus, novel antimalarials with unique mechanisms of action are needed to circumvent resistance. The primary active sites (also known as orthosteric site) of antimalarial target proteins has historically been the main target of competitive treatment efforts for malaria [249]. However, the emergence of mutations and their ability to alter the shape of the orthosteric pocket, which in turn leads to the loss of favourable hydrophobic interactions, subsequently lowers the binding affinity of inhibitors and highlights the need for an alternative avenue in drug discovery [250].

## 4.2 Allosteric regulation of proteins

Allosteric regulation of proteins using small molecules that bind on distal sites other than the main active site has revolutionalised the drug discovery process [251,252]. Allosteric regulation is a strategy for controlling cellular processes by regulating the affinities of biomolecules [253]. It is the process by which biomolecules (primarily proteins) transfer the effect of binding at one site (often distal) to another (functional site), thus, regulating its activity [254]. The orthosteric site is the unique binding site each receptor possesses for its endogenous ligand(s), and this site is highly conserved across receptor subtypes [255]. Allosteric sites differ structurally and functionally from their corresponding receptor orthosteric sites [255]. Compared to orthosteric sites, targeting allosteric sites has some advantages, such as target specificity, high levels of selectivity, and low toxicity. [256]. Allosteric ligands can exhibit a variety of pharmacological modes, including positive allosteric modulators (PAMs) that increase agonist-mediated receptor response and negative allosteric modulators (NAMs) that lower agonist-mediated receptor response in a non-competitive manner [255]. The effects of PAMs or NAMs can be inhibited by silent allosteric modulators (SAMs) or neutral allosteric ligands (NALs), which bind to allosteric sites but have no impact on the responses to orthosteric

ligands [255]. Many essential proteins use allosteric regulations to control their activities [250]. Among the best examples of allosteric regulation is the negative allosteric modulation of hemoglobin by 2,3-bisphosphoglycerate (2,3-BPG) [257]. Hemoglobin (Heme + Globin) is a molecule that transports nearly all of the oxygen in the red blood cells (RBCs) of all vertebrates [258]. It is made up of four subunits, each containing a heme group and a globin chain [258]. 2,3-BPG binds to hemoglobin, reducing its affinity for oxygen and increasing the efficiency of oxygen transport [253]. Oxygen, on the other hand, positively allosterically modulates hemoglobin [253]. When oxygen binds to one subunit of hemoglobin, a conformational change occurs, increasing the oxygen affinity of the remaining active sites and enabling hemoglobin to carry more oxygen [253]. Allosteric control also play a leading role in cell signalling, enzyme activities, and biomolecule transport [253,259].

In Chapter 1, we established the importance of FP-2 and FP-3 in the erythrocytic cycle of malaria parasite, which is responsible for the clinical manifestation of the disease. Due to the importance of falcipains in the parasite's development and metabolism, multiple active site inhibitors including have been E-64 [56,260], falstatin (an endogenous macromolecular inhibitor) [261,262], leupeptin [56,263], peptidyl fluoromethyl ketones [264,265], and vinyl sulfones [85,266,267] has characterized [195]. Research has extensively focused on inhibiting the active site of falcipains [195]. However, recent research on the identification and characterization of allosteric inhibitors of FP-2 has emerged [268,269]. Studies associated with other diseases have shown that small molecule inhibitors and peptides can potentially target exosites with high selectivity and potency [195]. When an inhibitor occupying the ligand binding site (distinct from the enzymatic active site) of human amyloid precursor protein cleaving enzyme (BACE) is present, it inhibits proteolysis for amyloid precursor protein (APP)-related, protein-based substrates of BACE [270]. The allosteric sites of FP-2 and FP-3 in the presence and absence of mutations have not been explored for inhibitory design; hence this research will provide a novel insight into a search for potential allosteric compounds effective against the WT and mutant systems.

## 4.3 Proposed work

In chapters 2 and 3, we identified six potential allosteric pockets in FP-2 and FP-3, respectively, using various computational methods. Pocket 1 was one of the six putative allosteric pockets deemed druggable in FP-2 and FP-3. To find potential allosteric modulators that might bind to pocket 1, we screened 2089 FDA compounds retrieved from DrugBank. In this chapter, pocket

1 of FP-2 and FP-3 in the presence and absence of mutations were explored for an allosteric inhibitory design using high-throughput virtual screening combined with MD simulations and DRN analysis. A subset of 2089 compounds from the DrugBank database was screened against the WT of FP-2 and FP-3. Subsequently, the stability of selected compounds was first explored via all-atom MD simulations. In the next step, compounds that showed high levels of stability during simulations were pooled and evaluated against all mutant systems. Post-MD analysis, including comparative essential dynamics and DRN analysis, was utilized to evaluate the effects of the ligand on the respective protein systems.

## 4.4 Methodology

### 4.4.1 Data Retrieval and protein preparation

The protein template of FP-2 [PDB ID: 2OUL] [196] and FP-3 [PBD ID: 3BWK] [85] was retrieved from the RCSB PDB [197,198] and modeled as described in chapters 2 and 3, respectively. Through the PlasmoDB version 9.3 [102], we identified 11 mutations in the catalytic domain of FP-2 and five in the catalytic domain of FP-3. Each mutation was introduced to its respective protein structure using the mutate option in the Discovery Studio software package [271]. All mutant structures were protonated to pH 5.5 using the PROPKA tool from PDB2PQR (version 2.1.1) [208,209] to correspond to the acidic environment in the food vacuole. The crystal structures of the catalytic domain of the human cathepsins Cat K [PDB ID: 3OVZ] [272] and Cat L [PDB ID: 3OF8] [273] were retrieved from the RCSB PDB [197,198] and prepared as previously described by Musyoka et al. (2016) [87]. In Chapters 2 and 3, we identified six potential allosteric pockets (Pockets 1 - 6) in FP-2 and FP-3; however, only Pocket 1 will be explored in this chapter. We also identified potential allosteric pockets in Cat K and L using Allosite [210], FTMAP [211], and SiteMap [213]. In an effort to discover allosteric compounds specific to the *Plasmodium* protein, compounds binding to the active and allosteric pockets of the human homologs will be excluded.

### 4.4.2 Preparation of compound library

The compounds employed for molecular docking in this chapter were prepared by Olivier Sheik Amamuddy and Rita Afriyie Boateng [274]. A library of 2705 FDA-approved compounds was downloaded in smile format from the DrugBank database [275]. The metal-bound complexes and halogen-containing compounds were excluded from the downloaded

dataset. The resulting 2089 compounds were protonated and constructed via the RDKit library [276].

### 4.4.3 High-throughput virtual screening

A prerequisite for molecular docking is the preparation of the protein structures. Gasteiger-Hückel protocol in AutoDockTools (ADT) [277] assigned partial charges, atom types, and polar hydrogen to the WT proteins (FP-2, FP-3, Cat K, and Cat L). The entire surface of the protein structures was subjected to high-throughput virtual screening. For FP-2, a docking box size of 80 × 85 × 65 Å with a grid spacing of 0.375 Å was estimated using the AutoDock Vina plugin [278] in PyMOL [199] and centred at (x=121.18, y=83.12, and z=-191.00) coordinates. For FP-3, a docking box size of 80 × 85 × 65 Å with a grid spacing of 0.375 Å was estimated centred at (x=4.25, y=-10.34, and z=42.33) coordinates. For Cat K, a docking box size of 80 × 85 × 65 Å with a grid spacing of 0.375 Å was estimated centred at (x=6.09, y=-23.91, and z=-13.17) coordinates. For Cat L, a docking box size of 80 × 85 × 65 Å with a grid spacing of 0.375 Å was estimated centred at (x=22.53, y=19.04, and z=0.62) coordinates. In addition, an exhaustiveness of 320 was used, and the maximum number of docking poses was set at nine. Using the QuickVina-W program [279], blind docking simulations were performed at the Center for High Performance Computing (CHPC).

### 4.4.4 Post High-throughput virtual screening filter

After docking, nine poses per ligand were generated as a single file, which was then split into separate poses using the *vina_split* scripts. The best pose with the lowest binding energy was retained for further analysis. Ligands binding within the potential allosteric pocket and interacting with allosteric residues were filtered using an *in-house* Python script. The selected ligand poses were visualized using Discovery Studio visualizer [271] and PyMOL [199]. To identify ligands specific to the *Plasmodium* protein, we excluded ligands binding in the active site, allosteric pockets, and human homologs (Cat K and Cat L). Using the above criteria, eight compounds were shortlisted for FP-2, and five compounds were shortlisted for FP-3. These compounds were further subjected to molecular dynamics simulation to assess their stability.

### 4.4.5 Molecular dynamics

100 ns all-atom MD simulations were conducted on the respective WT and ligand complexes using the GROMACS 5.1.2 package [217]. The stability of the selected ligands in the potential allosteric pocket was first assessed via MD simulation. Ligands that showed high levels of stability across simulations were then pooled and evaluated against all mutant systems. Protein and ligand topology input files were generated using the AMBER99SB-ILDN force field [139] and AnteChamber PYthon Parser interfacE (ACPYPE) tools [280], respectively. The generated topologies were then solvated using the TIP3P water model [218] inside a triclinic box with 1.75 nm spacing between protein structures and edges. The system charges were neutralized by adding 0.15 M NaCl. To minimize the systems, we employed the steepest descent algorithm (emstep) with a preliminary energy step (nsteps) of 0.01nm without constraints until a tolerance limit of < 1000 kj $mol^{-1}$ $nm^{-1}$ was reached. Upon convergence, each system was equilibrated using the modified Berendsen thermostat [23] and Parrinello-Rahman barostat [24] algorithms, respectively. MD simulations of 100 ns with 2 fs (femtoseconds) timesteps were performed. Post-MD analyses using GROMACS in-built tools were performed to remove periodic boundary conditions (PBC) and calculate the Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), and Radius of Gyration (Rg).

### 4.4.6 Comparative essential dynamics

Understanding the differences in conformational sampling between the WT and mutant systems in the presence of ligands provide insights into the behavioural patterns induced by the ligands. To investigate the different conformational samplings obtained during MD simulation and extract the essential dynamic motions, we performed comparative essential dynamics using the *compare_essential_dynamics.py* script in MDM-TASK-web [163]. Based on covariance matrix calculations, the method aligns specific regions of each WT ligand-binding system with the corresponding mutant ligand-binding system. Herein, we specified the potential allosteric pocket residues. Regarding system dynamics, the first principal component (PC1) represents the largest amplitude motion, known as the essential dynamics. The distribution of the potential allosteric pocket is depicted by scatter plots based on the first and second principal components (PC1 and PC2).

### 4.4.7 Dynamic residue network analysis

In order to analyze the effect of ligand binding in the presence and absence of mutations, as well as the residues dynamics throughout MD simulations, DRN was performed using the MDM-TASK-web [163]. DRN was applied to the entire trajectories of the ligand-bound apo and mutant systems. The concept of DRN is explained in Chapter 2. Herein, a cut-off distance of 6.7 Å and a step size of 10 frames were used.

## 4.5 Results and Discussion

### 4.5.1 Revisiting allosteric pocket 1

In FP-2 and FP-3, pocket 1 is located within the nose region loop. The nose region connects the L and R domains and facilitates the proper folding of FP-2 and FP-3 in the absence of a prodomain [281]. In FP-2, pocket 1 comprises 23 residues (**Table 2.2** and **Figure 4.1A**). In FP-3, pocket 1 comprises 23 residues (**Table 3.2** and **Figure 4.1B**).

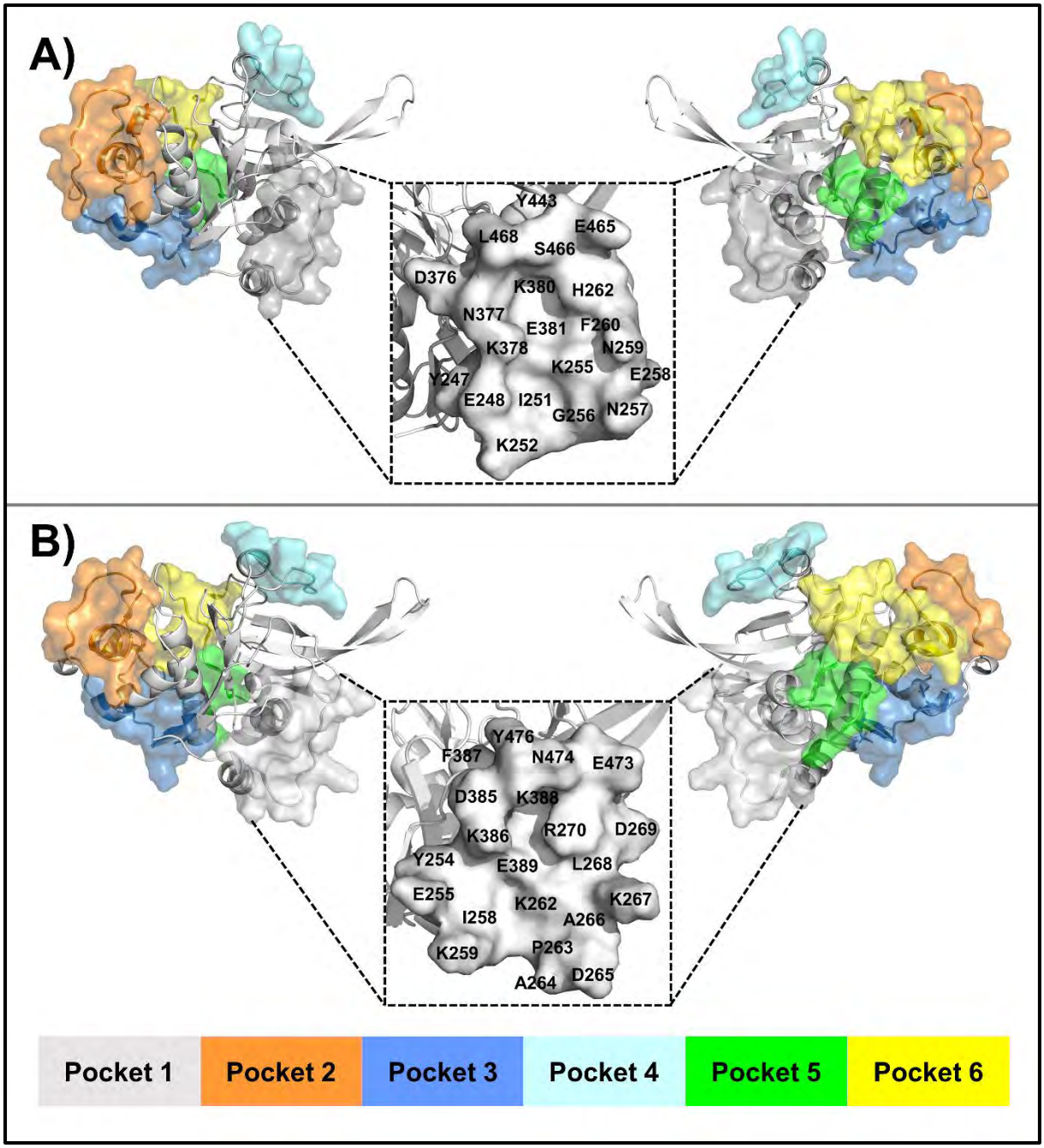



**Figure 4.1: Structural representation of the potential allosteric pockets of FP-2 and FP-3.** Pockets are shown in surface, and residues in Pocket 1 (pocket of interest) are highlighted in bold.

### 4.5.2 Identification of allosteric modulators against Pocket 1

In this chapter, 2089 FDA-approved compounds were examined via high-throughput virtual screening for allosteric inhibitory studies. Blind docking was conducted on the entire surface of FP-2, FP-3, Cat K and L. Compounds binding to pocket 1 of FP-2 and FP-3 were filtered using an in-house bash script. To ensure specificity to FP-2 and FP-3, compounds binding to the active site and allosteric sites of the human homologs (Cat K and L) were eliminated.

#### 4.5.2.1 Falcipain 2

Of the 2089 compounds screened against FP-2, 22 compounds successfully bound to pocket 1. A somewhat similar docking orientation was observed for all 22 compounds, in which each compound binds to the groove in pocket 1. The additional criteria of screening out compounds that bind to Cat K and L resulted in 14 compounds (DB00117, DB00123, DB00130, DB00142, DB00149, DB00151, DB00160, DB03088, DB03929, DB04398, DB06775, DB09394, DB09499, and DB11091) being excluded from further analysis. The resulting eight compounds (DB00128, DB00312, DB00766, DB00951, DB02893, DB03754, DB13972, and DB14159) were selected as potential allosteric hits based on their potential allosteric interactions with the *Plasmodium* protein (**Figure 4.2**). The selected compounds exhibited at least three conventional hydrogen bonds with allosteric residues in pocket 1 (**Figure 4.2**). The hydrogen bonds formed with the selected are listed as follows: DB00128 (N377, K378, E381, and S466), DB00312 (N377, K378, K380, E381, and S466), DB00766 (N377, K378, K380, E381, and S466), DB00951 (K380, E381, and S466), DB02893 (E381, and S466), DB03754 (H262, N377, E381, and S466), DB13972 (N377, K380, E381, and S466), and DB14159 (N377, K378, K380, and E381) (**Figure 4.2**). The hydrogen bond interactions with these compounds were consistent with those observed in two recent studies [268,269]. In research conducted by Hernández González et al. 2021, two compounds (ZINC72290660 and ZINC03225317) identified as noncompetitive inhibitors of FP-2 formed stable hydrogen bonds with the residues in site 6 (labelled as Pocket 1 in this study) [268]. ZINC72290660 formed hydrogen bonds with N134, K137, and E138 (N377, K380, and E381 whole protein numbering) while ZINC03225317 formed hydrogen bonds with D133, K137, and Y200 (D376, K380, and Y443 whole protein numbering) [268]. An additional study by Hernández González et al. 2022, identified the hydrogen bonds formed between Tetracycline derivative (Methacycline) and E138, K135, and K137 (E381, K378, and K380 whole protein numbering) in site 6 (labelled as Pocket 1 in this study) [269]. The chemical structures and medical uses of these compounds are highlighted in **Table 4.1**. To understand the effect of these compounds on the global motion of FP-2, we performed an all-atom molecular dynamics simulation.

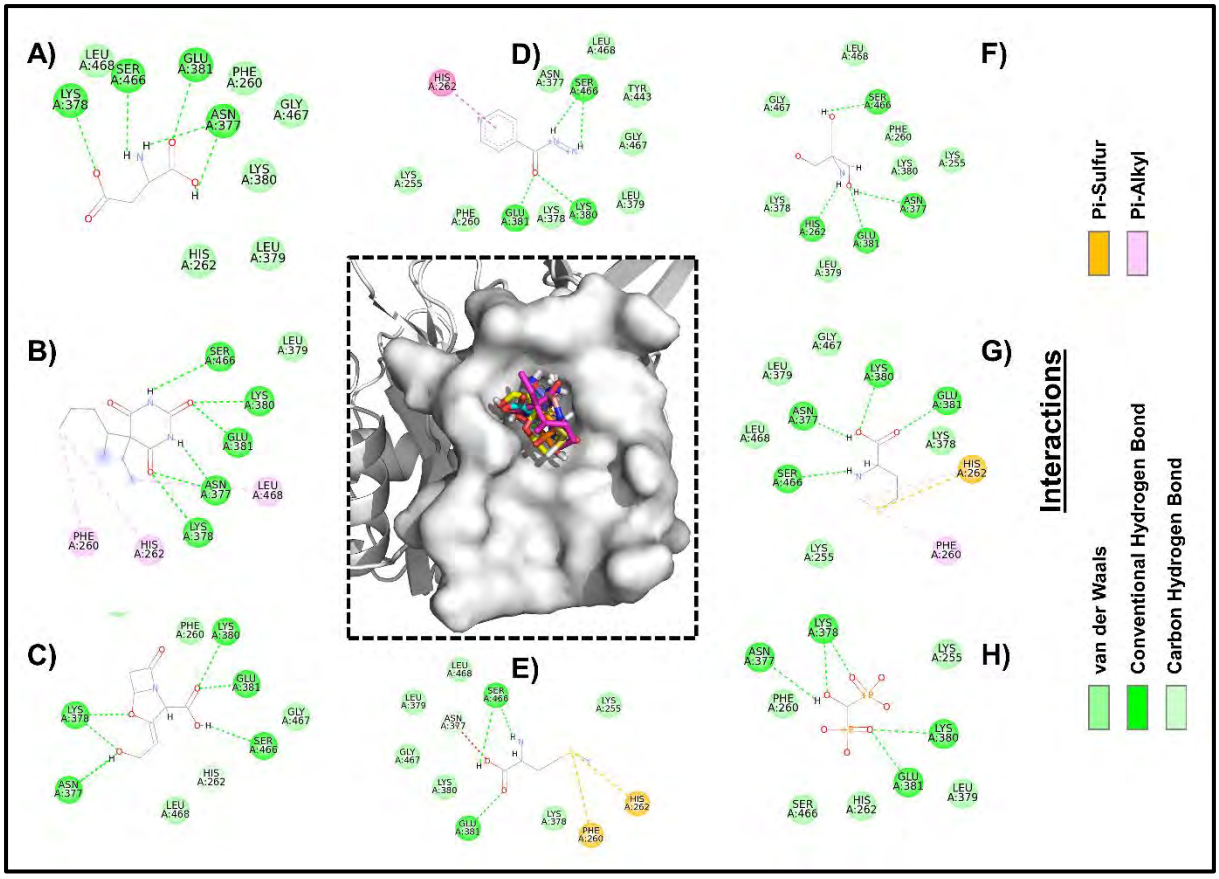
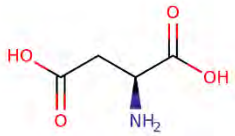
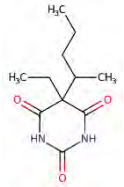
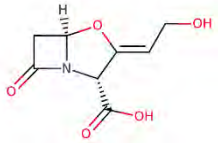
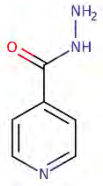
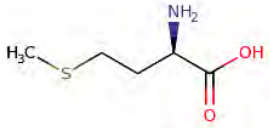
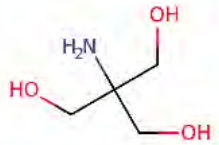
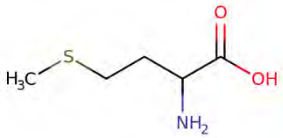
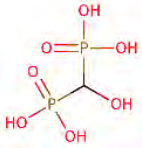



**Figure 4.2: Binding modes and interactions of selected allosteric hits in FP-2**. 2D representation of A) DB00128, B) DB00312, C) DB00766, D) DB00951, E) DB02893, F) DB03754, G) DB13972, and H) DB14159.

**Table 4.1: The chemical structure, name, and medical use of selected allosteric from DrugBank (DB)** [275].

| DB ID | Chemical structure | Name | Medical use |
|---|---|---|---|
| DB00128 |  | Aspartic acid | Amino acid supplementation |
| DB00312 |  | Pentobarbital | Sedative and hypnotic (`but not as an anti-anxiety) agent |
| DB00766 |  | Clavulanic acid | Beta-lactamase inhibitor |
| DB00951 |  | Isoniazid | Antibacterial agent used primarily as a tuberculostatic |
| DB02893 |  | D-Methionine | Nutrient supplementation |

| | | | |
|---|---|---|---|
| DB03754 |  | Tromethamine | Synthesis of surface-active agents and pharmaceuticals |
| DB13972 |  | Racemethionine | Nutrient supplementation |
| DB14159 |  | Oxidronic acid | Diagnostic skeletal imaging agent |

#### 4.5.2.2 Falcipain 3

From molecular docking results, 15 compounds successfully bound to pocket 1. As observed in FP-2, a similar docking orientation for all 15 compounds, in which each compound binds to the groove in pocket 1. Out of the 15 compounds, ten compounds (DB00233, DB00339, DB00936, DB01291, DB01956, DB03209, DB05381, DB06823, DB09543, and DB11100) bound to Cat K and L; thus, these compounds were excluded from further analysis. The resulting five compounds (DB00853, DB00951, DB01613, DB04173, and DB09419) were selected as potential allosteric hits based on their potential allosteric interactions with the *Plasmodium* protein (FP-3) (**Figure 4.3**). The only potential allosteric compound identified by both FP-2 and FP-3 was DB00951 (isoniazid). A minimum of two conventional hydrogen bonds were present with the allosteric residues in pocket 1 in the selected compounds (**Figure 4.3**). Unfavourable donor-donor interactions were observed in DB00853 and DB04173. The binding of a ligand alters the energy level of a system, disrupting several interactions, such as hydrogen bonds and van der Waals interactions, which could result in unfavourable contributions among the proteins and ligands [282]. As the formation of unfavourable donor-donor interactions between protein-ligand complexes reduces the stability of the complex, we performed molecular dynamics simulations to assess the stability of each ligand and its impact on FP-3's global motion. The chemical structures and medical uses of these compounds are highlighted in **Table 4.2**.

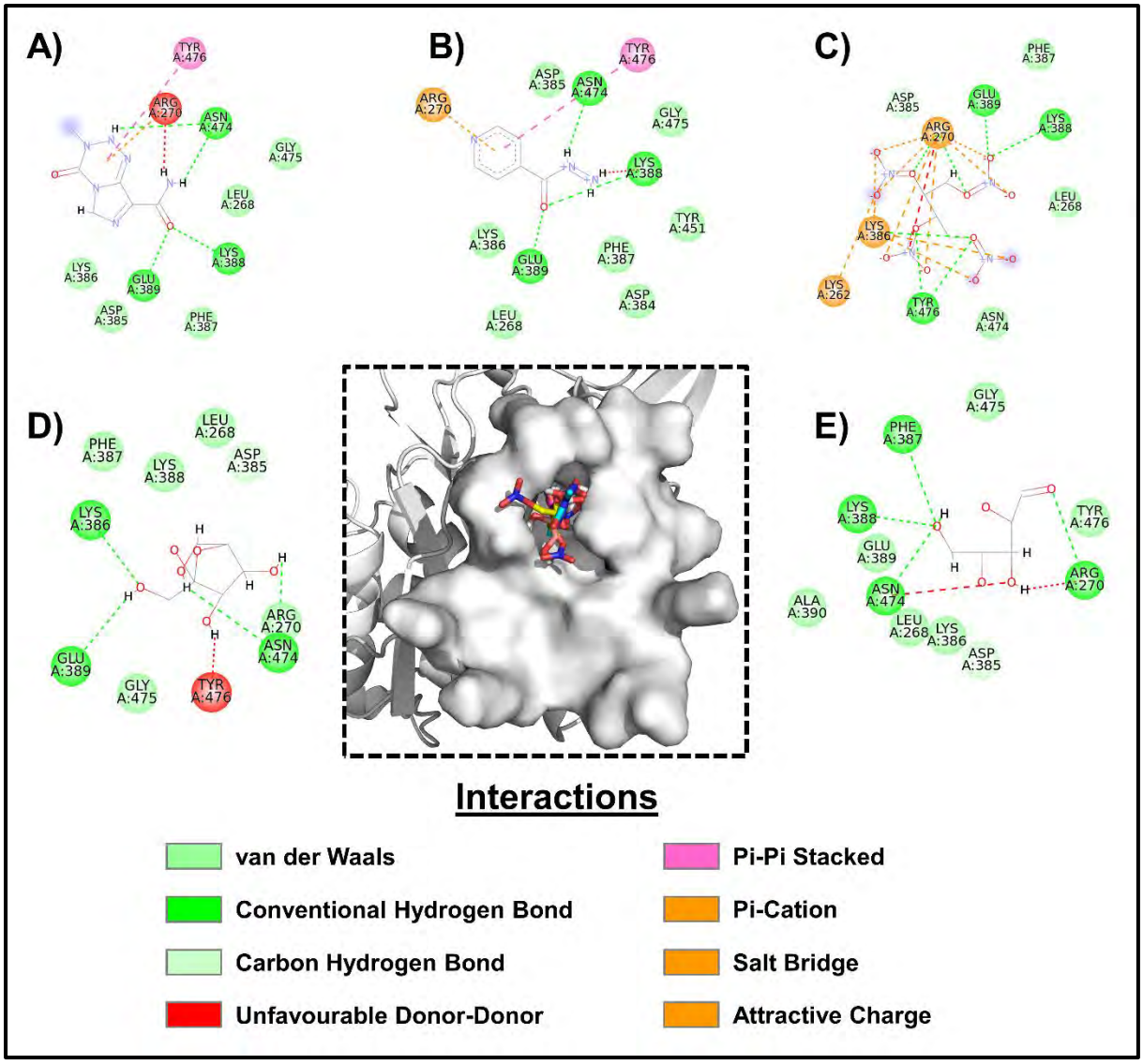
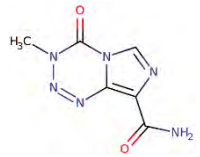
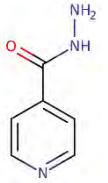
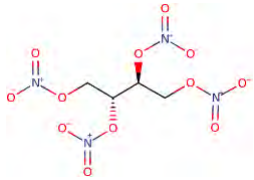
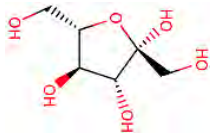
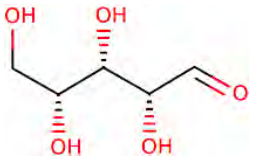



**Figure 4.3: Binding modes of selected allosteric hits**. 2D representation of A) DB00853, B) DB00951, C) DB01613, D) DB04173, and E) DB09419.

**Table 4.2: The chemical structure, name, and medical use of selected allosteric from DrugBank (DB)** [275]

| DB ID | Chemical structure | Name | Medical use |
|---|---|---|---|
| DB00853 |  | Temozolomide | Alkylating agent used to treat glioblastoma multiforme and refractory anaplastic astrocytoma |
| DB00951 |  | Isoniazid | Antibacterial agent used primarily as a tuberculostatic |
| DB01613 |  | Erythrityl tetranitrate | Vasodilator with general properties similar to nitroglycerin |
| DB04173 |  | Fructose | Treatment of nausea and stomach upset |
| DB09419 |  | Xylose | Diagnostic agent to observe malabsorption.<br>Common food additives or sweeteners used in place of regular sugars |

### 4.5.3 Stability of allosteric modulators via molecular dynamics simulation

MD simulations of the respective protein-ligand complexes were conducted for 100 ns. The stability of each complex was analyzed using RMSD calculation with respect to the binding pocket, followed by a visual inspection of the ligand via VMD [231].

#### 4.5.3.1 Falcipain 2

From the ligand RMSD plots (**Figure 4.4**), DB00128, DB00766, DB02893, DB03754, and DB13972 were classified as unstable. Visual inspection of these ligands using VMD [231] showed that the compounds exited the binding pocket (pocket 1) during the simulation. Therefore, these compounds were then discarded. DB00312, DB00951, and DB14159 showed stability throughout the MD simulation (**Figure 4.5A**). **Figure 4.5B** shows that each ligand exhibited a unimodal distribution indicating the presence of a single dominant conformation. Visual inspection of these compounds showed that the compounds remained in the binding pocket throughout the MD simulation. Among the stable compounds is isoniazid (DB00951), an antibacterial agent primarily used to treat tuberculosis. Hence these compounds were retained for further studies.

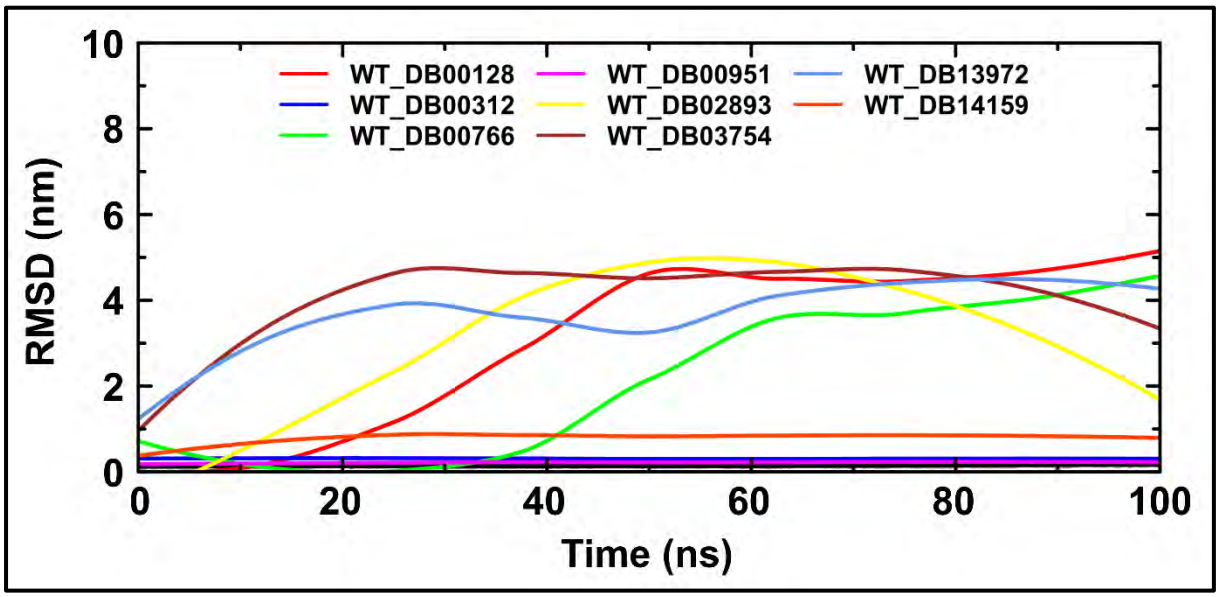



**Figure 4.4: Representation of the ligand RMSD of the eight selected ligands binding to pocket 1 of FP-2.** Ligands that showed stability during the simulation include DB00312, DB00951, and DB14159. Unstable ligands include DB00128, DB00766, DB02893, DB03754, and DB13972.

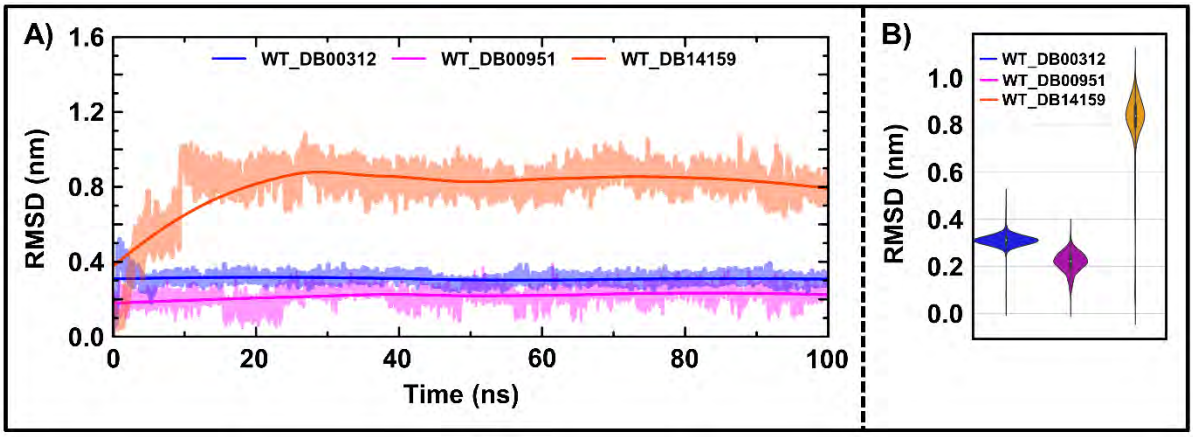



**Figure 4.5: Line and Violin representation of the stable ligand RMSD in FP-2.**

#### 4.5.3.2 Falcipain 3

The stability of the five potential allosteric inhibitors was investigated during MD simulation. According to the ligand RMSD plots (**Figure 4.6**), DB01613, DB04173, and DB09419 are unstable. A visual inspection of these ligands using VMD [231] revealed that their compounds left the binding pocket (pocket 1) during simulation. DB01613 was stable and remained in pocket 1 during the first 60 ns of the simulation but exited the pocket. The unstable compounds were excluded from further analysis. DB00853 and DB00951 remained stable throughout the MD simulation (**Figure 4.7A**).

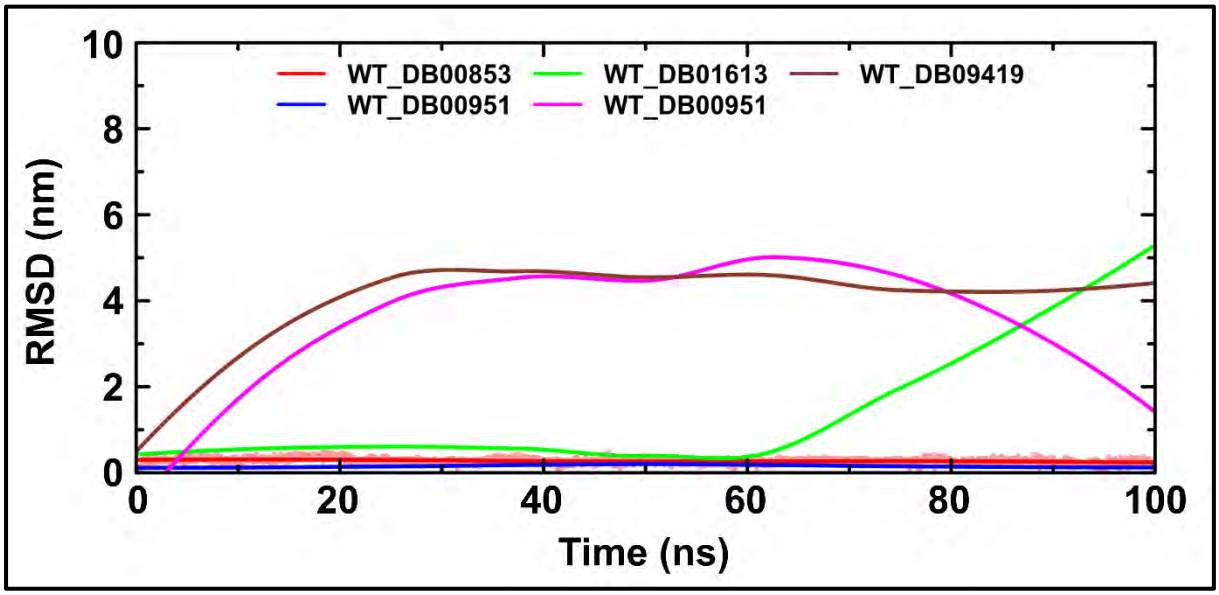



**Figure 4.6: Representation of the ligand** RMSD **of the eight selected ligands binding to pocket 1 of FP-3.** Ligands that showed stability during the simulation include DB00853 and DB00951. Unstable ligands include DB01613, DB04173, and DB09419.

The Ligand RMSD violin plot revealed that DB00853 exhibited unimodal distribution while DB00951 exhibited bimodal distribution. Despite the bimodal distribution observed in DB00951, visual inspection of both compounds revealed that they remained in the binding pocket throughout the MD simulation. These compounds were retained for further analysis. Interestingly, DB00951 (isoniazid) was stable when bound to FP-2 and FP-3.

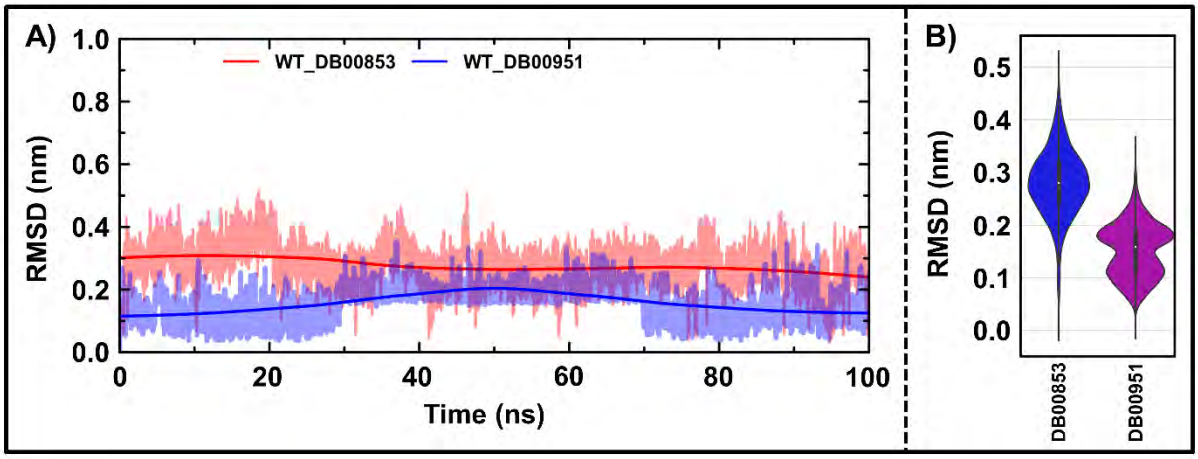



**Figure 4.7: Line and violin representation of the stable ligand RMSD** in A) line plot and B) violin plots.

### 4.5.4 Understanding the stability of allosteric modulators in the presence of mutations

In chapters 2 and 3, we identified 11 mutations occurring in the catalytic domain of FP-2 and five in the catalytic domain of FP-3. Due to the emerging threat of mutations against drug discovery, examining the effects of identified mutations on potential allosteric hits is pertinent. The selected compounds DB00312, DB00951, and DB14159 (in FP-2) and DB00853 and DB00951 (in FP-3) may function as allosteric modulators due to their stability in pocket 1; however, mutations in FP-2 and FP-3 could hamper their effectiveness. Hence, it is important to understand the effect of the mutations on the selected compounds.

#### 4.5.4.1 Falcipain 2

The behaviour of three potential allosteric modulators was examined in the presence of 11 mutations in FP-2. In total, we calculated the ligand stability of 33 systems (3 ligands x 11 mutant proteins) via RMSD. Ligand RMSD line plots depicting the stability are shown in **Figure 4.8**.

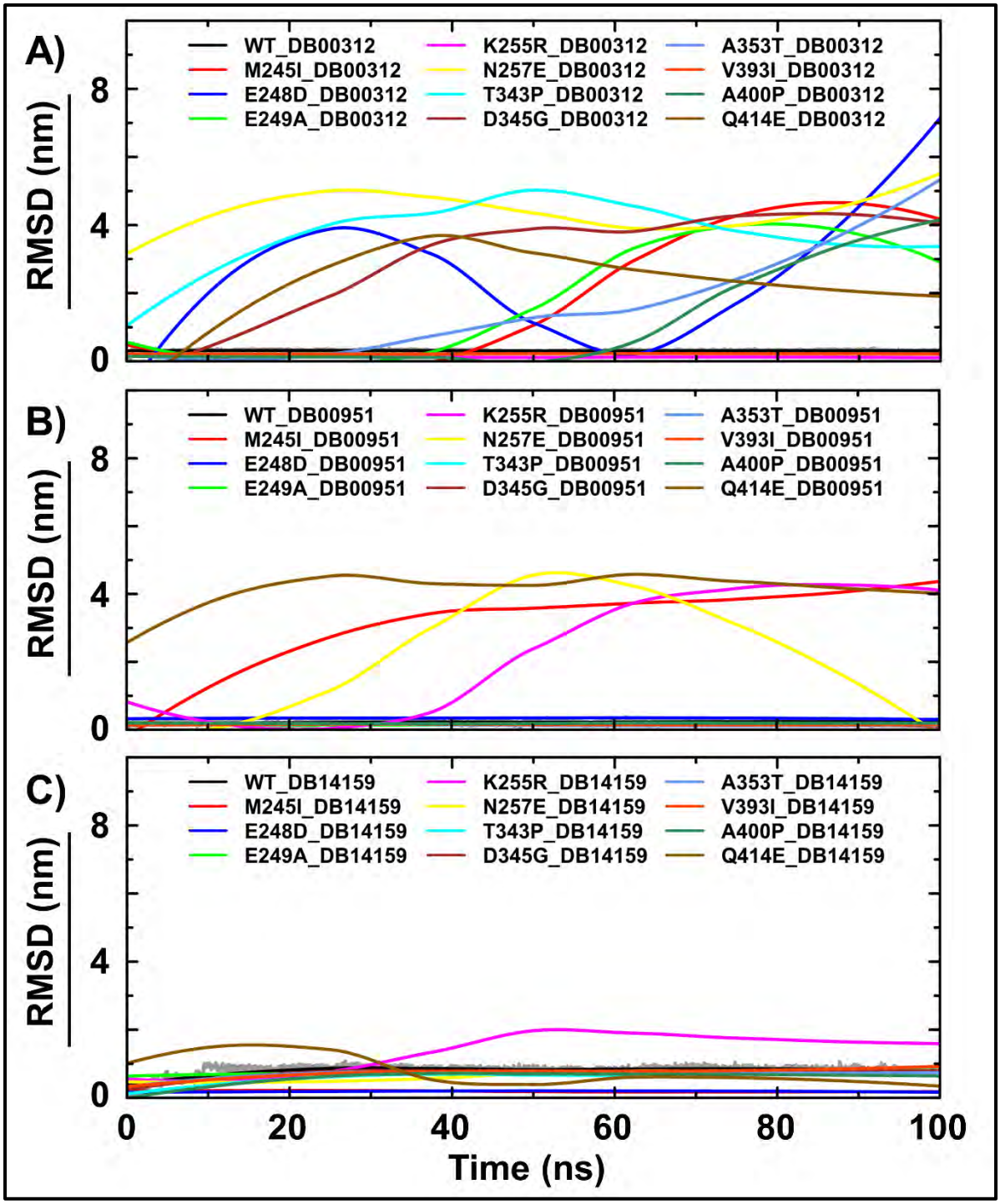



**Figure 4.8: Ligand RMSD line plots of the WT and mutant systems.** A) DB00312 (Pentobarbital), B) DB00951 (Isoniazid) and C) DB14159 (Oxidronic acid) binding to WT and mutant proteins of FP-2.

DB00312 (Pentobarbital) was stable in two mutant systems: K255R and V393I (**Figure 4.9A**), and unstable in the rest of the mutant systems. DB00951 (Isoniazid) was stable seven mutant systems: E248D, E249A, T343P, D345G, A353T, V393I, A400P (**Figure 4.9B**), whereas

unstable in M245I, N257E, A400P, and Q414E. DB14159 (Oxidronic acid) had the most stability in mutant systems (**Figure 4.9C**). It remained stable in all mutant systems except K255R and Q414E. Using VMD [231], the various ligands were visually examined in their respective protein systems to confirm their stability and instability. Interestingly, all compounds remained stable in only V393I. DB00951 and DB14159 were stable in E248D, E249A, T343P, D345G, A353T, and A400P. All compounds remained unstable in Q414E. Overall, DB00951 and DB14159 showed stability in most mutant systems.

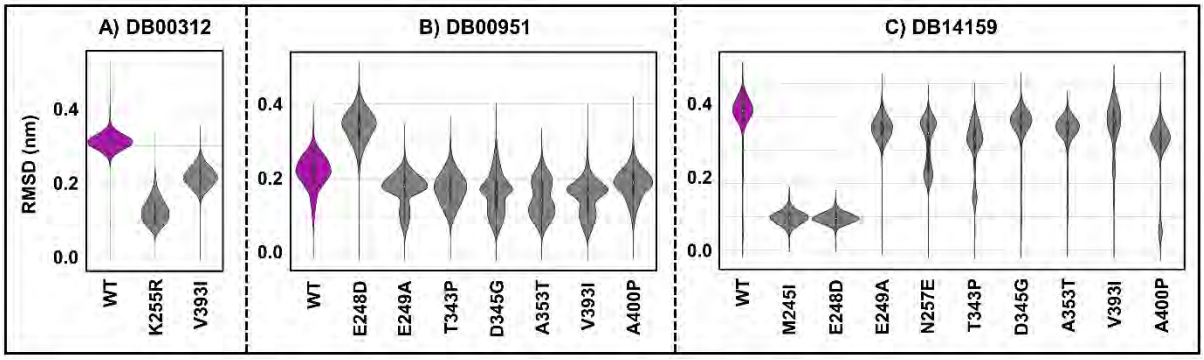



**Figure 4.9: Violin distribution plots of stable ligand RMSD values of the WT and mutant systems of FP-2.** A) DB00312, B) DB00951, and C) DB14159. WT is coloured magenta, and mutant systems are coloured grey.

### 4.5.4.2 Falcipain 3

In FP-3, we estimated the ligand stability of 10 systems (2 x 5 mutant systems) via RMSD. Ligand RMSD line plots depicting the stability are shown in **Figure 4.10**. DB00853 (Temozolomide) was stable in two mutant systems: R411K and E446G (**Figure 4.11A**), and unstable in N371K, A422T, and N468Y. Interestingly, DB00951 (Isoniazid) was stable in all mutant systems (**Figure 4.11B**). Although DB00951 explored more than one conformation during the simulation, visual inspection of the trajectories revealed that the ligand remained stable in the pocket throughout the simulation.

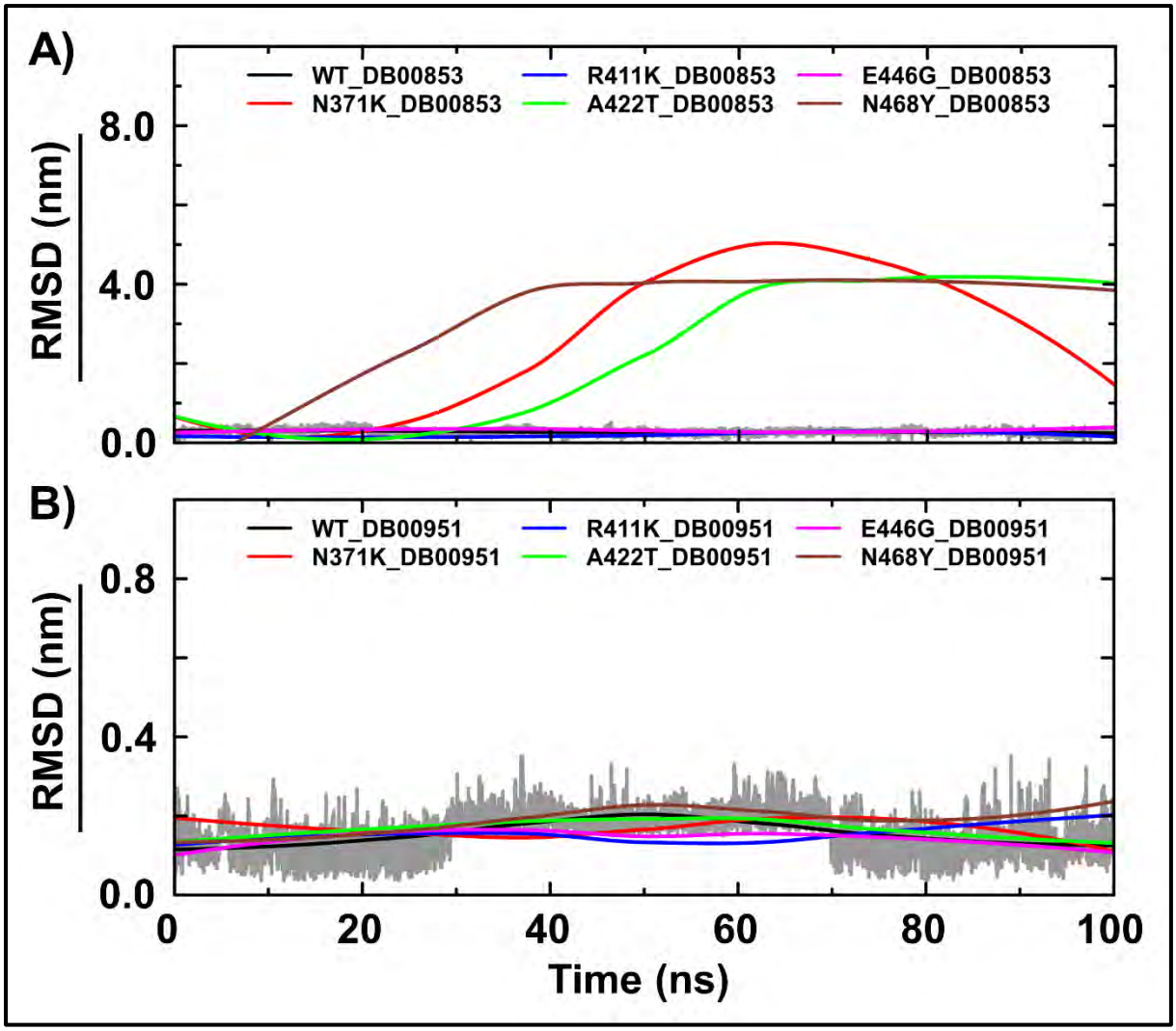



**Figure 4.10: Ligand RMSD line plots of the WT and mutant systems of FP-3.** A) DB00853 (Temozolomide) and B) DB00951 (Isoniazid) binding to WT and mutant proteins of FP-3.

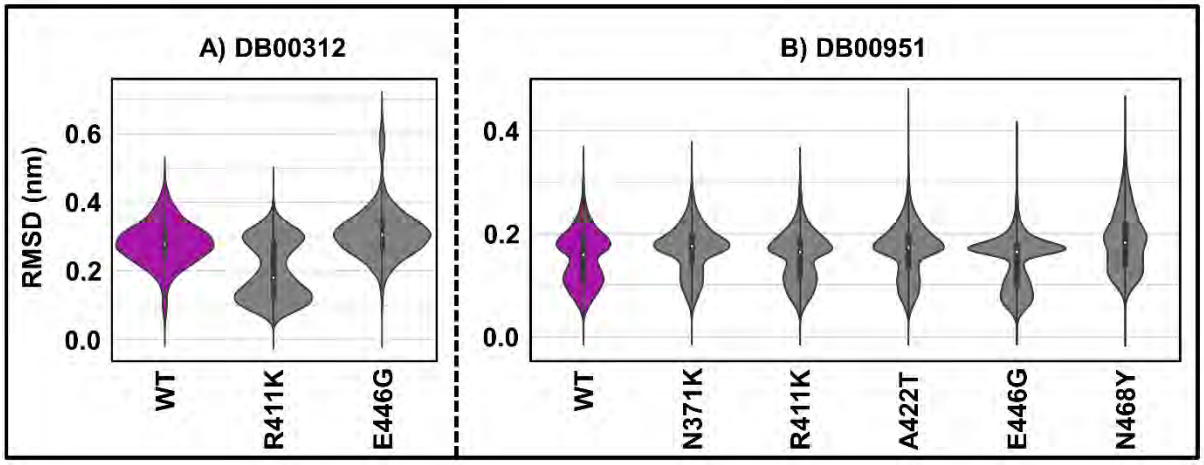



**Figure 4.11: Violin distribution plots of stable ligand RMSD of the WT and mutant systems of FP-3.** A) DB00312 and B) DB00951. WT is coloured magenta, and mutant systems are coloured grey.

### 4.5.5 Effect of ligands on the structural dynamics of the WT and mutant systems via RMSD and Rg

In order to analyze the difference in conformation between mutant-bound systems, we computed the RMSD and Rg for each system.

#### 4.5.5.1 Falcipain 2

Backbone RMSD analysis of the WT and mutant systems bound to DB00312 revealed no differences in the conformational distribution of the systems, as all systems exhibited a unimodal distribution (**Figure 4.12A**). However, the mutant systems displayed lower RMSD values in comparison to the WT. Rg analysis revealed the compactness of the WT and mutant systems bound to DB00312. As seen in (**Figure S4.1A**), a similar conformational distribution was observed in all systems. Backbone RMSD analysis of the WT and mutant systems bound to DB00951 revealed a unimodal distribution across all systems. Mutant system T343P displayed the highest RMSD values (**Figure 4.12B**). Regarding Rg analysis, all systems remained compact in the presence of DB00951, with K255R displaying the highest Rg values and E249A with the lowest Rg values (**Figure S4.1B**). Backbone RMSD analysis of the WT and mutant systems bound to DB14159 displayed the widest conformational range among the ligand-bound systems. WT and the majority of the mutations exhibited a unimodal distribution. Mutant systems M245I, V393I, and A400P displayed a bimodal distribution, indicating that the mutant-bound systems evolved during the simulation (**Figure 4.12C**). Similarly, Rg analysis revealed the unimodal distribution of the WT and mutant systems. The highest Rg values were observed in M245I, V393I, and A400P (**Figure S4.1C**). Overall, DB14159 displayed the highest effect on the mutant systems.

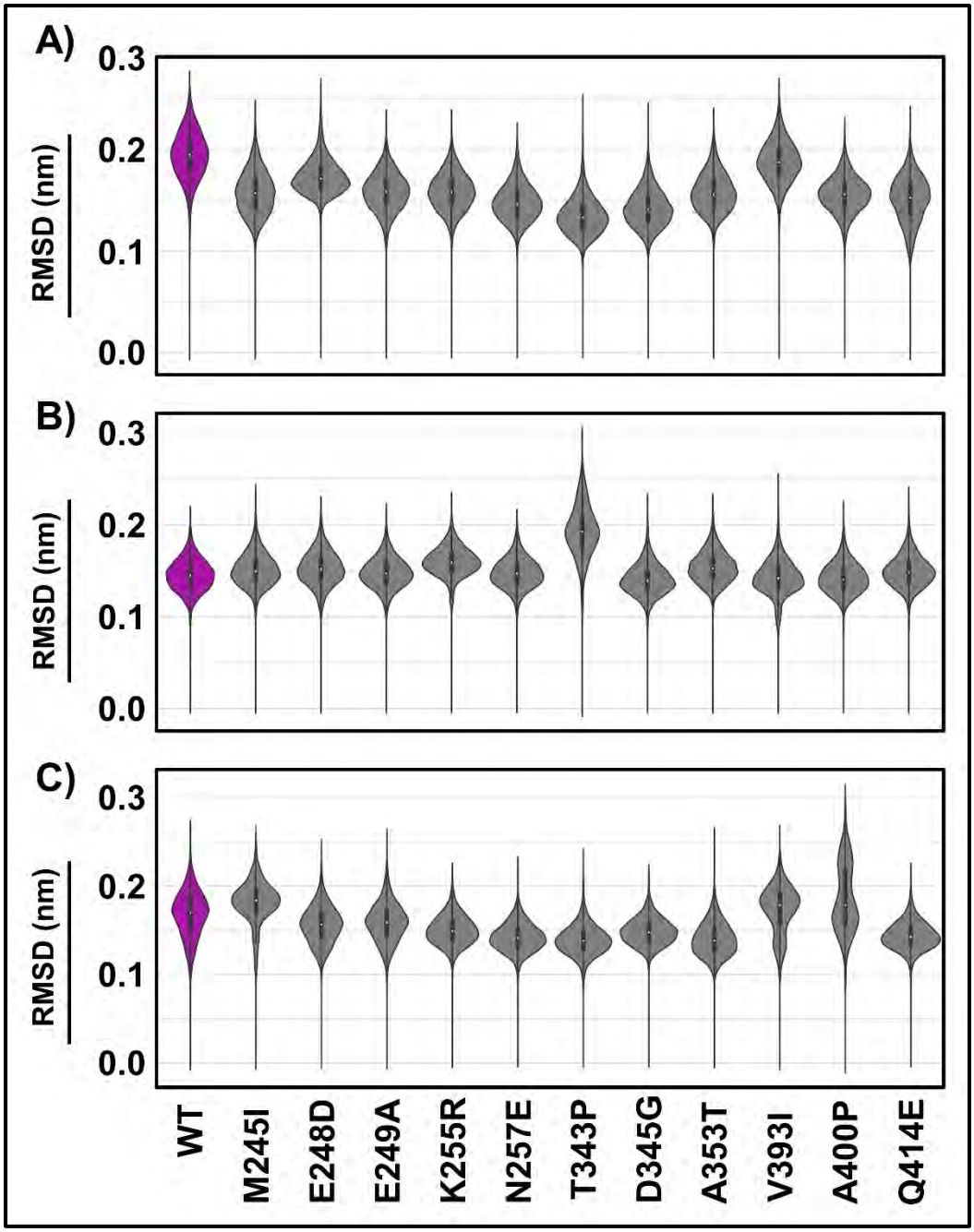



**Figure 4.12: Violin distribution of the backbone** RMSD **of the WT and mutant-bound systems of FP-2.** A) WT and mutant systems bound to DB00312, B) WT and mutant systems bound to DB00951, and C) WT and mutant systems bound to DB14159. WT is coloured magenta, and mutant systems are coloured grey.

#### 4.5.5.2 Falcipain 3

Backbone RMSD analysis of the WT and mutant systems bound to DB00853 revealed the conformational distribution. WT and mutant systems A422T, E446G, and N468Y displayed a unimodal distribution (**Figure 4.13A**). Mutant systems N371K and R411K exhibited a somewhat bimodal distribution. Rg analysis showed that all systems, with the exception of

E446G, exhibited a unimodal distribution (**Figure S4.2A**). Backbone RMSD analysis of the WT and mutant systems bound to DB00951 revealed that all systems, with the exception of N468Y, exhibited a unimodal distribution (**Figure 4.13B**). N468Y displayed a bimodal distribution. According to the Rg analysis, all systems remained compact during the simulation, with N468Y displaying the highest Rg values (**Figure S4.2B**).

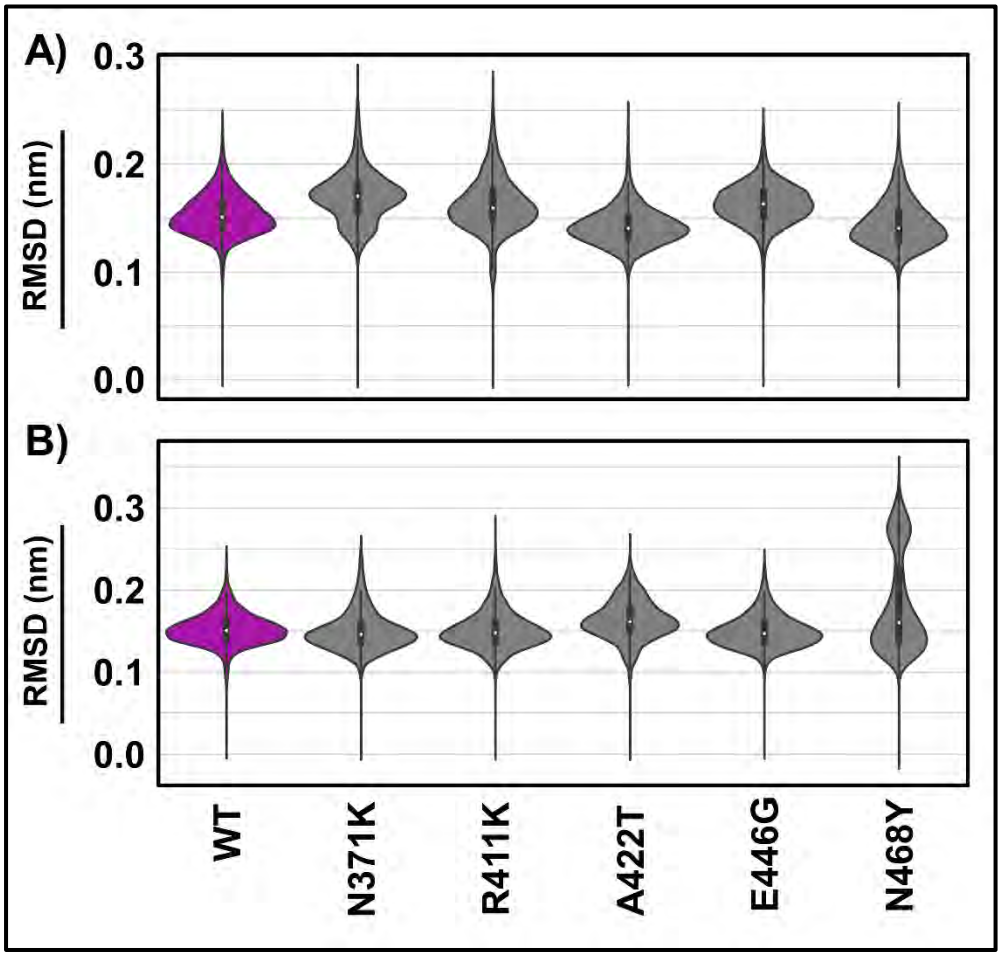



**Figure 4.13: Violin distribution of the backbone** RMSD **of the WT and mutant-bound systems of FP-3.** A) WT and mutant systems bound to DB00853, and B) WT and mutant systems bound to DB00951. WT is coloured magenta, and mutant systems are coloured grey.

### 4.5.6 Local per-residue flexibility analysis using RMSF

The root mean square fluctuation (RMSF) of the C$\alpha$ atoms of the WT and mutant bound systems revealed the influence of ligands on overall flexibility.

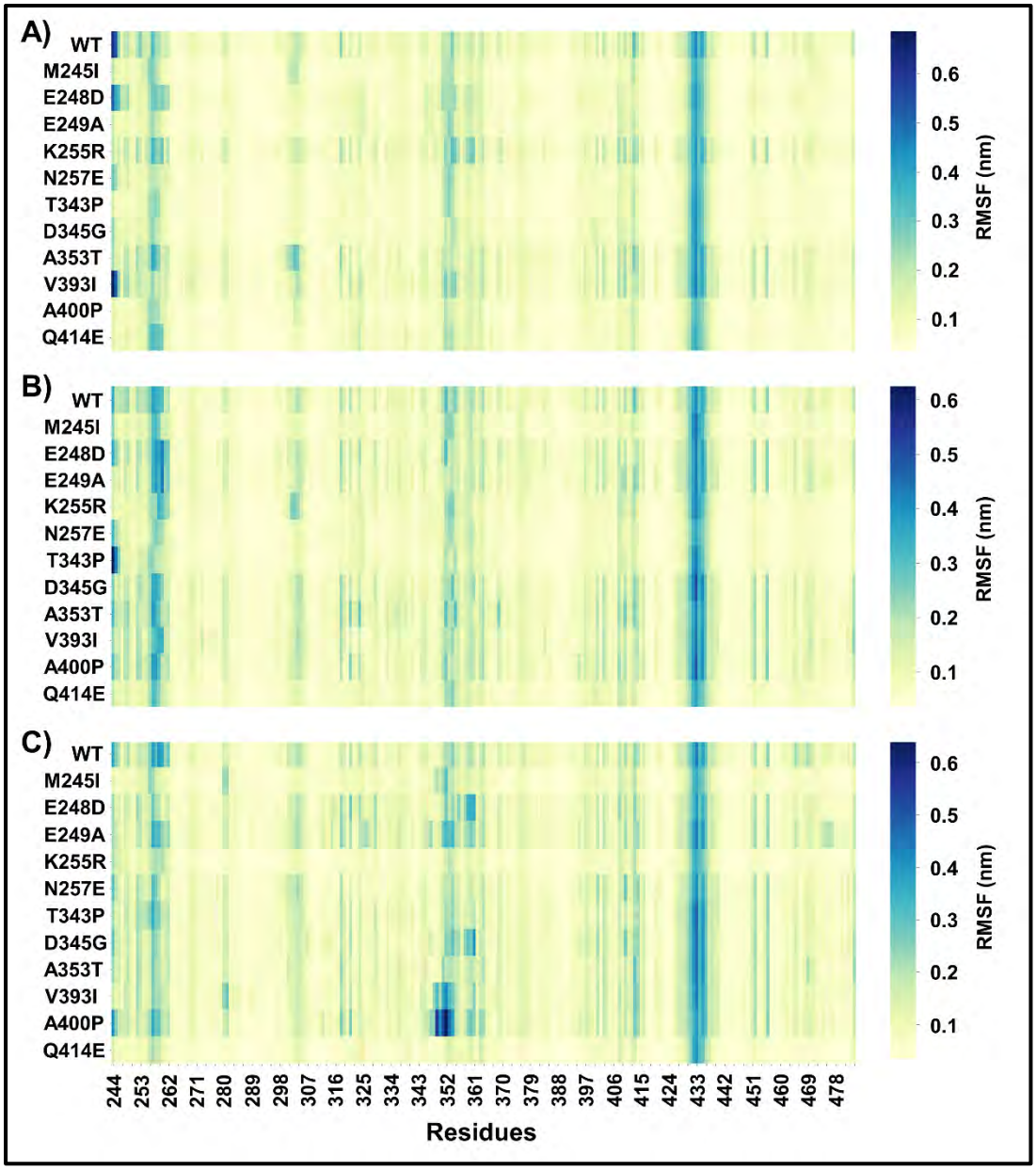



**Figure 4.14: Heatmap plots of the C$\alpha$ RMSF of the WT and mutant-bound systems of FP-2**. A) WT and mutant systems bound to DB00312, B) WT and mutant systems bound to DB00951, and C) WT and mutant systems bound to DB14159

#### 4.5.6.1 Falcipain 2

In all ligand-bound systems, high fluctuations were observed around the nose region, the loop region between $\alpha$5 and $\alpha$7, and the $\beta$-arm region (**Figure 4.14**). Additional fluctuations were DB14159 bound systems. In A400P-DB14159 bound system, high fluctuation was observed in residues 346 – 355 (**Figure 4.14C**). Interestingly, these residues make up pocket 2. In Chapter 2, we correlated high fluctuations in pocket 2 to the presence of a cryptic pocket; further analyses were carried out to investigate the presence of these high fluctuations in pocket 2.

#### 4.5.6.2 Falcipain 3

In FP-3, all ligand-bound systems exhibited high fluctuations around the $\beta$-arm region. Additional fluctuations were observed in some mutant-bound systems (**Figure 4.15**). In N371K-DB00853 and E446G-DB00853, fluctuations were observed in residues 358 – 367. Interestingly, these residues make up pocket 2. In N468Y-DB00951, high fluctuations were observed in residues 250 – 253, which are part of the nose region.

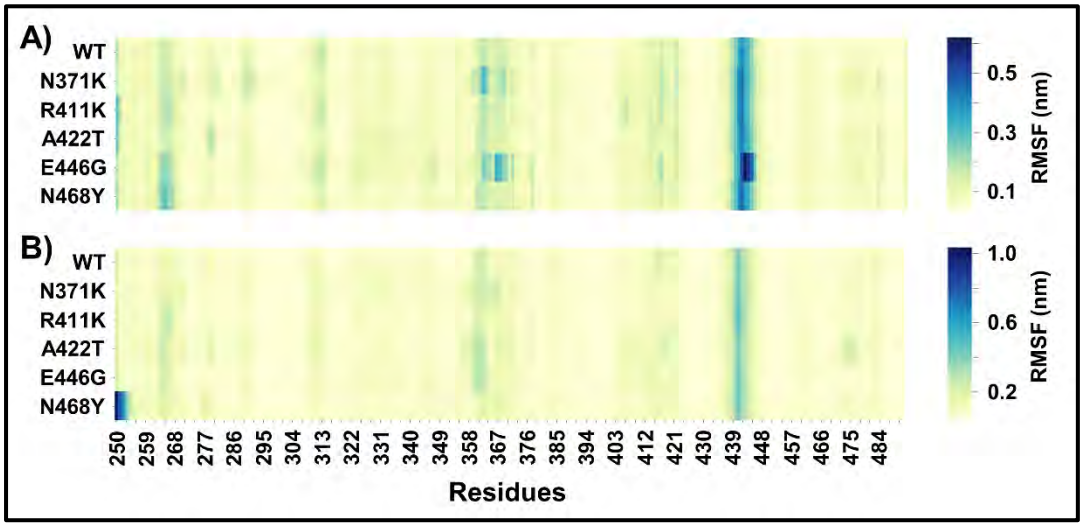



**Figure 4.15: Heatmap plots of the Cα RMSF of the WT and mutant-bound systems of FP-3**. A) WT and mutant systems bound to DB00853, and B) WT and mutant systems bound to DB00951.

### 4.5.7 Effect of ligands on all potential allosteric pockets

Ligand binding can induce conformational changes in all or part of the protein structure. To characterize the effect of ligands binding in pocket 1 on other pockets (pockets 2 – 6), we calculated the Rg of each pocket. The radius of gyration indicates the distribution of atoms of a protein around its axis, and it can be used to predict the structural activity of proteins [283].

#### 4.5.7.1 Falcipain 2

The effect of DB00312 on the binding pocket and the other pockets revealed subtle changes in their topologies. In pocket 1, all systems, with the exception of E248D and Q414E, exhibited a unimodal distribution (**Figure S4.3A**). Mutant systems E248D and Q414E exhibited a bimodal distribution. Visual inspection revealed subtle changes in the topology of the pocket. In comparison with the other pockets, pocket 2 exhibited a higher degree of variability. WT, E249A, K255R, T343P, A353T, V393I, and Q414E exhibited a unimodal distribution, while M245I, E248D, N257E, D345G, and A400P exhibited bimodal distribution (**Figure S4.3B**). Visual inspection of the trajectories revealed the conformational sampling exerted on the pocket. In pocket 3, all systems exhibited unimodal distribution, with A353T exhibiting the highest Rg values (**Figure S4.3C**). In Pocket 4, WT, M245I, N257E, T343P, A400P, and Q414E exhibited a unimodal distribution. E248D, E249A, K255R, D345G, D345G, A353T, and V393I exhibited bimodal distribution (**Figure S4.3D**). Pocket 5 had a similar distribution to pocket 3, where all mutant systems exhibited a unimodal distribution with the exception of A353T and A400P (**Figure S4.3E**). Lastly, all systems, with the exception of M245I and K255R, exhibited a unimodal distribution. Upon visual inspection of the trajectories, the pocket had undergone conformational sampling (**Figure S4.3F**).

Results showed that the binding of DB00951 in pocket 1 revealed subtle changes in the topologies of pockets 2 – 6. In pocket 1, all systems, with the exception of E248D, E249A, K255R, and N257E, exhibited a unimodal distribution (**Figure S4.4A**). In Pocket 2, all systems exhibited a unimodal conformation with a flatter distribution (**Figure S4.4B**). Among the systems in pocket 3, only M245I and K255R showed a bimodal distribution, while the remainder exhibited unimodal distribution (**Figure S4.4C**). In pocket 4, AT, M245I, E248D, E249A, N257E, D345G, and V393I remained compact and maintained a unimodal distribution, while K255R, T343P, A353T, A400P, and Q414E exhibited bimodal distribution (**Figure S4.4D**). In pocket 5, all systems, with the exception of K255R, maintained a unimodal distribution (**Figure S4.4E**). In pocket 6, all systems remained compact and maintained a unimodal distribution. (**Figure S4.4F**).

A more diverse range of changes in the topologies of the binding pocket and other pockets was observed when bound to DB14159 (**Figure 4.16**). In pocket 1, the WT system explored various conformations when bound to DB14159, while all mutant systems remained compact and maintained a unimodal distribution (**Figure 4.16A**). A more diverse distribution was observed

in pocket 2. WT, K255R, N257E, T343P, A353T, and Q414E exhibited a single conformation (**Figure 4.16B**). M245I, E248D, and D345G exhibited a single conformation with a flatter distribution. E249A, V393I, and A400P sampled different conformations and exhibited a bimodal distribution. Visual inspection of these trajectories revealed the conformational differences associated with the different mutations when bound to DB14159. In pocket 3, all systems maintained rigidity throughout the simulation (**Figure 4.16C**). In pocket 4, all systems exhibited a single conformation with M245I, E249A, N257E, and D345G, maintaining a flatter distribution (**Figure 4.16D**). In pocket 5, all systems maintained rigidity throughout the simulation (**Figure 4.16E**). Lastly, in pocket 6, all systems, with the exception of V393I and A400P, maintained rigidity throughout the simulation (**Figure 4.16F**). Visual inspection of M245I revealed an opening of pocket 6 and movement of residues in pocket 2 into pocket 6. In V393I, there was an opening and closing of pocket 6 and a slight movement of pocket 2 into pocket 6. Additionally, in A400P mutant protein indicated a unique opening pocket 6. This unique opening is seen slightly in mutant M245I and E249A but not seen in the WT and other mutant systems, indicating the presence of a cryptic pocket. We performed comparative essential dynamics to probe this unique opening.

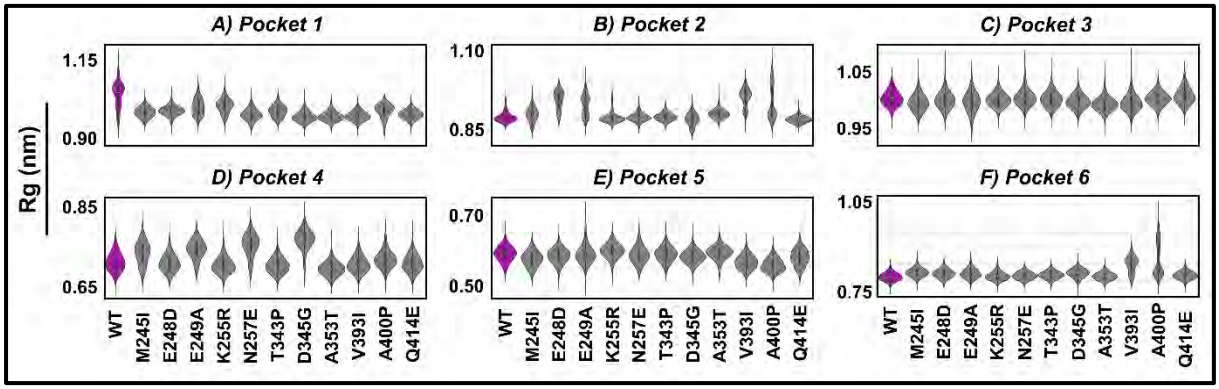



**Figure 4.16: Violin distribution of the Rg values for** Pockets **1 – 6 (A – F) of WT and mutant systems bound to DB14159 of FP-2.** The purple ensemble in the violin plots represents WT.

### 4.5.7.2 Falcipain 3

The binding of DB00853 to pocket 1 of the WT and mutant systems exhibited a very subtle effect on the other pockets. In Pocket 1, all systems maintained a single conformation with N371K, A422T and N468Y exhibiting a flatter distribution (**Figure S4.5A**). In pocket 2, WT, R411K, A422T, and N468Y exhibited unimodal distribution, while N371K and E446G sampled more than one conformation (**Figure S4.5B**). In pocket 3, mutant systems N371K,

A422T, and N468Y maintained rigidity throughout the simulation, while the WT, R411K, and E446G sampled a wider conformation (**Figure S4.5C**). In pockets 4 and 5, all systems maintained a unimodal distribution with a flatter distribution (**Figure S4.5D and E**). In pocket 6, all systems, with the exception of N371K and N468Y, exhibited a unimodal conformation (**Figure S4.5F**).

Examining the binding effect of DB00951 on its binding pocket and other pockets revealed a more diverse result. In pocket 1, WT and E446G maintained rigidity, while N371K and R411K sampled a single conformation with a flatter distribution (**Figure S4.6A**). A422T and N468Y sampled a wider conformation while displaying a bimodal distribution. In pocket 2, N371K and A422T exhibited a bimodal distribution, while the remaining systems maintained a unimodal distribution (**Figure S4.6B**). In pockets 3, 5, and 6, all systems sampled a single conformation, with pocket 5 of A422T displaying a higher Rg value (**Figure S4.6C, E, and F**). In pocket 4, N371K, R411K, and N468Y maintained a single conformation (**Figure S4.6D**). The WT and E446G mutant system sampled a wider conformation. A442T exhibited a bimodal distribution. Visual inspection of A422T trajectory revealed the conformational variability of pocket 4. Topology assessment revealed an opening in pocket 4. Overall, A422T exhibited the most diverse results.

### 4.5.8 Comparative essential dynamics (CED) confirms the effect of the binding of DB14159 in FP-2 on the other pockets

Based on the Rg analysis, we discovered changes associated with allosteric pockets 2 and 6 in mutant systems bound to DB14159. Our next step was to investigate the topological differences between the allosteric pockets of the WT and mutant systems using the comparative essential dynamics (CED) approach implemented in MD-TASK-web [163]. We investigated the extreme projections of the first and second principal components (PC1 and PC2) of the WT-DB14159-bound and mutant DB14159-bound systems, which are responsible for the dominant motions of the protein structures.

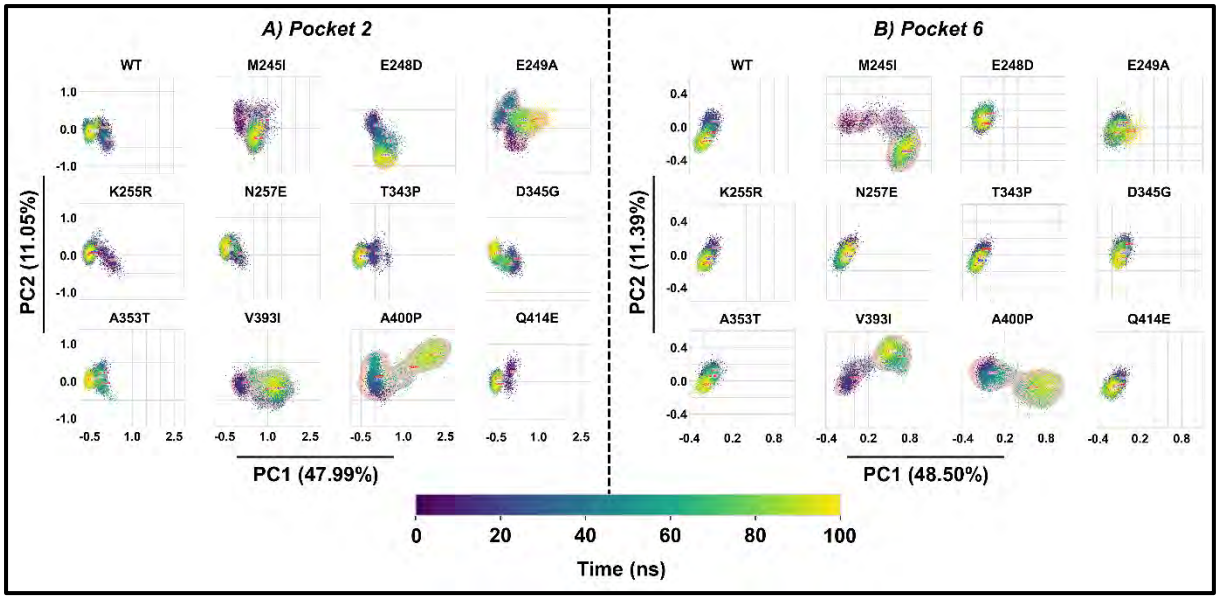



**Figure 4.17: Comparative essential dynamic analysis of pockets 2 and 6 of the WT and mutant systems of FP-2 bound to DB14159 along PC1 and PC2.** The colour code (dark= start, yellow = end) represents the simulation time (ns). The timestamp indicated blue on each PCA plot when the structure was most stable.

Analysis of the comparative essential dynamics of pockets 1, 3, 4, and 5 are highlighted in (**Figure S4.7**) and correlated with Rg results (**Figure 4.16**). PC1 and PC2 accounted for 53.42% and 15.91% of the total variance in pocket 1. As seen in **Figure S4.7A,** the mutant systems remained compact compared to the WT. In pocket 3 (PC1: 19.44% and PC2: 13.53%) and pocket 5 (PC1: 30.83% and PC2: 20.05%), all systems varied less throughout the simulation (**Figure S4.7B and D**). In pocket 4, where PC1 and PC2 accounted for 51.78% and 13.78% of the total variance, respectively, the conformational variability was observed in M245I, E249A, N257E, and D345G (**Figure S4.7C**). Interestingly, pockets 2 and 6 showed the most divergence; therefore, we focused on them using comparative essential dynamics. In pocket 2, PC1 and PC2 accounted for 47.99% and 11.05% of the total variance, respectively. Mutant systems M245I, E249A, V393I, and A400P exhibited a larger conformational distribution compared to the WT and other mutant systems. (**Figure 4.17A**). In pocket 6, PC1 and PC2 accounted for 48.50% and 11.39% of the total variance, respectively (**Figure 4.17B**). M245I, V393I, and A400P exhibited more conformational diversity compared to WT and other mutant systems. Visual inspection of E249A showed a conformational sampling of pocket 2. According to our findings, DB14159 alters the pocket's topology when bound to E249A mutant systems.

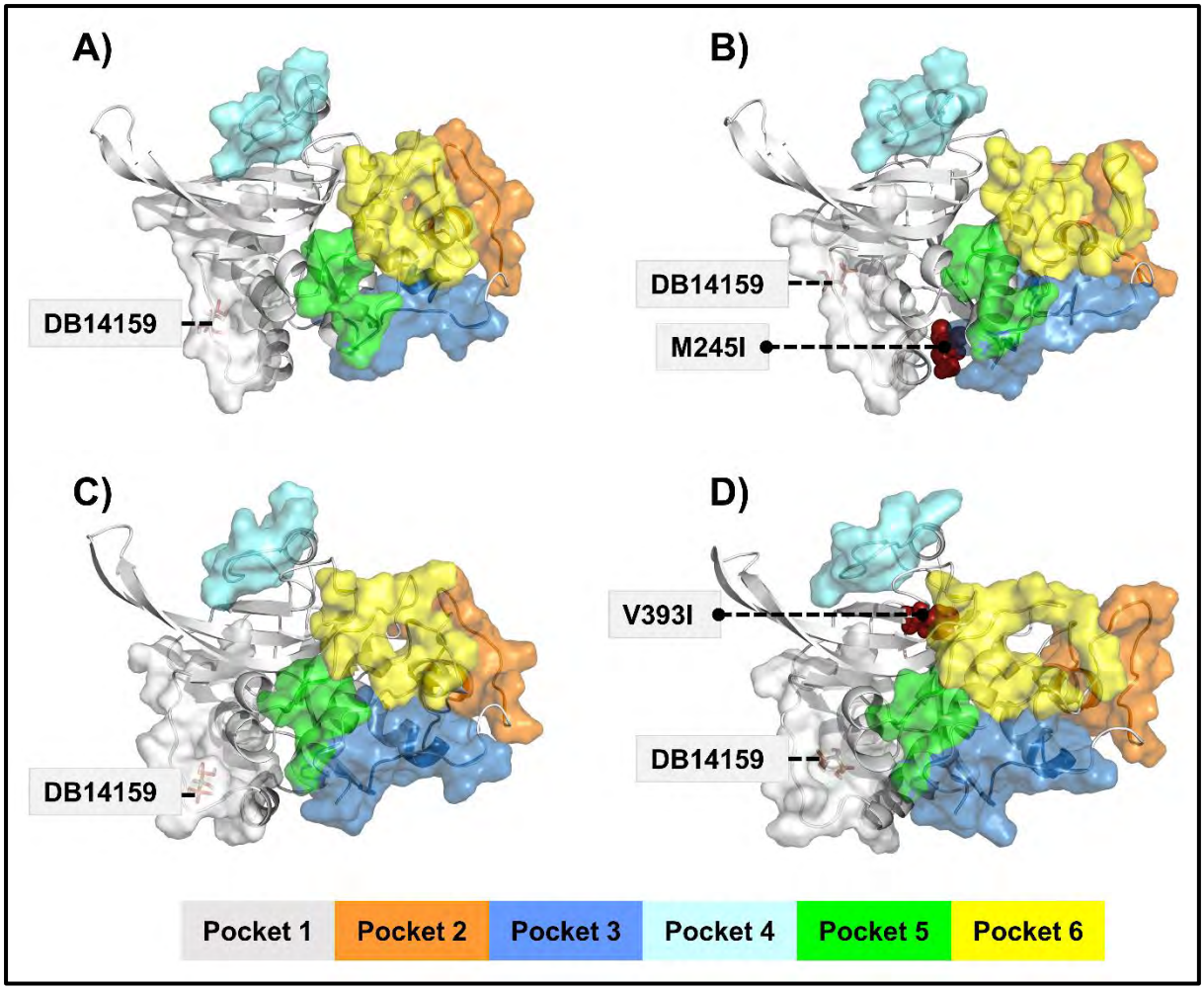



**Figure 4.18: Cartoon representation of the WT and mutant systems M245I and V393I extracted at various timestamps.** A) WT and B) M245I extracted at 52180 ps. C) WT and D) V393I extracted at 42910 ps. Identified potential allosteric pockets are represented in surface and coloured, respectively. DB14159 is represented in sticks and labelled accordingly.

Visual inspection of M245I revealed an opening of pocket 6 and movement of residues in pocket 2 into pocket 6. This movement was captured and extracted from the MD simulation frame at 52180 ps (**Figure 4.18A**). It is noteworthy that in chapter 2, in which we analyzed the effect of mutants on FP-2, we found a cryptic pocket in mutant M245I. The results confirm that M245I has an effect on FP-2 regardless of ligand presence or absence. Additionally, V393I also showed an opening and closing of pocket 6 and a slight movement of pocket 2 into pocket 6. This movement was captured and extracted from the MD simulation frame at 42910 ps (**Figure 4.18B**). An examination of the pose extracted by the k-means algorithm at 94560 ps of the A400P mutant protein MD simulation indicated that pocket 6 has a unique opening (**Figure 4.19**). This unique opening is seen slightly in mutant M245I and E249A but not seen in the WT and other mutant systems, indicating the presence of a cryptic pocket. Upon

visualization of the trajectory, residues from pocket 2 moved into pocket 6, resulting in their fusion into a single groove (**Video S1**). According to our findings, the binding of DB14159 in pocket 1 affected pockets 2 and 6 as well.

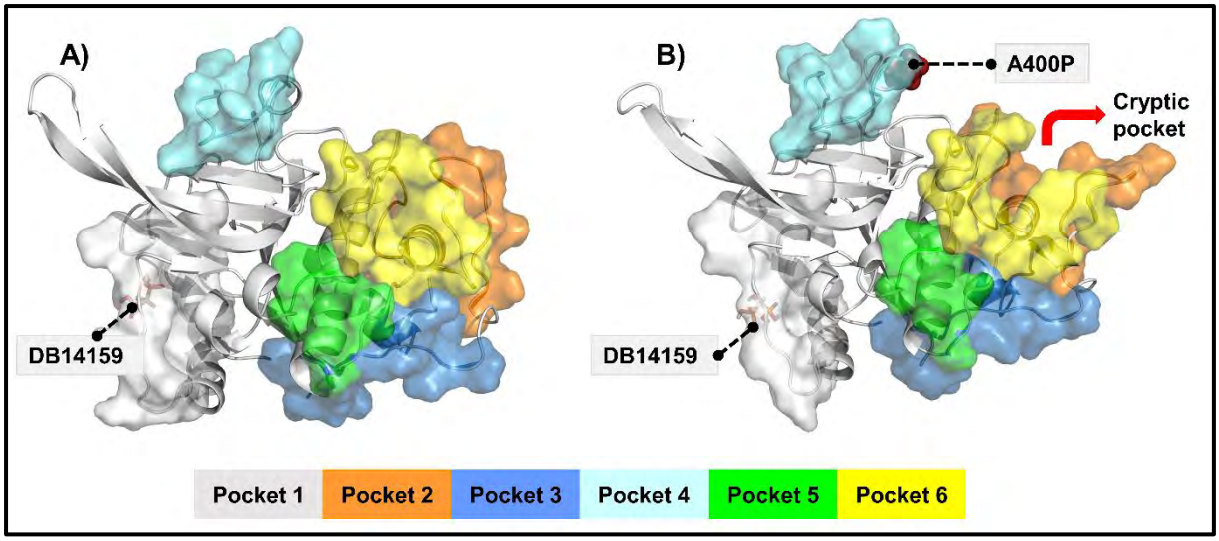



**Figure 4.19: Cartoon representation of A) WT and B) A400P extracted at 94560 ps**. Identified potential allosteric pockets are represented in surface and coloured, respectively. DB14159 is represented in sticks and labelled accordingly.

### 4.5.9 Binding of ligands affects the communication network of WT and mutant systems

Following the observation of ligand binding to potential allosteric pockets and the entire protein structure, a further analysis was conducted to identify conserved residue communication changes. Dynamic residue analysis (DRN) plays a crucial role in determining how ligand binding and mutations affect the communication in a protein ensemble. In previous chapters, the DRN methodology has been described in detail; however, we utilized the entire trajectory of the protein structure. We implemented five different centrality metrics (*BC*, *CC*, *DC*, *EC*, *KC*) in MDM-TASK-web [163] server to perform DRN on the stable WT and mutant-bound systems. In the previous chapters, we defined "*centrality hub*" as any node in the set of nodes with the highest centrality. In this case, hubs are defined as nodes with the highest centrality based on all metrics calculated. A shortlist of the top 5% nodes from each system was made based on raw average values for each system. This list was irrespective of the centrality metric used. As with the previous chapters, we have introduced the following terminology: (1) *Persistent hub*: a hub that persists throughout the entire system. (2) *Distinctive hub*: Hub absent in WT-bound complex but present in the protein-bound complex. DB00312, DB00951, and DB14159 were stable in the WT and some mutant systems of FP-2. DB00853 and DB00951

were stable in the WT and some mutant systems of FP-3. Hence, we investigated the effect of ligands on the stable complexes.

#### 4.5.9.1 Falcipain 2

***DB00312***

DB00312 was stable in the WT and two mutant systems (K255R and V393I). K255R is one of the ART-resistant linked missense mutations, and V393I occurs within the active pocket. We investigated the global top 5%, which accounts for residues with the highest centrality metrics. The results are presented as heat maps in **Figure 4.20**. Persistent hubs across all systems are highlighted in **Table 4.3**. By applying *BC* metric, we determine how frequently residues are used in network navigation. There is evidence that residues with high *BC* values play a critical role in protein-ligand [284], protein-protein [285], or allosteric communication [286,287]. *BC* persistent hubs across the WT and mutant systems are C285, F288, L383, I389, M420, Y441, Y443, I445, and I480 (**Figure 4.20A**). Residue C285 forms part of the catalytic triad that forms part of the trench-like binding pockets. Visual inspection and mapping of the hubs on the protein structure show that most hubs are located within the binding site of FP-2 (**Figure S4.8**). I291 and G292 were distinctive in both mutant systems. In the WT system, V294 and active site residues N447 were distinctive and absent in the mutant systems indicating that the usage of these residues in the communication network was lost in the mutant systems. *CC* persistent hubs across the WT and mutant systems are S289, S290, G292, S293, P388, I389, S390, I391, V419, M420, L421, and I445 (**Figure 4.20B**). *DC* persistent hubs across the WT and mutant systems are S284, W286, S289, S296, I328, I389, I391, and M420 (**Figure 4.20C**). The communication importance of residues N320 and I368 in the WT were lost in the mutant systems. *Distinctive hubs* observed in the mutant systems: K255R (D334, I341, and I461), and V393I (I341, C411, I461). *EC* persistent hubs across the WT and mutant systems are S286, A287, S289, S290, I291, S293, I328, A331, S390, and M420 (**Figure 4.20D**). The usage of residues F288 and G292 in communication was lost in all mutant systems. *Distinctive hubs* I389 and I391 were observed in the mutant systems K255R and V393I. *KC* persistent hubs across the WT and mutant systems are S286, A287, S289, S290, I291, S293, I328, I389, S390, M420 (**Figure 4.20E**). The usage of residue S284 in communication was lost in all mutant systems. *Distinctive hubs* I391 were observed in the mutant systems K255R and V393I.

**Table 4.3: Persistent hubs in all WT and mutant-bound systems of FP-2**.

| Metric | Persistent hubs | |
|---|---|---|
| *Betweenness Centrality (BC)* | DB00312 | 285, 288, 383, 389, 420, 441, 443, 445, 480 |
| | DB00951 | 288, 293, 383, 389, 420, 442, 445, 480 |
| | DB14159 | 288, 389, 420, 441, 445 |
| *Closeness Centrality (CC)* | DB00312 | 289, 290, 292, 293, 388, 389, 390, 391, 419, 420, 421, 445 |
| | DB00951 | 289, 290, 292, 293, 388, 389, 390, 391, 419, 420, 421, 445 |
| | DB14159 | 289, 290, 292, 293, 388, 389,390, 391, 419, 420, 421, 445 |
| *Degree Centrality* (*DC*) | DB00312 | 284, 286, 289, 296, 328, 389, 391, 420 |
| | DB00951 | 286, 289, 293, 296, 328, 389, 420 |
| | DB14159 | 296, 328, 389 |
| *Eigenvector Centrality* (*EC*) | DB00312 | 286, 287, 289, 290, 291, 293, 328, 331, 390, 420 |
| | DB00951 | 286, 289, 290, 293, 328, 390, 420 |
| | DB14159 | 286, 289, 290, 293, 328, 390, 420 |
| *Katz Centrality* (*KC*) | DB00312 | 286, 287, 289, 290, 291, 293, 328, 389, 390, 420 |
| | DB00951 | 286, 289, 290, 293, 328, 389, 390, 420 |
| | DB14159 | 286, 289, 290, 293, 389, 390, 420 |

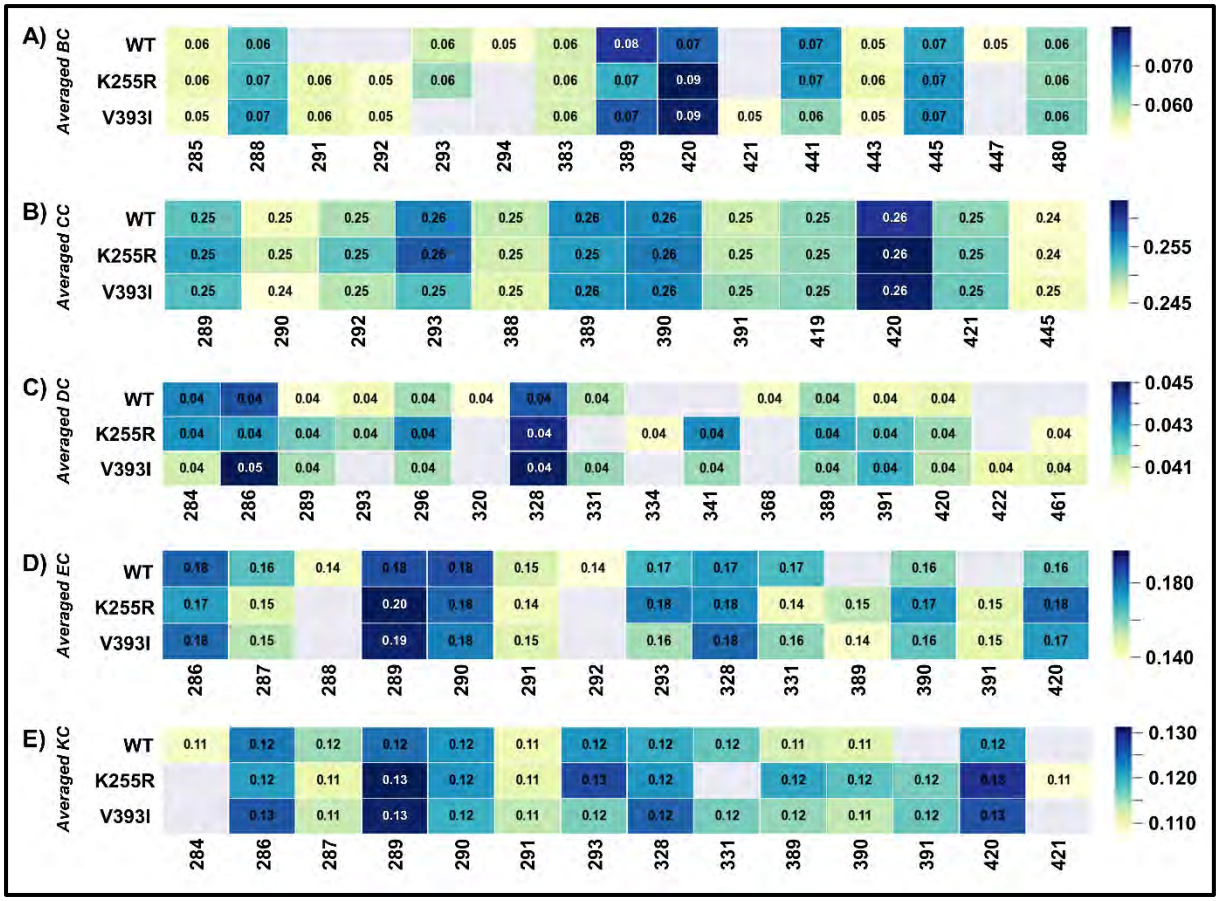



**Figure 4.20: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems of FP-2 bound to DB00312.** Hubs are shown along the x-axis, and WT and mutant complexes along the y-axis. The detected hubs are annotated with a centrality value, and the homologous residues of the alternate samples are not annotated. Low to high central values are colour-coded from yellow, green, and blue.

### *DB00951*

DB00951 was stable in the WT and six mutant systems (E248D, E249A, D345G, A353T, A400P, and V393I). *BC persistent hubs* across the WT and mutant systems are F288, S293, L383, I389, M420, Y442, I445, and I480 (**Figure 4.21A**). Visual inspection and mapping of the hubs on the protein structure show that most hubs are located within the binding site of FP-2 (**Figure S4.9**). A number of *distinctive hubs* were observed in the mutant systems; E248D (I291 and Y443), E249A (V419 and Y443), D345G (V419 and Y443), A400P (Y443), and V393I (I291). *CC persistent hubs* across the WT and mutant systems are S289, S290, G292, S293, P388, I389, S390, I391, V419, M420, L421, and I445 (**Figure 4.21B**). *DC persistent hubs* across the WT and mutant systems are W286, S289, S293, S296, I328, I389, and M420 (**Figure 4.21C**). Active site residue A394 was observed as a *distinctive hub* in E248D, E249A,

and V393I. Additionally *distinctive hubs* were observed in A353T (S290, I445, G457, and I461) and A400P (A331, I368, V422, and I461).

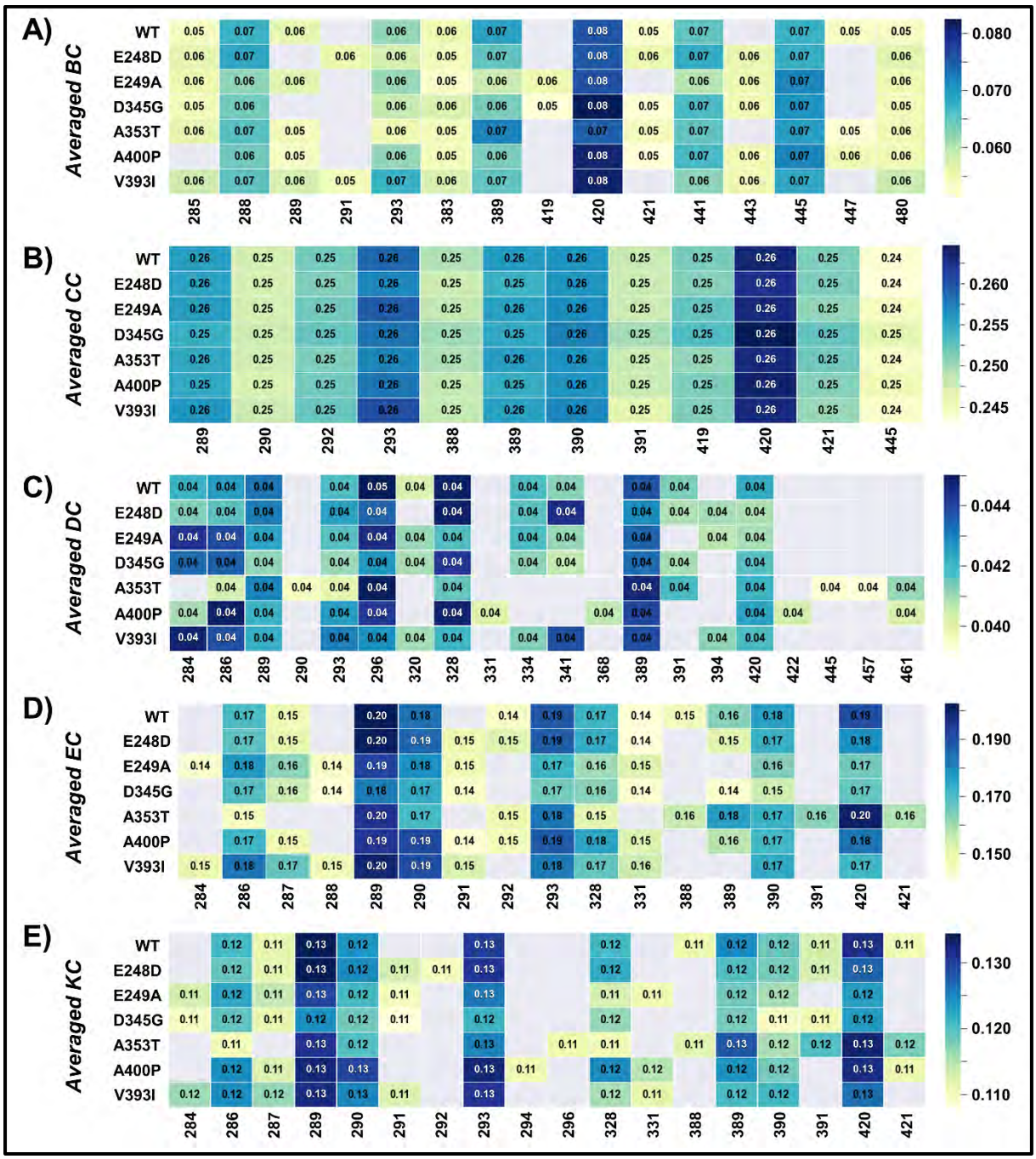



**Figure 4.21: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems bound to DB00951.** Hubs are shown along the x-axis, and WT and mutant complexes along the y-axis. The detected hubs are annotated with a centrality value, and the homologous residues of the alternate samples are not annotated. Low to high central values are colour-coded from yellow, green, and blue.

*EC persistent hubs* across the WT and mutant systems are W286, S289, S290, S293, I328, S390, and M420 (**Figure 4.21D**). Visual inspection and mapping of the *EC* hubs on the protein

structure show that most hubs are located within the binding site of FP-2 (**Figure S4.10**). Each hub is coloured according to whether it occurs only in the mutant-bound systems or is shared with the WT-bound system. We identified a semi-allosteric path consisting of three *distinct hubs* (I291-F288-S284) that are connected to each other in E249A (**Figure 4.22B**) and V393I (**Figure 4.22C**) bound systems. *KC persistent hubs* across the WT and mutant systems are W286, S289, S290, S293, I328, I389, S390, and M420 (**Figure 4.21E**). Several distinctive hubs were observed in the mutant systems; E248D (I291 and G292), E249A (S284, I291, and A331), D345G (S284 and I291), A353T (S296), A400P (V294), and V393I (S284, I291, and A331).

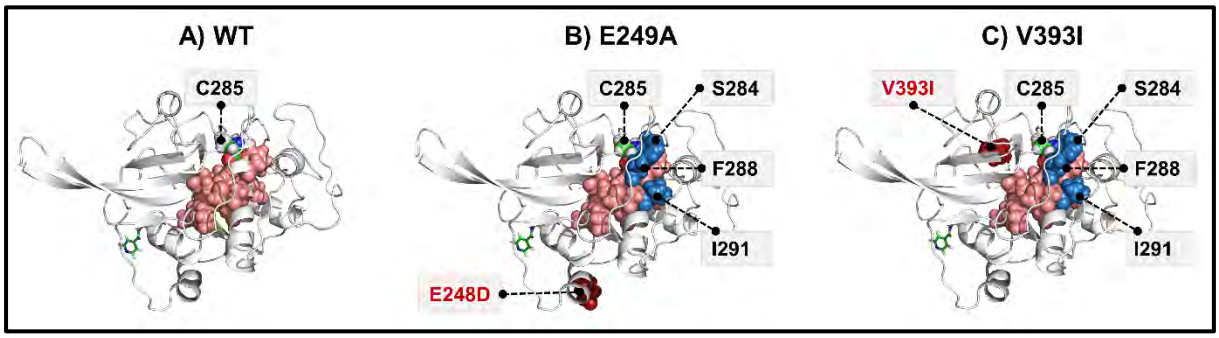



**Figure 4.22: Cartoon representation of the averaged *EC* hubs of the WT and mutant systems.** A) WT with the distribution of averaged *EC* hubs depicted as salmon spheres. B) and C) E249A and V393I with their distinct *EC* hubs depicted in skyblue spheres and the one common to the WT hubs (salmon).

### *DB14159*

DB14159 was stable in the WT and nine mutant systems (M245I, E248D, E249A, N257E, T343P, D345G, A353T, V393I, and A400P) (**Figure 4.23A**). *BC* persistent hubs across the WT and mutant systems are F288, I389, M420, Y441, and I445. Visual inspection and mapping of the hubs on the protein structure show that most hubs are located within the binding site of FP-2 (**Figure S4.11**). Active site residue N447 increased usage in the majority of the mutant systems (M245I, E249A, T343P, D345G, and A353T). I291 also appeared as a *distinctive hub* in the majority of the mutant systems (N257E, T343P, D345G, V393I, and A400P). Additional *distinctive hubs* present in the mutant systems include M245I (L383), E248D (L383 and I391), E249A (V294 and L383), N257E (I328). *CC* persistent hubs across the WT and mutant systems are S289, S290, G292, S293, P388, I389, S390, I391, V419, M420, L421, and I445 (**Figure 4.23B**). *DC* persistent hubs across the WT and mutant systems are S296, I328, and I389 (**Figure 4.23C**). *Distinctive hubs* observed in mutant systems include M245I (L313 and I368), E248A (S293, I368, and G457), E249A (I368, V374, I461, and I480), N257E (A331, I341,

G457, and I461), T343P (S293, I341, A394 and G457), D345G (S293, I341, and I461), A353T (A331, I368, G457, and I461), V393I (A394), and A400P (S293, A331, and A394).

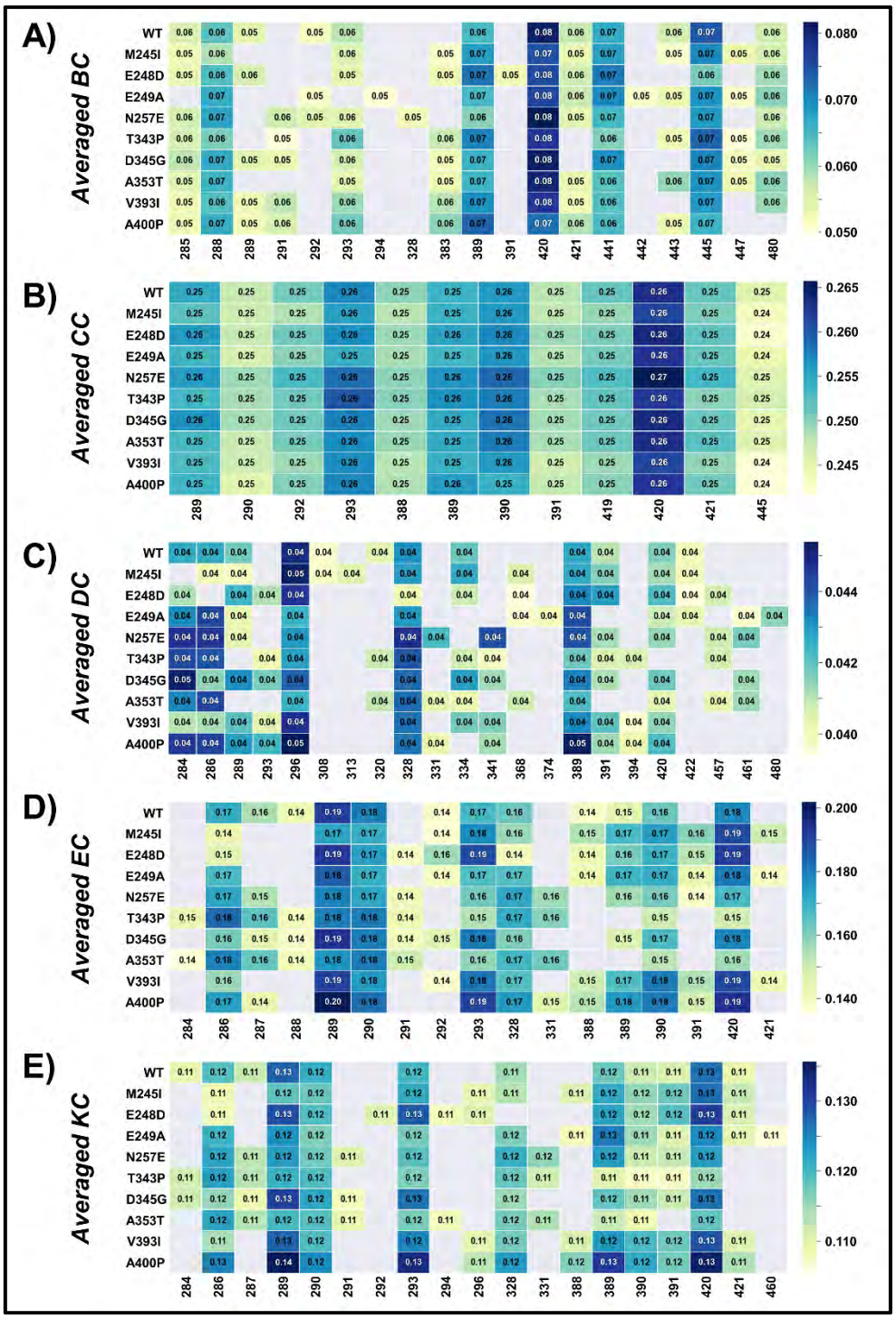



**Figure 4.23: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems of FP-2 bound to DB14159.** Hubs are shown along the x-axis, and WT and mutant complexes along the y-axis. The detected hubs are annotated with a centrality value, and the homologous residues of the alternate samples are not annotated. Low to high central values are colour-coded from yellow, green, and blue.

*EC* persistent hubs across the WT and mutant systems are W286, S289, S290, S293, I328, S390, and M420 (**Figure 4.23D**). *Distinctive hubs* in mutant systems include M245I (I391 and L421), E248D (I291 and I391), E249A (I391 and L421), N257E (I291, A331, and I391), T343P (S284 and I291), D345G (I291), A353T (S284, I291, and A331), V393I (I391 and L421), and A400P (A331, and I391). *KC* persistent hubs across the WT and mutant systems are W286, S289, S290, S293, I389, S390, and M420. *Distinctive hubs* in mutant systems include M245I (S296 and P388), E248D (G292 and V294), E249A (P388 and N460), N257E (I291 and A331), T343P (A331), D345G (I291), A353T (I291, V294, and A331), V393I (S296 and P388), and A400P (S296 and P388).

#### 4.5.9.2 Falcipain 3

***DB00853***

DB00853 was stable in WT and two mutant systems (R411K and E446G). Persistent hubs across all systems are highlighted in **Table 4.3**. *BC* persistent hubs across the WT and mutant systems are F296, S301, I397, I428, L429, I453, and V488 (**Figure 4.24A**). In the WT system, G300 and Y451 were distinctive and absent in the mutant systems indicating the loss of communication in the presence of mutations. *Distinctive hubs* observed in mutant system R411K are S297, V302, and L391. *Distinctive hubs* observed in mutant system E446G are S297, V302, L391, and F448. *CC* persistent hubs across the WT and mutant systems are S297, S298, G300, S301, P396, I397, S398, I399, V427, I428, and L429 (**Figure 4.24B**). Active site residue I336 appeared as a *distinctive hub* in E446G. *DC* persistent hubs across the WT and mutant systems are S292, W294, S297, S301, S304, N328, I336, I397, and I428 (**Figure 4.24C**). The usage of residue D342 and active site residue A402 were lost in the presence of the mutant systems. *Distinctive hubs* observed in mutant system R411K are F296 and V430. *Distinctive hubs* observed in mutant system E446G are V302, V322 and V430. *EC* persistent hubs across the WT and mutant systems are W294, A295, S297, S298, G300, S301, I336, S398, and I428 (**Figure 4.24D**). *Distinctive hubs* were only noticed in mutant system E446G, including S304, P396, and I397. *KC* persistent hubs across the WT and mutant systems are W294, A295, S297, S298, S301, I336, I397, S398, and I428 (**Figure 4.24E**). The usage of residues G300 and A339 were lost in the presence of the mutant systems. *Distinctive hubs* observed in mutant system R411K are F296 and S304. *Distinctive hubs* observed in mutant system E446G are S292, S304, and P396.

**Table 4.4: Persistent hubs in all WT and mutant-bound systems of FP-3**.

| Metric | Persistent hubs | |
|---|---|---|
| *Betweenness Centrality (BC)* | DB00853 | 296, 301, 397, 428, 429, 453, 488 |
| | DB00951 | 296, 397, 428, 429, 453, 488 |
| *Closeness Centrality (CC)* | DB00853 | 297, 298, 300, 301, 396, 397, 398, 399, 427, 428, 429 |
| | DB00951 | 297, 298, 300, 301, 396, 397, 398, 399, 427, 428, 429 |
| *Degree Centrality* (*DC*) | DB00853 | 292, 294, 297, 301, 304, 328, 336, 397, 428 |
| | DB00951 | 292, 294, 301, 304, 336, 397, 401, 430 |
| *Eigenvector Centrality* (*EC*) | DB00853 | 294, 295, 297, 298, 300, 301, 336, 398, 428 |
| | DB00951 | 294, 297, 298, 300, 301, 336, 398, 428 |
| *Katz Centrality* (*KC*) | DB00853 | 294, 295, 297, 298, 301, 336, 397, 398, 428 |
| | DB00951 | 294, 297, 298, 301, 336, 398, 428 |

***DB00951***

Interestingly, DB00951 was stable in WT and all mutant systems. *BC* persistent hubs across the WT and mutant systems are F296, I397, I428, L429, I453, and V488 (**Figure 4.25A**). *Distinctive hubs* observed in some mutant systems include N371K (S297), R411K (S297), E446G (S297 and L391), and N468Y (G300, I399, and V427). C293 (catalytic thiol), I399, and I401 (active site residue) were *distinctive hubs* in A422T. *CC* persistent hubs across the WT and mutant systems are S297, S298, G300, S301, P396, I397, S398, I399, V427, I428, and L429 (**Figure 4.25B**). Active site residue I336 appeared as a *distinctive hub* in mutant systems A422T and N468Y. Additionally, I453 appeared as a *distinctive hub* in E446G. *DC* persistent hubs across the WT and mutant systems are S292, W294, S301, S304, I336, I397, and I428 (**Figure 4.25C**). The usage of residue G465 were lost in the presence of the mutant systems. Additionally, active site residue A402 was lost in the presence of all mutant systems except A422T. ). *Distinctive hubs* observed in some mutant systems include N371K (W275, I428, and V488), R411K (W275 and I428), A422T (W275 and D342), E446G (F296 and I428), and N468Y (S317, A339, and L349). *EC* persistent hubs across the WT and mutant systems are W294, S297, S298, G300, S301, I336, S398, and I428 (**Figure 4.25D**). The usage of residue S292 were lost in the presence of the mutant systems. *Distinctive hubs* observed in some mutant systems include N371K (P396), R411K (S304, A339, and P396), A422T (A339), E446G (S304, P396, and I397). *KC* persistent hubs across the WT and mutant systems are W294, S297,

S298, S301, I336, S398, and I428 (**Figure 4.25E**). S304 appeared as a *distinctive hub* in all mutant systems except A422T. Additional distinctive hubs were observed in some mutant systems, which include N371K (G300), R411K (V430), A422T (F296), E446G (V302 and L429), and N468Y (A339).

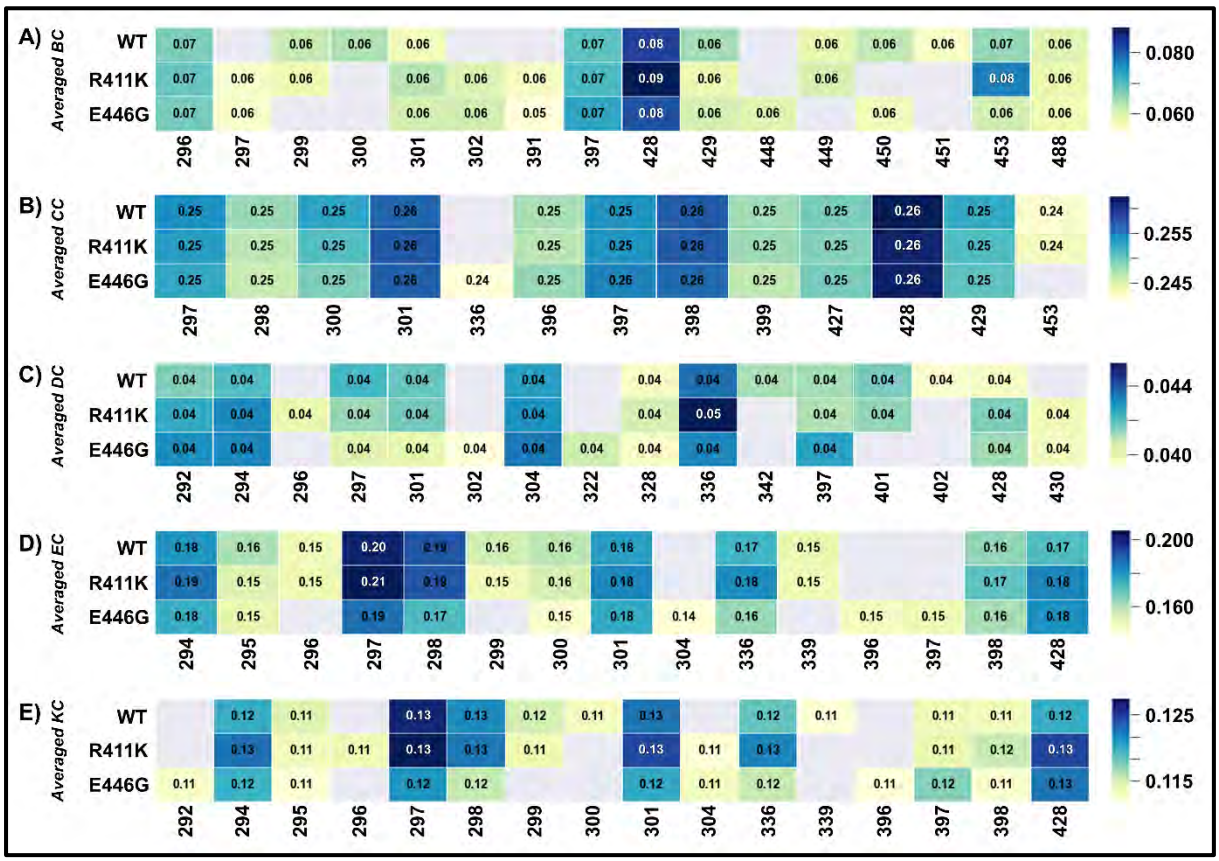



**Figure 4.24: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems of FP-3 bound to DB00853.** Hubs are shown along the x-axis, and WT and mutant complexes along the y-axis. The detected hubs are annotated with a centrality value, and the homologous residues of the alternate samples are not annotated. Low to high central values are colour-coded from yellow, green, and blue.

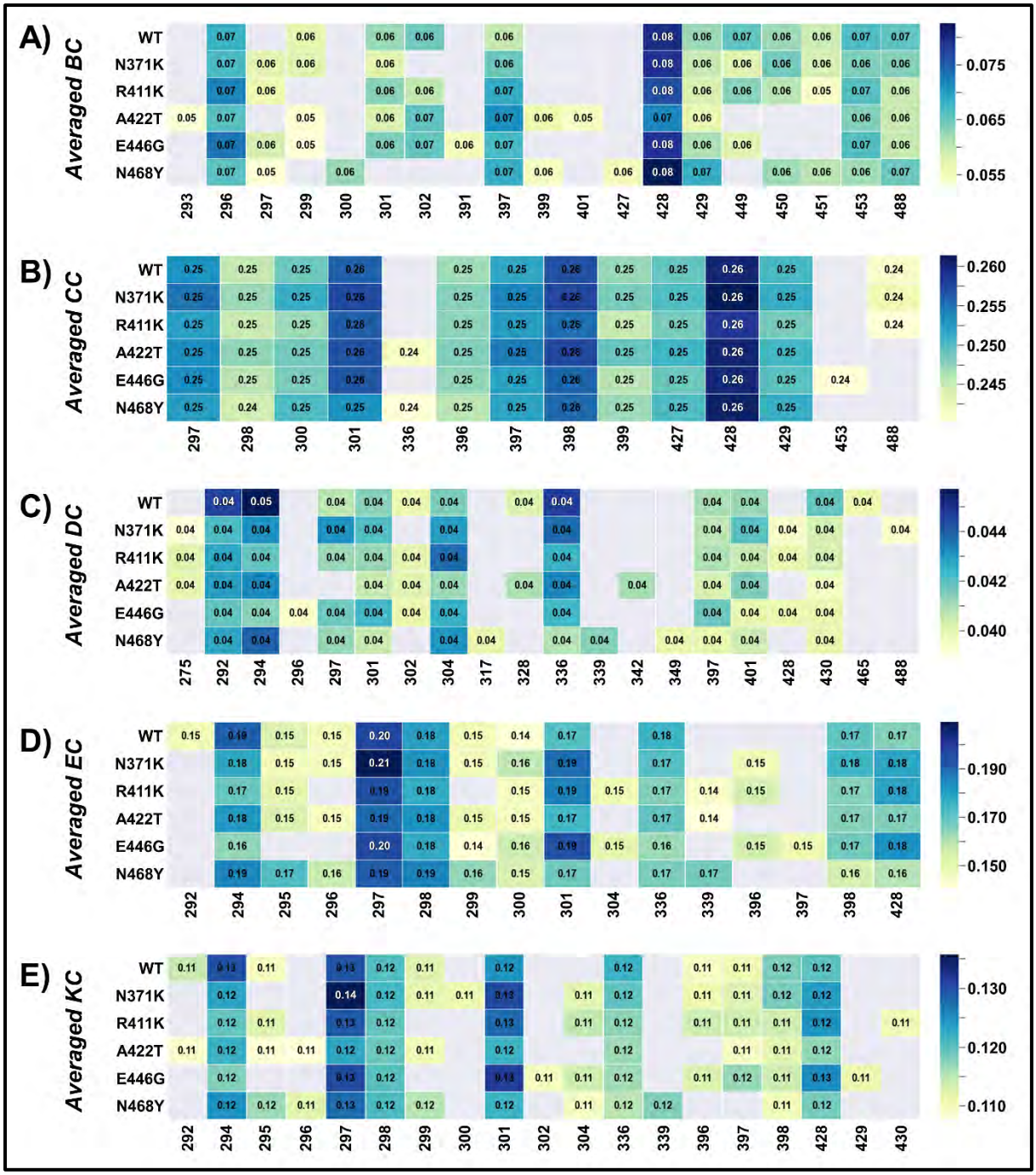



**Figure 4.25: Heatmap representation of the global top 5% for each centrality metric of the WT and mutant systems of FP-3 bound to DB00951.** Hubs are shown along the x-axis, and WT and mutant complexes along the y-axis. The detected hubs are annotated with a centrality value, and the homologous residues of the alternate samples are not annotated. Low to high central values are colour-coded from yellow, green, and blue.

## 4.6 Conclusion

In recent years, allosteric modulators have received considerable attention because they offer several advantages over classical orthosteric modulators [249]. Thus, this chapter aims to identify potential allosteric modulators targeting FP-2 and FP-3 using *in silico* approaches such as high-throughput virtual screening, molecular dynamics, comparative essential dynamics, and dynamic residue network (DRN) analysis. In Chapters 2 and 3, we identified six potential allosteric pockets in FP-2 and FP-3 using various computational tools. Of the six potential allosteric pockets, pocket 1 was considered druggable in both FP-2 and FP-3; hence we screened 2089 FDA compounds obtained from DrugBank to identify potential allosteric modulators binding to pocket 1. High-throughput virtual screening identified 22 and 15 compounds successfully bound to FP-2 and FP-3, respectively. In order to isolate compounds specific to *Plasmodium* proteins, compounds binding to human homologs (Cat K and Cat L) were excluded from further analysis. A total of eight compounds were shortlisted for FP-2, and five were shortlisted for FP-3 based on the above criteria. In both FP-2 and FP-3, DB00951 (Isoniazid) was identified as a potential hit. These compounds were subjected to 100 ns to evaluate their stability on the respective WT systems. Three compounds (DB00312, DB00951, and DB14159) in FP-2 and two (DB00853 and DB00951) in FP-3 were stable and further subjected to MD simulation to assess their stability and effect on the respective mutant systems. Post-MD analysis and comparative essential dynamics revealed the effect of the potential hits on the mutant systems. We determined the effects of the potential hit on the six potential allosteric pockets. The presence of a cryptic pocket was observed in A400P mutant system when bound to DB14159. DRN analysis identified the changes in communication networks in the WT and mutant systems when bound to the potential hits. Overall, DB00312, DB00951, and DB14159 in FP-2 and DB00853 and DB00951 in FP-3 have been identified as potential allosteric hits that will be further examined in vitro for their allosteric properties.

# Chapter 5

## 5.1 Conclusion and Future Perspectives

Antimalarial drugs are an integral part of global malaria mitigation and elimination efforts. However, the parasite's resistance to all currently available antimalarial drugs is the biggest challenge toward malaria treatment and eradication. Mutations in genes encoding key plasmodial enzymes confer antimalarial resistance. Determining the consequences of mutations on protein structure and function continues to be a critical method for understanding the underlying resistance mechanisms. With increasing technological advancements in the post-genomics era, numerous computational approaches for elucidating mutation effects have been established. This thesis provided insights into the conserved mutational effects in FP-2 and FP-3. Both enzymes are synthesized as zymogens and subsequently activated during the erythrocytic stage and play integral roles in nutrient acquisition and mediation of ART therapeutic action in *P. falciparum*. Thus, deciphering underlying ART resistance mechanisms is paramount with the increased detection and spread of ART resistance Pf strains.

The objective of Chapter 2 was to analyse the effects of mutations on the structure and dynamics of FP-2. Mutations were identified from the PlasmoDB [101,102]. With the exception of V393I and A400P, all other mutations are distal to the orthosteric site. WT and mutant systems were predicted using homology modeling, and these structures were validated using Verify3D, ProSA, and PROCHECK. We identified six putative pockets using a combination of structure-based tools, and the effects of mutations on these pockets were also characterized. We performed MD simulations and observed the conformational changes associated with each mutation via post-MD analysis, such as RMSD, Rg, and RMSF. Using comparative essential dynamics analysis, we analyzed the correlation in dynamics of the different systems (WT and mutant protein systems). The analysis was based on the whole protein, central core residues, binding pocket residues, and potential allosteric pockets. Observations from the global level reveal the effects of mutations on the conformation of the partial zymogen complex and the catalytic domain of FP-2. We zoomed in on the potential allosteric pockets to analyse the effects. In the catalytic domain, M245I, located in the nose region, a structural characteristic unique to FP-2 and other related Plasmodium homolog, exhibited the highest conformational changes compared to the WT. This resulted in the formation of a cryptic pocket where pocket 6 opened up, and residues in pocket 2 migrated into it. Interestingly, in the partial zymogen complex of FP-2, mutant systems A159V, M245I, and

E249A had the same effect (formation of the cryptic pocket) on their respective systems. Overall, the global analysis revealed the conformational changes associated with mutations. Application of local analysis based on the centrality and weighed residue contact maps presents a reliable approach to identifying subtle changes associated with mutations, especially in proteins showing conformational variability. Using DRN analysis, we examined the effects of mutations on the residue communication network and identified key residues responsible for the alterations in the conformational variability of the mutant systems. Overall, in silico analysis provided insights into the effects of mutations on the partial zymogen complex and the catalytic domain of FP-2.

The objectives in Chapter 3 mirror that of Chapter 2, and in this case, we characterize the effects of mutations on FP-3. FP-2 and FP-3 share a sequence similarity of 68%, so we sought to understand the effects of mutations on FP-3 and establish any similarity with FP-2. Mutations occurring in the partial zymogen complex and catalytic domain were identified from the *Pf* genomic resource database (PlasmoDB) [101,102]. Mutant systems were built using the WT as a template, and these systems were validated using Verify3D, ProSA, and PROCHECK. We also identified six putative pockets using a combination of structure-based tools, and the effects of mutations on the whole protein, central core, binding site residues, and the potential allosteric pockets were characterized using long MD simulations. The conformational sampling, global compactness, and residue contribution to variations were evaluated using RMSD, Rg, and RMSF combined with comparative essential dynamics analysis. Observation of the combined global analysis revealed the conformational variability associated with mutations. In the presence of mutation A422T, a cryptic pocket was observed. A422T occurs around the active site and has various conformational effects on various parts of its system. Local analysis revealed the changes in the communication network associated with mutations. Interestingly, a correlation was observed between the DRN analysis, weighted contact map, and alanine scanning. This local analysis revealed the communication changes in the interface residues and the general protein network.

Finally, based on the potential allosteric pockets identified in chapters 2 and 3, the most druggable pocket (pocket 1) in FP-2 and FP-3 was screened in an effort to uncover allosteric modulators. High-throughput virtual screening (HTVS) was applied to this course. The compound library used in this study contained 2089 FDA compounds culled from the DrugBank database. HTVS identified 22 compounds binding to FP-2 and 15 to FP-3. Certain

criteria were set to ensure target specificity, including binding within the potential allosteric site and interaction with allosteric residues. A further criterion to ensure specificity to *Plasmodium* protein is the exemption of ligands binding to the human homolog (Cat K and Cat L). Using the highlighted criteria, eight compounds were shortlisted for FP-2 and five for FP-3. Interestingly, Isoniazid (DB00951) was shortlisted for both FP-2 and FP-3. The selected compounds were then taken to 100 all-atom MD simulations. The criteria for selecting compounds were based on their stability in the respective WT systems. One of Chapter 4's aims was to analyze the effect of selected compounds in the presence of mutations. So stable compounds in the WT systems were selected for further mutation analysis studies. The following compounds have been identified as potential allosteric hits: DB00312, DB00951, and DB14159 in FP-2; DB00853 and DB00951 in FP-3; these compounds will be further investigated for their allosteric properties *in vitro*.

In conclusion, we demonstrated how different *in silico* tools and methods can be used as a robust and efficient way to elucidate the role of mutations in conferring antimalarial drug resistance. As a complement to existing web servers using static information (3D structures) to determine mutation effects, our hybrid approach incorporating global and local dynamic analysis may be useful in various protein systems. Additionally, this framework could be applied to other targets of interest. The current findings provide a novel understanding of the possible effects of missense mutations which may ultimately alter the catalytic efficiency of FP-2 and FP-3. Allosteric drugs targeting FP-2 and FP-3 are receiving increased attention in the antimalarial drug discovery process; this study offers a fresh exploration of allosteric modulators and the effects of mutations on these modulators. As a first-ever study, this study provides important insights which may help in the discovery of new drugs against these hemoglobinases.

# Appendix

**Table S1:** Quality assessment of protein modeled structures.

| Domain | Model | z-DOPE score | Verify 3D score (%) | ProSA | PROCHECK (%) | | |
|---|---|---|---|---|---|---|---|
| | | | | | Ramachandran (residue location) | | |
| | | | | | Favored | Allowed | Disallowed |
| Catalytic domain | WT | -1.26 | 91.29 | -8.29 | 90.40 | 9.10 | 0.50 |
| | M245I | -1.16 | 92.53 | -8.11 | 92.80 | 6.70 | 0.50 |
| | E248D | -1.17 | 87.97 | -8.31 | 91.90 | 7.60 | 0.50 |
| | E249A | -1.14 | 85.89 | -8.25 | 92.80 | 6.70 | 0.50 |
| | K255R | -1.16 | 92.95 | -8.22 | 92.30 | 7.20 | 0.50 |
| | N257E | -1.13 | 90.04 | -8.12 | 92.30 | 7.20 | 0.50 |
| | T343P | -1.19 | 90.04 | -8.10 | 92.30 | 7.20 | 0.50 |
| | D345G | -1.20 | 91.29 | -8.17 | 92.80 | 6.70 | 0.50 |
| | A353T | -1.18 | 90.87 | -8.19 | 92.80 | 6.70 | 0.50 |
| | V3931 | -1.18 | 90.46 | -8.16 | 91.90 | 7.60 | 0.50 |
| | A400P | -1.19 | 88.80 | -8.08 | 92.30 | 7.20 | 0.50 |
| | Q414E | -1.15 | 88.80 | -8.21 | 92.30 | 7.20 | 0.50 |

**Table S2:** Quality assessment of protein modeled structures.

| Domain | Model | z-DOPE score | Verify 3D score (%) | ProSA | PROCHECK (%) | | |
|---|---|---|---|---|---|---|---|
| | | | | | Ramachandran (residue location) | | |
| | | | | | Favored | Allowed | Disallowed |
| Partial zymogen complex | WT | -1.05 | 88.48 | -8.27 | 88.90 | 10.40 | 0.70 |
| | A159V | -0.98 | 80.30 | -8.15 | 91.20 | 8.10 | 0.70 |
| | F165I | -0.97 | 82.73 | -8.08 | 91.20 | 8.10 | 0.70 |
| | M167T | -1.00 | 80.61 | -8.23 | 91.60 | 7.70 | 0.70 |
| | N197K | -0.98 | 80.30 | -8.18 | 90.90 | 8.40 | 0.70 |
| | N204K | -0.96 | 80.91 | -8.19 | 90.90 | 8.40 | 0.70 |
| | N224S | -0.98 | 84.55 | -8.15 | 90.90 | 8.40 | 0.70 |
| | S228T | -0.98 | 80.30 | -8.22 | 90.90 | 8.40 | 0.70 |
| | M245I | -0.96 | 80.91 | -8.16 | 90.20 | 9.50 | 0.30 |
| | E248D | -1.00 | 84.55 | -8.25 | 91.20 | 8.10 | 0.70 |
| | E249A | -0.98 | 80.61 | -8.13 | 91.20 | 8.10 | 0.70 |
| | K255R | -1.01 | 80.91 | -8.33 | 90.90 | 8.40 | 0.70 |
| | N257E | -0.96 | 82.12 | -8.13 | 90.90 | 8.40 | 0.70 |
| | T343P | -1.00 | 80.61 | -8.15 | 90.90 | 8.40 | 0.70 |
| | D345G | -1.01 | 81.21 | -8.08 | 90.90 | 8.40 | 0.70 |
| | A353T | -1.01 | 84.55 | -8.24 | 90.90 | 8.40 | 0.70 |
| | V393I | -0.99 | 80.91 | -8.34 | 90.20 | 9.10 | 0.70 |
| | A400P | -1.00 | 84.55 | -8.04 | 91.50 | 7.80 | 0.70 |
| | Q414E | -1.00 | 80.30 | -8.25 | 90.60 | 8.70 | 0.70 |

**Table S3:** Druggability potential of the predicted putative allosteric pockets as determined by AutoLigand and SiteMap.

| Pockets | D-score | Druggability |
|---|---|---|
| 1 | 0.63 | Moderate |
| 2 | 0.45 | Weak |
| 3 | 0.54 | Weak |
| 4 | 0.43 | Weak |
| 5 | 0.50 | Weak |
| 6 | 0.44 | Weak |

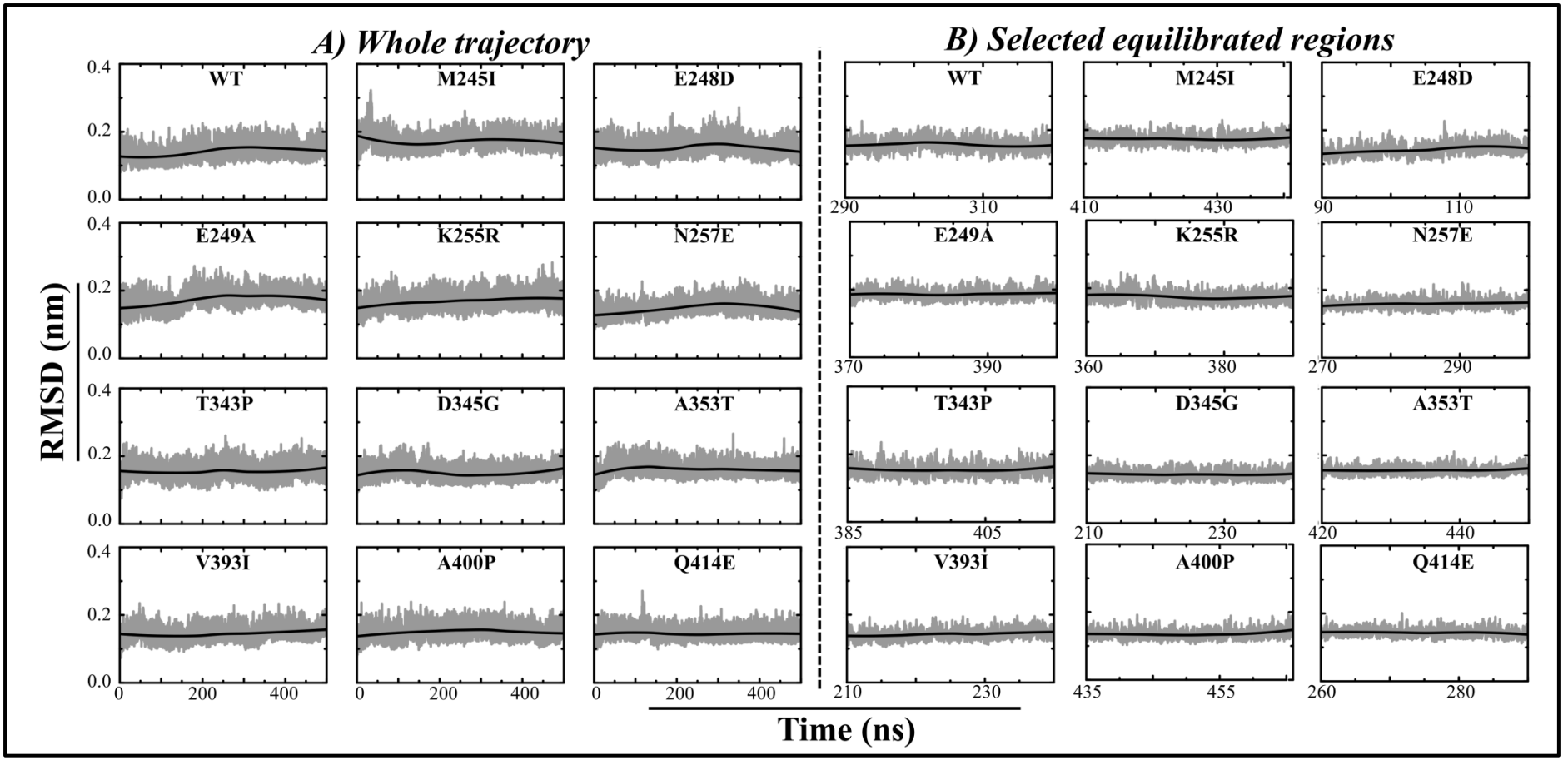



**Figure S2.1:** RMSD line plots of the whole trajectory (500 ns) of the WT and mutant systems of the catalytic domain of FP-2 showing conformational changes during the simulation and selected equilibrated regions (30 ns) used for DRN analysis. This figure is produced with permission from Okeke et al. 2021 [118].

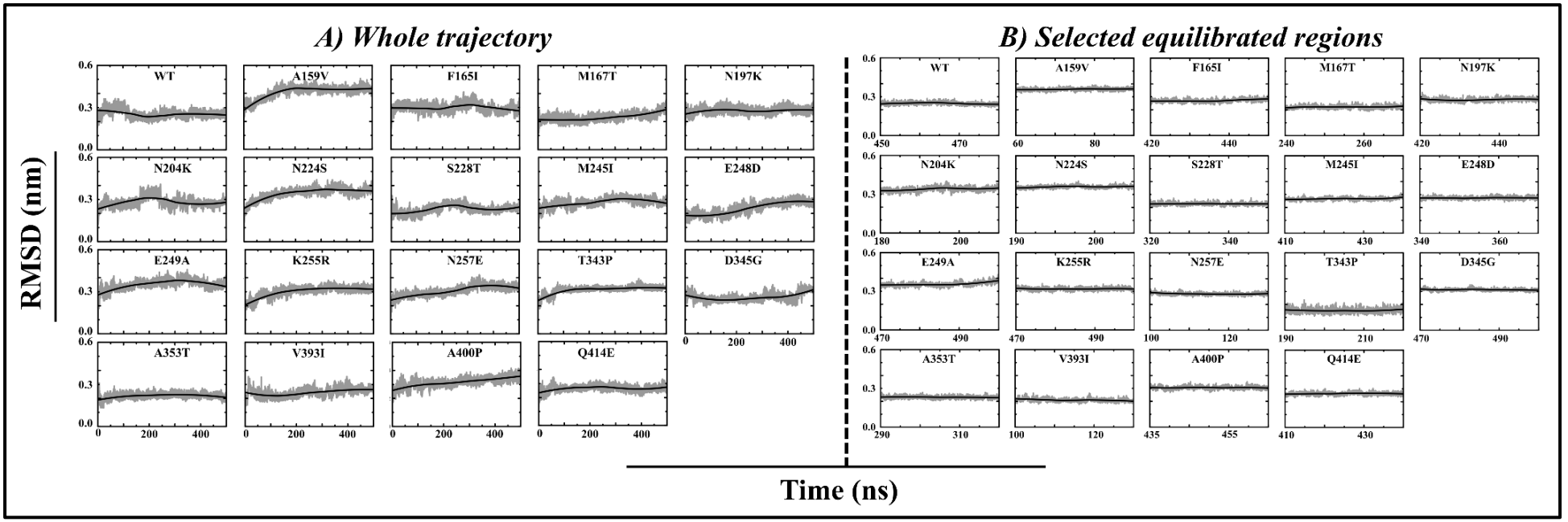



**Figure S2.2:** RMSD line plots of the whole trajectory (500 ns) of the WT and mutant systems of the partial zymogen complexes of FP-2 showing conformational changes during the simulation and selected equilibrated regions (30 ns) used for DRN analysis.

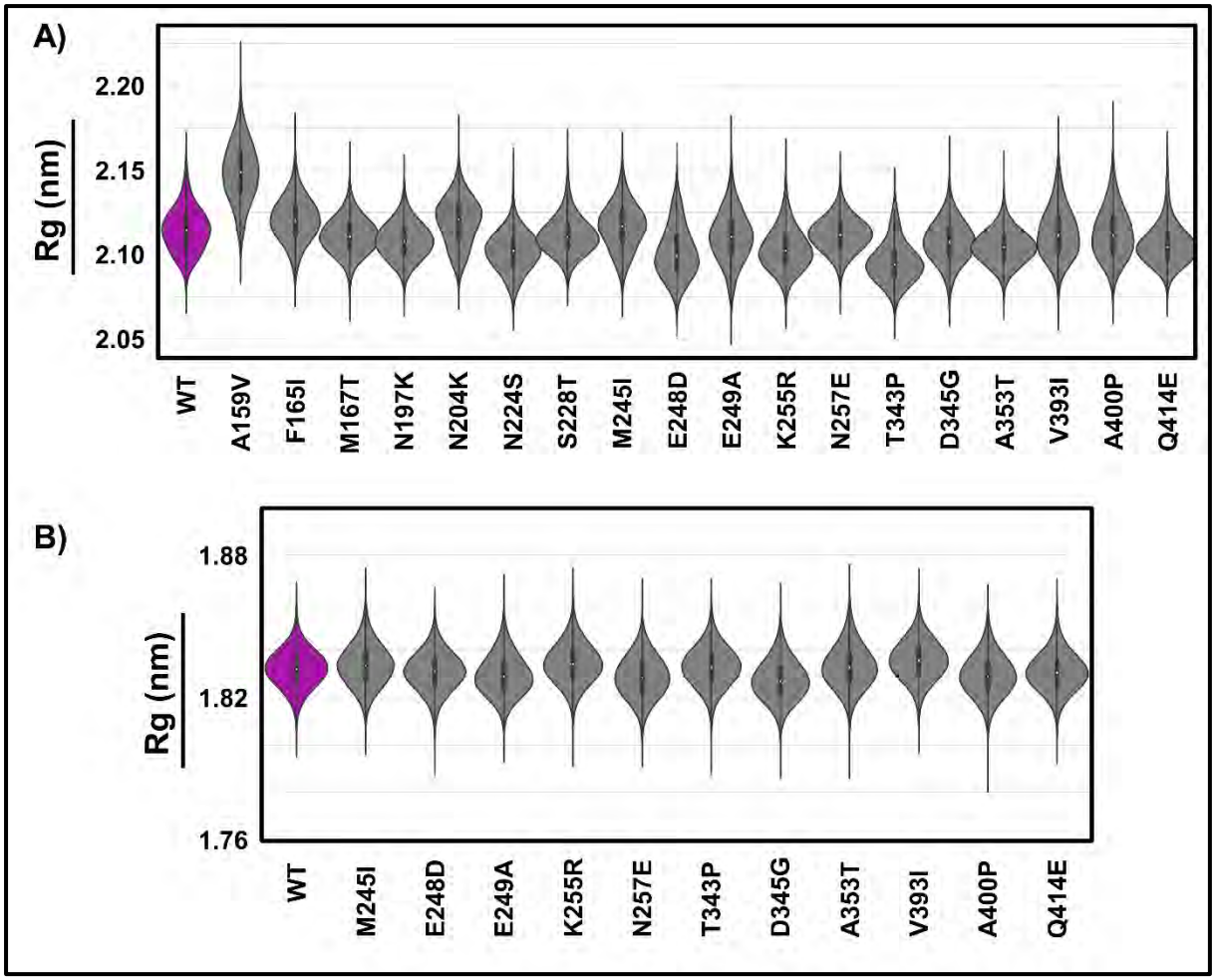



**Figure S2.3**: Rg violin distribution plots of the WT and mutant systems (whole protein) of A) partial zymogen and B) catalytic domain of FP-2.

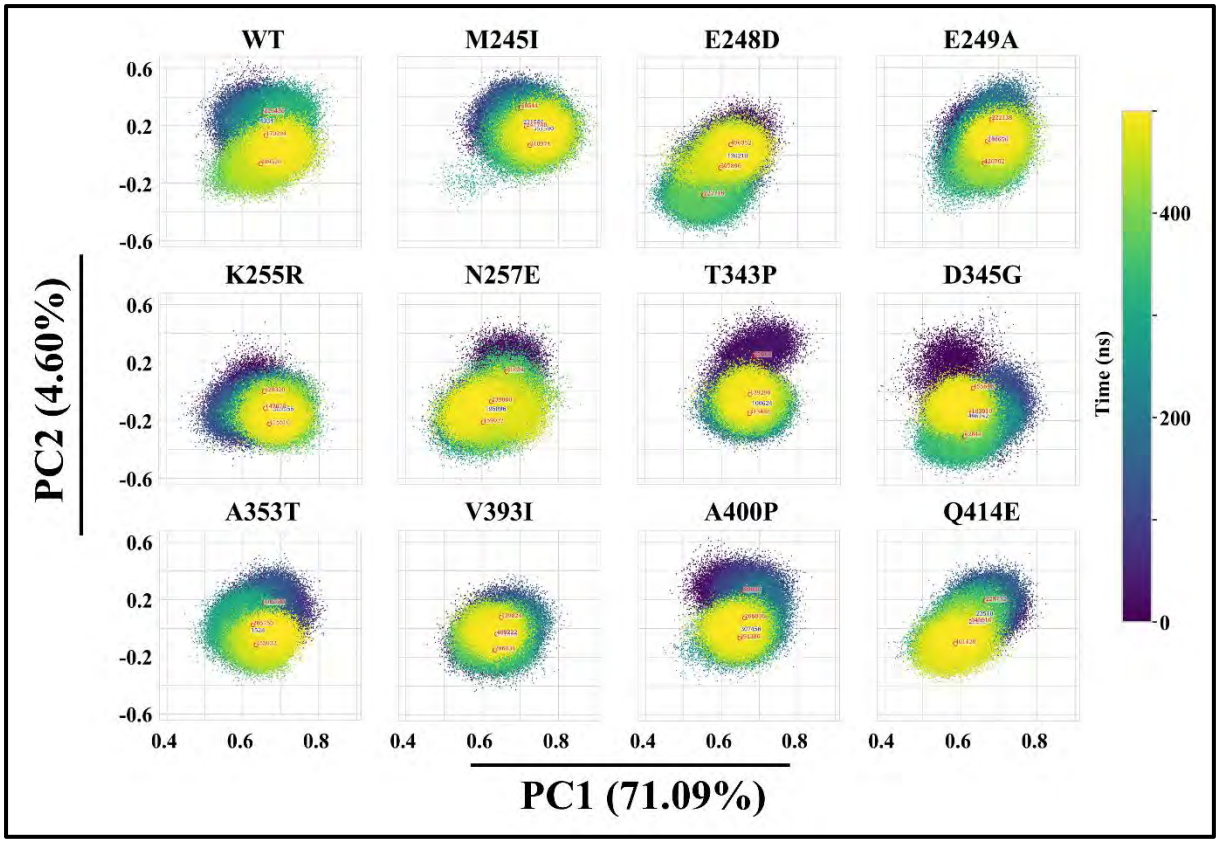



**Figure S2.4:** Comparative essential dynamic analysis of active site pocket residues of the WT and mutant systems of the catalytic domain of FP-2 along PC1 and PC2. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue. This figure is produced with permission from Okeke et al. 2021 [118].

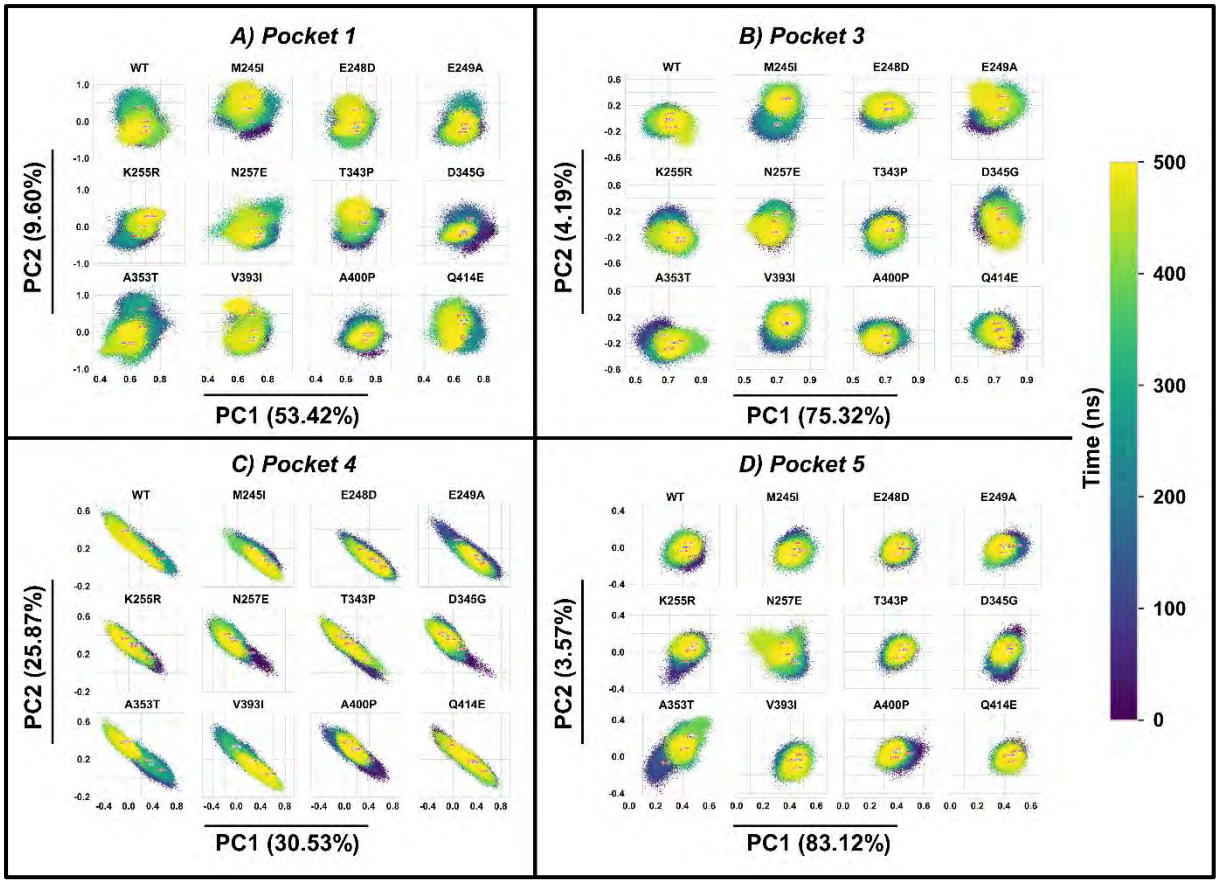



**Figure S2.5:** Comparative essential dynamic analysis of pockets 1, 3, 4, and 5 along PC1 and PC2 of the WT and mutant systems of the catalytic domain of FP-2. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue. This figure is produced with permission from Okeke et al. 2021 [118].

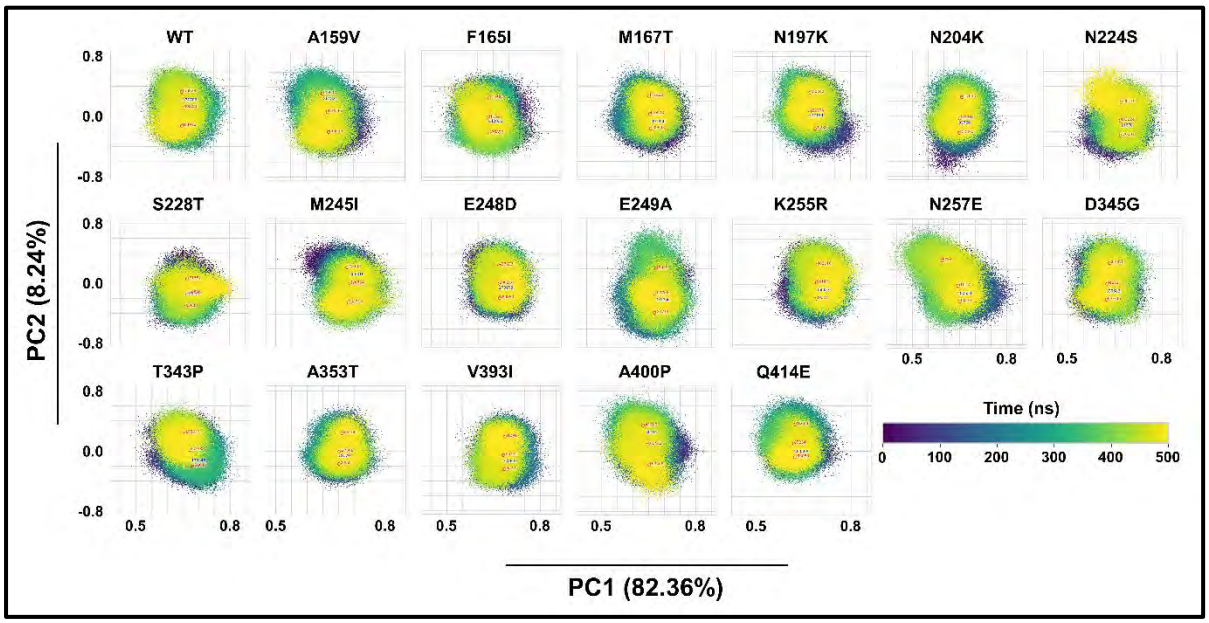



**Figure S2.6:** Comparative essential dynamic analysis of active site pocket residues of the WT and mutant systems of the partial zymogen complexes of FP-2 along PC1 and PC2. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

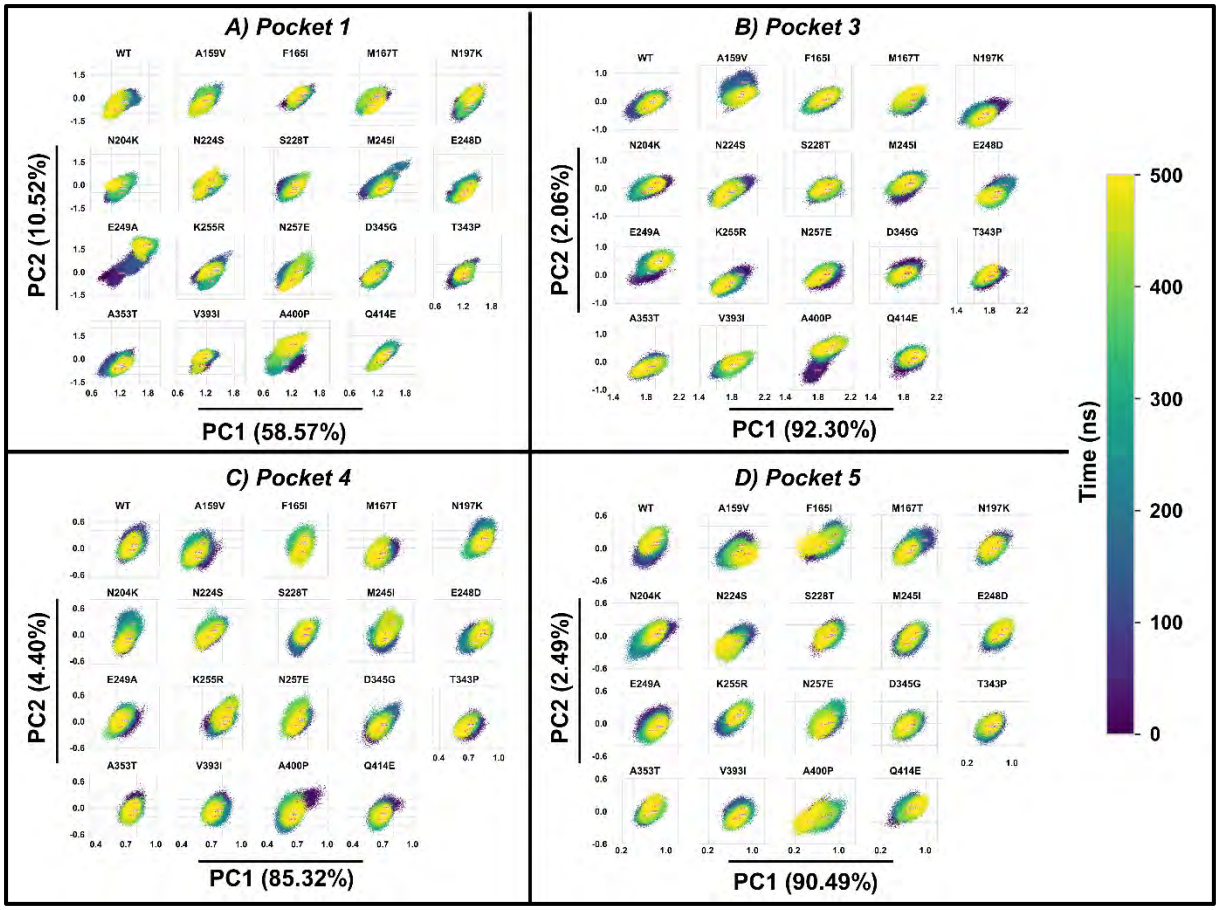



**Figure S2.7:** Comparative essential dynamic analysis of pockets 1, 3, 4, and 5 along PC1 and PC2 of the WT and mutant system of the partial zymogen complex of FP-2. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

**Table S4:** Quality assessment of the WT and mutant modeled structures of the partial zymogen complex and catalytic domain, respectively.

| Domain | Model | z-DOPE score | VERIFY 3D score (%) | ProSA | PROCHECK (%) | | |
|---|---|---|---|---|---|---|---|
| | | | | | Ramachandran (residue location) | | |
| | | | | | Favored | Allowed | Disallowed |
| Catalytic | WT | -1.41 | 89.67 | -7.35 | 85.60 | 13.90 | 0.50 |
| | N371K | -1.27 | 90.08 | -7.25 | 91.90 | 7.60 | 0.50 |
| | R411K | -1.27 | 92.98 | -7.28 | 92.30 | 7.20 | 0.50 |
| | A422T | -1.33 | 94.63 | -7.10 | 90.90 | 8.60 | 0.50 |
| | E446G | -1.26 | 98.35 | -7.27 | 91.30 | 8.20 | 0.50 |
| | N468Y | -1.29 | 90.08 | -7.42 | 91.40 | 8.10 | 0.50 |
| Partial zymogen | WT | -0.94 | 84.64 | -7.84 | 89.70 | 10.30 | 0.00 |
| | E177Q | -0.95 | 83.73 | -7.96 | 90.10 | 9.60 | 0.30 |
| | I193L | -0.95 | 83.13 | -7.73 | 91.50 | 8.20 | 0.30 |
| | N371K | -0.93 | 90.36 | -7.83 | 91.20 | 8.50 | 0.30 |
| | R411K | -0.91 | 89.16 | -7.77 | 91.20 | 8.50 | 0.30 |
| | A422T | -0.97 | 83.43 | -7.74 | 90.80 | 8.90 | 0.30 |
| | E446G | -0.93 | 84.64 | -7.84 | 91.10 | 8.60 | 0.30 |
| | N468Y | -0.94 | 81.33 | -7.85 | 91.20 | 8.50 | 0.30 |

**Table S5:** Druggability potential of the predicted putative allosteric pockets as determined by AutoLigand and SiteMap.

| Pockets | D-score | Druggability |
|---|---|---|
| 1 | 0.84 | Druggable |
| 2 | 0.76 | Moderate |
| 3 | 0.43 | Weak |
| 4 | 0.68 | Moderate |
| 5 | 0.51 | Weak |
| 6 | 0.51 | Weak |

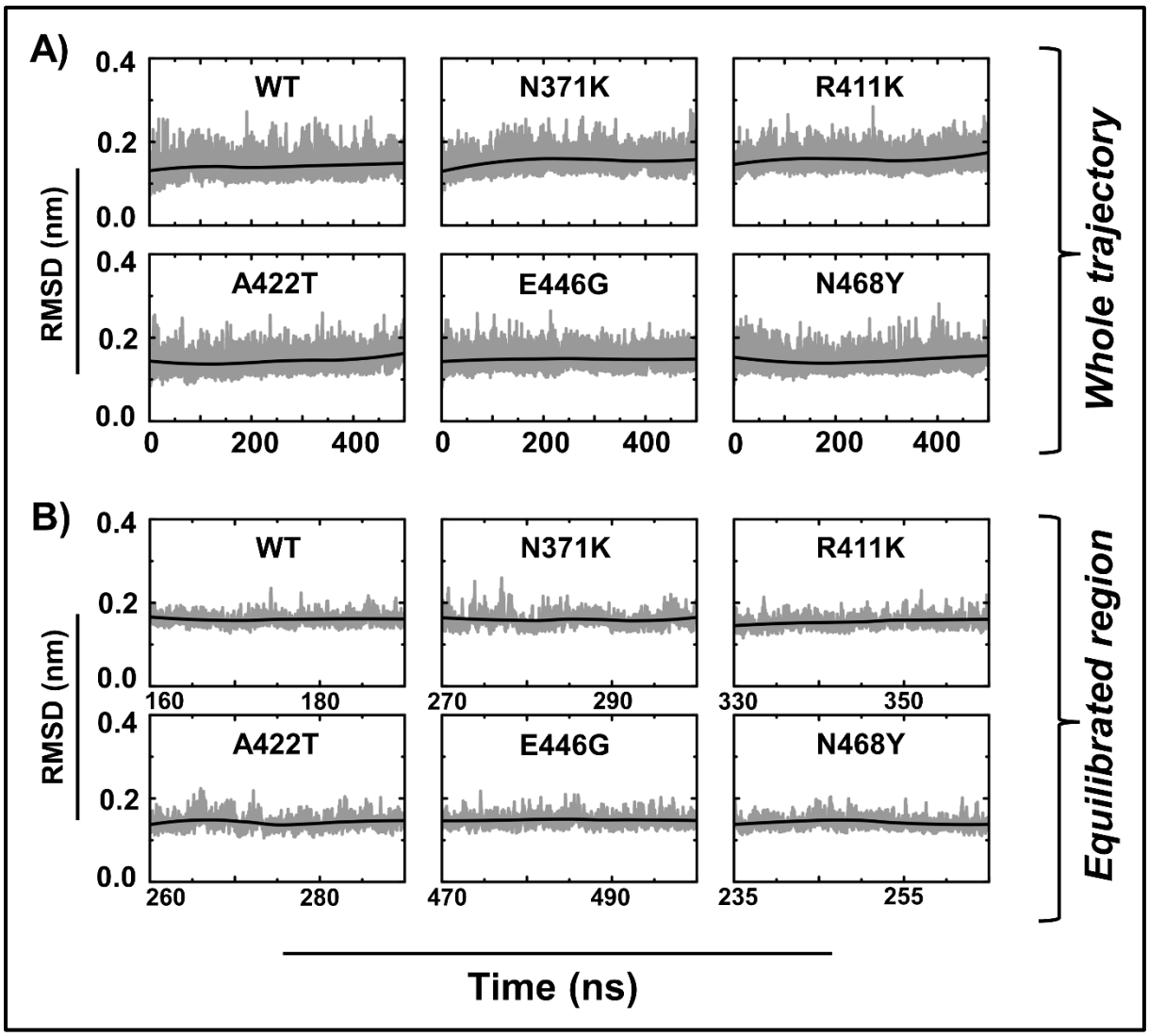



**Figure S3.1:** RMSD line plots of the whole trajectory (500 ns) of the WT and mutant systems of the catalytic domain of FP-3 showing conformational changes during the simulation and selected equilibrated regions (30 ns) used for DRN analysis.

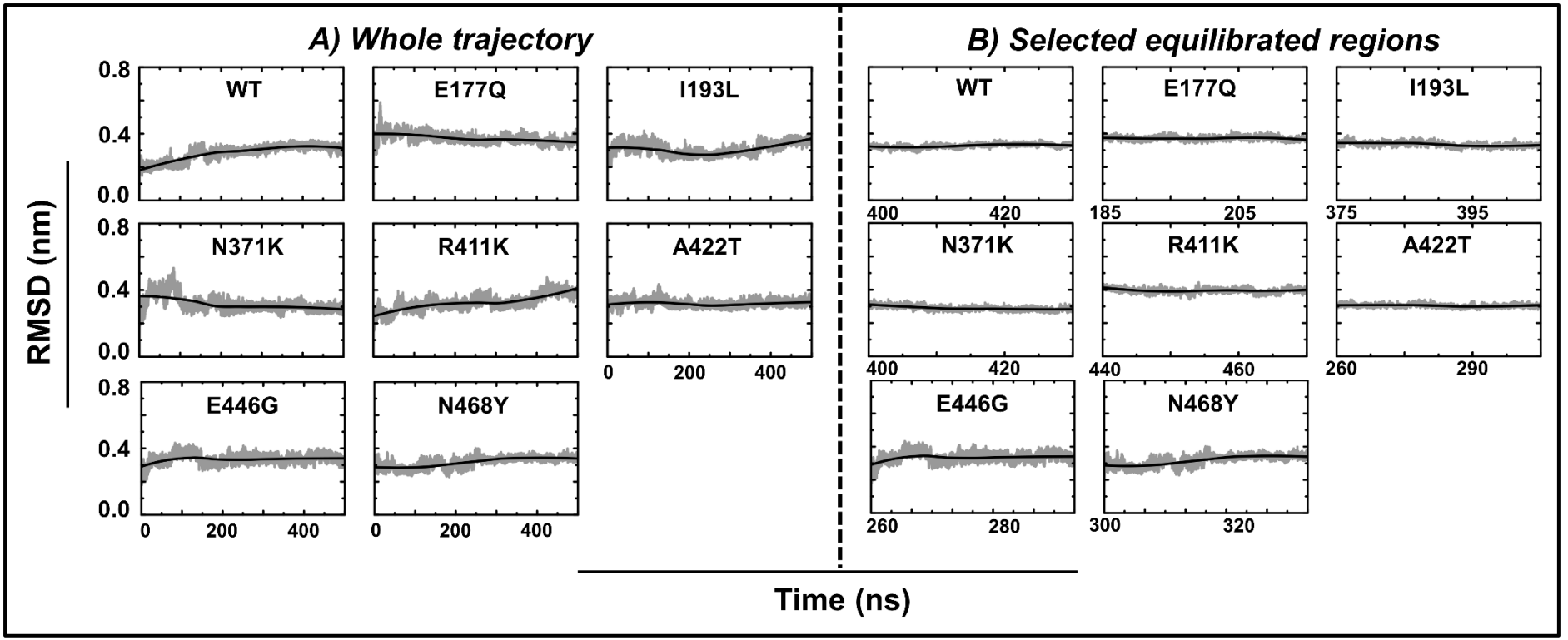



**Figure S3.2:** RMSD line plots of the whole trajectory (500 ns) of the WT and mutant systems of the partial zymogen complex of FP-3 showing conformational changes during the simulation and selected equilibrated regions (30 ns) used for DRN analysis.

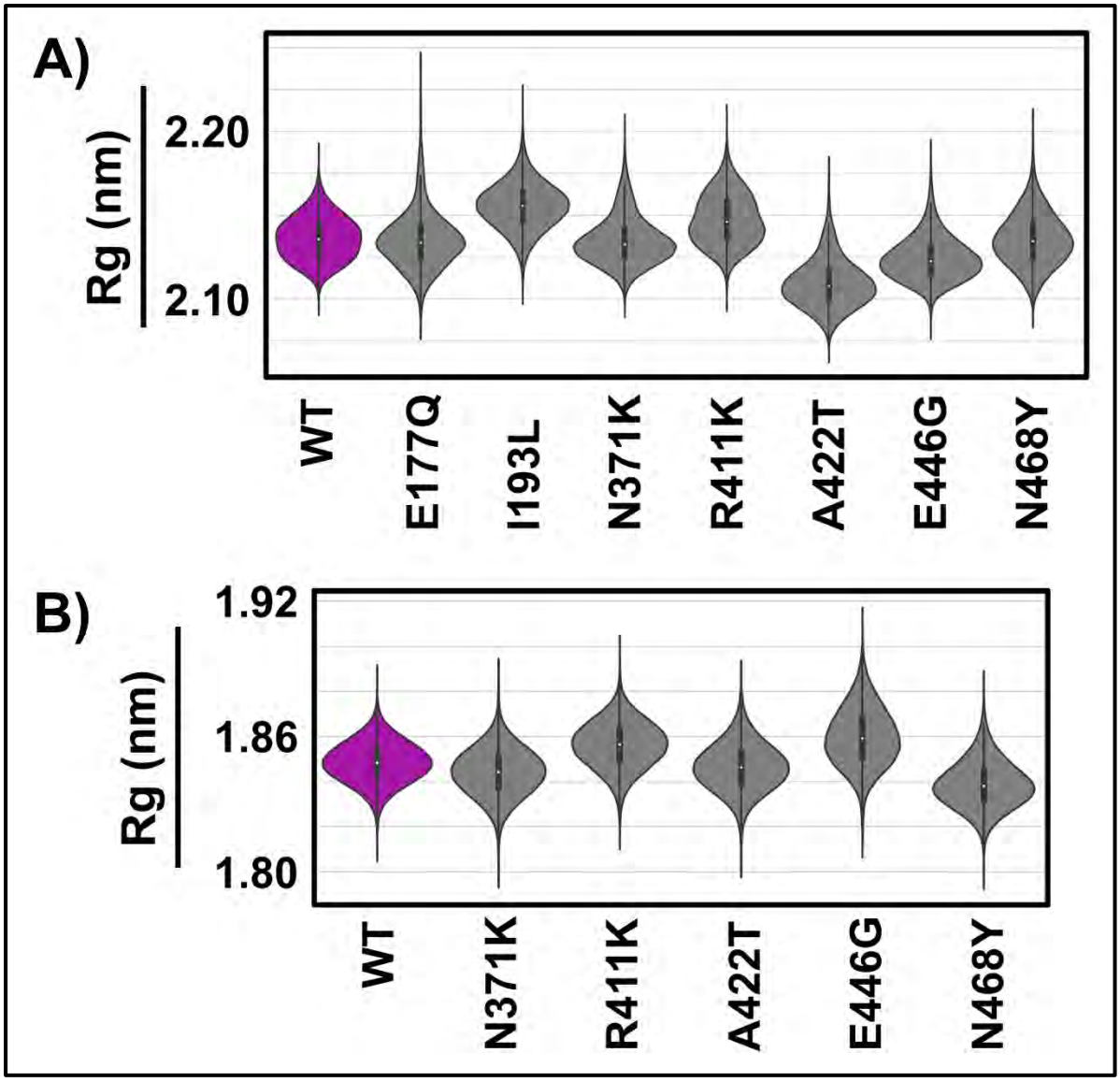



**Figure S3.3**: Rg violin distribution plots of the WT and mutant systems (whole protein) of A) partial zymogen and B) catalytic domain of FP-3.

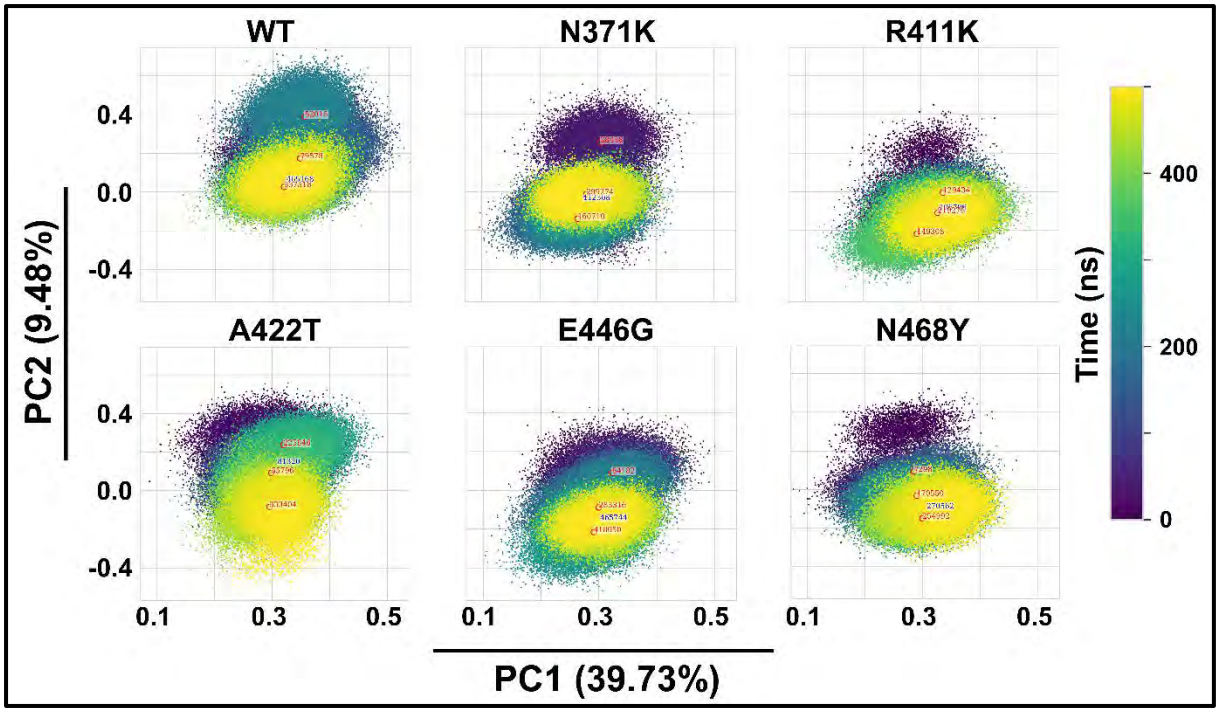



**Figure S3.4:** Comparative essential dynamic analysis of active site pocket residues of the WT and mutant systems of the catalytic domain of FP-3 along PC1 and PC2. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

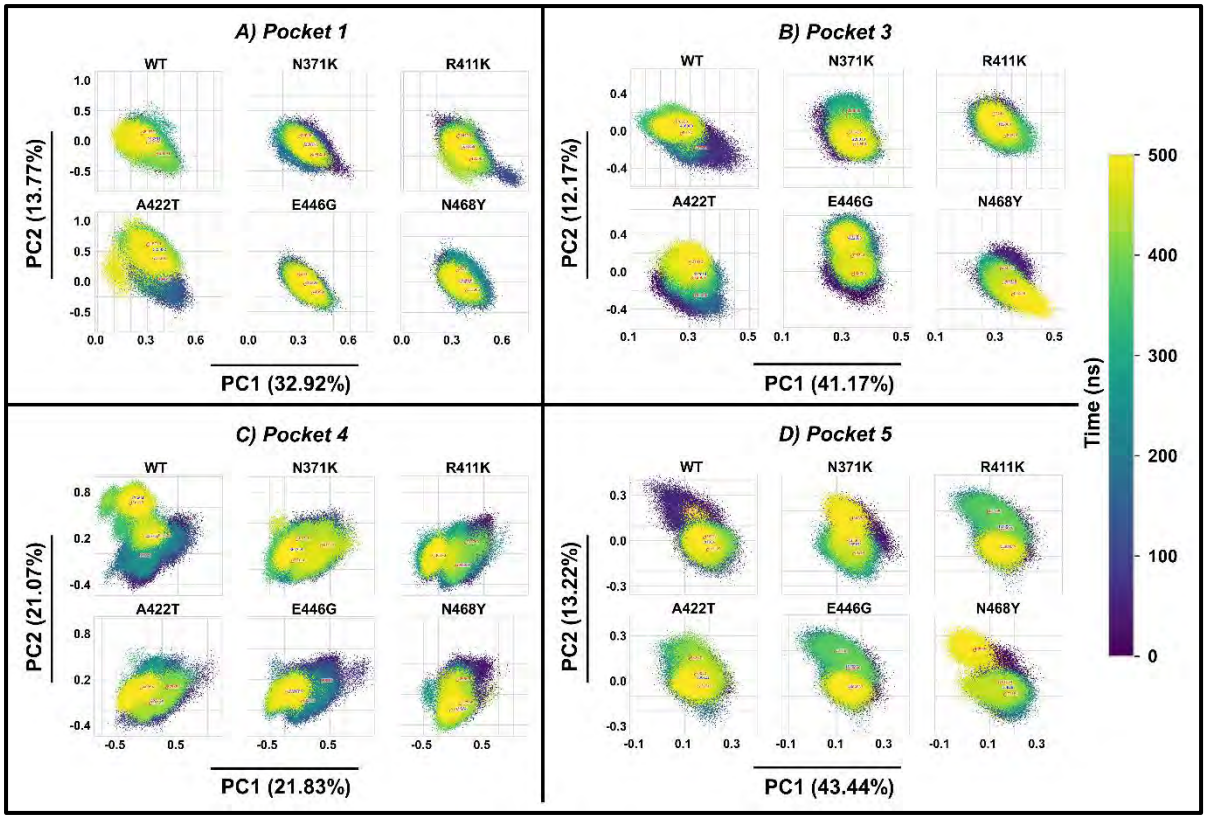



**Figure S3.5:** Comparative essential dynamic analysis of pockets 1, 3, 4, and 5 along PC1 and PC2 of the WT and mutant systems of the catalytic domain of FP-3. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

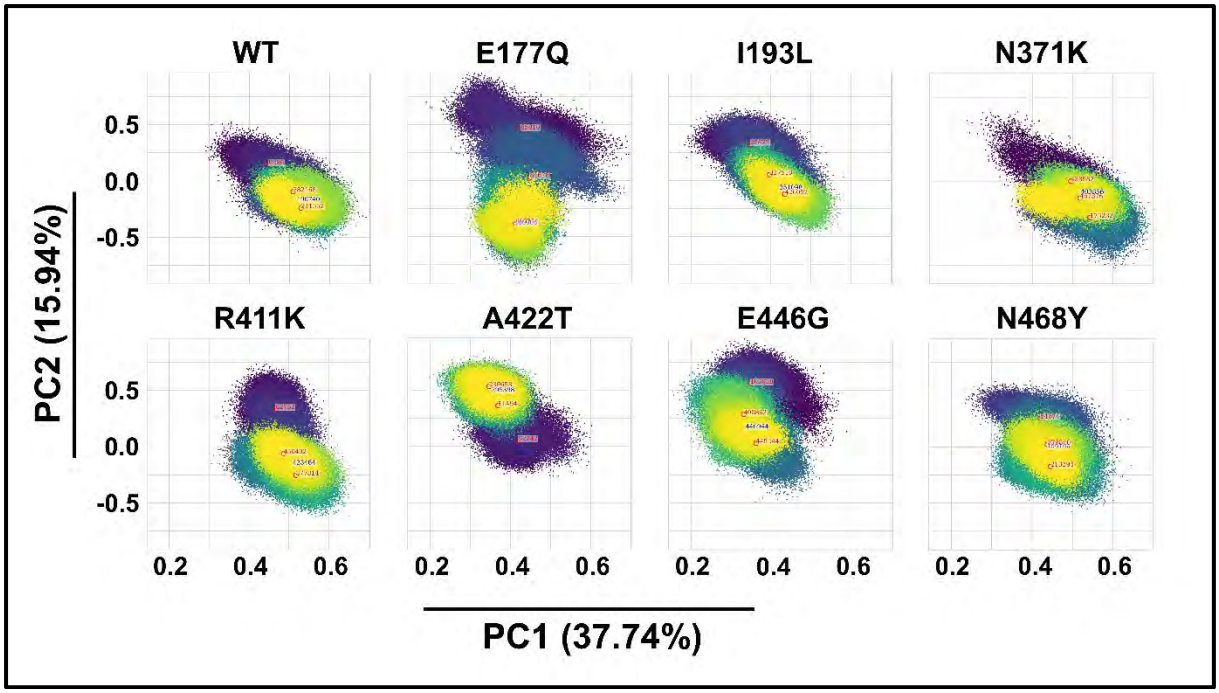



**Figure S3.6:** Comparative essential dynamic analysis of active site pocket residues of the WT and mutant systems of the partial zymogen complex of FP-3 along PC1 and PC2. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

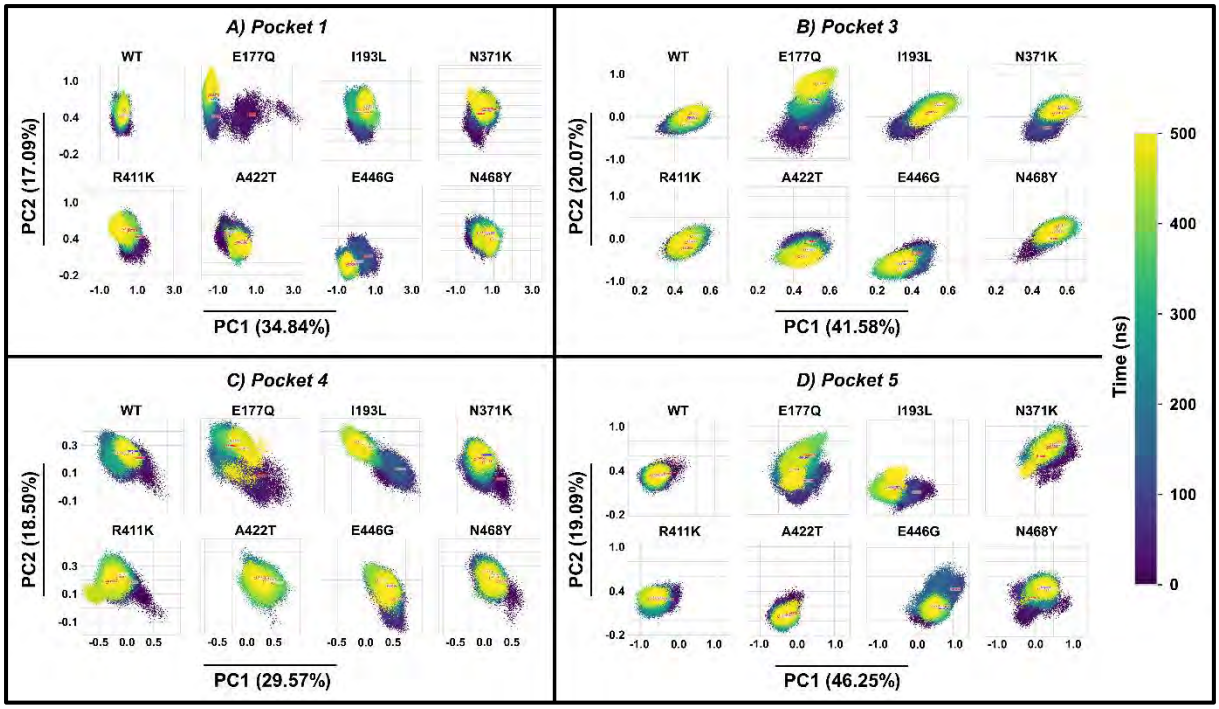



**Figure S3.7:** Comparative essential dynamic analysis of pockets 1, 3, 4, and 5 along PC1 and PC2 of the WT and mutant systems of the partial zymogen complex of FP-3. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

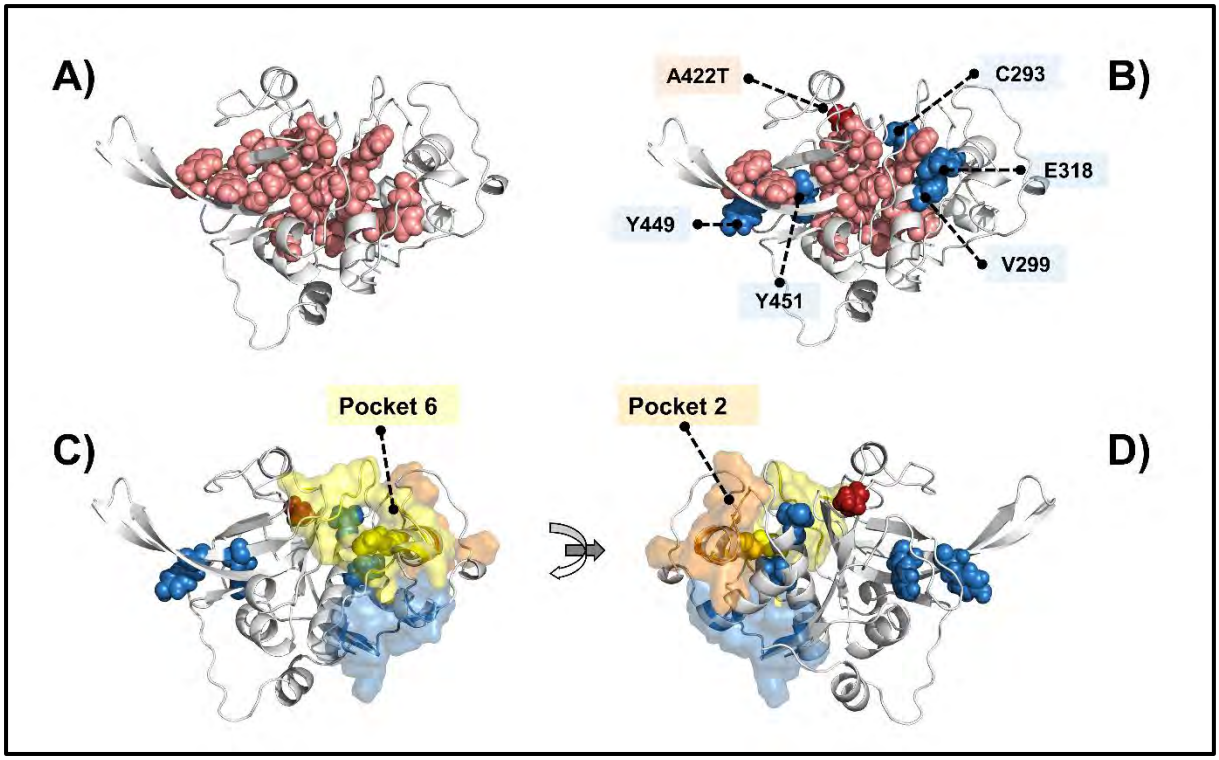



**Figure S3.8:** Cartoon representation of WT and A422T mutant protein of the catalytic domain of FP-3 with its BC hubs. (A) The distribution of the averaged BC hubs of WT (salmon spheres). (B) A422T with its distinctive hubs (sky-blue) and the ones common to the WT hubs (salmon). (C) and (D) show the allosteric communication path formed by the three distinctive hubs originating from A422T and going to the catalytic site, passing through the interface of three pockets.

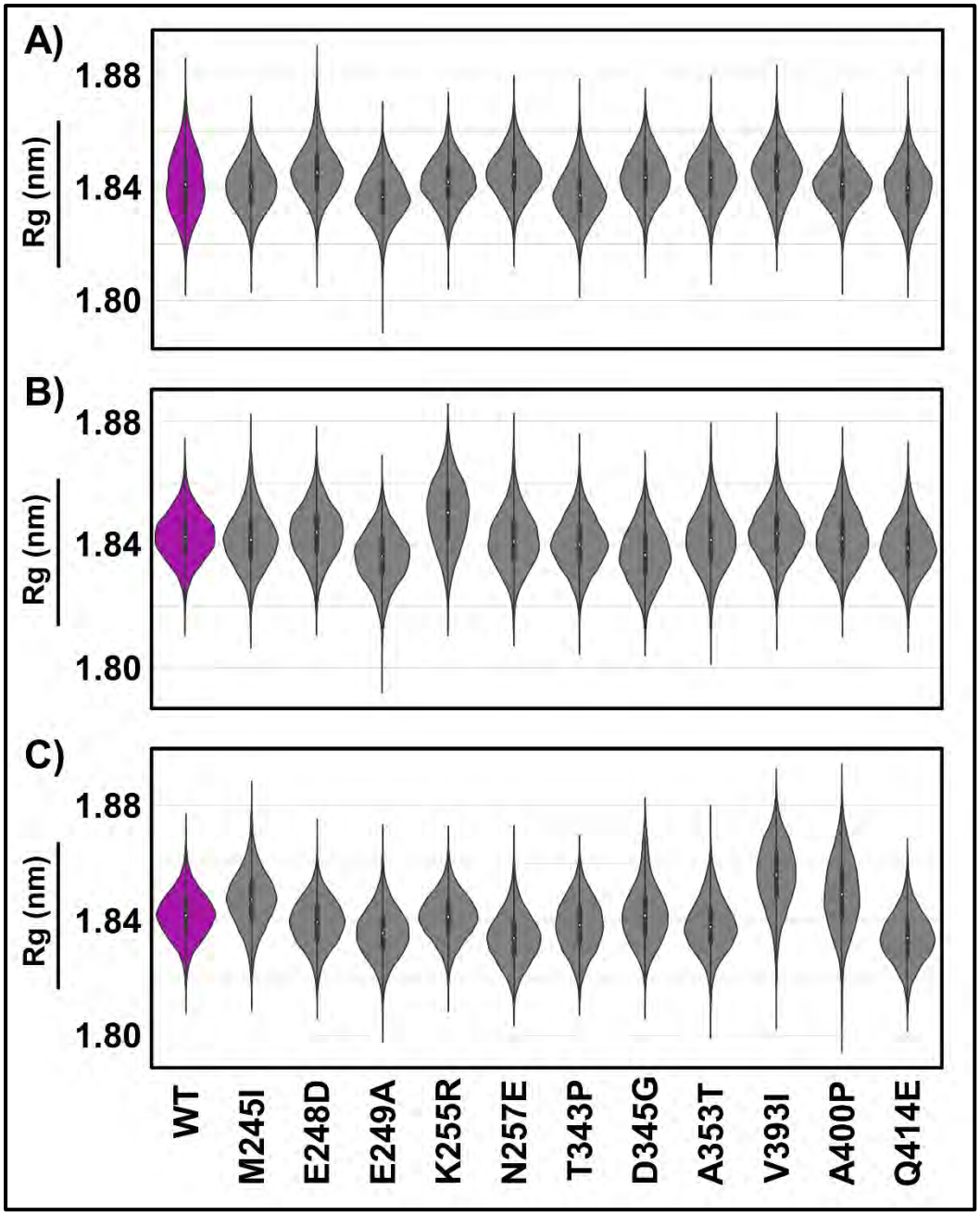



**Figure S4.1:** Violin distribution of the Rg of the WT and mutant-bound systems of FP-2. A) WT and mutant systems bound to DB00312, B) WT and mutant systems bound to DB00951, and C) WT and mutant systems bound to DB14159. WT is coloured magenta, and mutant systems are coloured grey.

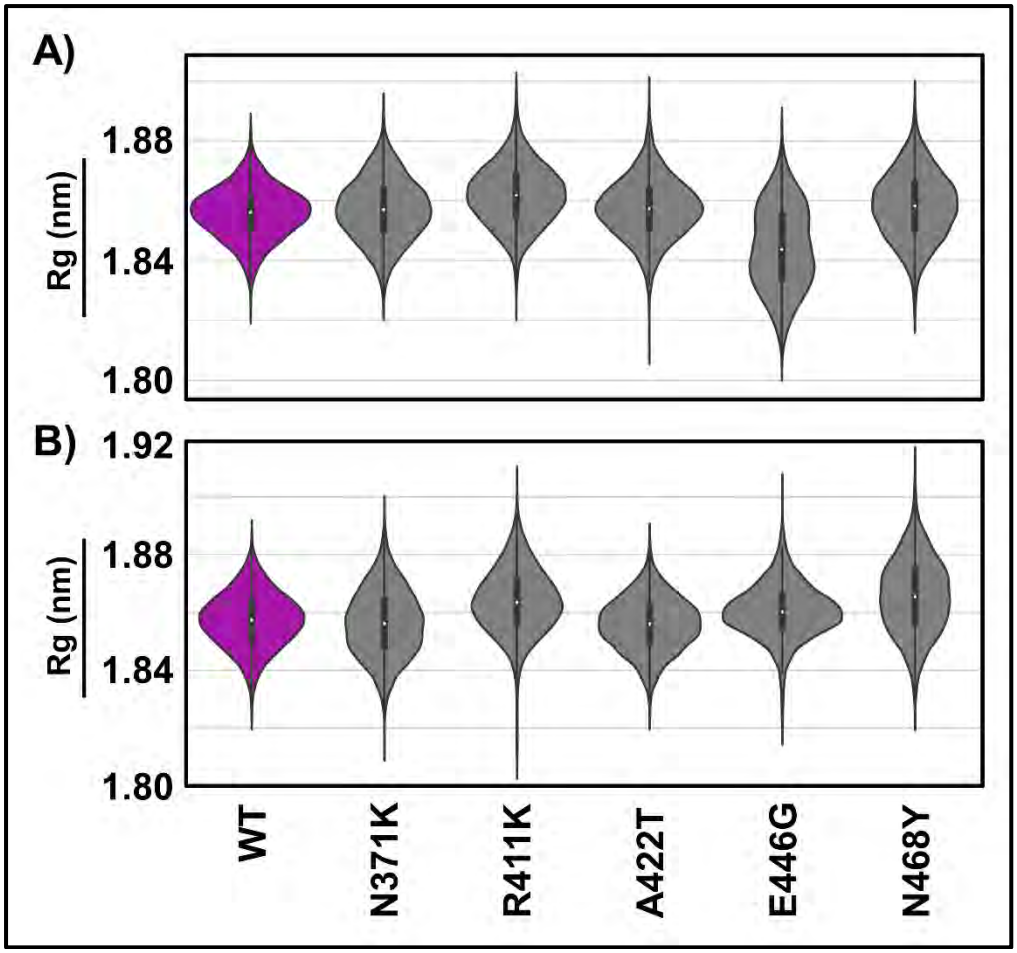



**Figure S4.2:** Violin distribution of the Rg of the WT and mutant-bound systems of FP-3. A) WT and mutant systems bound to DB00853, and B) WT and mutant systems bound to DB00951. WT is coloured magenta, and mutant systems are coloured grey.

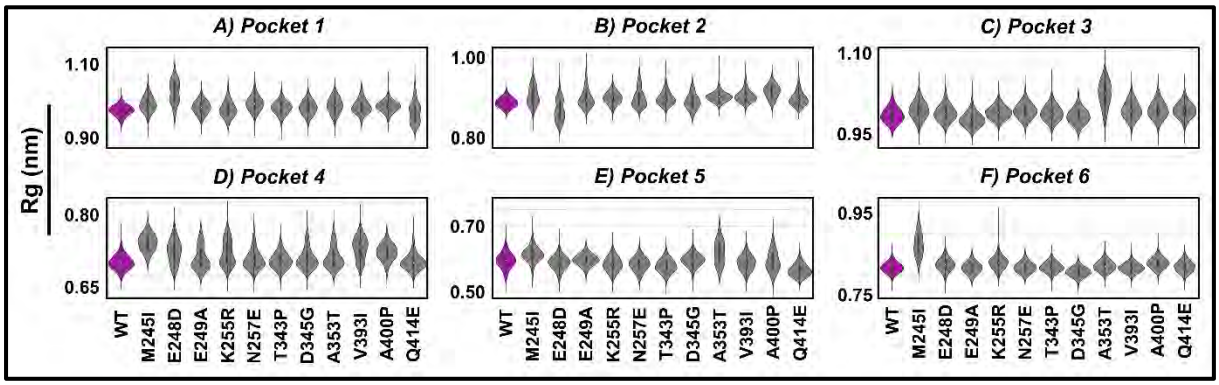



**Figure S4.3:** Violin distribution of the Rg values for Pockets 1 – 6 (A – F) of WT and mutant systems of the catalytic domain of FP-2 bound to DB00853. The purple in the violin plots represents WT.

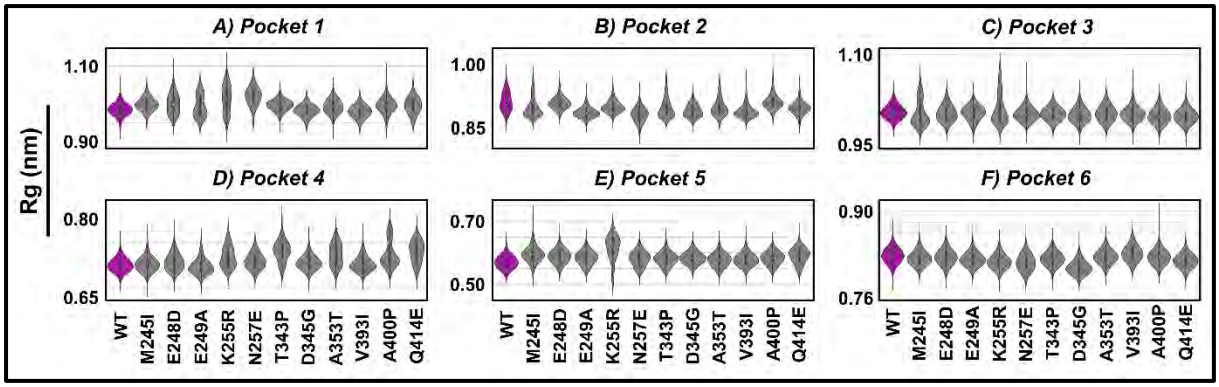



**Figure S4.4:** Violin distribution of the Rg values for Pockets 1 – 6 (A – F) of WT and mutant systems of the catalytic domain of FP-2 bound to DB00951. The purple in the violin plots represents WT.

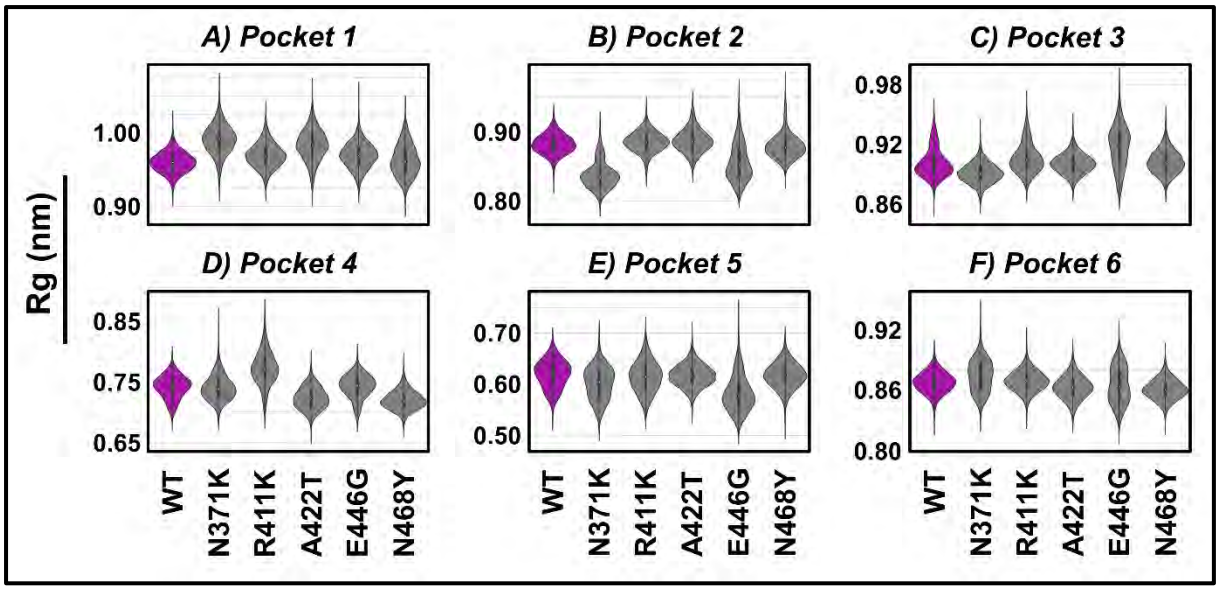



**Figure S4.5:** Violin distribution of the Rg values for Pockets 1 – 6 (A – F) of WT and mutant systems of the catalytic domain of FP-3 bound to DB00853. The purple in the violin plots represents WT.

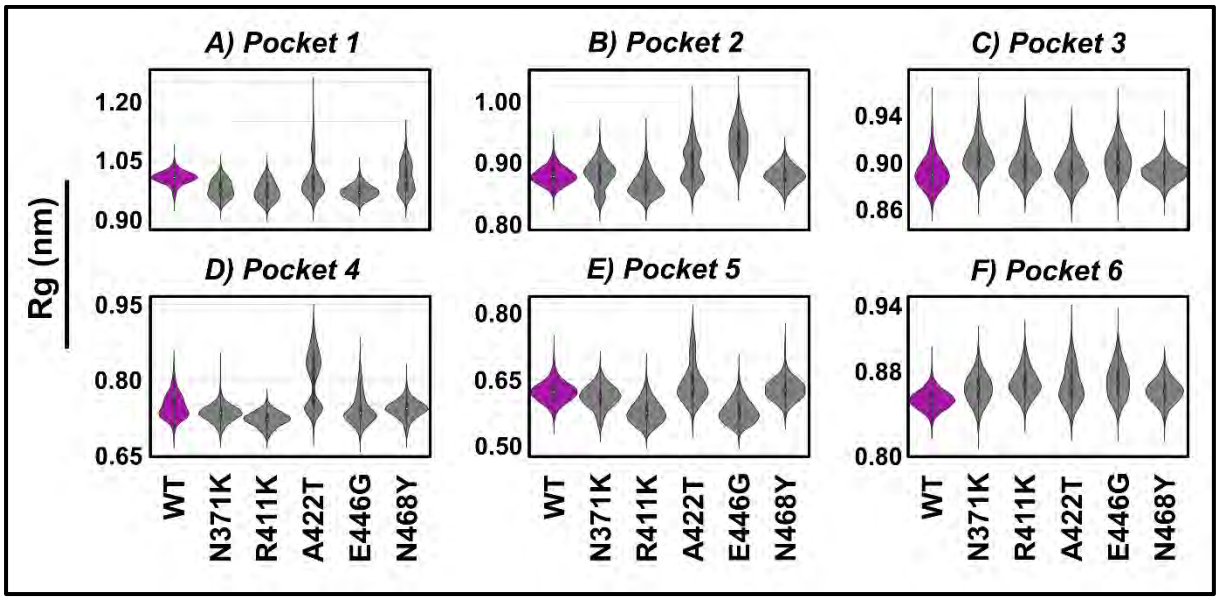



**Figure S4.6:** Violin distribution of the Rg values for Pockets 1 – 6 (A – F) of WT and mutant systems of the catalytic domain of FP-3 bound to DB00951. The purple in the violin plots represents WT.

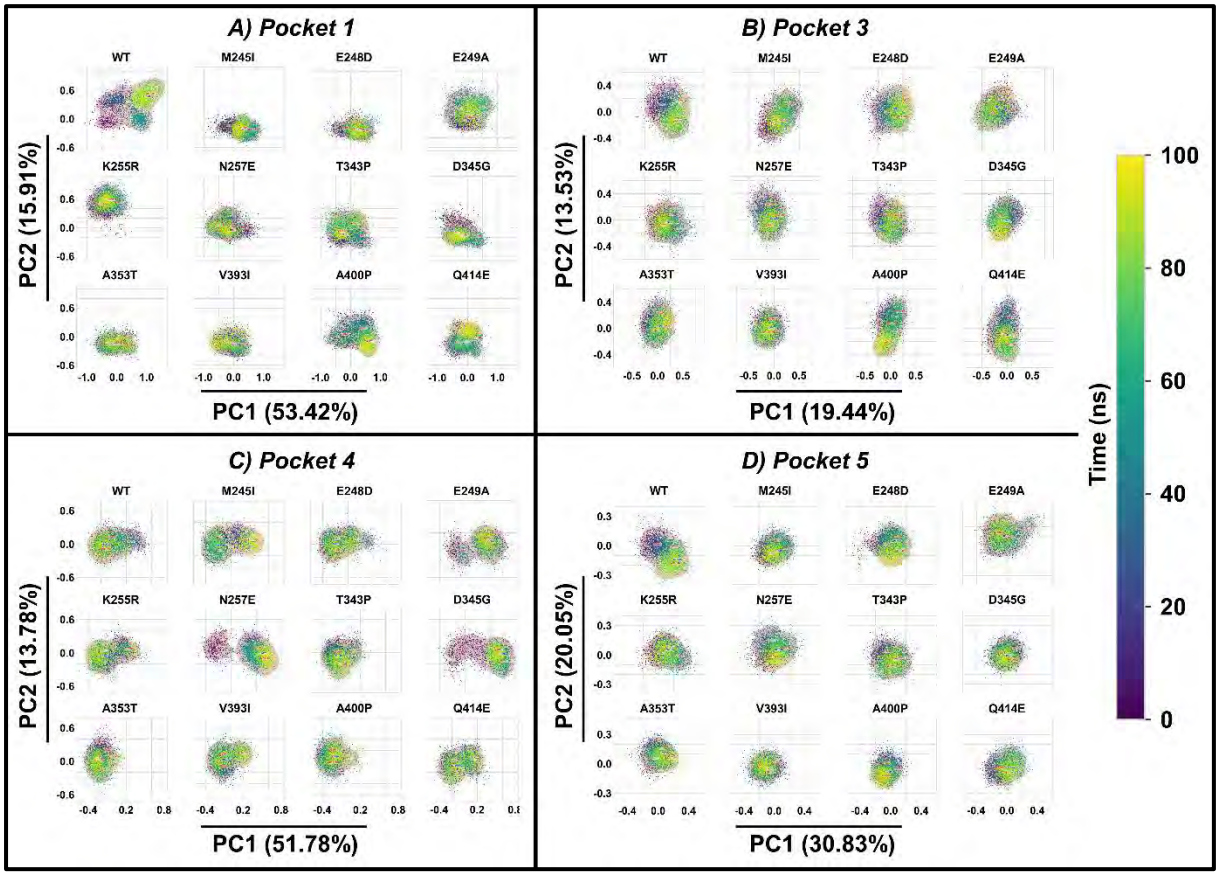



**Figure S4.7:** Comparative essential dynamic analysis of pockets 1, 3, 4, and 5 along PC1 and PC2 of the WT and mutant systems of the catalytic domain of FP-2 bound to DB14159. The simulation time (ns) is represented by a color code (black = start, yellow = end). In each PCA plot, the time stamp with the most stable structure is indicated in blue.

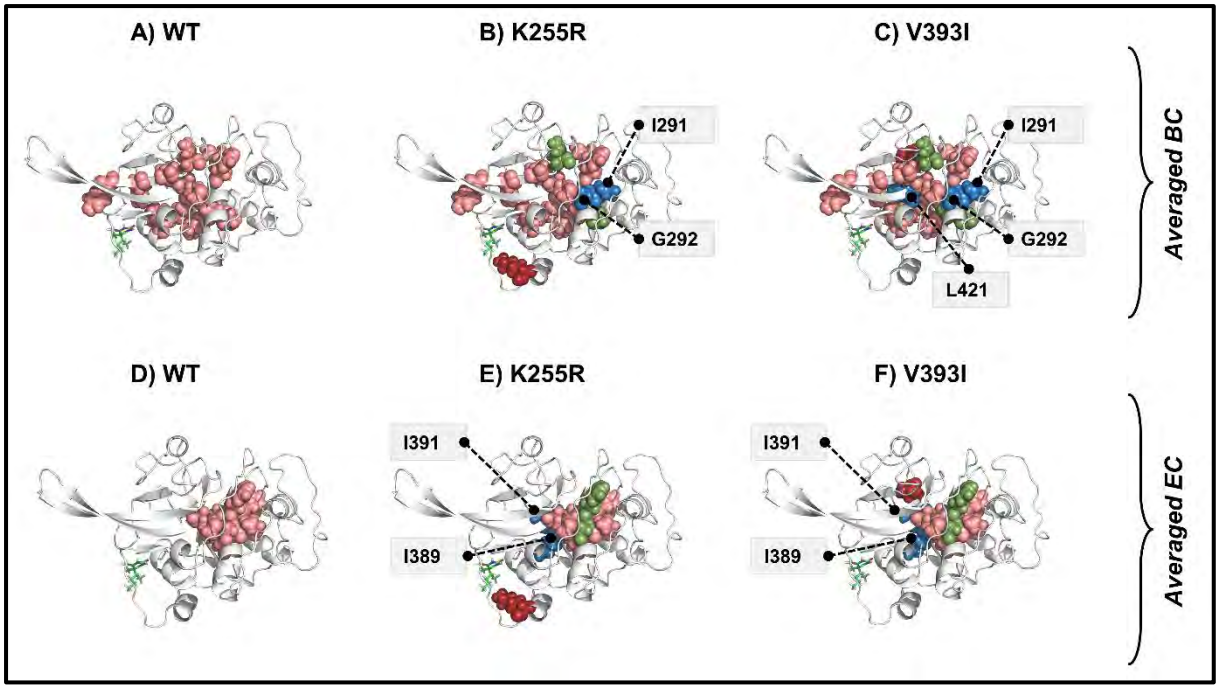



**Figure S4.8:** A cartoon illustration of the distribution of the averaged *BC* hubs unique to each mutant-DB00312 bound protein compared with the WT-DB00312 bound protein (skyblue spheres) in A) WT, B) K255R, and C) V393I. D) WT, E) K255R, and F) V393I show the distribution of the averaged *EC* hubs unique to each mutant protein compared with the WT. Hubs unique to the WT systems are represented in smudge spheres.

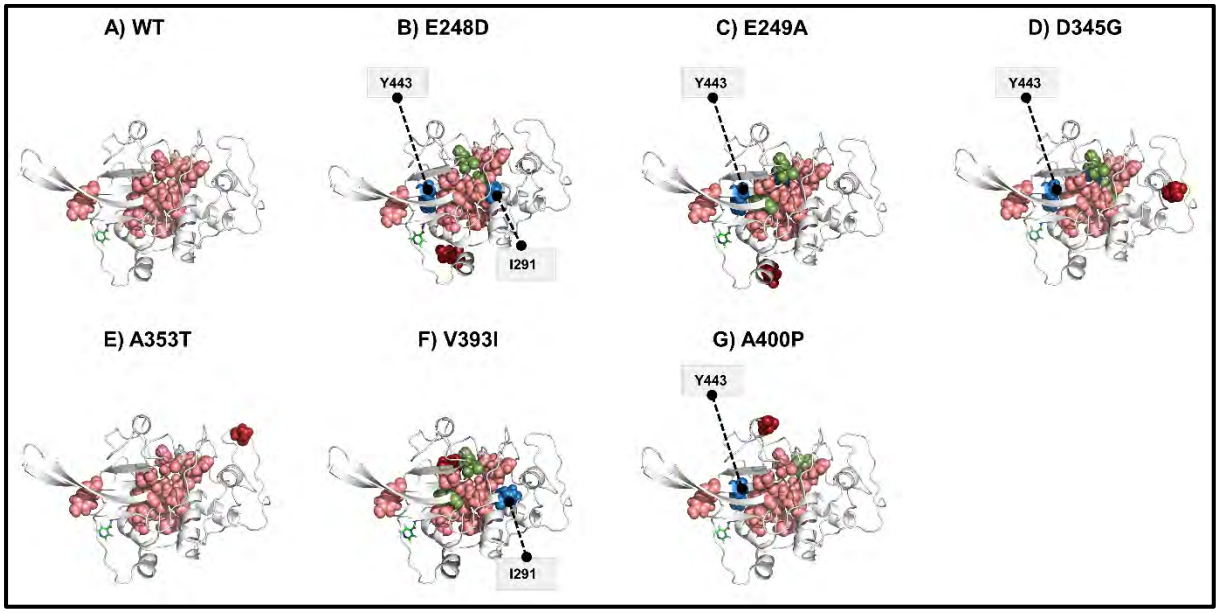



**Figure S4.9:** A cartoon illustration of the distribution of the averaged *BC* hubs unique to each mutant-DB00951 bound protein compared with the WT-DB00951 bound protein (skyblue spheres) in A) WT, B) E248D, C) E249A, D) D345G, E) A353T, F) A400P, and G) V393I. Hubs unique to the WT systems are represented in smudge spheres.

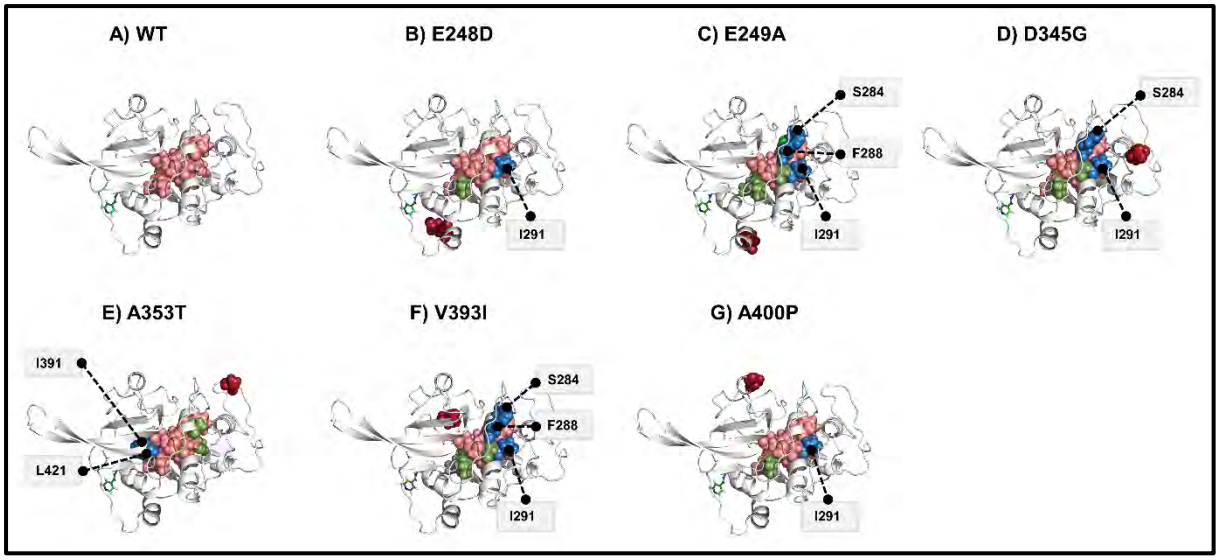



**Figure S4.10:** A cartoon illustration of the distribution of the *averaged EC hubs* unique to each mutant-DB00951 bound protein compared with the WT-DB00951 bound protein (skyblue spheres) in A) WT, B) E248D, C) E249A, D) D345G, E) A353T, F) A400P, and G) V393I. Hubs unique to the WT systems are represented in smudge spheres.

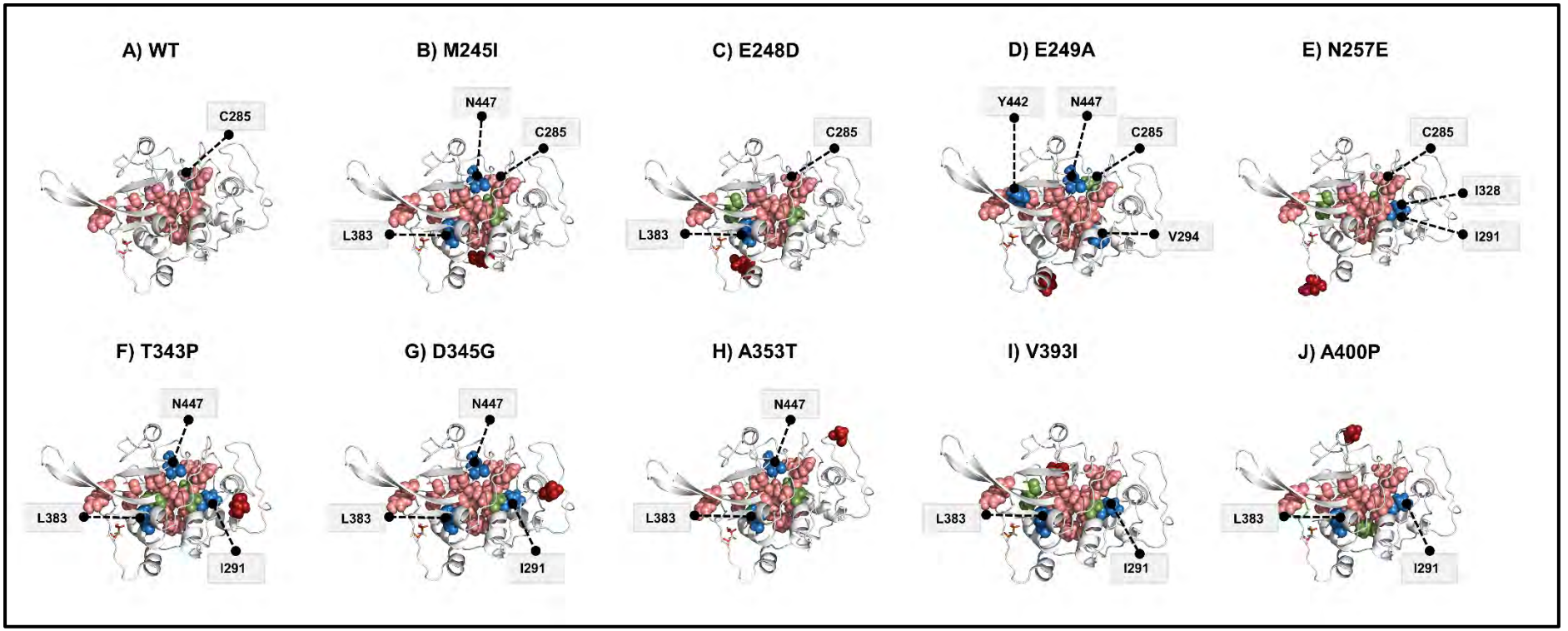



**Figure S4.11:** A cartoon illustration of the distribution of the *averaged BC hubs* unique to each mutant-DB14159 bound protein compared with the WT-DB14159 bound protein (skyblue spheres) in A) WT, B) M245I, C) E248D, D) E249A, E) N257E, F) T343P, G) D345G, H) A353T, I) V393I, and J) A400P. Hubs unique to the WT systems are represented in smudge spheres.

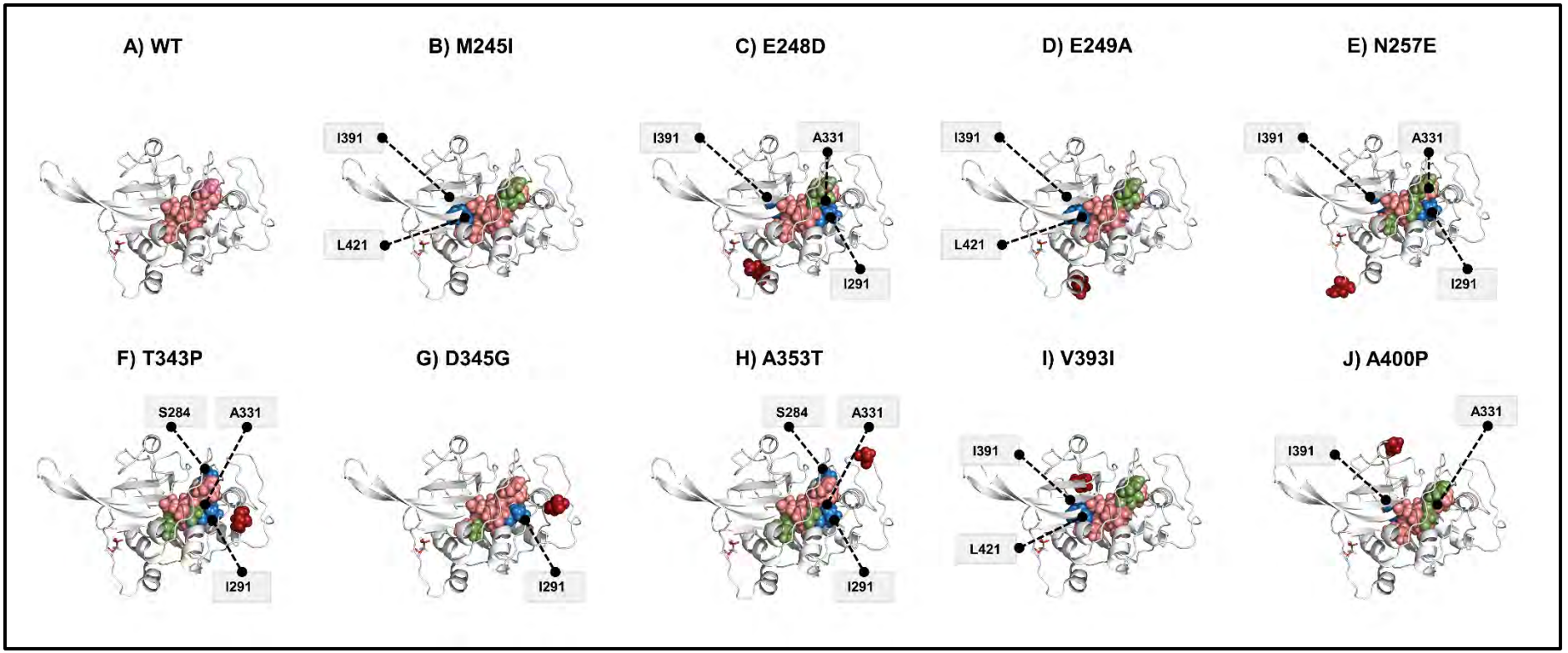



**Figure S4.12:** A cartoon illustration of the distribution of the *averaged EC hubs* unique to each mutant-DB14159 bound protein compared with the WT-DB14159 bound protein (skyblue spheres) in A) WT, B) M245I, C) E248D, D) E249A, E) N257E, F) T343P, G) D345G, H) A353T, I) V393I, and J) A400P. Hubs unique to the WT systems are represented in smudge spheres.

# References


1. Chapter 55. Drug Resistance. In *Disease Control Priorities in Developing Countries (2nd Edition)*; Jamison, D.T., Breman, J.G., Measham, A.R., Alleyne, G., Claeson, M., Evans, D.B., Jha, P., Mills, A., Musgrove, P., Eds.; World Bank Publications, 2006; pp. 1031–1052.

2. Ayukekbong, J.A.; Ntemgwa, M.; Atabe, A.N. The threat of antimicrobial resistance in developing countries: causes and control strategies. *Antimicrob. Resist. Infect. Control* **2017**, *6*, 47, doi:10.1186/s13756-017-0208-x.

3. Munita, J.M.; Arias, C.A. Mechanisms of Antibiotic Resistance. *Microbiol. Spectr.* **2016**, *4*, doi:10.1128/MICROBIOLSPEC.VMBF-0016-2015.

4. Tanner, M.; Greenwood, B.; Whitty, C.J.M.M.; Ansah, E.K.; Price, R.N.; Dondorp, A.M.; von Seidlein, L.; Baird, J.K.; Beeson, J.G.; Fowkes, F.J.I.I.; et al. Malaria eradication and elimination: views on how to translate a vision into reality. *BMC Med.* **2015**, *13*, 167, doi:10.1186/s12916-015-0384-6.

5. Sato, S. Plasmodium—a brief introduction to the parasites causing human malaria and their basic biology. *J. Physiol. Anthropol.* **2021**, *40*, doi:10.1186/S40101-020-00251-9.

6. Mishra, M.; Singh, V.; Singh, S. Structural Insights Into Key Plasmodium Proteases as Therapeutic Drug Targets. *Front. Microbiol.* **2019**, *10*, doi:10.3389/fmicb.2019.00394.

7. WHO Global *World malaria report 2019*; 2019; ISBN 9789241565721.

8. Elimination, T. malERA R.C.P. on T. for M. malERA: An updated research agenda for diagnostics, drugs, vaccines, and vector control in malaria elimination and eradication. *PLOS Med.* **2017**, *14*, e1002455, doi:10.1371/JOURNAL.PMED.1002455.

9. Cunha-Rodrigues, M.; Prudêncio, M.; Mota, M.M.; Haas, W. Antimalarial drugs - host targets (re)visited. *Biotechnol. J.* **2006**, *1*, 321–332, doi:10.1002/BIOT.200500038.

10. Winter, R.W.; Kelly, J.X.; Smilkstein, M.J.; Dodean, R.; Hinrichs, D.; Riscoe, M.K. Antimalarial Quinolones: Synthesis, potency, and mechanistic studies. *Exp. Parasitol.* **2008**, *118*, 487, doi:10.1016/J.EXPPARA.2007.10.016.

11. Severini, C.; Menegon, M. Resistance to antimalarial drugs: An endless world war against Plasmodium that we risk losing. *J. Glob. Antimicrob. Resist.* **2015**, *3*, 58–63,

doi:10.1016/j.jgar.2015.02.002.

12. Wellems, T.E. Plasmodium chloroquine resistance and the search for a replacement antimalarial drug. *Science (80-. ).* **2002**, *298*, 124–126, doi:10.1126/SCIENCE.1078167/ASSET/0C8EA1EA-D0ED-4BC7-BCE2-492568E309CE/ASSETS/GRAPHIC/SE3929012001.JPEG.

13. Klein, E.Y. Antimalarial drug resistance: a review of the biology and strategies to delay emergence and spread. *Int. J. Antimicrob. Agents* **2013**, *41*, 311–317, doi:10.1016/j.ijantimicag.2012.12.007.

14. Haldar, K.; Bhattacharjee, S.; Safeukui, I. Drug resistance in Plasmodium. *Nat. Rev. Microbiol.* **2018**, *16*, 156–170, doi:10.1038/nrmicro.2017.161.

15. Sidhu, A.B.S.; Valderramos, S.G.; Fidock, D.A. pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in Plasmodium falciparum. *Mol. Microbiol.* **2005**, *57*, 913–926, doi:10.1111/j.1365-2958.2005.04729.x.

16. Dondorp, A.M.; Yeung, S.; White, L.; Nguon, C.; Day, N.P.J.; Socheat, D.; Von Seidlein, L. Artemisinin resistance: Current status and scenarios for containment. *Nat. Rev. Microbiol.* **2010**, *8*, 272–280, doi:10.1038/nrmicro2331.

17. Greenwood, B.M.; Fidock, D.A.; Kyle, D.E.; Kappe, S.H.I.; Alonso, P.L.; Collins, F.H.; Duffy, P.E. Malaria: progress, perils, and prospects for eradication. *J. Clin. Invest.* **2008**, *118*, 1266, doi:10.1172/JCI33996.

18. Lindner, S.E.; Miller, J.L.; Kappe, S.H.I. Malaria Parasite Pre-erythrocytic Infection: Preparation Meets Opportunity. *Cell. Microbiol.* **2012**, *14*, 316, doi:10.1111/J.1462-5822.2011.01734.X.

19. Aly, A.S.I.; Vaughan, A.M.; Kappe, S.H.I. Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host. *Annu. Rev. Microbiol.* **2009**, doi:10.1146/annurev.micro.091208.073403.

20. Gerald, N.; Mahajan, B.; Kumar, S. Mitosis in the human malaria parasite plasmodium falciparum. *Eukaryot. Cell* **2011**, doi:10.1128/EC.00314-10.

21. Soulard, V.; Bosson-Vanga, H.; Lorthiois, A.; Roucher, C.; Franetich, J.F.; Zanghi, G.; Bordessoulles, M.; Tefit, M.; Thellier, M.; Morosan, S.; et al. Plasmodium falciparum

full life cycle and Plasmodium ovale liver stages in humanized mice. *Nat. Commun.* **2015**, doi:10.1038/ncomms8690.

22. Kappe, S.H.I.; Vaughan, A.M.; Boddey, J.A.; Cowman, A.F. That was then but this is now: Malaria research in the time of an eradication agenda. *Science (80-. ).* **2010**, *328*, 862–866, doi:10.1126/SCIENCE.1184785/ASSET/531A53F9-8EA5-46BB-BF70-384EC0B98943/ASSETS/GRAPHIC/328_862_F3.JPEG.

23. Moore, L.R.; Fujioka, H.; Williams, P.S.; Chalmers, J.J.; Grimberg, B.; Zimmerman, P.A.; Zborowski, M. Hemoglobin degradation in malaria-infected erythrocytes determined from live cell magnetophoresis. *FASEB J.* **2006**, *20*, 747, doi:10.1096/FJ.05-5122FJE.

24. Alam, A. Plasmodium proteases as therapeutic targets against malaria. *Proteases Hum. Dis.* **2017**, 69–90, doi:10.1007/978-981-10-3162-5_4/COVER.

25. Molina-Franky, J.; Cuy-Chaparro, L.; Camargo, A.; Reyes, C.; Gómez, M.; Salamanca, D.R.; Patarroyo, M.A.; Patarroyo, M.E. Plasmodium falciparum pre-erythrocytic stage vaccine development. *Malar. J.* **2020**, *19*, doi:10.1186/S12936-020-3141-Z.

26. Fidock, D.A.; Eastman, R.T.; Ward, S.A.; Meshnick, S.R. Recent highlights in antimalarial drug resistance and chemotherapy research. *Trends Parasitol.* **2008**, *24*, 537, doi:10.1016/J.PT.2008.09.005.

27. Machin, J.M.; Kantsadi, A.L.; Vakonakis, I. The complex of Plasmodium falciparum falcipain-2 protease with an (E)-chalcone-based inhibitor highlights a novel, small, molecule-binding site. *Malar. J.* **2019**, *18*, 388, doi:10.1186/s12936-019-3043-0.

28. Francis, S.E.; Sullivan, D.J.; Goldberg, D.E. Hemoglobin metabolism in the malaria parasite Plasmodium falciparum. *Annu. Rev. Microbiol.* **1997**, *51*, 97–123, doi:10.1146/ANNUREV.MICRO.51.1.97.

29. Olliaro, P. Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacol. Ther.* **2001**, *89*, 207–219, doi:10.1016/S0163-7258(00)00115-7.

30. Ismail, H.M.; Barton, V.; Phanchana, M.; Charoensutthivarakul, S.; Wong, M.H.L.; Hemingway, J.; Biagini, G.A.; O'Neill, P.M.; Ward, S.A. Artemisinin activity-based probes identify multiple molecular targets within the asexual stage of the malaria parasites Plasmodium falciparum 3D7. *Proc. Natl. Acad. Sci. U. S. A.* **2016**,

doi:10.1073/pnas.1600459113.

31. Heller, L.E.; Roepe, P.D. Artemisinin-Based Antimalarial Drug Therapy: Molecular Pharmacology and Evolving Resistance. *Trop. Med. Infect. Dis.* **2019**, *4*, 89, doi:10.3390/tropicalmed4020089.

32. Razzaq, A.; Shamsi, S.; Ali, A.; Ali, Q.; Sajjad, M.; Malik, A.; Ashraf, M. Microbial Proteases Applications. *Front. Bioeng. Biotechnol.* **2019**, *7*, 110, doi:10.3389/FBIOE.2019.00110.

33. Antalis, T.M.; Buzza, M.S. Extracellular: Plasma Membrane Proteases – Serine Proteases. *Encycl. Cell Biol.* **2016**, *1*, 650, doi:10.1016/B978-0-12-394447-4.10076-8.

34. McKerrow, J.H.; Caffrey, C.; Kelly, B.; Loke, P.; Sajid, M. PROTEASES IN PARASITIC DISEASES. *https://doi.org/10.1146/annurev.pathol.1.110304.100151* **2006**, *1*, 497–536, doi:10.1146/ANNUREV.PATHOL.1.110304.100151.

35. Alam, A. Serine proteases of malaria parasite plasmodium falciparum: Potential as antimalarial drug targets. *Interdiscip. Perspect. Infect. Dis.* **2014**, *2014*, doi:10.1155/2014/453186.

36. Drag, M.; Salvesen, G.S. Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov. 2010 99* **2010**, *9*, 690–701, doi:10.1038/nrd3053.

37. Melnikova, I. The anticoagulants market. *Nat. Rev. Drug Discov. 2009 85* **2009**, *8*, 353–354, doi:10.1038/nrd2851.

38. Melnikova, I. Hepatitis C therapies. *Nat. Rev. Drug Discov.* **2008**, *7*, 799–800, doi:10.1038/NRD2661.

39. Flexner, C.; Bate, G.; Kirkpatrick, P. Tipranavir. *Nat. Rev. Drug Discov.* **2005**, *4*, 955–956, doi:10.1038/NRD1907.

40. López-Otín, C.; Bond, J.S. Proteases: Multifunctional Enzymes in Life and Disease. *J. Biol. Chem.* **2008**, *283*, 30433, doi:10.1074/JBC.R800035200.

41. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632, doi:10.1093/NAR/GKX1134.

42. Deu, E. Proteases as antimalarial targets: strategies for genetic, chemical, and therapeutic validation. *FEBS J.* **2017**, *284*, 2604–2628, doi:10.1111/febs.14130.

43. Goldberg, D.E. Hemoglobin degradation. *Curr. Top. Microbiol. Immunol.* **2005**, *295*, 275–291, doi:10.1007/3-540-29088-5_11.

44. Ponsuwanna, P.; Kochakarn, T.; Bunditvorapoom, D.; Kümpornsin, K.; Otto, T.D.; Ridenour, C.; Chotivanich, K.; Wilairat, P.; White, N.J.; Miotto, O.; et al. Comparative genome-wide analysis and evolutionary history of haemoglobin-processing and haem detoxification enzymes in malarial parasites. *Malar. J.* **2016**, *15*, 1–14, doi:10.1186/S12936-016-1097-9/FIGURES/8.

45. Murata, C.E.; Goldberg, D.E. Plasmodium falciparum falcilysin: an unprocessed food vacuole enzyme. *Mol. Biochem. Parasitol.* **2003**, *129*, 123–126, doi:10.1016/S0166-6851(03)00098-7.

46. Klemba, M.; Gluzman, I.; Goldberg, D.E. A Plasmodium falciparum dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation. *J. Biol. Chem.* **2004**, *279*, 43000–43007, doi:10.1074/JBC.M408123200.

47. Istvan, E.S.; Dharia, N. V.; Bopp, S.E.; Gluzman, I.; Winzeler, E.A.; Goldberg, D.E. Validation of isoleucine utilization targets in Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1627–32, doi:10.1073/pnas.1011560108.

48. Liu, J.; Istvan, E.S.; Gluzman, I.Y.; Gross, J.; Goldberg, D.E. Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc. Natl. Acad. Sci.* **2006**, *103*, 8840–8845, doi:10.1073/pnas.0601876103.

49. Qidwai, T. Hemoglobin Degrading Proteases of Plasmodium falciparum as Antimalarial Drug Targets. *Curr. Drug Targets* **2015**, *16*, 1133–1141, doi:10.2174/1389450116666150304104123.

50. Rosenthal, P.J. Cysteine proteases of malaria parasites. *Int. J. Parasitol.* **2004**, *34*, 1489–1499, doi:10.1016/J.IJPARA.2004.10.003.

51. Bah, S.; Paulsen, B.S.; Diallo, D.; Johansen, H.T. Characterization of cysteine proteases in Malian medicinal plants. *J. Ethnopharmacol.* **2006**, *107*, 189–198, doi:10.1016/J.JEP.2006.03.008.

52. Barrett, A.J.; Rawlings, N.D. Evolutionary lines of cysteine peptidases. *Biol. Chem.* **2001**, *382*, 727–733, doi:10.1515/BC.2001.088.

53. Roy, K.K. Targeting the active sites of malarial proteases for antimalarial drug discovery: approaches, progress and challenges. *Int. J. Antimicrob. Agents* **2017**, *50*, 287–302, doi:10.1016/J.IJANTIMICAG.2017.04.006.

54. Mottram, J.C.; Helms, M.J.; Coombs, G.H.; Sajid, M. Clan CD cysteine peptidases of parasitic protozoa. *Trends Parasitol.* **2003**, *19*, 182–187, doi:10.1016/S1471-4922(03)00038-2.

55. Shenai, B.R.; Semenov, A. V.; Rosenthal, P.J. Stage-specific antimalarial activity of cysteine protease inhibitors. *Biol. Chem.* **2002**, *383*, 843–847, doi:10.1515/BC.2002.089.

56. Kerr, I.D.; Lee, J.H.; Pandey, K.C.; Harrison, A.; Sajid, M.; Rosenthal, P.J.; Brinen, L.S. Structures of falcipain-2 and falcipain-3 bound to small molecule inhibitors: Implications for substrate specificity. *J. Med. Chem.* **2009**, *52*, 852–857, doi:10.1021/jm8013663.

57. Rosenthal, P.J.; McKerrow, J.H.; Aikawa, M.; Nagasawa, H.; Leech, J.H. A malarial cysteine proteinase is necessary for hemoglobin degradation by Plasmodium falciparum. *J. Clin. Invest.* **1988**, *82*, 1560–1566, doi:10.1172/JCI113766.

58. Rosenthal, P.; Sijwali, P.; Singh, A.; Shenai, B. Cysteine proteases of malaria parasites: targets for chemotherapy. *Curr. Pharm. Des.* **2002**, *8*, 1659–1672, doi:10.2174/1381612023394197.

59. McKerrow, J.H.; Sun, E.; Rosenthal, P.J.; Bouvier, J. THE PROTEASES AND PATHOGENICITY OF PARASITIC PROTOZOA. *https://doi.org/10.1146/annurev.mi.47.100193.004133* **2003**, *47*, 821–853, doi:10.1146/ANNUREV.MI.47.100193.004133.

60. Rosenthal, P.J. Plasmodium falciparum: Effects of Proteinase Inhibitors on Globin Hydrolysis by Cultured Malaria Parasites. *Exp. Parasitol.* **1995**, *80*, 272–281, doi:10.1006/EXPR.1995.1033.

61. Debrabant, A.; Delplace, P. Leupeptin alters the proteolytic processing of P126, the major parasitophorous vacuole antigen of Plasmodium falciparum. *Mol. Biochem.*

*Parasitol.* **1989**, *33*, 151–158, doi:10.1016/0166-6851(89)90029-7.

62. Hadley, T.; Aikawa, M.; Miller, L.H. Plasmodium knowlesi: Studies on invasion of rhesus erythrocytes by merozoites in the presence of protease inhibitors. *Exp. Parasitol.* **1983**, *55*, 306–311, doi:10.1016/0014-4894(83)90027-9.

63. Lyon, J.A.; Haynes, J.D.; Diggs, C.L.; Chulay, J.D.; Pratt-Rossiter, J.M. Plasmodium falciparum antigens synthesized by schizonts and stabilized at the merozoite surface by antibodies when schizonts mature in the presence of growth inhibitory immune serum. *J. Immunol.* **1986**, *136*, 2252–2258, doi:10.4049/JIMMUNOL.136.6.2252.

64. Salmon, B.L.; Oksman, A.; Goldberg, D.E. Malaria parasite exit from the host erythrocyte: A two-step process requiring extraerythrocytic proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 271–276, doi:10.1073/PNAS.98.1.271/ASSET/36111C01-1386-4F98-9819-EA9733C9CC1A/ASSETS/GRAPHIC/PQ0114131006.JPEG.

65. Wickham, M.E.; Culvenor, J.G.; Cowman, A.F. Selective Inhibition of a Two-step Egress of Malaria Parasites from the Host Erythrocyte. *J. Biol. Chem.* **2003**, *278*, 37658–37663, doi:10.1074/JBC.M305252200.

66. Liuh, L.G.; Sim, T.S. Characterization of amino acid variation at strategic positions in parasite and human proteases for selective inhibition of falcipains in Plasmodium falciparum. *Biochem. Biophys. Res. Commun.* **2005**, *335*, 762–770, doi:10.1016/J.BBRC.2005.07.147.

67. Eksi, S.; Czesny, B.; Greenbaum, D.C.; Bogyo, M.; Williamson, K.C. Targeted disruption of Plasmodium falciparum cysteine protease, falcipain 1, reduces oocyst production, not erythrocytic stage growth. *Mol. Microbiol.* **2004**, *53*, 243–250, doi:10.1111/J.1365-2958.2004.04108.X.

68. Shenai, B.R.; Sijwali, P.S.; Singh, A.; Rosenthal, P.J. Characterization of Native and Recombinant Falcipain-2, a Principal Trophozoite Cysteine Protease and Essential Hemoglobinase ofPlasmodium falciparum. *J. Biol. Chem.* **2000**, *275*, 29000–29010, doi:10.1074/JBC.M004459200.

69. Nielsen, K.M.; Kasper, J.; Choi, M.; Bedford, T.; Kristiansen, K.; Wirth, D.F.; Volkman, S.K.; Lozovsky, E.R.; Hartl, D.L. Gene Conversion as a Source of Nucleotide Diversity in Plasmodium falciparum. *Mol. Biol. Evol.* **2003**, *20*, 726–734,

doi:10.1093/MOLBEV/MSG076.

70. S. SIJWALI, P.; R. SHENAI, B.; GUT, J.; SINGH, A.; J. ROSENTHAL, P. Expression and characterization of the Plasmodium falciparum haemoglobinase falcipain-3. *Biochem. J.* **2001**, *360*, 481, doi:10.1042/0264-6021:3600481.

71. Singh, N.; Sijwali, P.S.; Pandey, K.C.; Rosenthal, P.J. Plasmodium falciparum: Biochemical characterization of the cysteine protease falcipain-2′. *Exp. Parasitol.* **2006**, *112*, 187–192, doi:10.1016/J.EXPPARA.2005.10.007.

72. Pandey, K.C.; Sijwali, P.S.; Singh, A.; Na, B.K.; Rosenthal, P.J. Independent Intramolecular Mediators of Folding, Activity, and Inhibition for the Plasmodium falciparum Cysteine Protease Falcipain-2. *J. Biol. Chem.* **2004**, *279*, 3484–3491, doi:10.1074/JBC.M310536200.

73. Rosenthal, P.J. Falcipains and other cysteine proteases of malaria parasites. In *Advances in Experimental Medicine and Biology*; Springer, Boston, MA, 2011; Vol. 712, pp. 30–48 ISBN 9781441984135.

74. Sijwali, P.S.; Rosenthal, P.J. Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 4384–9, doi:10.1073/pnas.0307720101.

75. Greenbaum, D.C.; Baruch, A.; Grainger, M.; Bozdech, Z.; Medzihradszky, K.F.; Engel, J.; DeRisi, J.; Holder, A.A.; Bogyo, M. A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **2002**, *298*, 2002–2006, doi:10.1126/SCIENCE.1077426.

76. Sijwali, P.S.; Brinen, L.S.; Rosenthal, P.J. Systematic Optimization of Expression and Refolding of the Plasmodium falciparum Cysteine Protease Falcipain-2. *Protein Expr. Purif.* **2001**, *22*, 128–134, doi:10.1006/PREP.2001.1416.

77. Sajid, M.; McKerrow, J.H. Cysteine proteases of parasitic organisms☆. *Mol. Biochem. Parasitol.* **2002**, *120*, 1–21, doi:10.1016/S0166-6851(01)00438-8.

78. Dahl, E.L.; Rosenthal, P.J. Biosynthesis, localization, and processing of falcipain cysteine proteases of Plasmodium falciparum. *Mol. Biochem. Parasitol.* **2005**, *139*, 205–212, doi:10.1016/j.molbiopara.2004.11.009.

79. Pandey, K.C.; Dixit, R. Structure-Function of Falcipains: Malarial Cysteine Proteases.

*J. Trop. Med.* **2012**, *2012*, 1–11, doi:10.1155/2012/345195.

80. Sundararaj, S.; Singh, D.; Saxena, A.K.; Vashisht, K.; Sijwali, P.S.; Dixit, R.; Pandey, K.C. The Ionic and Hydrophobic Interactions Are Required for the Auto Activation of Cysteine Proteases of Plasmodium falciparum. *PLoS One* **2012**, *7*, e47227, doi:10.1371/journal.pone.0047227.

81. Huete-Pérez, J.A.; Engel, J.C.; Brinen, L.S.; Mottram, J.C.; McKerrow, J.H. Protease trafficking in two primitive eukaryotes is mediated by a prodomain protein motif. *J. Biol. Chem.* **1999**, *274*, 16249–16256, doi:10.1074/jbc.274.23.16249.

82. Brömme, D.; Nallaseth, F.S.; Turk, B. Production and activation of recombinant papain-like cysteine proteases. *Methods* **2004**, *32*, 199–206, doi:10.1016/S1046-2023(03)00212-3.

83. Wang, S.X.; Pandey, K.C.; Somoza, J.R.; Sijwali, P.S.; Kortemme, T.; Brinen, L.S.; Fletterick, R.J.; Rosenthal, P.J.; McKerrow, J.H. Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease. *Proc. Natl. Acad. Sci.* **2006**, *103*, 11503–11508, doi:10.1073/pnas.0600489103.

84. Hogg, T.; Nagarajan, K.; Herzberg, S.; Chen, L.; Shen, X.; Jiang, H.; Wecke, M.; Blohmke, C.; Hilgenfeld, R.; Schmidt, C.L. Structural and functional characterization of Falcipain-2, a hemoglobinase from the malarial parasite Plasmodium falciparum. *J. Biol. Chem.* **2006**, *281*, 25425–25437, doi:10.1074/JBC.M603776200.

85. Kerr, I.D.; Lee, J.H.; Farady, C.J.; Marion, R.; Rickert, M.; Sajid, M.; Pandey, K.C.; Caffrey, C.R.; Legac, J.; Hansell, E.; et al. Vinyl Sulfones as Antiparasitic Agents and a Structural Basis for Drug Design. *J. Biol. Chem.* **2009**, *284*, 25697–25703, doi:10.1074/jbc.M109.014340.

86. Rosenthal, P.J. Falcipain cysteine proteases of malaria parasites: An update. *Biochim. Biophys. Acta - Proteins Proteomics* **2020**, *1868*, 140362, doi:10.1016/j.bbapap.2020.140362.

87. Musyoka, T.M.; Kanzi, A.M.; Lobb, K.A.; Tastan Bishop, Ö. Analysis of non-peptidic compounds as potential malarial inhibitors against Plasmodial cysteine proteases via integrated virtual screening workflow. *J. Biomol. Struct. Dyn.* **2016**, *34*, 2084–2101, doi:10.1080/07391102.2015.1108231.

88. Nureye, D.; Nureye, D. Malaria: Introductory Concepts, Resistance Issues and Current Medicines. *Plasmodium Species Drug Resist.* **2021**, doi:10.5772/INTECHOPEN.98725.

89. Imwong, M.; Suwannasin, K.; Kunasol, C.; Sutawong, K.; Mayxay, M.; Rekol, H.; Smithuis, F.M.; Hlaing, T.M.; Tun, K.M.; van der Pluijm, R.W.; et al. The spread of artemisinin-resistant Plasmodium falciparum in the Greater Mekong subregion: a molecular epidemiology observational study. *Lancet. Infect. Dis.* **2017**, *17*, 491, doi:10.1016/S1473-3099(17)30048-8.

90. WHO report - Strategy to respond to antimalarial drug resistance in Africa Available online: https://www.malariaworld.org/blogs/who-report-strategy-to-respond-to-antimalarial-drug-resistance-in-africa (accessed on Mar 15, 2023).

91. Ikeda, M.; Kaneko, M.; Tachibana, S.I.; Balikagala, B.; Sakurai-Yatsushiro, M.; Yatsushiro, S.; Takahashi, N.; Yamauchi, M.; Sekihara, M.; Hashimoto, M.; et al. Artemisinin-Resistant Plasmodium falciparum with High Survival Rates, Uganda, 2014–2016. *Emerg. Infect. Dis.* **2018**, *24*, 718, doi:10.3201/EID2404.170141.

92. Lu, F.; Culleton, R.; Zhang, M.; Ramaprasad, A.; von Seidlein, L.; Zhou, H.; Zhu, G.; Tang, J.; Liu, Y.; Wang, W.; et al. Emergence of Indigenous Artemisinin-Resistant Plasmodium falciparum in Africa . *N. Engl. J. Med.* **2017**, *376*, 991–993, doi:10.1056/NEJMC1612765/SUPPL_FILE/NEJMC1612765_DISCLOSURES.PDF.

93. Balikagala, B.; Fukuda, N.; Ikeda, M.; Katuro, O.T.; Tachibana, S.-I.; Yamauchi, M.; Opio, W.; Emoto, S.; Anywar, D.A.; Kimura, E.; et al. Evidence of Artemisinin-Resistant Malaria in Africa. *N. Engl. J. Med.* **2021**, *385*, 1163–1171, doi:10.1056/NEJMOA2101746/SUPPL_FILE/NEJMOA2101746_DISCLOSURES.PDF.

94. Fairhurst, R.M.; Dondorp, A.M. Artemisinin-Resistant Plasmodium falciparum Malaria . *Microbiol. Spectr.* **2016**, *4*, 409–429, doi:10.1128/microbiolspec.ei10-0013-2016.

95. Klonis, N.; Crespo-Ortiz, M.P.; Bottova, I.; Abu-Bakar, N.; Kenny, S.; Rosenthal, P.J.; Tilley, L. Artemisinin activity against Plasmodium falciparum requires hemoglobin uptake and digestion. *Proc. Natl. Acad. Sci.* **2011**, *108*, 11405–11410, doi:10.1073/pnas.1104063108.

96. Siddiqui, F.A.; Liang, X.; Cui, L. Plasmodium falciparum resistance to ACTs:

Emergence, mechanisms, and outlook. *Int. J. Parasitol. Drugs Drug Resist.* **2021**, *16*, 102, doi:10.1016/J.IJPDDR.2021.05.007.

97. Siddiqui, F.A.; Cabrera, M.; Wang, M.; Brashear, A.; Kemirembe, K.; Wang, Z.; Miao, J.; Chookajorn, T.; Yang, Z.; Cao, Y.; et al. Plasmodium falciparum Falcipain-2a Polymorphisms in Southeast Asia and Their Association With Artemisinin Resistance. *J. Infect. Dis.* **2018**, *218*, 434–442, doi:10.1093/infdis/jiy188.

98. Xie, S.C.; Dogovski, C.; Hanssen, E.; Chiu, F.; Yang, T.; Crespo, M.P.; Stafford, C.; Batinovic, S.; Teguh, S.; Charman, S.; et al. Haemoglobin degradation underpins the sensitivity of early ring stage Plasmodium falciparum to artemisinins. *J. Cell Sci.* **2016**, *129*, 406–416, doi:10.1242/jcs.178830.

99. Sijwali, P.S.; Koo, J.; Singh, N.; Rosenthal, P.J. Gene disruptions demonstrate independent roles for the four falcipain cysteine proteases of Plasmodium falciparum. *Mol. Biochem. Parasitol.* **2006**, *150*, 96–106, doi:10.1016/J.MOLBIOPARA.2006.06.013.

100. Ariey, F.; Witkowski, B.; Amaratunga, C.; Beghain, J.; Langlois, A.-C.; Khim, N.; Kim, S.; Duru, V.; Bouchier, C.; Ma, L.; et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature* **2014**, *505*, 50–55, doi:10.1038/nature12876.

101. Bahl, A. PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic Acids Res.* **2003**, *31*, 212–215, doi:10.1093/nar/gkg081.

102. Aurrecoechea, C.; Brestelli, J.; Brunk, B.P.; Dommer, J.; Fischer, S.; Gajria, B.; Gao, X.; Gingle, A.; Grant, G.; Harb, O.S.; et al. PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Res.* **2009**, *37*, D539–D543, doi:10.1093/nar/gkn814.

103. Kulshreshtha, S.; Chaudhary, V.; Goswami, G.K.; Mathur, N. Computational approaches for predicting mutant protein stability. *J. Comput. Aided. Mol. Des.* **2016**, *30*, 401–412, doi:10.1007/S10822-016-9914-3.

104. Ben-Tal, N.; Kihara, D.; Pandurangan, A.P. Editorial: Computational Approaches to Study the Impact of Mutations on Disease and Drug Resistance. *Front. Mol. Biosci.* **2021**, *8*, 813552, doi:10.3389/fmolb.2021.813552.

105. Hao, G.F.; Yang, G.F.; Zhan, C.G. Structure-based methods for predicting target mutation-induced drug resistance and rational drug design to overcome the problem. *Drug Discov. Today* **2012**, *17*, 1121–1126, doi:10.1016/J.DRUDIS.2012.06.018.

106. Chen, Y.; Lu, H.; Zhang, N.; Zhu, Z.; Wang, S.; Li, M. PremPS: Predicting the impact of missense mutations on protein stability. *PLoS Comput. Biol.* **2020**, *16*, doi:10.1371/JOURNAL.PCBI.1008543.

107. Brown, D.K.; Sheik Amamuddy, O.; Tastan Bishop, Ö. Structure-Based Analysis of Single Nucleotide Variants in the Renin-Angiotensinogen Complex. *Glob. Heart* **2017**, *12*, 121, doi:10.1016/j.gheart.2017.01.006.

108. Sheik Amamuddy, O.; Bishop, N.T.; Tastan Bishop, Ö. Characterizing early drug resistance-related events using geometric ensembles from HIV protease dynamics. *Sci. Rep.* **2018**, *8*, 17938, doi:10.1038/s41598-018-36041-8.

109. Barozi, V.; Musyoka, T.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**, *7*, 13313–13332, doi:10.1021/ACSOMEGA.2C01036/ASSET/IMAGES/MEDIUM/AO2C01036_M007.GIF.

110. Barozi, V.; Edkins, A.L.; 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. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 4562–4578, doi:10.1016/J.CSBJ.2022.08.015.

111. Monama, M.Z.; Olotu, F.; Tastan Bishop, Ö. Investigation of Multi-Subunit Mycobacterium tuberculosis DNA-Directed RNA Polymerase and Its Rifampicin Resistant Mutants. *Int. J. Mol. Sci.* **2023**, *24*, 3313, doi:10.3390/IJMS24043313/S1.

112. Sanyanga, T.A.; Nizami, B.; Tastan Bishop, Ö. Mechanism of Action of Non-Synonymous Single Nucleotide Variations Associated with α-Carbonic Anhydrase II Deficiency. *Molecules* **2019**, *24*, 3987, doi:10.3390/molecules24213987.

113. Amusengeri, A.; Tata, R.B.; Tastan Bishop, Ö. Understanding the Pyrimethamine Drug Resistance Mechanism via Combined Molecular Dynamics and Dynamic Residue

Network Analysis. *Molecules* **2020**, *25*, 904, doi:10.3390/molecules25040904.

114. Sanyanga, T.A.; Bishop, Ö.T. Structural Characterization of Carbonic Anhydrase VIII and Effects of Missense Single Nucleotide Variations to Protein Structure and Function. *Int. J. Mol. Sci. 2020, Vol. 21, Page 2764* **2020**, *21*, 2764, doi:10.3390/IJMS21082764.

115. Sheik Amamuddy, O.; Musyoka, T.M.; Boateng, R.A.; Zabo, S.; Tastan Bishop, Ö. Determining the unbinding events and conserved motions associated with the pyrazinamide release due to resistance mutations of Mycobacterium tuberculosis pyrazinamidase. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1103–1120, doi:10.1016/j.csbj.2020.05.009.

116. Sheik Amamuddy, O.; Verkhivker, G.M.; Tastan Bishop, Ö. Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 M pro. *J. Chem. Inf. Model.* **2020**, *60*, 5080–5102, doi:10.1021/acs.jcim.0c00634.

117. Chebon-Bore, L.; Sanyanga, T.A.; Manyumwa, C.V.; Khairallah, A.; Bishop, Ö.T.; Tastan Bishop, Ö. Decoding the Molecular Effects of Atovaquone Linked Resistant Mutations on Plasmodium falciparum Cytb-ISP Complex in the Phospholipid Bilayer Membrane. *Int. J. Mol. Sci. 2021, Vol. 22, Page 2138* **2021**, *22*, 2138, doi:10.3390/IJMS22042138.

118. Okeke, C.J.; Musyoka, T.M.; Sheik Amamuddy, O.; Barozi, V.; Tastan Bishop, Ö. Allosteric pockets and dynamic residue network hubs of falcipain 2 in mutations including those linked to artemisinin resistance. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 5647–5666, doi:10.1016/j.csbj.2021.10.011.

119. Sheik Amamuddy, O.; Afriyie Boateng, R.; Barozi, V.; Wavinya Nyamai, D.; Tastan Bishop, Ö. 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.* **2021**, *19*, 6431–6455, doi:10.1016/J.CSBJ.2021.11.016.

120. Peng, Y.; Alexov, E.; Basu, S. Structural Perspective on Revealing and Altering Molecular Functions of Genetic Variants Linked with Diseases. *Int. J. Mol. Sci.* **2019**, *20*, doi:10.3390/IJMS20030548.

121. Kapetanovic, I.M. COMPUTER-AIDED DRUG DISCOVERY AND DEVELOPMENT (CADDD): in silico-chemico-biological approach. *Chem. Biol.*

*Interact.* **2008**, *171*, 165, doi:10.1016/J.CBI.2006.12.006.

122. Ferreira, L.G.; Dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, *20*, 13384, doi:10.3390/MOLECULES200713384.

123. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. The Shape and Structure of Proteins. **2002**.

124. Muhammed, M.T.; Aki-Yalcin, E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem. Biol. Drug Des.* **2019**, *93*, 12–20, doi:10.1111/CBDD.13388.

125. Di Luccio, E.; Koehl, P. A quality metric for homology modeling: The H-factor. *BMC Bioinformatics* **2011**, *12*, 1–19, doi:10.1186/1471-2105-12-48/TABLES/1.

126. Venselaar, H.; Joosten, R.P.; Vroling, B.; Baakman, C.A.B.; Hekkelman, M.L.; Krieger, E.; Vriend, G. Homology modelling and spectroscopy, a never-ending love story. *Eur. Biophys. J.* **2010**, *39*, 551, doi:10.1007/S00249-009-0531-0.

127. Lionta, E.; Spyrou, G.; Vassilatis, D.; Cournia, Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr. Top. Med. Chem.* **2014**, doi:10.2174/1568026614666140929124445.

128. Dhasmana, A.; Raza, S.; Jahan, R.; Lohani, M.; Arif, J.M. High-Throughput Virtual Screening (HTVS) of Natural Compounds and Exploration of Their Biomolecular Mechanisms: An In Silico Approach. *New Look to Phytomedicine Adv. Herb. Prod. as Nov. Drug Leads* **2019**, 523–548, doi:10.1016/B978-0-12-814619-4.00020-3.

129. Cerqueira, N.M.F.S.A.; Gesto, D.; Oliveira, E.F.; Santos-Martins, D.; Brás, N.F.; Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Receptor-based virtual screening protocol for drug discovery. *Arch. Biochem. Biophys.* **2015**, *582*, 56–67, doi:10.1016/J.ABB.2015.05.011.

130. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* **2011**, doi:10.2174/157340911795677602.

131. McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of folded proteins. *Nat. 1977 2675612* **1977**, *267*, 585–590, doi:10.1038/267585a0.

132. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652, doi:10.1038/NSB0902-646.

133. Hollingsworth, S.A.; Dror, R.O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143, doi:10.1016/j.neuron.2018.08.011.

134. Salo-Ahen, O.M.H.; Alanko, I.; Bhadane, R.; Alexandre, A.M.; Honorato, R.V.; Hossain, S.; Juffer, A.H.; Kabedev, A.; Lahtela-Kakkonen, M.; Larsen, A.S.; et al. Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes* **2021**, *9*, 1–63, doi:10.3390/PR9010071.

135. Takemura, K.; Kitao, A. Molecular dynamics. *Plasma Membr. Shap.* **2022**, 431–443, doi:10.1016/B978-0-323-89911-6.00024-8.

136. De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061, doi:10.1021/acs.jmedchem.5b01684.

137. Guvench, O.; MacKerell, A.D. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* **2008**, *443*, 63–88, doi:10.1007/978-1-59745-177-2_4.

138. Kollman, P.A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Acc. Chem. Res.* **1996**, *29*, 461–469, doi:10.1021/ar9500675.

139. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 1950–1958, doi:10.1002/prot.22711.

140. Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.M.; Mittal, J.; Feig, M.; MacKerell, A.D.; Jr; RB, B.; X, Z.; et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257, doi:10.1021/CT300400X.

141. Scott, W.R.P.; Hünenberger, P.H.; Tironi, I.G.; Mark, A.E.; Billeter, S.R.; Fennen, J.; Torda, A.E.; Huber, T.; Krüger, P.; Van Gunsteren, W.F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607,

doi:10.1021/JP984217F.

142. Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236, doi:10.1021/JA9621760/SUPPL_FILE/JA11225.PDF.

143. Hospital, A.; Goñi, J.R.; Orozco, M.; Gelpí, J.L. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinforma. Chem.* **2015**, *8*, 37–47, doi:10.2147/AABC.S70333.

144. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688, doi:10.1002/JCC.20290.

145. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, doi:10.1002/jcc.540040211.

146. Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, doi:10.1016/0010-4655(95)00042-E.

147. Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312, doi:10.1006/JCPH.1999.6201.

148. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802, doi:10.1002/jcc.20289.

149. Maruyama, Y.; Igarashi, R.; Ushiku, Y.; Mitsutake, A. Analysis of Protein Folding Simulation with Moving Root Mean Square Deviation. *J. Chem. Inf. Model.* **2022**, *63*, 1529–1541, doi:10.1021/ACS.JCIM.2C01444/ASSET/IMAGES/LARGE/CI2C01444_0014.JPEG.

150. Aier, I.; Varadwaj, P.K.; Raj, U. Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. *Sci. Reports 2016 61* **2016**, *6*, 1–10,

doi:10.1038/srep34984.

151. Abdalla, M.; Eltayb, W.A.; El-Arabey, A.A.; Singh, K.; Jiang, X. Molecular dynamic study of SARS-CoV-2 with various S protein mutations and their effect on thermodynamic properties. *Comput. Biol. Med.* **2022**, *141*, 105025, doi:10.1016/J.COMPBIOMED.2021.105025.

152. Lobanov, M.Y.; Bogatyreva, N.S.; Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* **2008**, *42*, 623–628, doi:10.1134/S0026893308040195/METRICS.

153. David, C.C.; Jacobs, D.J. Principal component analysis: A method for determining the essential dynamics of proteins. In *Methods in Molecular Biology*; NIH Public Access, 2014; Vol. 1084, pp. 193–226 ISBN 9781627036573.

154. Brylinski, M.; Skolnick, J. What is the relationship between the global structures of apo and holo proteins? *Proteins Struct. Funct. Genet.* **2008**, doi:10.1002/prot.21510.

155. Ha, J.H.; Loh, S.N. Protein Conformational Switches: From Nature to Design. *Chemistry* **2012**, *18*, 7984, doi:10.1002/CHEM.201200348.

156. Kitao, A. Principal Component Analysis and Related Methods for Investigating the Dynamics of Biological Macromolecules. *J 2022, Vol. 5, Pages 298-317* **2022**, *5*, 298–317, doi:10.3390/J5020021.

157. David, C.C.; Singam, E.R.A.; Jacobs, D.J. JED: A Java Essential Dynamics Program for comparative analysis of protein trajectories. *BMC Bioinformatics* **2017**, *18*, 1–9, doi:10.1186/S12859-017-1676-Y/FIGURES/5.

158. Maisuradze, G.G.; Liwo, A.; Scheraga, H.A. Principal Component Analysis for Protein Folding Dynamics. *J. Mol. Biol.* **2009**, *385*, 312–329, doi:10.1016/J.JMB.2008.10.018.

159. Meyer, T.; Ferrer-Costa, C.; Pérez, A.; Rueda, M.; Bidon-Chanal, A.; Luque, F.J.; Laughton, C.A.; Orozco, M. Essential dynamics: A tool for efficient trajectory compression and management. *J. Chem. Theory Comput.* **2006**, *2*, 251–258.

160. Bakan, A.; Bahar, I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14349, doi:10.1073/PNAS.0904214106.

161. Nizami, B.; Sydow, D.; Wolber, G.; Honarparvar, B. Molecular insight on the binding of NNRTI to K103N mutated HIV-1 RT: molecular dynamics simulations and dynamic pharmacophore analysis. *Mol. Biosyst.* **2016**, *12*, 3385–3395, doi:10.1039/C6MB00428H.

162. Tastan Bishop, Ö.; Musyoka, T.M.; Barozi, V. Allostery and Missense Mutations as Intermittently Linked Promising Aspects of Modern Computational Drug Discovery. *J. Mol. Biol.* **2022**, *434*, 167610, doi:10.1016/J.JMB.2022.167610.

163. Sheik Amamuddy, O.; Glenister, M.; Tshabalala, T.; Tastan Bishop, Ö. MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 5059–5071, doi:10.1016/j.csbj.2021.08.043.

164. Duron, C. Heatmap centrality: A new measure to identify super-spreader nodes in scale-free networks. *PLoS One* **2020**, *15*, doi:10.1371/JOURNAL.PONE.0235690.

165. Brown, D.K.; Penkler, D.L.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A.R.; Atilgan, C.; Tastan Bishop, Ö. MD-TASK: A software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **2017**, doi:10.1093/bioinformatics/btx349.

166. Özgür, A.; Vu, T.; Erkan, G.; Radev, D.R. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* **2008**, *24*, i277–i285, doi:10.1093/bioinformatics/btn182.

167. Hansen, D.L.; Shneiderman, B.; Smith, M.A. Analyzing Social Media Networks With NodeXL. *Anal. Soc. Media Networks with NodeXL* **2011**, doi:10.1016/C2009-0-64028-9.

168. Zhan, J.; Guidibande, V.; Parsa, S.P.K. Identification of top-K influential communities in big networks. *J. Big Data* **2016**, *3*, 1–28, doi:10.1186/S40537-016-0050-7/TABLES/19.

169. Adhikari, B.; Cheng, J. Protein Residue Contacts and Prediction Methods. *Methods Mol. Biol.* **2016**, *1415*, 463, doi:10.1007/978-1-4939-3572-7_24.

170. Monastyrskyy, B.; D'Andrea, D.; Fidelis, K.; Tramontano, A.; Kryshtafovych, A. New encouraging developments in contact prediction: Assessment of the CASP11 results. *Proteins Struct. Funct. Bioinforma.* **2016**, *84*, 131–144, doi:10.1002/PROT.24943.

171. Simkovic, F.; Ovchinnikov, S.; Baker, D.; Rigden, D.J. Applications of contact

predictions to structural biology. *urn:issn:2052-2525* **2017**, *4*, 291–300, doi:10.1107/S2052252517005115.

172. Luttrell, J.; Liu, T.; Zhang, C.; Wang, Z. Predicting protein residue-residue contacts using random forests and deep networks. *BMC Bioinformatics* **2019**, *20*, 115–127, doi:10.1186/S12859-019-2627-6/FIGURES/2.

173. Ray, S.; Das, S.; Suar, M. Molecular Mechanism of Drug Resistance. In *Drug Resistance in Bacteria, Fungi, Malaria, and Cancer*; Springer International Publishing: Cham, 2017; pp. 47–110 ISBN 9783319486833.

174. Revie, N.M.; Iyer, K.R.; Robbins, N.; Cowen, L.E. Antifungal drug resistance: evolution, mechanisms and impact. *Curr. Opin. Microbiol.* **2018**, *45*, 70–76, doi:10.1016/j.mib.2018.02.005.

175. Mason, S.; Devincenzo, J.P.; Toovey, S.; Wu, J.Z.; Whitley, R.J. Comparison of antiviral resistance across acute and chronic viral infections. *Antiviral Res.* **2018**, *158*, 103–112, doi:10.1016/J.ANTIVIRAL.2018.07.020.

176. Cowman, A.F.; Morry, M.J.; Biggs, B.A.; Cross, G.A.M.; Foote, S.J. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. *Proc. Natl. Acad. Sci.* **1988**, *85*, 9109–9113, doi:10.1073/pnas.85.23.9109.

177. Djimdé, A.; Doumbo, O.K.; Cortese, J.F.; Kayentao, K.; Doumbo, S.; Diourté, Y.; Coulibaly, D.; Dicko, A.; Su, X.Z.; Nomura, T.; et al. A molecular marker for chloroquine-resistant falciparum malaria. *N. Engl. J. Med.* **2001**, *344*, 257–63, doi:10.1056/NEJM200101253440403.

178. Reed, M.B.; Saliba, K.J.; Caruana, S.R.; Kirk, K.; Cowman, A.F. Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum. *Nature* **2000**, *403*, 906–9, doi:10.1038/35002615.

179. Antony, H.; Parija, S. Antimalarial drug resistance: An overview. *Trop. Parasitol.* **2016**, *6*, 30, doi:10.4103/2229-5070.175081.

180. Akhoon, B.A.; Singh, K.P.; Varshney, M.; Gupta, S.K.; Shukla, Y.; Gupta, S.K. Understanding the mechanism of atovaquone drug resistance in Plasmodium falciparum cytochrome b mutation Y268S using computational methods. *PLoS One* **2014**, *9*,

e110041, doi:10.1371/journal.pone.0110041.

181. Sun, H.; Li, Y.; Li, D.; Hou, T. Insight into crizotinib resistance mechanisms caused by three mutations in ALK tyrosine kinase using free energy calculation approaches. *J. Chem. Inf. Model.* **2013**, *53*, 2376–89, doi:10.1021/ci400188q.

182. Chen, J. Drug resistance mechanisms of three mutations V32I, I47V and V82I in HIV-1 protease toward inhibitors probed by molecular dynamics simulations and binding free energy predictions. *RSC Adv.* **2016**, *6*, 58573–58585, doi:10.1039/C6RA09201B.

183. Palzkill, T. Structural and Mechanistic Basis for Extended-Spectrum Drug-Resistance Mutations in Altering the Specificity of TEM, CTX-M, and KPC β-lactamases. *Front. Mol. Biosci.* **2018**, *5*, doi:10.3389/fmolb.2018.00016.

184. Rehman, A.U.; Khan, M.T.; Liu, H.; Wadood, A.; Malik, S.I.; Chen, H.-F. Exploring the Pyrazinamide Drug Resistance Mechanism of Clinical Mutants T370P and W403G in Ribosomal Protein S1 of Mycobacterium tuberculosis. *J. Chem. Inf. Model.* **2019**, *59*, 1584–1597, doi:10.1021/acs.jcim.8b00956.

185. Moonsamy, S.; Bhakat, S.; Walker, R.C.; Soliman, M.E.S. Single Active Site Mutation Causes Serious Resistance of HIV Reverse Transcriptase to Lamivudine: Insight from Multiple Molecular Dynamics Simulations. *Cell Biochem. Biophys.* **2016**, *74*, 35–48, doi:10.1007/s12013-015-0709-2.

186. Ragland, D.A.; Nalivaika, E.A.; Nalam, M.N.L.; Prachanronarong, K.L.; Cao, H.; Bandaranayake, R.M.; Cai, Y.; Kurt-Yilmaz, N.; Schiffer, C.A. Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. *J. Am. Chem. Soc.* **2014**, *136*, 11956–63, doi:10.1021/ja504096m.

187. Henes, M.; Lockbaum, G.J.; Kosovrasti, K.; Leidner, F.; Nachum, G.S.; Nalivaika, E.A.; Lee, S.-K.; Spielvogel, E.; Zhou, S.; Swanstrom, R.; et al. Picomolar to Micromolar: Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance. *ACS Chem. Biol.* **2019**, *14*, 2441–2452, doi:10.1021/acschembio.9b00370.

188. Ren, J.; Stammers, D.K. Structural basis for drug resistance mechanisms for non-nucleoside inhibitors of HIV reverse transcriptase. *Virus Res.* **2008**, *134*, 157–170, doi:10.1016/j.virusres.2007.12.018.

189. Bastys, T.; Gapsys, V.; Walter, H.; Heger, E.; Doncheva, N.T.; Kaiser, R.; de Groot, B.L.; Kalinina, O. V. Non-active site mutants of HIV-1 protease influence resistance and sensitisation towards protease inhibitors. *Retrovirology* **2020**, *17*, 13, doi:10.1186/s12977-020-00520-6.

190. Naganathan, A.N. Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Curr. Opin. Struct. Biol.* **2019**, *54*, 1–9, doi:10.1016/j.sbi.2018.09.004.

191. Mitternacht, S.; Berezovsky, I.N. Binding leverage as a molecular basis for allosteric regulation. *PLoS Comput. Biol.* **2011**, *7*, 1002148, doi:10.1371/journal.pcbi.1002148.

192. Guarnera, E.; Berezovsky, I.N. On the perturbation nature of allostery: sites, mutations, and signal modulation. *Curr. Opin. Struct. Biol.* **2019**, *56*, 18–27, doi:10.1016/j.sbi.2018.10.008.

193. Pandey, K.C.; Barkan, D.T.; Sali, A.; Rosenthal, P.J. Regulatory Elements within the Prodomain of Falcipain-2, a Cysteine Protease of the Malaria Parasite Plasmodium falciparum. *PLoS One* **2009**, *4*, doi:10.1371/JOURNAL.PONE.0005694.

194. Pandey, K.C.; Wang, S.X.; Sijwali, P.S.; Lau, A.L.; McKerrow, J.H.; Rosenthal, P.J. The Plasmodium falciparum cysteine protease falcipain-2 captures its substrate, hemoglobin, via a unique motif. *Proc. Natl. Acad. Sci.* **2005**, *102*, 9138–9143, doi:10.1073/pnas.0502368102.

195. Pant, A.; Kumar, R.; Wani, N.A.; Verma, S.; Sharma, R.; Pande, V.; Saxena, A.K.; Dixit, R.; Rai, R.; Pandey, K.C. Allosteric Site Inhibitor Disrupting Auto-Processing of Malarial Cysteine Proteases. *Sci. Reports 2018 81* **2018**, *8*, 1–15, doi:10.1038/s41598-018-34564-8.

196. Wang, S.X.; Pandey, K.C.; Scharfstein, J.; Whisstock, J.; Huang, R.K.; Jacobelli, J.; Fletterick, R.J.; Rosenthal, P.J.; Abrahamson, M.; Brinen, L.S.; et al. The Structure of Chagasin in Complex with a Cysteine Protease Clarifies the Binding Mode and Evolution of an Inhibitor Family. *Structure* **2007**, *15*, 535–543, doi:10.1016/j.str.2007.03.012.

197. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N. The Protein Data Bank (www.rcsb.org). *Nucleic Acids Res.* **2000**,

doi:10.1093/nar/28.1.235.

198. Burley, S.K.; Berman, H.M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L.; Christie, C.; Dalenberg, K.; Duarte, J.M.; Dutta, S.; et al. RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* **2019**, *47*, D464–D474, doi:10.1093/nar/gky1004.

199. DeLano, W.L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **2002**.

200. Bateman, A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515, doi:10.1093/nar/gky1049.

201. Šali, a MODELLER: A Program for Protein Structure Modeling Release 9.12, r9480. *Rockefeller Univ.* **2013**, 779–815.

202. Shen, M.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **2006**, doi:10.1110/ps.062416606.

203. Eisenberg, D.; Lüthy, R.; Bowie, J.U. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **1997**, *277*, 396–404, doi:10.1016/S0076-6879(97)77022-8.

204. Wiederstein, M.; Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410, doi:10.1093/NAR/GKM290.

205. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291, doi:10.1107/S0021889892009944.

206. Musyoka, T.M.; Njuguna, J.N.; Tastan Bishop, Ö. Comparing sequence and structure of falcipains and human homologs at prodomain and catalytic active site for malarial peptide based inhibitor design. *Malar. J.* **2019**, doi:10.1186/s12936-019-2790-2.

207. Tina, K.G.; Bhadra, R.; Srinivasan, N. PIC: Protein Interactions Calculator. *Nucleic Acids Res.* **2007**, *35*, W473–W476, doi:10.1093/nar/gkm423.

208. Dolinsky, T.J.; Czodrowski, P.; Li, H.; Nielsen, J.E.; Jensen, J.H.; Klebe, G.; Baker,

N.A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522–W525, doi:10.1093/nar/gkm276.

209. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667, doi:10.1093/nar/gkh381.

210. Huang, W.; Lu, S.; Huang, Z.; Liu, X.; Mou, L.; Luo, Y.; Zhao, Y.; Liu, Y.; Chen, Z.; Hou, T.; et al. Allosite: A method for predicting allosteric sites. *Bioinformatics* **2013**, doi:10.1093/bioinformatics/btt399.

211. Kozakov, D.; Grove, L.E.; Hall, D.R.; Bohnuud, T.; Mottarella, S.E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* **2015**, doi:10.1038/nprot.2015.043.

212. Panjkovich, A.; Daura, X. PARS: a web server for the prediction of Protein Allosteric and Regulatory Sites. **2014**, *30*, doi:10.1093/bioinformatics/btu002.

213. Halgren, T.A.A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389, doi:10.1021/ci800324m.

214. Harris, R.; Olson, A.J.; Goodsell, D.S. Automated prediction of ligand-binding sites in proteins. *Proteins Struct. Funct. Bioinforma.* **2008**, *70*, 1506–1517, doi:10.1002/PROT.21645.

215. Huang, B.; Schroeder, M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.* **2006**, *6*, 19, doi:10.1186/1472-6807-6-19.

216. Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857, doi:10.1021/JM00145A002.

217. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindah, E.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25, doi:10.1016/j.softx.2015.06.001.

218. Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960, doi:10.1021/JP003020W.

219. Lemak, A.S.; Balabaev, N.K. On The Berendsen Thermostat. *http://dx.doi.org/10.1080/08927029408021981* **2006**, *13*, 177–187, doi:10.1080/08927029408021981.

220. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190, doi:10.1063/1.328693.

221. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **1997**, doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

222. Petersen, H.G. Accuracy and efficiency of the particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 3668–3679, doi:10.1063/1.470043.

223. RStudio Team RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. URL http://www.rstudio.com/. *RStudio, Inc.* **2015**.

224. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95, doi:10.1109/MCSE.2007.55.

225. Van der Walt, S.; Colbert, S.C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30, doi:10.1109/MCSE.2011.37.

226. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the Proceedings of the 9th Python in Science Conference; SciPy, 2010; pp. 56–61.

227. Waskom, M.; Botvinnik, O.; Ostblom, J.; Gelbart, M.; Lukauskas, S.; Hobson, P.; Gemperline, D.C.; Augspurger, T.; Halchenko, Y.; Cole, J.B.; et al. mwaskom/seaborn: v0.10.1 (April 2020). *Zenodo* **2020**.

228. Nguyen, H.; Roe, D.R.; Swails, J.; Case, D.A. PYTRAJ: Interactive data analysis for molecular dynamics simulations. *New Brunswick, NJ Rutgers Univ.* **2016**, doi:10.5281/ZENODO.44612.

229. Kuzmanic, A.; Zagrovic, B. Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors. *Biophys. J.* **2010**, *98*, 861,

doi:10.1016/J.BPJ.2009.11.011.

230. Banerjee, V.; Kar, R.K.; Datta, A.; Parthasarathi, K.; Chatterjee, S.; Das, K.P.; Bhunia, A. Use of a Small Peptide Fragment as an Inhibitor of Insulin Fibrillation Process: A Study by High and Low Resolution Spectroscopy. *PLoS One* **2013**, *8*, e72318, doi:10.1371/journal.pone.0072318.

231. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, doi:10.1016/0263-7855(96)00018-5.

232. Surpeta, B.; Sequeiros-Borja, C.E.; Brezovsky, J. Dynamics, a Powerful Component of Current and Future in Silico Approaches for Protein Design and Engineering. *Int. J. Mol. Sci.* **2020**, *21*, 2713, doi:10.3390/ijms21082713.

233. Kumar, C.V.; Swetha, R.G.; Anbarasu, A.; Ramaiah, S. Computational analysis reveals the association of threonine 118 methionine mutation in PMP22 resulting in CMT-1A. *Adv. Bioinformatics* **2014**, *2014*, 1–10, doi:10.1155/2014/502618.

234. Foutch, D.; Pham, B.; Shen, T. Protein conformational switch discerned via network centrality properties. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3599, doi:10.1016/J.CSBJ.2021.06.004.

235. Coterón, J.M.; Catterick, D.; Castro, J.; Chaparro, M.J.; Díaz, B.; Fernández, E.; Ferrer, S.; Gamo, F.J.; Gordo, M.; Gut, J.; et al. Falcipain Inhibitors: Optimization Studies of the 2-Pyrimidinecarbonitrile Lead Series †. *J. Med. Chem.* **2010**, *53*, 6129–6152, doi:10.1021/jm100556b.

236. Domínguez, J.N.; León, C.; Rodrigues, J.; Gamboa de Domínguez, N.; Gut, J.; Rosenthal, P.J. Synthesis and Evaluation of New Antimalarial Phenylurenyl Chalcone Derivatives. *J. Med. Chem.* **2005**, *48*, 3654–3658, doi:10.1021/jm058208o.

237. Rudrapal, M.; Chetia, D.; Singh, V. Novel series of 1,2,4-trioxane derivatives as antimalarial agents. *J. Enzyme Inhib. Med. Chem.* **2017**, *32*, 1159–1173, doi:10.1080/14756366.2017.1363742.

238. Himangini; Pathak, D.P.; Sharma, V.; Kumar, S. Designing novel inhibitors against falcipain-2 of Plasmodium falciparum. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 1566–1569, doi:10.1016/j.bmcl.2018.03.058.

239. Sabnis, Y.A.; Desai, P. V; Rosenthal, P.J.; Avery, M.A. Probing the structure of

falcipain-3, a cysteine protease from Plasmodium falciparum: Comparative protein modeling and docking studies. *Protein Sci.* **2003**, *12*, 501–9, doi:10.1110/ps.0228103.

240. Fiser, A.; Šali, A. MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. *Methods Enzymol.* **2003**, doi:10.1016/S0076-6879(03)74020-8.

241. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **2016**, *54*, 5.6.1, doi:10.1002/CPBI.3.

242. Ross, C.; Nizami, B.; Glenister, M.; Sheik Amamuddy, O.; Atilgan, A.R.; Atilgan, C.; Tastan Bishop, Ö. MODE-TASK: large-scale protein motion tools. *Bioinformatics* **2018**, *34*, 3759–3763, doi:10.1093/bioinformatics/bty427.

243. Kortemme, T.; Kim, D.E.; Baker, D. Computational alanine scanning of protein-protein interfaces. *Sci. STKE* **2004**, *2004*, doi:10.1126/STKE.2192004PL2.

244. Katoh, K.; Misawa, K.; Kuma, K.I.; Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066, doi:10.1093/NAR/GKF436.

245. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.A.; Clamp, M.; Barton, G.J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191, doi:10.1093/bioinformatics/btp033.

246. Kuzmanic, A.; Pritchard, R.B.; Hansen, D.F.; Gervasio, F.L. Importance of the Force Field Choice in Capturing Functionally Relevant Dynamics in the von Willebrand Factor. *J. Phys. Chem. Lett.* **2019**, *10*, 1928–1934, doi:10.1021/ACS.JPCLETT.9B00517/ASSET/IMAGES/LARGE/JZ-2019-00517A_0005.JPEG.

247. Atilgan, A.R.; Akan, P.; Baysal, C. Small-World Communication of Residues and Significance for Protein Dynamics. *Biophys. J.* **2004**, doi:10.1016/S0006-3495(04)74086-2.

248. Patel, P.; Bharti, P.K.; Bansal, D.; Ali, N.A.; Raman, R.K.; Mohapatra, P.K.; Sehgal, R.; Mahanta, J.; Sultan, A.A.; Singh, N. Prevalence of mutations linked to antimalarial resistance in Plasmodium falciparum from Chhattisgarh, Central India: A malaria elimination point of view. *Sci. Reports 2017 71* **2017**, *7*, 1–8, doi:10.1038/s41598-017-

16866-5.

249. Wenthur, C.J.; Gentry, P.R.; Mathews, T.P.; Lindsley, C.W. Drugs for Allosteric Sites on Receptors. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 165, doi:10.1146/ANNUREV-PHARMTOX-010611-134525.

250. Louis, J.M.; Ishima, R.; Torchia, D.A.; Weber, I.T. HIV-1 Protease: Structure, Dynamics, and Inhibition. *Adv. Pharmacol.* **2007**, *55*, 261–298, doi:10.1016/S1054-3589(07)55008-8.

251. Bruder, M.; Polo, G.; Trivella, D.B.B. Natural allosteric modulators and their biological targets: Molecular signatures and mechanisms. *Nat. Prod. Rep.* **2020**, *37*, 488–514, doi:10.1039/c9np00064j.

252. Guarnera, E.; Berezovsky, I.N. Allosteric drugs and mutations: chances, challenges, and necessity. *Curr. Opin. Struct. Biol.* **2020**, *62*, 149–157, doi:10.1016/j.sbi.2020.01.010.

253. Cheng, X.; Jiang, H. Allostery in drug development. *Adv. Exp. Med. Biol.* **2019**, *1163*, 1–23, doi:10.1007/978-981-13-8719-7_1/COVER.

254. Motlagh, H.N.; Wrabl, J.O.; Li, J.; Hilser, V.J. The ensemble nature of allostery. *Nature* **2014**, *508*, 331–9, doi:10.1038/nature13001.

255. Han, B.; Salituro, F.G.; Blanco, M.J. Impact of Allosteric Modulation in Drug Discovery: Innovation in Emerging Chemical Modalities. *ACS Med. Chem. Lett.* **2020**, *11*, 1810–1819, doi:10.1021/ACSMEDCHEMLETT.9B00655/ASSET/IMAGES/MEDIUM/ML9B00655_0009.GIF.

256. Huang, W.; Nussinov, R.; Zhang, J. Computational tools for allosteric drug discovery: Site identification and focus library design. In *Methods in Molecular Biology*; NIH Public Access, 2017; Vol. 1529, pp. 439–446 ISBN 978-1-4939-6635-6.

257. Monod, J.; Wyman, J.; Changeux, J.P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12*, 88–118, doi:10.1016/S0022-2836(65)80285-6.

258. Pittman, R.N. Regulation of Tissue Oxygenation. *Regul. Tissue Oxyg.* **2011**, *3*, 1–100.

259. Nussinov, R.; Tsai, C.J.; Liu, J. Principles of allosteric interactions in cell signaling. *J. Am. Chem. Soc.* **2014**, *136*, 17692–17701,

doi:10.1021/JA510028C/ASSET/IMAGES/LARGE/JA-2014-10028C_0005.JPEG.

260. Kadenbach, B.; Jarausch, J.; Hartmann, R.; Merle, P.; Kagi, J.H.R.; Kojima, Y.; Knox, R.; Kohler, G.O.; Palter, R.; &walker, H.G.; et al. Plasmodium falciparum: Differential Sensitivity In Vitro to E-64 (Cysteine Protease Inhibitor) and Pepstatin A (Aspartyl Protease Inhibitor). *J. Protozool.* **1992**, *39*, 593–599, doi:10.1111/J.1550-7408.1992.TB04856.X.

261. Pandey, K.C.; Singh, N.; Arastu-Kapur, S.; Bogyo, M.; Rosenthal, P.J. Falstatin, a Cysteine Protease Inhibitor of Plasmodium falciparum, Facilitates Erythrocyte Invasion. *PLOS Pathog.* **2006**, *2*, e117, doi:10.1371/JOURNAL.PPAT.0020117.

262. Sundararaj, S.; Saxena, A.K.; Sharma, R.; Vashisht, K.; Sharma, S.; Anvikar, A.; Dixit, R.; Rosenthal, P.J.; Pandey, K.C. Cross-Talk between Malarial Cysteine Proteases and Falstatin: The BC Loop as a Hot-Spot Target. *PLoS One* **2014**, *9*, e93008, doi:10.1371/JOURNAL.PONE.0093008.

263. Barreiro, G.; De Alencastro, R.B.; Da, J.D.; Neto, M. A Semiempirical Study on Leupeptin: An Inhibitor of Cysteine Proteases. *Int J Quant Chem* **1997**, *65*, 11251134, doi:10.1002/(SICI)1097-461X(1997)65:6.

264. Shaw, E.; Angliker, H.; Rauber, P.; Walker, B.; Wikstrom, P. Peptidyl fluoromethyl ketones as thiol protease inhibitors. *Biomed. Biochim. Acta* **1986**, *45*, 1397–1403.

265. Olson, J.E.; Lee, G.K.; Semenov, A.; Rosenthal, P.J. Antimalarial effects in mice of orally administered peptidyl cysteine protease inhibitors. *Bioorg. Med. Chem.* **1999**, *7*, 633–638, doi:10.1016/S0968-0896(99)00004-8.

266. Palmer, J.T.; Rasnick, D.; Klaus, J.L.; Brömme, D. Vinyl Sulfones as Mechanism-Based Cysteine Protease Inhibitors. *J. Med. Chem.* **1995**, *38*, 3193–3196, doi:10.1021/JM00017A002/ASSET/JM00017A002.FP.PNG_V03.

267. Rosenthal, P.J.; Olson, J.E.; Lee, G.K.; Palmer, J.T.; Klaus, J.L.; Rasnick, D. Antimalarial effects of vinyl sulfone cysteine proteinase inhibitors. *Antimicrob. Agents Chemother.* **1996**, *40*, 1600–1603, doi:10.1128/aac.40.7.1600.

268. Hernández González, J.E.; Salas-Sarduy, E.; Hernández Alvarez, L.; Barreto Gomes, D.E.; Pascutti, P.G.; Oostenbrink, C.; Leite, V.B.P. In silico identification of noncompetitive inhibitors targeting an uncharacterized allosteric site of falcipain-2. *J.*

*Comput. Aided. Mol. Des.* **2021**, *35*, 1067–1079, doi:10.1007/s10822-021-00420-7.

269. Hernández González, J.E.; Alberca, L.N.; Masforrol González, Y.; Reyes Acosta, O.; Talevi, A.; Salas-Sarduy, E. Tetracycline Derivatives Inhibit Plasmodial Cysteine Protease Falcipain-2 through Binding to a Distal Allosteric Site. *J. Chem. Inf. Model.* **2022**, *62*, 159–175, doi:10.1021/acs.jcim.1c01189.

270. Kornacker, M.G.; Lai, Z.; Witmer, M.; Ma, J.; Hendrick, J.; Lee, V.G.; Riexinger, D.J.; Mapelli, C.; Metzler, W.; Copeland, R.A. An inhibitor binding pocket distinct from the catalytic active site on human β-APP cleaving enzyme. *Biochemistry* **2005**, *44*, 11567–11573, doi:10.1021/BI050932L/ASSET/IMAGES/MEDIUM/BI050932LN00001.GIF.

271. BIOVIA Discovery Studio - BIOVIA - Dassault Systèmes® Available online: https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/ (accessed on Apr 28, 2022).

272. Cai, J.; Robinson, J.; Belshaw, S.; Everett, K.; Fradera, X.; Van Zeeland, M.; Van Berkom, L.; Van Rijnsbergen, P.; Popplestone, L.; Baugh, M.; et al. Trifluoromethylphenyl as P2 for ketoamide-based cathepsin S inhibitors. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 6890–6894, doi:10.1016/J.BMCL.2010.10.012.

273. Shenoy, R.T.; Sivaraman, J. Structural basis for reversible and irreversible inhibition of human cathepsin L by their respective dipeptidyl glyoxal and diazomethylketone inhibitors. *J. Struct. Biol.* **2011**, *173*, 14–19, doi:10.1016/J.JSB.2010.09.007.

274. Boateng, R.A. Diversified Computational Approaches for the Identification of Orthosteric Drugs , Allosteric Modulators and Unveiling Drug Resistance Mechanisms ; Application to Infectious Diseases by. **2021**, 0–2.

275. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: a major update to the DrugBank database for 2018. **2018**, *46*, doi:10.1093/NAR/GKX1037.

276. Bento, A.P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L.J.; De Veij, M.; Leach, A.R. An open source chemical structure curation pipeline using RDKit. *J. Cheminform.* **2020**, *12*, 1–16, doi:10.1186/s13321-020-00456-1.

277. El-Hachem, N.; Haibe-Kains, B.; Khalil, A.; Kobeissy, F.H.; Nemer, G. AutoDock and AutoDockTools for Protein-Ligand Docking: Beta-Site Amyloid Precursor Protein

Cleaving Enzyme 1(BACE1) as a Case Study. *Methods Mol. Biol.* **2017**, *1598*, 391–403, doi:10.1007/978-1-4939-6952-4_20.

278. Seeliger, D.; De Groot, B.L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided. Mol. Des.* **2010**, doi:10.1007/s10822-010-9352-6.

279. Trott, O.; Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **2010**, *31*, 455, doi:10.1002/JCC.21334.

280. Sousa Da Silva, A.W.; Vranken, W.F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5*, 1–8, doi:10.1186/1756-0500-5-367/FIGURES/3.

281. Sijwali, P.S.; Shenai, B.R.; Rosenthal, P.J. Folding of the Plasmodium falciparum Cysteine Protease Falcipain-2 Is Mediated by a Chaperone-like Peptide and Not the Prodomain. *J. Biol. Chem.* **2002**, *277*, 14910–14915, doi:10.1074/JBC.M109680200.

282. Du, X.; Li, Y.; Xia, Y.L.; Ai, S.M.; Liang, J.; Sang, P.; Ji, X.L.; Liu, S.Q. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17*, doi:10.3390/IJMS17020144.

283. Falsafi-Zadeh, S.; Karimi, Z.; Galehdari, H. VMD DisRg: New User-Friendly Implement for calculation distance and radius of gyration in VMD program. *Bioinformation* **2012**, *8*, 341, doi:10.6026/97320630008341.

284. Liu, R.; Hu, J. Computational prediction of heme-binding residues by exploiting residue interaction network. *PLoS One* **2011**, *6*, e25560, doi:10.1371/journal.pone.0025560.

285. del Sol, A.; Fujihashi, H.; O'Meara, P. Topology of small-world networks of protein-protein complex structures. *Bioinformatics* **2005**, *21*, 1311–1315, doi:10.1093/bioinformatics/bti167.

286. Penkler, D.L.; Atilgan, C.; Tastan Bishop, Ö. Allosteric Modulation of Human Hsp90α Conformational Dynamics. *J. Chem. Inf. Model.* **2018**, *58*, 383–404, doi:10.1021/acs.jcim.7b00630.

287. Liu, Y.; Bahar, I. Toward understanding allosteric signaling mechanisms in the atpase domain of molecular chaperones. In *Pacific Symposium on Biocomputing 2010, PSB 2010*; WORLD SCIENTIFIC, 2010; pp. 269–280 ISBN 9814295299.